IEEE Signal Processing MAGAZINE
Volume 34 | Number 3 | May 2017

CONTENT ECOSYSTEM

Serving Diverse Interests in Our Community

Signal Processing for Finance

Perceptual Spatial Audio

High-Performance Depth Sensing

SP in Engineering Outreach

IEEE Signal Processing Society

IEEE

WORLD'S FIRST 10K Best Paper Award

IEEE ICME 2017

Diamond Best Paper

Knowledge-Guided Recurrent Neural Network Learning for Task-oriented Action Prediction

Sun Yat-sen University, China


Liang Lin


Lili Huang


Tianshui Chen


Yukang Gan


Hui Cheng

Platinum Best Papers

Analyzing the Group Sparsity Based on The Rank Minimization Methods

Zhiyuan Zha, Xin Liu, Xiaohua Huang, Henglin Shi, Yingyue Xu, Qiong Wang, Lan Tang, Xinggan Zhang

Nanjing University, China
University of Oulu, Finland

Efficient Graph-based Matrix Completion on Incomplete Animated Models

Evangelos Vlachos, Aris S. Lalos, Konstantinos Moustakas, Kostas Berberidis

University of Patras, Greece

Attribute Hashing for Zero-Shot Image Retrieval

Yahui Xu, Yang Yang, Fumin Shen, Xing Xu, Yuxuan Zhou, Heng Tao Shen

University of Electronic Science and Technology of China, China

FORCETAB: Visuo-Haptic Interaction with a Force-Sensitive Actuated Tablet

Jens Maiero, Ernst Kruijff, André Hinkenjann, Gheorghita Ghinea

Bonn-Rhein-Sieg University of Applied Sciences, Germany
Brunel University, UK

Intrinsic Decomposition From a Single RGB-D Image With Sparse and Non-Local Priors

Yujie Wang, Kun Li, Jingyu Yang, Xinchen Ye

Tianjin University, China
Dalian University of Technology, China

Evolving Boxes for Fast Vehicle Detection

Li Wang, Yao Lu, Hong Wang, Yingbin Zheng, Hao Ye, Xiangyang Xue

Fudan University, China
Shanghai Advanced Research Institute, CAS, China
University of Washington, USA

http://www.icme2017.org

Contents

Volume 34 | Number 3 | May 2017

FEATURES

ON THE COVER

This issue presents feature articles showcasing the significant and versatile roles that signal processing has been playing in quantitative finance, spatial audio, and visual depth sensing. The "SP Education" column presents systematic engineering outreach efforts at precollegiate levels, for which signal processing is a key building block in many projects engaging young students. The "Life Sciences" column presents velocity estimation in medical ultrasound images. The "From the Editor" column highlights an effort of "content ecosystem" that the magazine is exploring with other Society initiatives.

COVER IMAGE: ©ISTOCKPHOTO.COM/RZOZE19


PG. 112

COLUMNS


PG. 69

IEEE SIGNAL PROCESSING MAGAZINE (ISSN 1053-5888) (ISPREG) is published bimonthly by the Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, 17th Floor, New York, NY 10016-5997 USA (+1 212 419 7900). Responsibility for the contents rests upon the authors and not the IEEE, the Society, or its members. Annual member subscriptions included in Society fee. Nonmember subscriptions available upon request. **Individual copies:** IEEE Members US$20.00 (first copy only), nonmembers US$241.00 per copy. Periodicals postage paid at New York, NY, and at additional mailing offices. **Postmaster:** Send address changes to IEEE Signal Processing Magazine, IEEE, 445 Hoes Lane, Piscataway, NJ 08854 USA. Canadian GST #125634188 **Printed in the U.S.A.**

Digital Object Identifier 10.1109/MSP.2017.2678378

DEPARTMENTS

©GRAPHIC STOCK

ICME 2017 will be held in Hong Kong 10–14 July to discuss the latest development in multimedia technologies and related fields.

Digital Object Identifier 10.1109/MSP.2017.2678379

SCOPE: *IEEE Signal Processing Magazine* publishes tutorial-style articles on signal processing research and applications as well as columns and forums on issues of interest. Its coverage ranges from fundamental principles to practical implementation, reflecting the multidimensional facets of interests and concerns of the community. Its mission is to bring up-to-date, emerging and active technical developments, issues, and events to the research, educational, and professional communities. It is also the main Society communication platform addressing important issues concerning all members.

Min Wu | Editor-in-Chief | minwu@umd.edu

Content Ecosystem

Serving diverse interests in our community

I am writing this editorial just days after coming back from New Orleans, Louisiana, where the 42nd International Conference on Acoustics, Speech, and Signal Processing (ICASSP) was successfully held. It was an action-packed week, filled with technical presentations and exchanges as well as a broad range of discussions on our Society and community developments. About 50 committees and boards of the IEEE Signal Processing Society (SPS) met at ICASSP, engaging hundreds of volunteers from around the world.

My discussions with a number of colleagues during the conference were about our *IEEE Signal Processing Magazine (SPM)*, either to explore ideas for potential articles as inspired by technical talks and panels or to brainstorm how the magazine can help support the initiatives or activities of the IEEE Signal Processing Society (SPS). For this latter role, a notion of "content ecosystem" was coined in my meeting with chief editors of the SigPort Repository, Dr. Yan Lindsay Sun, and of the Resource Center (formerly SigView), Dr. John McAllister, respectively, and the SPS staff members on public outreach and publications.

We have been seeing an increase in the diverse needs from our members, and SPS and IEEE have been developing multiple products and tools to address these needs. For example, SigPort offers an increasingly popular platform to archive slides and posters from the latest conferences as well as student theses and reports viewable by anyone free of charge, while the Resource Center hosts keynote videos and tutorial webinars from curated sources, such as ICASSP plenary talks, as an SPS member benefit and may be available for a fee for nonmembers (see page 109 for more information about the Resource Center). As the Resource Center expands its content, where could a constant stream of candidates come from systematically?

Our magazine's tutorial articles provide a pool of candidates for the Resource Center—authors can be approached and encouraged to create a multimedia companion to their article, which allows content to reach a bigger audience. In the meantime, the slides and posters provided by SigPort users include download statistics that shine a light on the community's interests toward more recent technical work and grassroots educational material. SigPort thus can offer an additional set of candidates for the Resource Center to explore. If a healthy mechanism is developed and coordinated well, the contents on all three platforms can work hand in hand as an ecosystem.

Another example of coordinating multiple platforms and developing a content ecosystem was piloted in 2016 when we solicited input on student hands-on projects, and this is currently being expanded in collaboration with the SPS Education Committee. More specifically, SigPort provides a handy platform for students to archive their project reports, which is perhaps the only worldwide repository for such grassroots material. Editors from *SPM* and the Education Committee can evaluate these archived documents on a rolling basis and periodically create a curated summary and a set of highlights to share with the community, for example, through the magazine's Inside Signal Processing e-Newsletter and/or through a column in *SPM*. Some of the selected projects may also feed into the content pool for the Resource Center. These recognitions and visibilities provide positive feedback and encourage the students to engage in interesting projects and share the projects with the community at large.

For researchers and frequent authors of original research publications, a repository to host and broadly share their preprints is attractive. In addition to SigPort, quite some signal processing authors are archiving their preprints on arXiv.org, which traces its roots to the science community. For various historical reasons, arXiv does not have an electrical engineering topic branch; topics such as information theory and control theory are put either as part of computer science or applied mathematics, and there was no explicit topic category of signal processing.

The SPS Board of Governors had a long discussion on this issue last fall on approaching the arXiv leadership to create a signal processing topic category. Although it is perhaps easier to create a signal processing category under the existing computer science or applied math categories, several board members,

(continued on page 13)

Digital Object Identifier 10.1109/MSP.2017.2678380
Date of publication: 26 April 2017

PRESIDENT'S MESSAGE

Rabab Ward | SPS President | rababw@ece.ubc.ca

Diversity Through Adversity

The IEEE Signal Processing Society (SPS) aims to serve a global community of signal processors and stimulate collaboration among a diverse mix of scientists, academics, inventors, and industries. Our work thrives based on the diversity of our members—students, professors, engineers, researchers, and more in terms of cultural background, religion, gender identity, sexual orientation, profession, and age—whatever our defining characteristics, one common feature among signal processors is that we are problem solvers. We thrive on challenges, and those challenges include overcoming cultural and personal adversities. Recent news seems to have increased those challenges.

Engineers and scientists are not always notable activists, and we are certainly not immune to personal and cultural bias, which can adversely impact the development of technological breakthroughs. For many centuries, the greatest minds in science, including Galileo and Newton, were confounded by the problem of finding an accurate measurement for latitude, which left sailors lost at sea and curbed exploration. While scientists looked for the answer in the skies, through astronomy, an English carpenter named John Harrison set his mind to developing a seaworthy clock, spending decades refining his successful invention, while facing ridicule from stargazing scientists. Imagine how many shipwrecks might have been prevented if the scientists had partnered with the carpenter?

Perhaps the greatest modern example of a fruitful partnership that crossed virtually every cultural divide was that between Cambridge University don, Prof. Hardy, and the young, self-taught Indian mathematician, Srinivasa Ramanujan. Ramanujan's work was dismissed by British scientists, but fortunately Hardy recognized his genius. Their collaboration inspired the 2016 movie *The Man Who Knew Infinity* and the 2007 play, *A Disappearing Number*, which elegantly depicted the ways that diverging and converging series in mathematics become a metaphor for the struggles of nonconforming diasporas. Hardy was an atheist and a disciple of scientific rigor, while Ramanujan was driven by intuition and, as a devout Hindu, equated the meaning of equations to the hand of God.

Their complex relationship beautifully illustrates that our personal, political, religious, geographic, and cultural differences can stimulate and propel science. The task of problem solving has no single language, nationality, color, gender, faith, or personality type. Consider the most basic equation in our electrical engineering discipline: voltage = current × resistance, which is a product of scientists from different countries. Voltage is named after Alessandro Volta, an Italian physicist and chemist. Current is measured in amperes, named after the Frenchman André Marie Ampère, an early innovator in experimental physics. Resistance was named after the German physicist George Ohm, who for many years faced resistance by academics in the field. Taking this equation one step further, if you multiply voltage by current, you will get power, measured in watts, after the self-taught Scot instrument maker James Watt, whose early attempts to commercialize his inventions were often stymied by financial hardship and academic barriers. Tolerance, acceptance, and open communication are imperative for successful science. Closing channels for the exchange of ideas, knowledge, and information is not only scientifically disadvantageous but an affront to forward-thinking, solutions-based progress of the engineering field.

Let's also remember that the values of these quantities are expressed in Arabic numerals, and these include the number zero, which was developed in India in the third century and spread to the Middle East within 200 years. But it was absent from Western mathematics until the 1600s, in part due to political and racial bias.

Sometimes, we have to literally dig through history to find the proof of a breakthrough's origins and set the scientific record straight. Until recently, 14th century European scholars were believed to have cracked the code for predicting the movement of objects in space–time. However, a German astro-archeologist who has devoted many years attempting to decipher (Iraq's) Babylonian cuneiform tablets, has recently found a missing piece in the puzzle, receiving photographs from a Viennese astrophysicist of uncatalogued tablets from the British Museum in London. It turned out that ancient Babylonians had already mapped

Digital Object Identifier 10.1109/MSP.2017.2678381
Date of publication: 26 April 2017

planetary trajectories to calculate their future position using advanced geometrical calculations.

Many scientists have had to solve problems while also facing and overcoming personal life challenges and social stigmas.

For much of Alan Turing's early life, he had absentee parents and spent his formative years at boarding schools, where his natural gift and passion for math was often overlooked. He later endured homophobia, all the while developing a general-purpose computer and a code-breaking device that would ultimately save many lives.

The beautiful mind of John Nash wrestled with mental illness as well as its stigma and its draconian treatments, enduring insulin shock therapy, yet he later developed the Nash equilibria and game theory.

Sometimes our founding interests, theories, and inventions are seeded by personal challenges. My own life and career have been enriched by research collaborators who have overcome their own personal adversities, such as Charles Laszlo, a biomedical engineer. Laszlo started his long struggle with hearing loss at the age of 21 during his military service and then went on to develop many hearing-assistance devices and technologies that help people with hearing problems. My research collaboration with Laszlo was on analyzing the signals of a crying baby to understand *why* the baby is crying, for the benefit of those parents who are deaf. I have also worked closely with Gary Birch on brain–computer interface research. At 18, Birch became a quadriplegic. Yet within five years, he became a medal-winning Paralympian, while also studying engineering. He went on to receive a doctorate in electrical engineering and turned to researching brain–computer interfaces, a fascinating area of research that taps into the brain's signals to enrich the lives of people with disabilities. These colleagues have both received the highest award in Canada, The Order of Canada, in recognition for their tremendous contributions to society.

These colleagues have not only stimulated my career, they have become dear friends. Life cannot evolve and advance if we live and work in isolation. Technological breakthroughs are the result of diverse collaborations, and it's the goal of the SPS to facilitate and enhance these collaborations, with each other and by connecting with the world's current and future innovators.

You may have noticed that I didn't mention the important, barrier-breaking contributions of women scientists and engineers that have developed breakthrough technologies. I've had the good fortune to work with many women engineers and signal processors. To use the cliché, we've come a long way. Yet society still has a way to go to bridge the gender gap. I will talk about this fascinating subject in a future column. Meanwhile, I look forward to hearing any of your ideas about ways in which we can strengthen diversity and break barriers, both inside and outside the global signal processing community.

SP

Top Downloads in IEEE *Xplore*

Each "Reader's Choice" column focuses on a different publication of the IEEE Signal Processing Society (SPS). This month we are highlighting articles in *IEEE Signal Processing Letters*.

IEEE Signal Processing Letters is a monthly, archival publication designed to provide rapid dissemination of original, cutting-edge ideas and timely, significant contributions in signal, image, speech, language, and audio processing. Papers published in *IEEE Signal Processing Letters* can be presented within one year of their appearance in signal processing conferences such as the IEEE International Conference on Acoustics, Speech, and Signal Processing; IEEE Global Conference on Signal and Information Processing; and IEEE International Conference on Image Processing and also in several workshop organized by the SPS.

This issue's "Reader's Choice" column lists the top ten articles most downloaded for the past two years at the time of the print deadline. Your suggestions and comments are welcome and should be sent to Associate Editor Chungshui Zhang (zcs@mail.tsinghua.edu.cn).

A Universal Image Quality Index

Wang, Z.; Bovik, A.C.

In this paper, a new universal objective image quality index, which is easy to calculate and applicable to various image processing applications, is proposed. The proposed index is designed by modeling any image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion.

March 2002

Reversible Image Data Hiding with Contrast Enhancement

Wu, H.-T.; Dugelay, J.-L.; Shi, Y.-Q.

This paper proposes a novel reversible data hiding (RDH) algorithm for digital images. The proposed algorithm enhances the contrast of a host image to improve its visual quality. The authors claimed that it was the first algorithm that achieved image contrast enhancement by RDH.

January 2015

On the Performance of Non-Orthogonal Multiple Access in 5G Systems with Randomly Deployed Users

Ding, Z.; Yang, Z.; Fan, P.; Poor, H.V.

In this letter, the performance of nonorthogonal multiple access (NOMA) is investigated in a cellular downlink scenario with randomly deployed users. The developed analytical results show that NOMA can achieve superior performance in terms of ergodic sum rates.

Digital Object Identifier 10.1109/MSP.2017.2668518
Date of publication: 26 April 2017

However, the outage performance of NOMA depends critically on the choices of the users' targeted data rates and allocated power.

December 2014

Fairness for Non-Orthogonal Multiple Access in 5G Systems

Timotheou, S.; Krikidis, I.

The authors study nonorthogonal multiple access from a fairness standpoint and investigate power allocation techniques that ensure fairness for the downlink users under 1) instantaneous channel state information (CSI) at the transmitter and 2) average CSI. They have developed low-complexity polynomial algorithms for these nonconvex problems that yield the optimal solution in both cases considered.

October 2015

Robust Edge-Stop Functions for Edge-Based Active Contour Models in Medical Image Segmentation

Pratondo, A.; Chui, C.-K.; Ong, S.-H.

In this paper, authors propose a framework to construct a group of edge-stop functions (ESFs) for edge-based active contour models to segment objects with poorly defined boundaries. In this framework, which incorporates gradient information as well as probability scores from a standard classifier, the ESF can be constructed from any classification algorithm and applied to any edge-based model using a level set method.

February 2016

Discrete Anamorphic Transform for Image Compression

Asghari, M.H.; Jalali, B.

The authors introduce a physics-based transform that enables image compression by increasing the spatial coherency. They also present the stretched modulation distribution, a new density function that provides the recipe for the proposed image compression.

July 2014

An Adaptive Motion Model for Person Tracking with Instantaneous Head-Pose Features

Baxter, R.H.; Leach, M.J.V.; Mukherjee, S.S.; Robertson, N.M.

This letter presents novel behavior-based tracking of people in low resolution using instantaneous priors mediated by head pose. The authors extend the Kalman filter to adaptively combine motion information with an instantaneous prior belief about where the person will go based on where they are currently looking.

May 2015

Empirical Mode Decomposition as a Filter Bank

Flandrin, P.; Rilling, G.; Goncalves, P.

This paper reports on numerical experiments empirical mode decomposition (EMD) based on fractional Gaussian noise. In such a case, it turns out that EMD acts essentially as a dyadic filter bank resembling those involved in wavelet decompositions. It is also pointed out that the hierarchy of the extracted modes may be similarly exploited for getting access to the Hurst exponent.

February 2004

An Experimental Study on Speech Enhancement Based on Deep Neural Networks

Xu, Y.; Du, J.; Dai, L.-R.; Lee, C.-H.

This letter presents a regression-based speech enhancement framework using deep neural networks with a multiple-layer deep architecture. A series of pilot experiments were conducted under multicondition training with more than 100 h of simulated speech data, resulting in a good generalization capability even in mismatched testing conditions.

January 2014

Fast Matrix Inversion Updates for Massive MIMO Detection and Precoding

Rosário, F.; Monteiro, F.A.; Rodrigues, A.

In this letter, the authors propose an on-the-fly method to recompute the zero forcing filter when a user is added or removed from massive multiple-input, multiple-output system. They evaluate the recalculation of the inverse matrix after a new channel estimation is obtained for a given user. With fewer operations, the performance after an update also remains close to the initial one.

January 2016

SP

New Society Officers Elected for 2018

The Board of Governors (BoG) of the IEEE Signal Processing Society (SPS) elected three new officers who will start their terms on 1 January 2018: Ahmed H. Tewfik (University of Texas at Austin) will serve as 2018–2019 SPS president-elect. He succeeds Ali H. Sayed who has held the post of president-elect and will become SPS president in 2018. Fernando Pereira (Instituto de Telecomunicações, Portugal) will serve as 2018–2020 SPS vice president, conferences. He succeeds Carlo Regazzoni, who has held this position since January 2015. Sergios Theodoridis (University of Athens, Greece) will serve as 2018–2020 SPS vice president, Publications. He succeeds Thrasyvoulos N. Pappas who has held this position since January 2015.

Meet the Newly Elected Society Officers

2018–2019 SPS President-Elect

Ahmed H. Tewfik

Ahmed H. Tewfik received his B.Sc. degree from Cairo University, Egypt, in 1982 and his M.Sc., E.E., and Sc.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1984, 1985, and 1987, respectively. He is the Cockrell Family Regents Chair in Engineering and the chair of the Department of Electrical and Computer Engineering at the University of Texas at Austin. As chair, he is credited with substantially increasing the funding of the department; establishing public-private partnerships; and increasing the hands-on, industry, and entrepreneurial experiences of its students. He was the E.F. Johnson Professor of Electronic Communications with the Department of Electrical Engineering at the University of Minnesota until September 2010.

Dr. Tewfik worked at Alphatech, Inc. and served as a consultant to several companies, including MTS Systems, Inc., Eden Prairie, Minnesota; Emerson-Rosemount, Inc., Eden Prairie, Minnesota; CyberNova, Milipitas, California; Macrovision, Santa Clara, California; Visionaire Technology, Fremont, California; Ipsos, New York; InterDigital Communications, King of Prussia, Pennsylvania; Keyeye Communications, Sacramento, California; Transoma Medical, Arden Hills, Minnesota; and St. Jude Medical, Minnetonka, Minnesota. From August 1997 to August 2001, he was the president and chief executive officer of Cognicity, Inc., an entertainment marketing software tools publisher that he cofounded, while on partial leave of absence from the University of Minnesota. His current research interests are in technology for human cognitive augmentation and decision making, brain computing interfaces, bioelectronics, and mobility.

Dr. Tewfik is an IEEE Fellow. His professional activities include founding cochair, 2013 IEEE GlobalSIP Conference; SPS vice president, technical directions (2010–2012); general chair, 2009 IEEE Genomic Signal Processing and Statistics Conference; member-at-large, SPS BoG (2006–2008); president, SPS/IEEE Communications Society Minnesota Chapter (2003–2005); SPS Distinguished Lecturer (1998–1999); first editor-in-chief, *IEEE Signal Processing Letters* (1993–1999); associate editor, *IEEE Transactions on Signal Processing* (*TSP*); and guest editor of three special issues of *TSP* on wavelets and their applications and watermarking, special issue of *IEEE Transactions on Multimedia* on multimedia databases, and special issue of *IEEE Journal of Selected Topics in Signal Processing* on bioinformatics. He was awarded the IEEE Third Millennium Medal (2000), the E.F. Johnson Professorship of Electronic Communications at the University of Minnesota (1993), the Taylor Faculty Development Award from the Taylor Foundation (1992), and a National Science Foundation Research Initiation Award (1990). He has given several plenary and keynote lectures as well as taught tutorials on bioinformatics, ultrawideband communications, watermarking, and wavelets at major IEEE conferences. He has also volunteered in technical and STEM activities at the IEEE Central Texas Society Chapter.

Digital Object Identifier 10.1109/MSP.2017.2685938
Date of publication: 26 April 2017

2018–2020 SPS Vice President, Conferences

Fernando Pereira

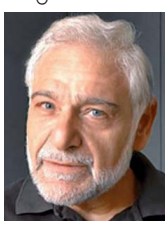
Fernando Pereira graduated with a degree in electrical and computer engineering from Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Portugal, in 1985. He received his M.Sc. and Ph.D. degrees in electrical and computer engineering from IST in 1988 and 1991, respectively. Prof. Pereira is currently with the Department of Electrical and Computer Engineering of IST and with the Instituto de Telecomunicações, Lisbon, Portugal. He is responsible for IST's participation in many national and international research projects.

He is a Fellow of the IEEE (2008), the Institution of Engineering and Technology (IET) (2015), and the European Association for Signal Processing (EURASIP) (2013). He was elected to serve on the European Signal Processing Society Board of Directors for the term 2015–2018. He has contributed more than 250 papers in international journals, conferences, and workshops and delivered multiple invited talks at conferences and workshops. His areas of interest are video analysis, coding, description and adaptation, and advanced multimedia services.

Prof. Pereira is/has been a member of the scientific and program committees of many IEEE international conferences and workshops, including: technical program cochair, IEEE International Conference on Image Processing (ICIP) 2010 and 2016; technical program chair, 2014 IEEE Workshop on Multimedia Signal Processing and 2008 and 2012 International Workshop on Image Analysis for Multimedia Interactive Services; and general chair, 2007 Picture Coding Symposium, he has assumed many other roles in IEEE and SPS conferences and workshops. He has also been involved in the following professional activities: member, SPS Awards Board (2017–2018); area editor, *Signal Processing: Image Communication Journal* (2017–present); member-at-large, SPS BoG (2012 and 2014–2016); ICIP 2016 representative, Technical Directions Board (2015–2016); and member, Publications Board (2013–2015). He has also been the editor-in-chief, *IEEE Journal of Selected Topics in Signal Processing* (2013–2015); Editorial Board member, *IEEE Signal Processing Magazine* (2009–2011); associate editor, *IEEE Transactions on Circuits and Systems for Video Technology, IEEE Transactions on Image Processing* (2002–2007), and *IEEE Transactions on Multimedia* (2009–2012); member, SPS Image, Video, and Multidimensional Signal Processing Technical Committee (2004–2009), SPS Multimedia Signal Processing Technical Committee (2006–2009 and 2011–2013), Circuits and Systems (CAS) Visual Signal Processing and Communications Technical Committee and CAS Multimedia Systems and Applications Technical Committee; and SPS Distinguished Lecturer (2005). In terms of international standardization efforts, he has been the MPEG Requirements chair from 2002 to 2007 and since February 2016, he has been the JPEG Requirements chair.

2018–2020 SPS Vice President, Publications

Sergios Theodoridis



Sergios Theodoridis is a professor of signal processing and machine learning in the Department of Informatics and Telecommunications of the University of Athens. He was the Otto Monstead guest professor, Technical University of Denmark, 2012, and holder of the Excellence Chair, Department of Signal Processing and Communications, University Carlos III, Madrid, Spain, 2011. His research interests lie in the areas of adaptive algorithms, distributed and sparsity-aware learning, machine learning and pattern recognition, and signal processing for audio processing and retrieval. He is the author of *Machine Learning: A Bayesian and Optimization Perspective* (Academic Press, 2015); the coauthor of *Pattern Recognition* (Academic Press, fourth edition, 2009) and *Introduction to Pattern Recognition: A MATLAB Approach* (Academic Press, 2010); the coeditor of *Efficient Algorithms for Signal Processing and System Identification* (Prentice Hall, 1993); and the coauthor of three books in Greek, two of them for the Greek Open University.

Prof. Theodoridis is the coauthor of seven papers that have received Best Paper Awards, including the 2014 IEEE Signal Processing Magazine Best Paper Award and the 2009 IEEE Computational Intelligence Society Transactions on Neural Networks Outstanding Paper Award. He is the recipient of the 2014 IEEE Signal Processing Society Education Award and the 2014 EURASIP Meritorious Service Award.

He is a Fellow of the IEEE, IET, and EURASIP and is a corresponding fellow of the Royal Society of Edinburgh. His professional activities include editor-in-chief, *IEEE Transactions on Signal Processing* and Signal Processing book series (Academic Press); coeditor-in-chief, *E-Reference Signal Processing Handbook* (Elsevier); Distinguished Lecturer, SPS (2009–2010) and CAS (2014–2016); chair, SPS Signal Processing Theory and Methods Technical Committee (2013–2015); member, SPS Fellows Committee (2014–2016); general chair, European Signal Processing Conference (1998) technical program cochair, the International Symposium on Circuits and Systems (ISCAS) 2006 and ISCAS 2013; cochair and cofounder, CIP Workshop on Cognitive Information Processing 2008 and cochair, CIP 2010; member, CAS Board of Governors (2008–2010); and member-at-large, SPS Board of Governors (2011–2013). He also served as a member, Greek National Council for Research and Technology; chair, SP advisory committee for the Edinburgh Research Partnership; vice chair, Greek Pedagogical Institute; and member, Board of Directors of COSMOTE (the Greek mobile phone operating company) for four years. He served as the president of EURASIP (2004–2006) and currently serves as chair, EURASIP Fellows Committee.

SP

John Edwards

New Directions in Navigation and Positioning

Signal processing-enabled technologies pinpoint people, places, and things

In an era of same-day product deliveries, interplanetary space probes, and autonomous vehicles, transporting something—or someone—from here to there quickly, directly, and precisely is becoming increasingly important.

An array of navigation and positioning technologies are now available to help guide and locate vehicles, people, and almost endless number of objects. The satellite-based global positioning system (GPS), for instance, now lies at the heart of an almost endless array of location, navigation, timing, mapping, and tracking tools. Real-time location system (RTLS) technologies, meanwhile, rely on resources such as GPS, Wi-Fi, Bluetooth, near-field communication (NFC), and radio-frequency identification (RFID) to detect the current location of a target, which may be anything from a vehicle to an item in a manufacturing plant to a person.

With navigation and positioning technologies continuing to fuel the development of innovative commercial, industrial, consumer, and scientific applications, researchers are turning to signal processing methods and approaches to tweak the performance of existing systems as well as to pioneer completely new tools and services.

A GPS alternative

Researchers at the University of California, Riverside (UCR), have developed a new navigation system that is based entirely on existing terrestrial signals, such as cellular and Wi-Fi, rather than GPS. The new technology, which the researchers claim is both highly reliable and accurate, can function as a standalone alternative to GPS or as an alternative to satellite signals to enable highly reliable, consistent, and tamper-proof navigation in autonomous systems, such as robots, driverless terrestrial vehicles, and unmanned aerial vehicles (UAVs).

"GPS is unreliable for anytime, anywhere navigation, including indoors and in deep urban canyons," says team leader Zak Kassas, an assistant professor of electrical and computer engineering in UCR's Bourns College of Engineering. He notes that GPS signals are also highly vulnerable to interference, jamming, and spoofing. "However, in most GPS-challenged environments, there are dozens of signals of opportunity (SOPs) that are available at various frequencies, geometry and transmission protocols, and whose received power is much higher than GPS," Kassas says.

By exploiting abundantly available SOPs, the new approach reduces the sensory payload that's typically used to compensate for GPS's shortcomings. "Current and future vehicles, whether manual, semiautonomous or fully autonomous, ground and aerial, would benefit from this research," Kassas says.

The system can be used by itself or to supplement inertial navigation system data in the event of GPS failure. The researchers' approach include theoretical analysis of SOPs in the environment, building specialized software-defined radios (SDRs) that can extract relevant timing and positioning information from SOPs, developing practical navigation algorithms and, finally, testing the system on ground vehicles and unmanned drones.

"We have designed state-of-the-art specialized SDRs for cellular code division multiple access and long-term evolution (LTE) signals," Kassas says. "We mounted our SDRs on ground vehicles and UAVs and demonstrated experimentally these vehicles navigating to an unprecedented level of accuracy only with cellular signals." The trajectories produced by the SDRs were within a few meters from a trajectory produced with traditional GPS receivers (Figure 1). "To my knowledge, we were the first to demonstrate UAVs navigating exclusively with cellular signals," Kassas remarks.

Kassas notes that the system uses signal processing in all of its stages. "We start by studying the SOPs and deriving theoretical signal models for what useful position-navigation-timing information we can extract from these signals," he says. "We then design SDRs that process these signals and output useful information [then] fuse the extracted information with signals from other sensors to achieve an accurate and robust navigation solution." The SDRs contain phase-locked loops, delay-locked loops, frequency-locked loops, fast Fourier transforms, inverse FFTs, and

Digital Object Identifier 10.1109/MSP.2017.2668526
Date of publication: 26 April 2017

FIGURE 1. The simulation results for an unmanned drone flying over downtown Los Angeles showing the true trajectory (red line) with GPS navigation only (yellow line) and GPS aided with cellular signals (blue line). (Figure used courtesy of UCR.)

numerically controlled oscillators. "Our integrated navigation filter uses extended Kalman filters," Kassas says.

SOPs have been extensively analyzed for communication applications, yet are still not well understood as potential PNT sources. "We have been deriving theoretical signal models capturing phenomena that were never discovered prior to our research, because these signals were not intended for PNT purposes," Kassas says. "For example, we discovered that the observed clock bias corresponding to different sectors in the same base transceiver station cell is not the same." The difference turned out to be on the order of nanoseconds. "While the difference is not harmful for communication purposes, it introduces positioning error of tens of meters if not modeled and accounted for appropriately," Kassas explains.

Kassas says that autonomous vehicles are likely to benefit most from his team's research. "Autonomous vehicles will inevitably result in a sociocultural revolution," he says. "Our overarching goal is to get these vehicles to operate with no human in the loop for prolonged periods of time, performing missions such as search, rescue, surveillance, mapping, farming, firefighting, package delivery, and transportation."

Space optical communication and navigation

In 2013, the U.S. National Aeronautics and Space Administration (NASA) demonstrated the Lunar Laser Communication Demonstration experiment, a laser communication prototype that achieved record-breaking data download and upload speeds between Earth and the moon. Now, a NASA optical physicist claims he can equal those speeds and create extremely precise distance and speed measurements using a single compact package (Figure 2).

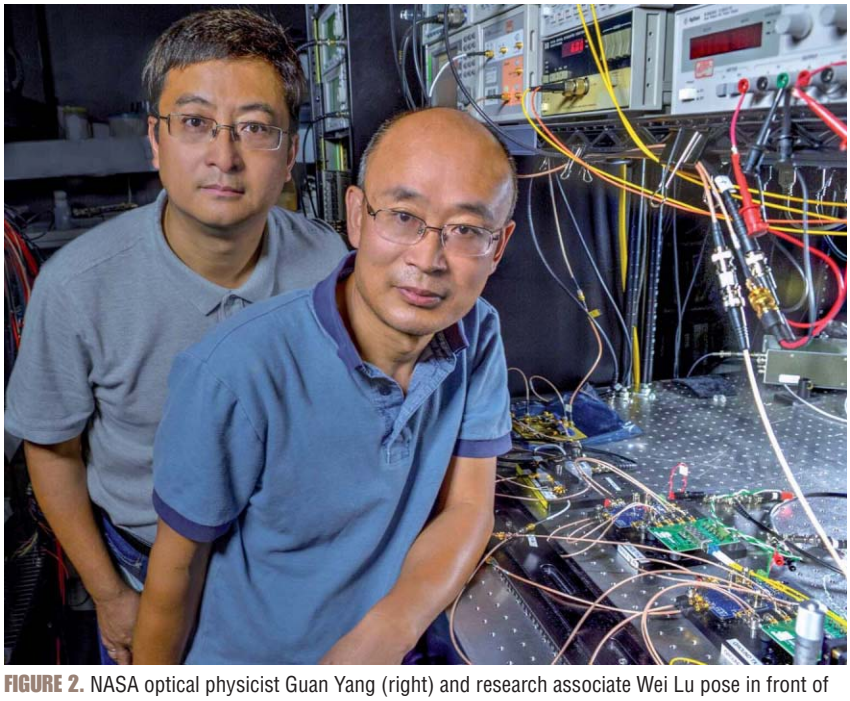


FIGURE 2. NASA optical physicist Guan Yang (right) and research associate Wei Lu pose in front of the lasercom breadboard they created to demonstrate high data-rate download and uplink speeds as well as highly precise distance and speed measurements all from the same, relatively small package. (Photo used courtesy of NASA.)

The new Space Optical Communication and Navigation System is a miniaturized lasercom transceiver comprising commercially available components that simulate both ground and space terminals. In recent laboratory tests, the system showed that it could provide micrometer-level distance and speed measurements over a 622 Mb/s laser communication link. "This technology decreases ranging and range-rate errors by orders of magnitude," says Guangning Yang, an optical physicist at NASA's Goddard Space Flight Center in Greenbelt, Maryland. "We can more precisely determine a spacecraft's orbit relative to an absolute location." Besides transmitting data at LLCD's record-breaking rate of 622 Mb/s, the new transceiver measured speed within a precision of less than 10 μm/s and distances within 20 μm. The system achieved the precise measurements by incorporating a Doppler frequency enabled by an FFT.

NASA's current space communication and navigation systems, such as the tracking data and relay satellite (TDRS), provide two major services—communication and satellite tracking. Tracking keeps tabs on a satellite's location, speed, and orbit. "We do this by continuously measuring the spacecraft's distance and speed relative to a fixed reference point with an RF communication satellite, like TDRS," Yang says. "With these measurements, we can calculate the spacecraft's speed, distance, and orbit." Space optical communications promise to provide similar services, but with higher data bandwidth and enhanced tracking precision.

"With significantly improved satellite location and speed information, we will enable better scientific-data collection and processing," Yang says. "This high-precision instrumentation, which is low mass, consumes less power, is relatively small, and will enable many other scientific instruments that require high-precision ranging, such as flying information with a constellation of satellites."

The system's precise measurement capability is tied directly to high-precision frequency synthesizing, low-noise implementation of frequency and timing detection. "The final instrumentation will be carried out in a digital domain with lots of digital signal processing," Yang says. "The types of digital signal processing we used include digital frequency synthesizing, digital-phase detecting, digital-phase lock loops, digital filters, and digital-time interval counters for ranging measurements," he explains. The measurements are processed in a software tool that relies on an extended Kalman filter to produce high-precision orbit-state estimates.

Yang says that the biggest challenge faced so far has been conducting high-precision frequency measurements of a Doppler-shifted clock signal within a digital domain.

Yang says that the biggest challenge faced so far has been conducting high-precision frequency measurements of a Doppler-shifted clock signal within a digital domain. "The obvious advantage with digital implementation is size and flexibility," Yang notes. "A single field-programmable gate array can accomplish much more in a small form factor."

Yang says that the project is moving ahead in two directions. "On one hand, we will try to implement the current version of this technology on an existing hardware platform, such as Goddard's NavCube (a spacecraft bus that is typically no larger than a shoebox), which is a powerful navigational technology," he remarks. "At the same time, we are pushing to further advance the technology into continuous optical-phase measurement."

Atomic gyroscope

U.S. National Institute of Science and Technology researchers have developed a compact, low-power atomic gyroscope design that promises to give precise navigational capabilities to spacecraft, submarines, and other vehicles hampered by size, weight, and power restrictions. The gyroscope can also simultaneously measure acceleration, enabling navigation by "dead reckoning" without reference to external landmarks or stars.

The gyroscope, an atom interferometer, is based on an expanding cloud of laser-cooled atoms, an approach first demonstrated at Stanford University. Traditional optical interferometry involves combining or "interfering" the electromagnetic waves in light and then extracting information about the original light paths from the resulting wave patterns. An atom interferometer leverages the ability of atoms to act as both particles and waves, interfering these waves to measure the forces exerted on atoms. When atoms speed up or rotate, their matter waves shift and interfere in predictable ways that are visible in interference patterns.

The basic concept behind the new gyroscope is similar to the principle underlying optical ring-laser gyroscopes, says Gregory Hoth, a postdoctoral research associate in NIST's Time and Frequency Division. "We take a wave and split it into two parts, he says. "We arrange for the two wave packets to travel along different paths and then recombine them and look at the amplitude of the wave that comes out." If the separated paths enclose an area, the output wave amplitude will depend on whether or not the device is rotating. "This is often called the Sagnac effect," Hoth says. For optical gyroscopes, the waves are light waves, and the amplitude corresponds to the intensity of the light. "For our gyroscope, the waves are matter-waves and the amplitude corresponds to the probability for an atom to occupy a specific energy state," Hoth says.

At the gyroscope's heart is a small glass chamber containing a sample of about 8 million cold rubidium atoms that are continuously trapped and released. As the atoms fall under gravity, a laser beam causes them to transition between two energy states. The process gives the atoms momentum and forces their matter waves to separate and later recombine to interfere. The cold atom cloud expands to as much as five times its initial size during the 50 ms (thousandths of a second) measurement sequence, which creates a correlation between each atom's speed and its final position. The interference effect on an atom depends on its speed, so rotations generate interfering bands of atoms across images of the final cloud (Figure 3).

The atoms are imaged by shining a second, weak laser beam through the cloud. Because atoms in different energy states absorb light of different frequencies, the final energy state of the atoms can be detected. The resulting images show interference bands of atom populations in the two different energy states. The rotation rate and axis are measured by analyzing the spacing and direction of the interference bands across the atom cloud. Acceleration is deduced from changes in the central band. The interferometer is sensitive to acceleration along the direction of the light and sensitive to rotations perpendicular to the light.

"The signal processing challenge in our experiment is to take images of the transition probability and estimate the wavelength of the fringe pattern," Hoth says. "To use the system as a gyroscope, you would use the wavelength of the fringe pattern to infer the unknown rotation rate." In experiments to date, the researchers' goal has been to quantify the relationship between the rotation rate and the wavelength of the fringe pattern so they can compare that relationship to theoretical predictions. "We do that by applying a known rotation rate and measuring the wavelength of the observed fringe pattern," Hoth says.

The process requires three images. "Each image has a fringe pattern with the same wavelength, but we vary the phase so that we see different parts of the fringe pattern," Hoth says. "By combining the three images, we can get the fourth image, which shows the spatial variation of the interferometer phase." The fringe pattern is equivalent to a slope or gradient in the interferometer phase. "By calculating the spatial phase, we solve both of our problems," Hoth continues. "The unwanted structure is suppressed and the hard problem of estimating a fringe wavelength has turned into the easy problem of estimating a best fit slope."

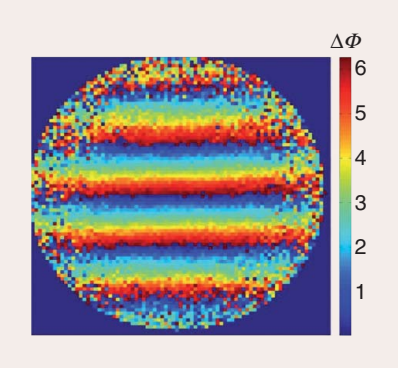


FIGURE 3. NIST's compact gyroscope measures rotation by analyzing patterns of interfering matter waves in an expanding cloud of atoms transitioning between two energy states. Each atom's speed determines both its final position in the cloud and the size of the rotational signal that shifts the interference patterns. Thus, rotations generate interfering bands of atoms across images of the final cloud. The color coding indicates how much the interference patterns shift in radians, the standard unit of angular measure. The orientation of the interfering bands (horizontal in the image) indicates the rotation axis. The rotation rate, determined by an analysis of the band spacing, is 44 milliradians/s. (Figure used courtesy of NIST.)

"The basic idea is that the part of the signal we're interested in changes when we modulate the phase, but the parts that we want to suppress stay the same," Hoth says. "So, by combining multiple images, we can separate the signal we want from the structure that we don't want."

Hoth goes on to say that the researchers are still experimenting with different ways of implementing the signal processing strategy. "It's mostly variations on the idea of modulating the phase and looking at the response," he says.

Although Stanford researchers were the first to demonstrate the technique of using an expanding cloud of laser-cooled atoms, they presented it in a 10-m-tall "atomic fountain" that was designed to be the world's most sensitive accelerometer. "In contrast to their work, our system was designed to be compact to open up the possibility of portable applications," Hoth says. The current experimental system is tabletop sized, but the researchers plan to eventually shrink the apparatus down to a portable cube approximately the size of a mini refrigerator. "There's a lot of very exciting work on atom interferometry being done all over the world," Hoth says.

Author

John Edwards (jedwards@johnedwardsmedia.com) is a technology writer based in the Phoenix, Arizona, area.

SP

FROM THE EDITOR

(continued from page 3)

myself included, argued for a strategy that interacts with arXiv to put signal processing under a more appropriate topic branch, which may also need to be created. Thanks to the efforts led by SPS Vice President for Membership Dr. Nicholas Sidiropoulos, we learned that the arXiv scientific board has agreed to work toward creating an electrical engineering topic branch under which signal processing, information theory, and control theory will be hosted. This would be a wonderful development as SPS enriches its content ecosystem. Ultimately, our hope is that you, our readers and members, will find the content ecosystem beneficial to your professional development and be a regular contributor to the ecosystem.

SP

Xiao-Ping (Steven) Zhang
and Fang Wang

Signal Processing for Finance, Economics, and Marketing

Concepts, framework, and big data applications

Economic data and financial markets are intriguing to researchers working on data and quantitative models. With rapid growth of and increasing access to data in digital form, finance, economics, and marketing data are poised to become one of the most important and tangible big data applications, owing not only to the relatively clean organization and structure of the data but also to clear application objectives and market demands. However, data-related economic studies often have different viewpoints from signal processing (SP). Also, many fundamental economics and business problems have been well formulated and studied in both theory and practice. The knowledge of foundational finance and economic theories will help SP and data researchers avoid reinventing the wheel and develop meaningful and useful research in these areas.

This tutorial article intends to introduce the mainstream foundational concepts and framework in finance, economics, and marketing research, elaborate on the relationships between the traditional economic research paradigm and SP methodology, and help SP researchers identify relevant research directions. The article aims to present a refreshing SP perspective of finance, economics, and marketing research as well as in-depth examples on SP applications in these fields. We hope to empower SP researchers to broaden their knowledge beyond their current areas of expertise and quickly grasp the right formulations of the research questions and related evaluation criteria in these fields.

Digital Object Identifier 10.1109/MSP.2017.2663138
Date of publication: 26 April 2017

Introduction

For SP researchers, comparing an economic system, either a financial market or a consumer market, with a familiar physical input–output system with an impulse–response relationship is intriguing. However, it is important to understand the different and unique perspectives of an economic system. An economic system is an open-loop system in which humans are active participants. System models to be used in analysis are not known a priori and for certain. In economic studies, researchers always work with assumptions or hypotheses that cannot be verified using physical or natural laws. In addition, it is difficult to perform controlled experiments as is usually done in SP systems.

Increasing amounts of digital market data are readily available, due to the rapid growth of electronic trading along with the development of broadband Internet. Electronic trading platforms can record every bid and ask as well as every high-frequency transaction. Indeed, finance, economics, and marketing data are poised to become one of the most important and tangible big data applications not only because of the relative clean organization and structure of the data but also because of clear application objectives and market demands.

For example, owing to its familiar appearance, financial time series has become a major interest for SP researchers, as is evident by recent special issues [1]–[3] on related topics.

Many SP researchers are eager to apply SP techniques to stock price prediction or profitable trading strategy by analyzing price time series. However, without economic justifications and rigorous out-of-sample tests, such analysis may easily fall into the category of technical analysis, which is not mainstream in finance academia or among financial professionals.

The knowledge of the major streams of finance theories is very useful for SP researchers to position and formulate their research before applying SP techniques to financial data analysis. Economic behaviors of people are often about individual choices. The foundational theory (or hypothesis) of economic choices is the expected utility theory, which explains the relationship between investment risk and expected return. Almost all asset-pricing models are based on the expected utility theory. To understand the aggregated market behavior, a fundamental and insightful theory in finance and economics research, the efficient market hypothesis (EMH) [4], was developed. The EMH assumes that all information is incorporated into price and that the market reaches equilibrium through rational decisions of market participants. Stock returns are compensation for risks investors take in markets. The fundamental asset-pricing model and portfolio theory are formulated on the basis of the EMH [5]. Overwhelming empirical evidence supports the EMH, suggesting that it holds most of the time.

Although the EMH is a foundational theory in academic finance research, most stock analyses used by financial practitioners, such as technical analysis [6], which is popular among individual investors, and fundamental analysis, which is popular among professional investors, contradict the EMH [4] because the current price does not reflect all available market information. Instead, these analyses share some common ground with behavioral economics theories [7]–[9]. Although powerful, behavioral economics theories rely on empirical psychological and qualitative arguments to explain economic events and have not led to fully developed asset-pricing models. In contrast with the EMH-based models, it is often difficult to develop rigorous mathematical formulations for psychological effects, making SP methodology difficult to apply to behavioral economics studies.

The knowledge of extant data techniques, results, and issues in finance research will greatly help SP researchers to exert useful research effort by identifying correct problem formulations more effectively. For example, SP researchers are often interested in market correlation structures, an area where modern asset-pricing and portfolio theories [5] have well formulated the correlation structure within stock markets under the normality assumption. In contrast with SP research in which controlled experiments or simulations are conducted and ground truth is often the basis for testing and verifying the models, model testing and verification in finance pose many traps and difficulties.

Nevertheless, with the tremendous increase in the amount of economic data in digital form, the demand for applying SP techniques to finance, economics, and marketing research is increasing, presenting huge opportunities to SP researchers. Many economics applications (including finance and marketing)

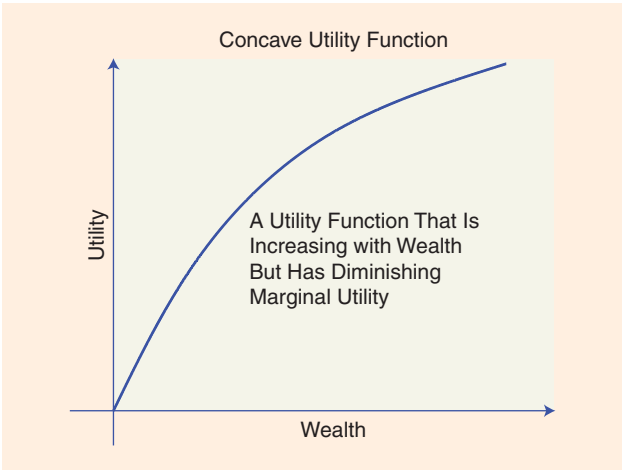


FIGURE 1. The expected utility theory assumes that people make choices according to the expected utility $\mathbb{E}[U(W)]$.

beyond stock price prediction, such as input–output relationships in certain economic systems [10]–[12], are available, and research questions await answers. The recent surge in high-frequency trading (HFT) practices and related theoretical studies [13] provides further opportunities for SP researchers to examine market microstructures and high-frequency system response. Indeed, many SP models and methods share common mathematical grounds with traditional econometric analysis [14] but present different analytical aspects. Therefore they can provide new tools for economic system modeling, analysis, and information extraction for massive finance and economic data.

This tutorial article intends to concisely introduce the mainstream foundational concepts and frameworks in finance, economics, and marketing research to SP researchers, with conceptual elaboration of the relationships between traditional economic research paradigms and SP methodology, and help SP researchers understand economics and business literature and identify relevant research directions with economic significance.

First, we introduce the risks and the foundational economics theory. Then, we formulate the fundamental market equilibrium asset-pricing model in finance within the parameters of modern portfolio theory. Equipped with a basic understanding of the minimum set of economics theories and principles, we introduce an economic viewpoint and basic hypotheses within the context of a free and competitive market—the EMH and competing behavioral economics along with the prospect theory. We elaborate on how different economics theories factor in the individual person—the center of any economic system—his or her choices, decision processes, and emotions. We present a philosophical analysis of the test on economic models with a data–joint hypothesis test. SP perspectives are provided across the article to help readers quickly grasp the differences.

We then move to introduce the fundamental econometrics models and time-series analysis in comparison to the parallel tools known in SP. We focus more on the basic concepts that are rarely present in SP but that are critical in analyzing time-series data in economics and business applications, such as unit roots and causality. We then briefly summarize the relationships between SP and econometric models so that SP researchers can apply their SP knowledge to quickly take advantage of econometrics models. These fundamental economics concepts and methodology that we will introduce have won Nobel Prizes from 1990 to 2013. They are by no means comprehensive, but they encompass a skeleton and basic set of building blocks for data-based economic studies.

We also provide a few detailed state-of-the-art examples in reference to applying SP to economics, finance, and marketing studies. Furthermore, we focus on a few illustrative formulations of economic and business problems. In addition, we give an empirical data analysis example to demonstrate the insight of economic systems that SP is poised to make significant contributions. Throughout this article, we use sidebars to present mathematical formulations and examples to further clarify and illustrate main concepts and ideas.

Risk, risk premium, portfolio optimization, and capital asset pricing

In this section, we introduce fundamental concepts that serve as language and building blocks for economics and finance theory. We begin by introducing the expected utility theory and risk premium (RP).

Expected utility theory and RP

The expected utility theory (or hypothesis) is a cornerstone for economics, game theory, and decision theory and pertains to people's preferences and choice. In economics, a utility function $U(w)$ is defined as a concave function of overall wealth w. An example is shown in Figure 1. Such a utility function assumes that 1) utility is increasing with wealth, i.e., $U(w)$ is monotonically increasing, and 2) wealth has diminishing marginal utility, represented by the concavity of $U(w)$. Therefore, $dU(w)/dw > 0$, and $d^2U(w)/dw^2 < 0$. The expected utility theory further assumes that people make rational choices according to the overall expected utility $\mathbb{E}[U(W)]$. Here, W represents a random variable.

Now, assume that a risk asset (e.g., a stock or an investment) has two possible outcomes of wealth with equal probability in the future, w_0 and w_1 and $w_0 < w_1$. The expected wealth is

$$\mathbb{E}[W] = \frac{1}{2}(w_0 + w_1). \tag{1}$$

The expected utility of this risk asset is

$$\mathbb{E}[U(W)] = \frac{1}{2}(\mathbb{E}[U(w_0)] + \mathbb{E}[U(w_1)]). \tag{2}$$

Given the concavity of the utility function, we have $\mathbb{E}[U(W)] < U(\mathbb{E}[W])$. That is, a rational person is risk averse and thus would prefer a riskless asset, such as cash, with guaranteed wealth of $\mathbb{E}[W]$, to the risk asset with expected wealth of the same $\mathbb{E}[W]$. The certainty equivalent (CE) represents guaranteed wealth whose utility is equivalent to that of the risk asset; i.e.,

$$CE = U^{-1}(\mathbb{E}[U(W)]). \tag{3}$$

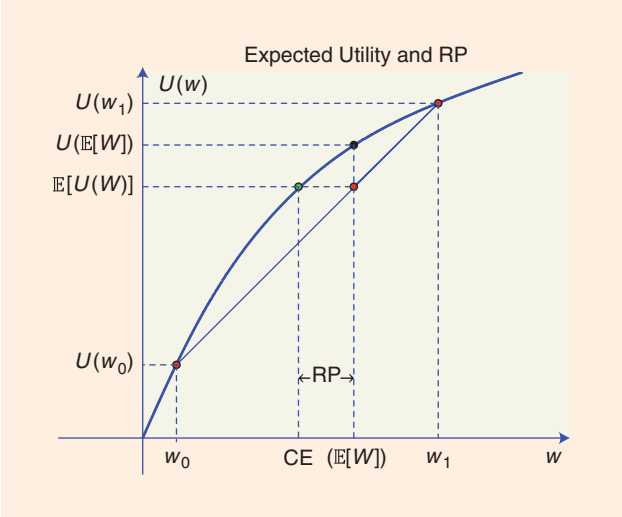


FIGURE 2. The relationship among expected wealth $\mathbb{E}[W]$, expected utility $\mathbb{E}[U(W)]$, CE, and RP.

The RP is therefore the amount of money that a rational person would demand for taking on a risk asset, or the amount of money that a rational person is willing to pay to eliminate risk. Thus,

$$RP = \mathbb{E}[W] - CE. \tag{4}$$

According to the expected utility theory, the source of the excess return of a risk asset, such as stock, is the RP. The RP is the reason people pay insurance premiums to insurance companies to remove risk, and it is also reflected by the interest rate of a loan. Figure 2 shows the relationship among expected wealth $\mathbb{E}[W]$, expected utility $\mathbb{E}[U(W)]$, CE, and RP. See "Take a Bet: Source of the Excess Return" for a gambling example for the risk-averse behavior and the RP.

Take a Bet: Source of the Excess Return

Assume you win the lottery and have the following two choices:

1) the bet: flip a coin (equal probability on both sides); heads, you get US$2,000, tails, you get US$0
2) cash US$1,000.

Would you take the bet or the cash? What if the cash amount were changed to US$900? US$800? Or more? The cash amount for which you are indifferent between choices 1) and 2) is the CE for the risky bet 1). The RP is then 1,000–CE, which is the extra money you need to be compensated for the risk you are taking in 1) or the price you are willing to pay to eliminate the risk in 1) in exchange for the certainty in 2). Are you risk averse? Do you believe the saying that a bird in the hand is worth two in the bush?

Asset correlations and portfolio optimization

For a security/asset i, the net return $R_{i,t}$ at time t is

$$R_{i,t} = \frac{p_{i,t} - p_{i,t-1} + d_{i,t}}{p_{i,t-1}}, \tag{5}$$

where $p_{i,t}$ is the price at time t and $d_{i,t}$ is the dividend during the period $t-1$ to t. [The logarithm of the total return $\log(1+R_{i,t})$ is often used because of asymmetry of the net return. It is easy to show that the log return and the net return are essentially the same when the net return is small. The log return is more commonly used in empirical research.]

Assume that $R_{i,t}$ is stationary and its return is represented by a random variable R_i. The expected return is $\mathbb{E}[R_i] = \mathbb{E}[R_{i,t}]$. If the utility function can be approximated by quadratic form [15], [16], the expected utility maximization becomes a mean-variance investment criterion: maximize the expected return for a given variance, or minimize the variance for a given expected return. The risk of an asset is represented by the variance $\sigma_i^2 = \mathrm{Var}(R_i)$ or the square root of variance σ_i, also called *volatility* in finance.

Using SP concepts, we can rephrase mean-variance criterion: because people are risk averse, they like the signal (expected return) but not the noise (risk). If we have a pool of securities, we can take advantage of the correlations among individual securities and improve the signal-to-noise ratio (SNR) by investing in a portfolio (i.e., a basket of n securities). The portfolio return is $R_p = \sum_{i=1}^{n} x_{ip} R_i$, where x_{ip} is the weight of the ith security, $\sum_i x_{ip} = 1$. The portfolio expected return is

$$\mathbb{E}[R_p] = \sum_{i=1}^{n} x_{ip} \mathbb{E}[R_i],$$

and the portfolio variance is

$$\sigma_{R_p}^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ip} x_{jp} \sigma_{ij} = \boldsymbol{x}_p^T \boldsymbol{\Sigma} \boldsymbol{x}_p,$$

where $\sigma_{ij} = \mathrm{Cov}(R_i, R_j)$ is the covariance of R_i and R_j and Σ is the covariance matrix. We can rewrite

$$\sigma_{R_p}^2 = \sum_{i=1}^{n} x_{ip} \left(\sum_{j=1}^{n} x_{jp} \sigma_{ij} \right) = \sum_{i=1}^{n} x_{ip} \mathrm{Cov}(R_i, R_p).$$

We see that the contribution of security i to the risk or variance of the return on portfolio p is $x_{ip}\mathrm{Cov}(R_i, R_p)$, i.e., the risk of security i in portfolio p or the weighted average of covariances. From the SP perspective, this risk is a projection of R_i on R_p. We can therefore formulate a portfolio optimization problem (see "Mean-Variance Portfolio") to find the best portfolio (weights) [17].

To find the optimal mean-variance portfolios (MVPs), we use the Lagrangian expression:

$$J = \sigma_{R_p}^2 + 2\lambda_e \left(r_e - \sum_{i=1}^{n} x_{ip} \mathbb{E}[R_i] \right) + 2\phi_e \left(1 - \sum_i x_{ip} \right), \tag{6}$$

where $2\lambda_e$ and $2\phi_e$ are the Lagrange multipliers. We then take derivative and set it to zero:

$$\frac{\partial \sigma_{R_e}^2}{\partial x_{ie}} - 2\lambda_e \mathbb{E}[R_i] - 2\phi_e = 0, \forall i = 1, \ldots, n,$$

Mean-Variance Portfolio

Mean-variance portfolios (MVPs) are the solution of the following mean-variance optimization (convex quadratic programming) criterion:

$$x_{ip}^{*} = \arg\min_{x_{ip}} \{J(x_{ip}) = \sigma_{R_p}^2 = \boldsymbol{x}_p^T \boldsymbol{\Sigma} \boldsymbol{x}_p\},$$

$$\text{s.t.} \quad \mathbb{E}[R_p] = \sum_{i=1}^{n} x_{ip} \mathbb{E}[R_i] = x_p^T \mathbb{E}[R] = r_e,$$

where $\sum_i x_{ip} = 1$. Here, r_e is a given target level of expected return. Such portfolios are called *mean-variance efficient portfolios*. All efficient portfolios with different r_e's constitute an efficient frontier.

where the subscript e represents the optimal (efficient) portfolio. Therefore,

$$\sum_{j=1}^{n} x_{je}\sigma_{ij} - \lambda_e \mathbb{E}[R_i] - \phi_e = 0, \forall i = 1 \cdots n. \tag{7}$$

In matrix form, $\boldsymbol{\Sigma}\boldsymbol{x}_e - \lambda_e \mathbb{E}[\boldsymbol{R}] - \phi_e = 0$. Along with $\boldsymbol{x}_e^T \mathbb{E}[\boldsymbol{R}] = r_e$, and $\mathbf{1}^T \boldsymbol{x}_e = 1$, we have

$$\begin{bmatrix} \boldsymbol{\Sigma} & \mathbb{E}[\boldsymbol{R}] & \mathbf{1} \\ \mathbb{E}[\boldsymbol{R}^T] & 0 & 0 \\ \mathbf{1}^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_e \\ -\lambda_e \\ -\phi_e \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ r_e \\ 1 \end{bmatrix}.$$

The optimal MVP can then be obtained by solving the linear equation.

Capital asset pricing model

Assume that market participants, or at least some of them, are rational in the sense of mean-variance optimal (by maximizing the expected utility). Then all securities in the market should be priced such that the market is in equilibrium. Note that we are indifferent to holding different assets given risk-compensated premia (in "Take a Bet: Source of the Excess Return," it means we are indifferent to taking the cash or the bet). Therefore, the market portfolio (e.g., the basket of all stocks in the U.S. stock market) must be an efficient portfolio e.

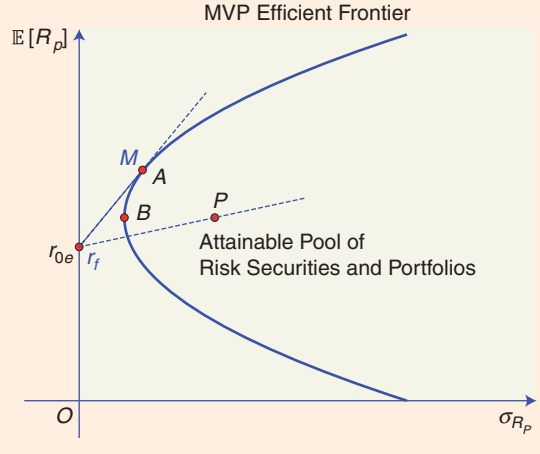


FIGURE 3. The efficient frontier of MVP.

Note that according to the envelope theorem [22], the Lagrange multiplier $2\lambda_e$ in (6) is the rate of change (the slope) of the optimal portfolio variance $\sigma_{R_e}^2$ as a function of desired expected portfolio return $\mathbb{E}[R_e]$; i.e.,

$$2\lambda_e = \frac{d\sigma_{R_e}^2}{d\mathbb{E}[R_e]},$$

i.e., the efficient frontier is parabolic, as shown in Figure 3. Figure 3 also shows that the efficient frontier is the boundary of all attainable portfolios and securities (within the efficient frontier). The portfolio B is the minimum-variance portfolio.

For an efficient portfolio A, the slope of the tangent line in Figure 3 is

$$S_e = \frac{d\mathbb{E}[R_e]}{d\sigma_{R_e}} = \frac{\lambda_e}{\sigma_{R_e}}.$$

We know that an efficient portfolio satisfies (7), implying that

$$\sum_{j=1}^{n} x_{je}\sigma_{kj} - \lambda_e \mathbb{E}[R_k] = \sum_{j=1}^{n} x_{je}\sigma_{ij} - \lambda_e \mathbb{E}[R_i].$$

Multiplying both sides by x_{ke} and then summing over k, we get

$$\sigma_{R_e}^2 - \lambda_e r_e = \sum_{j=1}^{n} x_{je}\sigma_{ij} - \lambda_e \mathbb{E}[R_i].$$

Note that the expected return of the efficient portfolio is $\mathbb{E}[R_e] = r_e$ by definition. Rearranging, we get

$$\mathbb{E}[R_i] - \mathbb{E}[R_e] = \frac{1}{\lambda_e}\left(\sum_{j=1}^{n} x_{je}\sigma_{ij} - \sigma_{R_e}^2\right).$$

Therefore,

$$\begin{aligned} \mathbb{E}[R_i] &= (\mathbb{E}[R_e] - S_e\sigma_{R_e}) + \frac{S_e}{\sigma_{R_e}}\text{Cov}(R_i, R_e) \\ &= r_{0e} + \frac{\mathbb{E}[R_e] - r_{0e}}{\sigma_{R_e}}\text{Cov}(R_i, R_e) \\ &= r_{0e} + \beta_{0e}(\mathbb{E}[R_e] - r_{0e}), \end{aligned} \tag{8}$$

where

$$\beta_{0e} = \frac{\text{Cov}(R_i, R_e)}{\sigma_{R_e}^2}, \tag{9}$$

and the quantity

$$r_{0e} \triangleq \mathbb{E}[R_e] - S_e\sigma_{R_e} \tag{10}$$

is the intercept with zero variance and represents a risk-free asset. The relationship (8) must hold in the market equilibrium between the return of a single security and the efficient portfolio. Therefore, there is no expected return reward for the part of security risk that is uncorrelated (zero covariance) with the efficient frontier.

CAPM

The price of a security i (or a portfolio) in a free market can be determined by the CAPM, also called the *Sharpe–Lintner model*, as follows:

$$\mathbb{E}[R_i] = r_f + \beta_{iM}(\mathbb{E}[R_M] - r_f), \forall i,$$

where

$$\beta_{iM} = \frac{\text{Cov}(R_i, R_M)}{\sigma^2(R_M)}.$$

Time-series regressions are used in practice to compute β. The following time-series regression is perhaps the most famous linear regression in the world:

$$R_{i,t} = r_{f,t} + \beta_{iM}(R_{M,t} - r_{f,t}) + \varepsilon_{i,t}.$$

Insights

- The CAPM is a market equilibrium that must hold.
- There is no expected return reward for idiosyncratic risk, the part of security risk that is uncorrelated with the market portfolio M.

The CAPM is the foundation of asset pricing. It is commonly used in corporate finance, such as merger-and-acquisition practice to evaluate a company and the cost of capital.

Terms Related to the CAPM and Asset Pricing

- *Risk asset*: assets that have uncertain returns, e.g., stocks, bonds, real estate, gold, crude oil. Their prices fluctuate over time. Almost all assets are in this category.
- *Risk-free asset*: assets that have fixed (certain) returns. Treasuries (especially T-bills) are considered to be risk free because they are backed by the U.S. government. The return on risk-free assets is very close to the current interest rate.
- *Excess return*: $\mathbb{E}[R_i] - r_f$ is the expected return in excess to the risk-free rate to reward the security risk, i.e., an RP.
- *Market RP*: $\mathbb{E}[R_M] - r_f$ is the excess expected return to reward market risk (approximately 7% annually in the U.S. market).
- *Systematic risk*: market risk that cannot be diversified and is rewarded by the expected return. Represented by β.
- *Idiosyncratic risk*: firm-specific risk that can be diversified, thus not rewarded by the expected return. Represented by $\varepsilon_{i,t}$. Note $\mathbb{E}[\varepsilon_{i,t}] = 0$. It looks like the noise term in SP, but it is not noise in economics.

Evaluating an Active Portfolio Manager—Is There Alpha?

The time-series CAPM regression for an investment portfolio is

$$R_{p,t} - r_{f,t} = \alpha + \beta_{pM}(R_{M,t} - r_{f,t}) + \varepsilon_{p,t},$$

where α is the intercept of the regression. It should be statistically insignificant according to the CAPM, i.e., $\alpha = 0$.

The skill of an investment manager should not be evaluated by the expected return or the excess return over the market portfolio but rather by α.

If the market risk-free rate is r_f, a market participant will hold a combination of the risk-free asset r_f and a risk portfolio or security within or on the efficient frontier. Apparently, the best risk portfolio to hold is on point A because the tangent line provides the best portfolios in terms of the mean-variance criterion. Indeed, the tangent line provides the best SNR from the SP perspective. [The Sharpe ratio of a portfolio p can be considered the SNR defined in finance: $(\mathbb{E}[R_p] - r_f)/(\sigma_{R_p})$.]

The efficient portfolio A is called the *tangency portfolio*. In market equilibrium, the market portfolio M (i.e., the portfolio consisting of all risk securities) must be the tangency portfolio A. A market index, such as the S&P 500 index, is often used as a proxy of the market portfolio M. All securities must have prices that satisfy (8). Such a price relationship (or correlations) among all risk securities in the market is described in the capital asset-pricing model (CAPM) [18]–[21] (see "CAPM" for details).

As we show, the CAPM is consistent with the expected utility theory and portfolio optimization. According to the CAPM, the stock return is an RP and is only related to its correlation with the market portfolio (i.e., the market risk a stock carries). The higher the β, the larger is the expected stock return. The diversifiable (idiosyncratic) risk a stock carries (i.e., the company-specific risk) does not have RP rewards.

Operationally, a higher expected return can always be achieved by leveraging the risk (i.e., by borrowing money to buy market portfolio). In Figure 3, this means moving along the tangency line to the right, thus generating higher risk and higher return with a fixed Sharpe ratio (or SNR).

The CAPM provides a way to evaluate an investment portfolio and attribute its performance (see "Evaluating an Active Portfolio Manager—Is There Alpha?"). "An Example CAPM Regression" shows a typical empirical regression study.

Fama–French three-factor model and multifactor asset pricing

It is difficult to have a significant alpha, i.e., to beat the market. People have long strived to find excess returns over the market premium. Many of the identified excess returns turned out to be statistically chance results, though a few proved to be persistent. The most prominent model identifying excess returns beyond the CAPM is the Fama–French (FF) three-factor

An Example CAPM Regression

Table S1 shows time-series regressions of excess stock portfolio returns (in percent), $R_{pt} - r_{f,t}$, on the excess stock-market return, $R_{M,t} - r_{f,t}$, from July 1963 to December 1991 (342 months) for 25 portfolios using Center for Research in Security Prices monthly return data of all U.S. market stocks. The 25 portfolios are constructed by dividing firms using market caps (size) and book-to-market (B/M) ratios into quintile buckets. The empirical results are extracted from [23]. The *t*-values and $s(\varepsilon)$ are in parentheses. The residual standard error $s(\varepsilon)$ is the root-mean-square error of the regression. As can be seen, only a few α coefficients are statistically significant, i.e., $|t(\alpha)| > 2$. All of the β coefficients are statistically significant.

Table S1. The time-series regressions from July 1963 through December 1991.

	B/M Ratio				
Size	**Low**	**2**	**3**	**4**	**High**
	α and *t*-value $t(\alpha)$				
Small	–0.22	0.15	0.30	0.42	0.54
	(–0.90)	(0.73)	(1.54)	(2.19)	(2.53)
2	–0.18	0.17	0.36	0.39	0.53
	(–1.00)	(1.05)	(2.35)	(2.79)	(3.01)
3	–0.16	0.15	0.23	0.39	0.50
	(–1.12)	(1.25)	(1.82)	(3.20)	(3.19)
4	–0.05	–0.14	0.12	0.35	0.57
	(–0.50)	(–1.50)	(1.20)	(2.91)	(3.71)
Big	–0.04	–0.07	–0.07	0.20	0.21
	(–0.49)	(–0.95)	(–0.70)	(1.89)	(1.41)

Table S1. The time-series regressions from July 1963 through December 1991.

	B/M Ratio				
Size	**Low**	**2**	**3**	**4**	**High**
	β and *t*-value $t(\beta)$				
Small	1.40	1.26	1.11	1.06	1.08
	(26.33)	(28.12)	(27.01)	(25.03)	(23.01)
2	1.42	1.15	1.12	1.02	1.13
	(26.33)	(28.12)	(27.01)	(25.03)	(23.01)
3	1.36	1.15	1.04	0.96	1.08
	(26.33)	(28.12)	(27.01)	(25.03)	(23.01)
4	1.24	1.14	1.03	0.95	1.10
	(26.33)	(28.12)	(27.01)	(25.03)	(23.01)
Big	1.03	0.99	0.89	0.84	0.89
	(26.33)	(28.12)	(27.01)	(25.03)	(23.01)
	Adjusted R^2 and Residual Standard Error $s(\varepsilon)$				
Small	0.67	0.70	0.68	0.65	0.61
	(–1.46)	(3.76)	(3.55)	(3.56)	(3.92)
2	0.79	0.79	0.76	0.76	0.71
	(3.34)	(2.96)	(2.85)	(2.59)	(3.25)
3	0.84	0.84	0.80	0.79	0.74
	(2.65)	(2.28)	(2.33)	(3.26)	(2.90)
4	0.89	0.90	0.87	0.80	0.76
	(2.01)	(1.73)	(1.84)	(2.21)	(2.83)
Big	0.89	0.91	0.54	0.79	0.69
	(1.66)	(1.35)	(1.73)	(1.95)	(2.69)

FF Three-Factor Model and Cross-Sectional Returns

The FF three-factor model is an empirical asset-pricing model featuring two cross-sectional return factors, the small-minus-big (SMB) factor and the high-minus-low (HML) factor:

$$\mathbb{E}[R - r_f] = \beta_M \mathbb{E}[R_M - r_f] + \beta_s \mathbb{E}[R_{\text{SMB}}] + \beta_h \mathbb{E}[R_{\text{HML}}].$$

- The SMB factor is a firm size factor, represented by the portfolio return of SMB (a portfolio of long small-market-cap stocks and short big-cap stocks. Firm size is measured by firm market capitalization (cap). Small-size firms have excess returns in addition to the market RP.
- The HML is a market-value factor, calculated by the portfolio return of HML (a portfolio of long high-value stocks and short low-value stocks). Firm value is measured by the B/M ratio or the price-to-earnings ratio. High B/M ratios represent value stocks and low ratios growth stocks. Value stocks have excess returns in addition to the market RP.

Insight

SMB (or small-market cap) and HML (or value) portfolios have significant nonzero alpha with respect to the CAPM, i.e., excess returns in addition to the market RP. Note that these two excess returns are predicted by the cross-sectional factors determined by firm accounting features. Cross-sectional means across various stocks/portfolios (sections).

MMV Portfolios and Intertemporal CAPM

Other than expected returns, do people have other tastes? That is, they may like aspects of a firm other than the expected return. For example, a potential explanation for the value factor is that many people like to hold growth stocks (e.g., unicorn firms) for the glory and therefore lower their return expectation.

With multiple tastes/preferences, the MMV portfolio optimization problem can be formulated as follows:

$$x_{ie} = \arg\min_{x_{ip}} \{J(x_{ip}) = \sigma_{R_p}^2 = \boldsymbol{x}_p^T \boldsymbol{\Sigma} \boldsymbol{x}_p\},$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_{ip} b_{is} = b_{es}, s = 1, \ldots, S, (i.e., \boldsymbol{x}_p^T \boldsymbol{b}_p = \boldsymbol{b}_e),$$

$$\mathbb{E}[R_p] = \sum_{i=1}^{n} x_{ip} \mathbb{E}[R_i] = \boldsymbol{x}_p^T \mathbb{E}[R] = r_e,$$

$$\sum_i x_{ip} = 1,$$

where r_e is the desired level of expected return. Note that s represents the tastes other than expected returns and b_{es} is the desired level of the sth taste.

The market equilibrium for the MMV criterion is a multifactor intertemporal CAPM [28]:

$$\mathbb{E}[R_i] - r_f = \beta_{ie}(\mathbb{E}[R_M] - r_f) + \sum_{s=1}^{S} \beta_{is}(E[R_s] - r_f), \forall i.$$

That is, there are multiple MMV-efficient market portfolios (factors).

Minimum-Variance Portfolio and Capon Beamformer

Consider the problem of finding the minimum-variance portfolio, i.e., point B in Figure 3, formulated as

$$R_i(t) = a_i \mathbb{E}[R_i] + \varepsilon_i(t), i = 1, \ldots, n.$$

Find weights $\boldsymbol{x}_{mv}$ such that

$$\boldsymbol{x}_{mv} = \arg\min_{\boldsymbol{x}_p} \boldsymbol{x}_p^T \boldsymbol{\Sigma}_R \boldsymbol{x}_p,$$

subject to the weight normalization:

$$\boldsymbol{x}_p^T \boldsymbol{a} = 1.$$

In array SP, this is exactly the minimum-variance distortionless response beamformer, also known as the *Capon beamformer*, to recover the signal sources $E[R_i]$ with minimum noise variance [29]:

$$\boldsymbol{x}_{mv} = \frac{\boldsymbol{\Sigma}_R \boldsymbol{a}}{\boldsymbol{a}^T \boldsymbol{\Sigma}_R^{-1} \boldsymbol{a}}.$$

Remarks

1) Many portfolio construction/analysis problems have similar forms to array SP problems.
2) The covariance matrix $\boldsymbol{\Sigma}_R$ can be a very large matrix in portfolio analysis because n is large. It is often difficult to estimate such a large matrix reliably. In addition, because of high correlations among assets, this matrix often has a very high conditional number (close to singular), and therefore it is difficult to compute its inverse.

model [24], [23] (see "FF Three-Factor Model and Cross-Sectional Returns"), which suggests that the size and firm value factors can generate excess returns.

Based on the FF three-factor model, prior empirical studies suggest that additional factors persistently produce excess returns. The most notable include the momentum factor, as in the four-factor model by Carhart [25], and the two additional factors, profitability and investment patterns, that Fama and French [26] added in their more recent five-factor model. Theoretically, multiple-factor models can be explained by multifactor minimum-variance (MMV) portfolios [27]. See "MMV Portfolios and Intertemporal CAPM." New research evaluating investment alpha needs to incorporate all known factors in the present multiple-factor models. Historical empirical analysis has shown that, when the four-factor model is considered, no skilled or informed mutual fund portfolio managers can generate persistent investment alpha [25].

The CAPM-related study is a mature area with established statistical methodology. Current studies in this area are mainly empirical research identifying new cross-sectional factors, e.g., [26]. Beta evaluation in traditional study is based on a windowed regression with typically three- to five-year windows, assuming stationarity within the window. The opportunity for SP might lie in the time-varying nonstationary models. The difficulty of any time-varying model is the evaluation criteria given that there are no ground-truth beta parameters.

Portfolio analysis and array SP

Consider the return series of an asset as a noisy signal. A portfolio can be considered a way to use diversity to improve SNR. This is quite similar to many SP formulations. See "Minimum-Variance Portfolio and Capon Beamformer" for the relationships between them in array SP.

Even if we cannot generate investment alpha in portfolio construction, there are other aspects of a portfolio with which we are concerned. For example, different people often have different preferences for systematic market risks represented by various risk factors. These systematic market risks cannot be eliminated by diversification, but investors can choose/analyze the amount of portfolio exposure to those market risks

Example SP Research on Portfolio Optimization and Related Risk Modeling

In [30], a subspace formulation of MVP optimization with risk-factor constraints and related toy examples is given. For example, a market-neutral portfolio requires $\boldsymbol{F}^T\boldsymbol{x}_p = 0$, where $\boldsymbol{F}$ is the loading weights of a single market factor, usually estimated by a time-series regression of all security return time series with a market index, and $\boldsymbol{x}_p$ represents the desired portfolio weights. Torun et al.'s article in *IEEE Signal Processing Magazine* [31] shows an eigenfiltering method to estimate the covariance matrix with a large number of securities. When the security pool is large, the raw covariance matrix is almost singular due to high correlations among securities. Note that the inverse of the covariance matrix, i.e., the precision matrix, is often necessary in MVP optimization. Therefore, a noise regularization term is added on the eigenfiltered covariance matrix to improve the robustness of covariance matrix estimation and its conditional number. The same authors [32] further use an autoregressive model to improve the stability and computational efficiency of estimation of the empirical covariance matrix of highly correlated securities. We also direct readers to an example in [33] that uses Bayesian and regularization methods for the estimation of the unknown observation covariance matrix and the related MVP optimization algorithm. Also, the authors in [34] propose to use smooth and monotone regularization to tackle the high correlation problem in covariance matrix estimation. The work in [35] summarizes the relative performance of different estimation strategies for minimum-variance portfolio optimization problems using the inversion of the estimated covariance matrix or the direct estimation of the precision matrix.

Remarks

We caution SP researchers that the risk and optimal portfolio obtained from the covariance matrix are only from a period of historical data, assuming that there is a stationary joint return distribution among the assets. It is, at most, a rough approximation of the ever-changing financial market. The out-of-sample backtesting is necessary and helpful, but it is by no means comprehensive and bullet proof in terms of representing the future distributions. Therefore, the economic reasoning and understanding of risk factors are always necessary in both practice and economics theoretical development.

(by changing the weight of the risk-free asset). The idiosyncratic risk represented by the regression residual (or noise) is the risk not rewarded (priced) by the market and therefore is the part of the risk that investors need to try to remove by diversification. A portfolio analysis and risk-factor attribution problem usually has the following formulation.

For a set of n portfolios, the time-series model of the portfolio excess return (with risk-free rate subtracted) is

$$R_i(t) = \sum_{s=1}^{K} \beta_{si} R_s(t) + \varepsilon_i(t), i = 1, \ldots, n,$$

where $R_s, s = 1, \ldots, K$ are risk factors. Note that risk factors are also portfolios.

The portfolio analysis or construction problems can be 1) finding common risk portfolios R_s given R_i; 2) given risk factors, which are often represented by different market or section indices, or cross-sectional portfolios, such as those in the FF three- or five-factor models, finding the risk loadings (betas), which reflect the risk exposures of a portfolio; or 3) covariance estimation, portfolio optimization or analysis, often subject to various constraints, such as long/short, borrowing, liquidity, and transaction cost. This is an area to which the SP research can directly contribute due to the similarity in problem formulation in statistical and array SP.

Specifically, practical portfolio construction and optimization usually has following steps.

- Identify portfolio constraints. First, identify the desired expected return or the tolerable risk represented by variance. Note that we can either minimize the variance given the expected return or maximize the expected return given the variance. Second, identify the investable security pool along with investment constraints, including short sell constraints, liquidity constraints, transaction costs, and risk-factor exposure constraints. In practice, these constraints are often determined by the requirements from the investment fund stakeholders, trading systems, and regulatory bodies.
- Identify historical data or proxy data that can be used to, first, estimate covariance and expected returns of a security pool and, second, conduct backtesting.
- Find a robust algorithm to estimate the covariance matrix (or precision matrix) and expected returns.
- Formulate a constrained optimization problem, and find a robust solution.
- Conduct backtesting, and evaluate the portfolio performance against benchmark models, usually known factor models.

In "Example SP Research on Portfolio Optimization and Related Risk Modeling," we show SP research in portfolio optimization. The nature of an economic system is different from that of an SP system. So far, all economic models and theories we introduced are hypotheses. In SP, system and signal models are often rooted in physical laws known to be sufficiently accurate descriptions of system mechanisms. SP researchers tend to be quite confident with their models and believe that the same physical law applies at all times (yesterday, today, and tomorrow).

Therefore, even when the mathematical representation in portfolio analysis, and in economic and business studies in general, is similar to that in SP [36], the treatment can be quite different. For example, in financial empirical studies, when certain risk factors are identified mathematically using the data at hand, it is actually not clear whether the identified risk factors are indeed true risk factors or a result of statistical

chance. Two commonly used methods can verify this: 1) conduct statistical significance analysis with out-of-sample data and 2) have reasonable economic explanations for the identified factor. Even then, the results are still confirmatory rather than conclusive because there are no certain physical laws in the market activities, it is unknown whether the same model will hold over time, and controlled experiments cannot be conducted to test the model. Because researchers always operate on hypotheses in economic and business studies, the burden of proof is on their side, especially when results contradict with conventional wisdom or well-recognized theories, such as the EMH.

Efficient market theory and behavioral economics

An efficient capital market is a market that is efficient in processing information: asset prices fully reflect available information. An efficient market does not necessarily mean that the stock price follows the random-walk model or any existing model.

EMH

Because of the similarity to many noisy signals in SP applications, the stock return series is fascinating to many SP researchers. It is tempting to look into the historical stock charts trying to identify money-making patterns and hope that an autoregressive moving average (ARMA) model can fit the chart. It is often all too easy for SP researchers to think that they may be able to predict stock price using sophisticated SP models to make free money.

However, as the expected utility theory suggests, the source of excess return is the risk that investors carry. In contrast with a physical system, two parties are involved in any financial transaction. In a free market, the two parties agree on a (fair) price for the transaction. In market equilibrium, the price should reflect all available information in the market, such that the transaction is merely an RP changing hands. For example, if both seller and buyer predict that the price of an asset will go up tomorrow, the price should go up today. A rational seller will not sell if the price is lower than what he or she predicts. The efficiency of information processing in the market is indeed a natural consequence of competition in a free market.

In the 1960s, Eugene Fama, a 2013 Nobel Laureate in Economics, was one of the first using computers to study stock prices. Based on his empirical study, he proposed that security prices at any time fully reflect all available information [4]. Such a market is called an *efficient market*. See "Efficient Capital Markets—Different Forms."

Holding the utility function constant, we can argue that the efficient market means that the stock price tomorrow is unpredictable, as all available information has been incorporated into today's price. The market only moves by new information, not by known information. See "Fundamental Analysis, Technical Analysis, and Quantitative Analysis" for different types of stock analysis.

Joint hypothesis testing and its implications in SP

To test the market efficiency, we need to have a market equilibrium model, i.e., an asset-pricing model, such as the CAPM, the FF three-factor model, or the random-walk model. We need to test whether the properties of expected returns implied by the market equilibrium asset-pricing model are observed in actual returns. The main difficulty in testing market efficiency indeed lies in the joint hypothesis test: when the test fails—i.e., we find credible anomaly in price behavior—we do not know whether the market is inefficient, a conclusion people often jump to, or the asset-pricing model used is wrong. Note that while the random-walk model is a simple (hypothetical) form for an efficient market (an idea popularized by a bestselling book [37]), an efficient market does not necessarily mean that the stock price follows the random-walk model or any existing model, as many people have assumed.

While all empirical scientific research is more or less subject to the joint hypothesis test, in SP, researchers do not pay much attention to 1) statistical significance and confidence of model estimation, 2) error distribution of the estimation

Efficient Capital Markets—Different Forms

An efficient capital market is a market that is efficient in processing information and asset prices fully reflect available information. Markets have different forms of efficiency.

- Weak form: current prices fully reflect all information in past prices. Technical analysis using past price patterns will not produce profits.
- Semistrong form: current prices fully reflect past prices and all publicly available information. Fundamental analysis (e.g., studying financial statements) will not produce profits.
- Strong form: current prices fully reflect all information, public and private. Insider trading will not produce profits.

Evidence

1) For the weak form, the random-walk stock price model has been empirically validated for short-term stock returns in many tests, i.e., the short-term stock return series is not statistically different from white noise and therefore is unpredictable.
2) For the semistrong form, a large number of event studies have compared the stock returns before and after corporate news events, such as stock split and earning announcements, showing that the information is incorporated into the stock price almost immediately.
3) For the strong form, insider trading makes profit, indicating that the strong-form market efficiency does not hold. However, mutual fund managers who have more private information than the general public do not outperform on a consistent basis.

Fundamental Analysis, Technical Analysis, and Quantitative Analysis

Fundamental analysts analyze financial statements, management and competitive advantages, competitors, and markets to find the intrinsic value of a firm. Fundamental analysis is the mainstream methodology for financial professionals.

- Pick a company with mispriced intrinsic/correct value
- Critique: subjective guesswork is required for future growth and risks.

Technical analysts (chartists) analyze price and volume patterns/charts to predict price movement and direction. Technical analysis is not mainstream among financial professionals.

- Buy low, sell high by analyzing price-chart patterns, e.g., head-and-shoulder or double top/bottom reversal patterns.
- Technical indicators: moving average, relative strength index, moving average convergence divergence, etc.
- Critique: the pattern definition is subjective and without rigorous statistical tests. Different chartists have different opinions.

Quantitative analysts (quants) analyze financial markets using complex mathematical and statistical modeling, measurement, and research of historical data.

- All quantitative risk modeling and algorithm trading, and increasingly more hedge funds, are based on quantitative analysis.
- Critique: many quantitative analyses lack economic justifications. History may not represent the future. The validity of the analysis is subject to various model and data mining risks.

beyond the Cramér–Rao bound, and 3) statistical tests against alternative models, because the confidence in the model used is established from the understanding of physical laws rather than purely from data. For example, to track a flying missile, we have basic understanding of its motion model based on Newton's laws and sources of noise. However, we do not have such understanding of the market price movement other than common sense. The statistical test is often the only quantitative tool on which we can rely.

Thus, for stock price prediction, researchers need to ask themselves, Are we confident that we have done enough tests to bet the money in the market? Are we confident that our price model beats the mainstream asset-pricing models and the collective information processing (crowdsourcing) of all other market participants? Note that market equilibrium is reached by competing arbitrage-seeking efforts in the market. Apparently, there is a large burden of proof against the EMH.

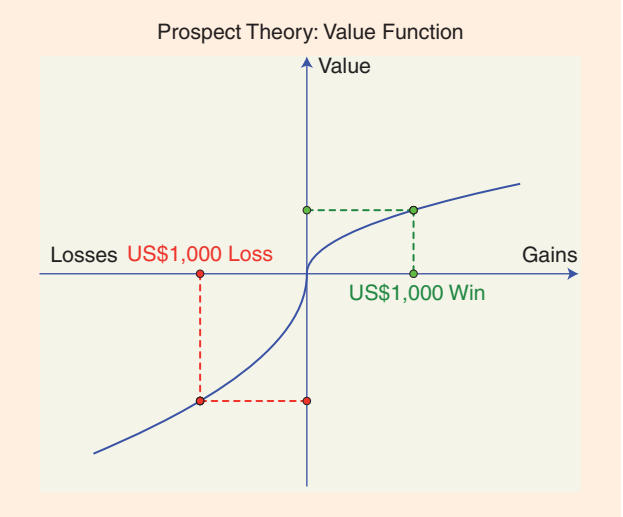


FIGURE 4. The prospect theory assumes that people make choices according to the value function. The amount of pain felt from losing US$1,000 is about double the amount of joy felt from winning the same amount.

Prospect theory and behavioral economics

Both the expected utility theory and the EMH assume that people, or at least some people, act rationally and optimally to exploit risk-free profit in the market to maximize the expected utility. Therefore, the only source of excess return is risk. The expected utility theory and the EMH were dominant before the 1980s in academia.

However, people frequently argue that it is often the irrational or emotional decisions that drive the market, generating bubbles and recessions. Such psychological overreaction or underreaction to information moves the market, generating mispricing and predictable market movement patterns, and thus possible risk-free profitable opportunities. This line of thinking has led to the development of behavioral economics, embraced by many people eager to justify price anomalies and patterns, both academics and (more) practitioners, as it provides a formidable (desirable) alternative to the EMH.

In 1979, the psychologists Daniel Kahneman and Amos Tversky developed prospect theory [7], which has become a foundational theory for behavioral economics. Daniel Kahneman won the 2002 Nobel Memorial Prize in Economic Sciences for his foundational work in behavioral economics (Amos Tversky died in 1996).

In contrast with the expected utility theory, prospect theory states that people make choices based on the value function illustrated in Figure 4. The value function states that people make financial decisions not from the expected utility of the absolute total wealth but from the psychological and emotional perceptions of joy and pain depending on the relative gain or loss. The function is asymmetric in that the amount of pain from losing US$1,000 is about double the amount of joy of winning the same amount. The gain part is concave, while the loss part is convex. A loss-aversion gambling example is shown in "Take a Bet Again—Intolerance of Loss?"

Take a Bet Again—Intolerance of Loss?

Assume you got a traffic ticket and have the following two choices:

1) the bet: flip a coin (equal probability on both sides); heads, you pay a US$2,000 fine, tails, you pay no fine
2) cash: you pay a US$1,000 fine in cash.

Would you take the bet or pay the US$1,000 cash fine? Are you loss averse and risk seeking?

The implication of prospect theory to economics, decision theory, and business management is profound. However, it is also all too easy for researchers to attribute any stock return anomaly found against established asset-pricing models to behavioral factors. Note that for a purely behaviorally caused market anomaly to persist over time, one must assume that even smart investors as a group do not learn over time (or generations) from their psychological mistakes.

It is important to note that most mathematical models in economics are based on the expected utility theory and EMH. For example, the Black–Scholes option pricing model [38] assumes an efficient market, in that stock price follows a random walk and there are no riskless arbitrage opportunities. (Both Robert C. Merton and Myron S. Scholes won the 1997 Nobel Prize in Economics for their contributions to option pricing.) Behavioral factors in the market are still difficult to quantify and have not led to fully developed asset-pricing models that can be tested and potentially rejected.

Long-term stock return predictability—efficient market or behavioral?

Robert Shiller, a prominent behavioral economist who won the 2013 Nobel Prize in Economics the same year as Fama, found that long-term stock return is predictable: a higher stock price (P) to dividend (D) ratio signals a lower expected return, and vice versa [8], [39]. The following time-series regression between the portfolio return and the portfolio firm dividend-to-price (D/P) ratio has significant nonzero coefficient b:

$$R_{t+1} = a + b(D/P)_t + \varepsilon_{t+1}.$$

The efficient market economist John H. Cochrane illustrated this well in an op-ed article in the *Wall Street Journal* at the end of 2008 [67], the height of the financial crisis, shown in Figure 5. In hindsight, the article turned out to be an accurate prediction of the stock market movement.

From the efficient market viewpoint [40], such predictable long-term stock returns are caused by time-varying risk premia. That is, the utility function is time varying. People are more risk averse in recessions because they do not have financial security and demand higher risk premia, i.e., lower stock prices and higher expected returns. By contrast, people are more tolerant of risks in economic booms, and therefore stock prices are high and expected returns (RP) are low. However, behavioral economists like Shiller believe that people overreact to economic situations, resulting in speculative bubbles in booms due to irrational exuberance [41] and recessions due to a lack of psychological confidence. Oftentimes, economists in both camps agree on the findings from data but dispute the cause: whether excess returns are indeed RP or psychological mispricing.

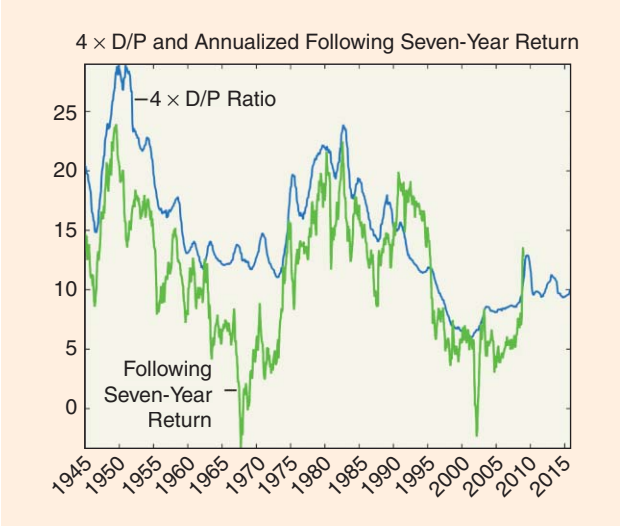


FIGURE 5. The D/P ratio can predict a future seven-year return. (Image courtesy of John H. Cochrane, a former president of the American Finance Association, updated from [40]. An earlier version of this figure also appeared in [67].)

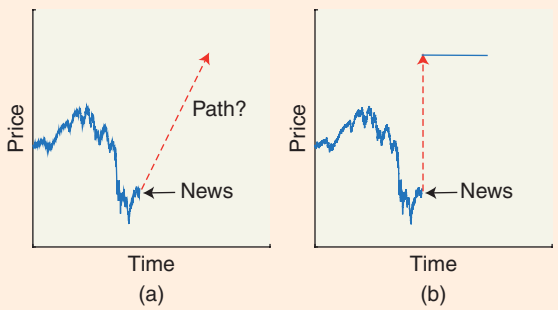


FIGURE 6. The information processing in the market: (a) it takes time for market participants to process complex information and for this information to reflect in the price, whereas in a mostly pure efficient market, information is processed immediately and fully reflected in the price (b). (This figure is modified from a figure in [42]. Image courtesy of John. H. Cochrane.)

Signal and information processing in efficient markets and traps in data

In the most purely efficient market, information is processed immediately and fully reflected in the price, as shown in Figure 6(b). That is, the market moves from one equilibrium to another equilibrium in no time. SP researchers are more familiar

with the information-processing process as shown in Figure 6(a). They know that it often takes time (however short) for market participants to process complex information, whether it comes from human behaviors or trading/communication systems. Such transition processes are not well studied quantitatively in finance literature, posing opportunities for SP researchers. It is possible to use SP models and methodologies to find impulse response to certain events, exploring the microstructure of trading systems, especially in HFT systems.

Without ground-truth models and controlled experiments/studies in modeling economic systems, the validity of SP models relies on data as well as economic explanations. We note several traps in dealing with the model and data as follows.

- Traps in data. First, survivorship bias: only data from successful companies or funds are used in analysis. For example, stock data downloaded from popular websites like Yahoo Finance do not include historical delisted companies. Second, forward-looking bias: future (test) data are used to estimate the model or inform the model construction. Third, selection bias: data are selectively reported in a database, e.g., only mutual funds/hedge funds performing well report results.
- Solution: use unbiased data sets, e.g., Center for Research in Security Prices (CRSP) data set, a gold standard in academic research on U.S. market daily stock returns. See [43] for more details.

When evaluating and interpreting empirical results, researchers need to be careful to select meaningful criteria and statistical test methodology.

- Test against alternative models, e.g., can you beat the FF three-factor or the random-walk model?
- Minimum mean-squared errors are not enough. Note that asset-pricing model residuals are often considered idiosyncratic risks or the unknown risks that are not captured by known factors. The important thing may be whether a model properly attributes different types of risks rather than minimizes the residual. The explanatory power of a model should be given more consideration depending on the applications.
- Conduct extensive out-of-sample tests, and prove with statistical significance that your results are not due to chance.

Again, we caution that out-of-sample backtesting is helpful as much as out-of-sample testing in SP and machine-learning applications to check model fitness and robustness for known data over multiple time periods. However, multiple models may have similar fit into a set of data, and the past data may not perfectly represent the future in an open economic system. There is generally no gold criterion to know the ground-truth model without controlled experiments. Indeed, we do not even know whether there exists a ground-truth model. The joint hypothesis problem always exists in discovering a model or a theory. Therefore, economic justifications are always an integral part of economic and business studies.

Basic econometric models, time-series analysis, HFT, and SP

SP shares many similar terms and mathematics with econometrics, even though there are few interactions between the two communities. Understanding basic concepts in econometrics will help propel SP researchers in their work and allow them to appreciate application issues in economics and business.

ARMA

In time-series econometrics [14], [44], the definitions of models have similar forms to those in SP. An autoregressive (AR) process with the first-order, AR(1) process, is defined as

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t,$$

AR(p) processes are defined as

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t,$$

and ARMA(p, q) processes are defined as

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i},$$

where ε_t is white noise with zero mean and variance σ^2 and $\mathbb{E}[\varepsilon_t \varepsilon_\tau] = 0$ when $t \neq \tau$. Note that ε_t are indeed unpredictable innovations (shocks).

In econometrics, in place of the Z transform used in SP, a lag operator L is used to represent time shift. Thus, the ARMA(p, q) process can be represented by lag operator polynomials as

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) Y_t = c + \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t.$$

When observing an economic system, researchers do not have control of the innovation or the unexpected shock, ε_t, which can be various events, such as a sudden decrease of crude oil prices. Meanwhile, there may be other exogenous variables that are not uncorrelated white noise, such as the advertising investment of a company or interest rates set by central banks. Thus, to analyze an economic system, the ARMA model needs to be generalized to the ARMA processes with the exogenous variables (ARMAX) model. ARMAX(p, q, r) can be defined as

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \sum_{i=1}^{r} \eta_i X_{t-i},$$

where X_t represents exogenous variables. A special case of the ARMAX model is AR distributed lag models (ARDL). An ARDL(p, r) model is defined as

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{r} \eta_i X_{t-i} + \varepsilon_t.$$

A vector AR (VAR) model can be used to examine the interaction of a set of n economic variables. VAR(p) is specified as

$$\boldsymbol{y}_t = \boldsymbol{c} + \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{y}_{t-i} + \varepsilon_t,$$

where $\boldsymbol{y}_t$, $\boldsymbol{c}$, and ε_t are $n \times 1$ vectors and $\boldsymbol{\Phi}_i$ is an $n \times n$ matrix.

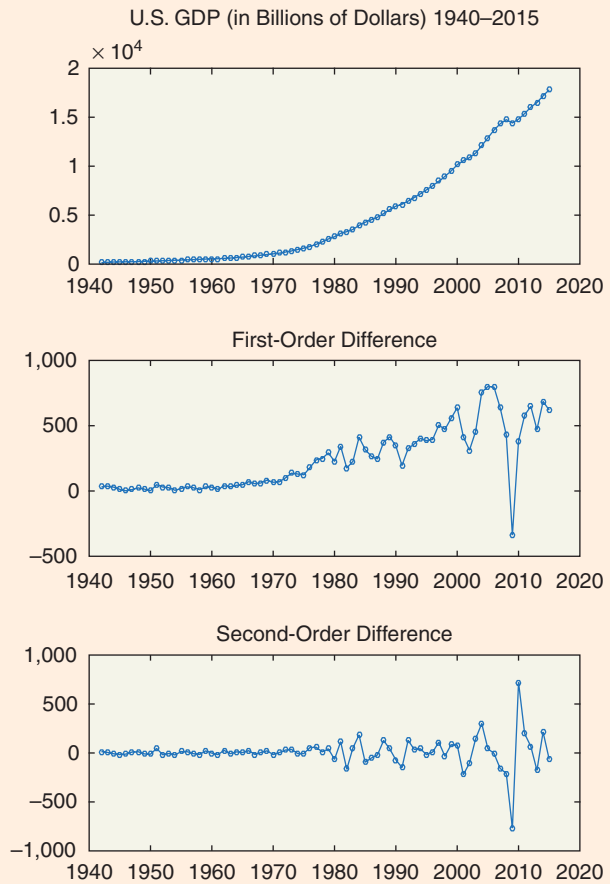


FIGURE 7. The U.S. GDP time series and its differences. This time series is an $I(2)$ process.

Nonstationary, integrated processes, and unit roots

The subtle differences in econometric time-series model definitions from SP reveal the different applications and assumptions in economics. A main difference is that in SP, the distinction between the system/model and the process/signal is clear because researchers can actively control or at least simulate the signal. In economics, the model is often a characteristic of the time series, which oftentimes can only be observed but not controlled or changed. In SP systems, there is no constant term c, as the output in a linear SP system is zero with zero input. In economic time series, the constant term c actually defines the stable status of the variable. For example, for AR(p) time series, we have

$$\mathbb{E}[Y_i] = \frac{c}{1 - \sum_{i=1}^{p} \phi_i}.$$

A more startling difference is that in SP, only stable processes and systems are investigated, i.e., the roots/poles must be within the unit circle. However, economic time series are often nonstable (the econometrics term is *nonstationary* or *evolving*). The nonstationarity (i.e., evolution) is often a desired property for economic variables. For example, a company hopes that its sales grow over time rather than die out over time, as does a country in terms of its gross domestic product (GDP). The analysis of nonstationary processes generates many research problems and tools.

A basic nonstationary evolving process is a random-walk process with drift:

$$Y_t = c + Y_{t-1} + \varepsilon_t.$$

Apparently, this process has one root on the unit circle, i.e., it has a unit root. Its first-order difference is stationary, and thus it is a process of integrated order 1, denoted by $I(1)$. An integrated process with order d, $I(d)$ process, is a nonstationary (unit-root) process whose dth-order difference is stationary. A unit-root process is also called an *evolving process*. See Figure 7 for the U.S. GDP time series and its differences.

The related AR integrated moving-average (ARIMA), ARIMA(p, d, q), model is defined as

$$\Delta^d Y_t = c + \sum_{i=1}^{p} \phi_i \Delta^d Y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j},$$

where Δ^d is the dth difference operator.

The first analysis step for an economic time series is often to determine whether it has a unit root, i.e., to conduct the unit-root test; see "Unit-Root Test."

Cointegration and causality

As we have shown, in contrast with the signals studied in SP, many economic time series are nonstationary integrated processes

Unit-Root Test

AR(1) unit-root (random-walk) test:

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2), \mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0, \forall t \neq s.$$

The null hypothesis is $H_0: \phi = 1$ (unit root). There is no analytical closed-form expression for the distribution of ϕ estimate. The Dickey–Fuller test [45] uses empirical distribution of the test statistics:

$$DF_t = \frac{\hat{\phi} - 1}{SE(\hat{\phi})},$$

where $\hat{\phi}$ is the OLS estimate.

The augmented Dickey–Fuller test tests unit root for the AR(p) process:

$$Y_t = c + \phi Y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta Y_{t-i} + \varepsilon_t,$$

for the null hypothesis $H_0: \phi = 1$ (unit root). The test statistics are the same as above.

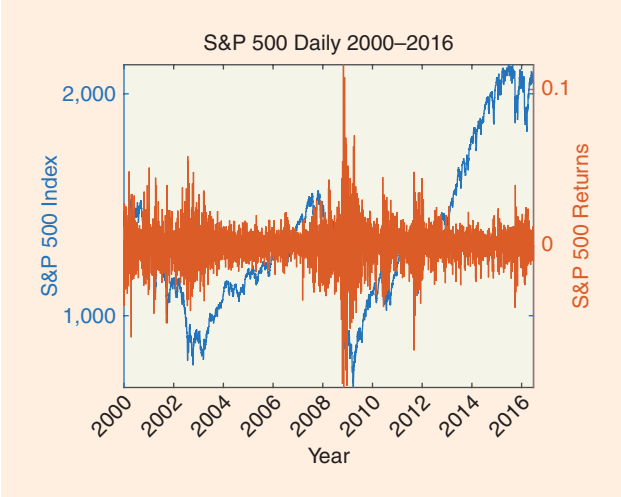


FIGURE 8. The S&P 500 daily prices and returns (2000–2016). When the price is low, the return volatility is high.

with at least one unit root. The integration order is a property of the process and represents a trend of the time series. In SP research, researchers do not care about this trend because they are working on stationary signals. However, a serious problem occurs when examining relationships of multiple nonstationary integrated economic time series.

Consider the following simple linear regression:

$$Y_t = c + \beta X_t + \varepsilon_t.$$

When X_t and Y_t are both unit-root processes with different orders, i.e., when ε_t is also a unit-root process, regression results and statistics become spurious or meaningless. Regression results are only meaningful when X_t and Y_t have a common trend (i.e., the same integration order) or are cointegrated. Clive W.J. Granger's finding of such spurious regressions [46] invalidated many empirical economic studies before the 1970s and, along with his work on cointegration [47], won him the 2003 Nobel Memorial Prize in Economic Sciences. For cointegrated time series, there must exist a linear combination of them that is stationary. The Engle–Granger test [47] applies the Dickey–Fuller unit-root test to examine the cointegration of multiple time series.

The time-series model helps capture correlations among time series but does not find causal relationships. In SP, inputs cause outputs because information flows are obvious in a physical system. However, in analyzing economic data or any data from nonphysical systems, such as a social network, causal relationships are not obvious and cannot be taken for granted. Yet identifying such causal relationships is of great importance to discover information hidden in the data and is necessary for decision making in many big data applications.

The causality relationship is always difficult to define and quantify. Granger gives a definition from the time-series perspective [48], [49]. If we agree that the cause must occur before the effect, the Granger noncausality (or strong exogeneity) can be defined if the following equation holds for the conditional mean:

$$\mathbb{E}[Y_{t+1}|Y_t, X_t, X(t-1), \cdots] = \mathbb{E}[Y_{t+1}|Y_t].$$

If we further agree that for X_t to be the cause of $Y(t+1)$, it must contain unique information about Y_{t+1}, then X_t is said to Granger-cause Y_{t+1} if for some A,

$$P(Y_{t+1} \in A\,|\Omega_t, X_t) = P(Y_{t+1} \in A|\,\Omega_t - X_t),$$

where P is the probability and Ω_t represents all knowledge in the universe available at time t. Note that Granger causality is only one of many definitions on causality, but it is statistically testable using a time-series model (VAR or ARDL model), making it instrumental in causality analysis.

Generalized AR conditional heteroskedasticity models

As we already discussed, volatility is a fundamental risk quantity that needs to be estimated in finance, especially in risk modeling and option pricing. As has been observed, the historical volatility of a time series changes over time. See Figure 8 for S&P 500 daily prices and returns.

The time-varying nature of volatility is called *heteroskedasticity*. Robert F. Engle invented the AR conditional heteroskedasticity (ARCH) model to capture the time-varying dynamics of volatility, winning the 2003 Nobel Memorial Prize in Economic Sciences. The qth order ARCH(q) model for a zero-mean normally distributed asset return time series, Y_t, with time-varying volatility σ_t is specified as

$$Y_t \sim N(0, \sigma_t^2),$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2.$$

It is indeed an *MA* model for time-varying variance.

Adding an AR term for σ_t, the generalized ARCH (GARCH) model [50], GARCH(p, q), is defined as

$$Y_t \sim N(0, \sigma_t^2),$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2 + \sum_{i=1}^{p} \alpha_i \sigma_{t-i}^2.$$

Many variations of the GARCH model have been subsequently developed and widely used in risk models, high-frequency volatility models, large-scale multivariate ARCH models, and derivative pricing models [51].

Relationships between SP and econometric models

The time-series analysis has achieved great success in economics, finance, and business studies. It is encouraging for SP researchers, as the time-series models and basic concepts are essentially similar to SP models. Indeed, the Granger causality concept was partly inspired by Nobert Wiener [48], [52]. Many other models used in econometrics [14], [53], including spectrum analysis, Kalman filtering, Markov models, maximum likelihood and Bayes methods, particle filtering, and

sequential Monte Carlo methods, are similar to those in SP as well.

However, researchers need to note major differences when applying these models to economic and finance applications.

1) As discussed previously, model specification is more important in economics and finance than in SP applications, because model specification is often not so obvious and testable in economics as it is in SP and the joint hypothesis problem always exists.
2) Statistical analysis and tests are always necessary for model verification and applications. Just an estimated number is not enough. Distributions and confidence intervals of parameter estimates need to be analyzed. Equally important is the distribution of the error terms, which often represent money and risks in financial decisions.
3) In economic applications, researchers often operate on nonstationary time series, and the causal relationships are not obvious and may be difficult to test.
4) In empirical studies, data are noisy and the SNR is low because there are often too many factors (interferences) in an economic system. For example, the Sharpe ratio of the U.S. market S&P 500 index is around 0.4, which roughly translates to an SNR of -8 dB.
5) The data traps discussed previously always exist in economic studies. Application-dependent conceptual justifications, alternative model testing, and out-of-sample tests are always necessary because, theoretically, multiple models could fit finite data well.
6) Although stock price forecasting and trading strategies using SP tools are attractive, these are probably the most difficult tasks because the price itself is, after all, a signal for resource allocation and is sensitive to all market actions, trading strategies, and information. The efficient market concept is a logically good approximation of information processing in the financial market. Conversely, many other economic and business time series and processes, such as production, sales, earnings, investments, and so on, may be easier to forecast and model.

Examples of applying SP to market data

In this section, we provide some examples of applying SP to economics, finance, and marketing studies based mainly on our own experiences. As such, they are only illustrative and by no means comprehensive. These examples, in our view, represent a few broader cutting-edge directions to which SP can contribute significantly.

Market evolution analysis using SP models

In financial markets and general economic systems, market participants can only passively observe the system input–output dynamics but are not able to manage input factors, as we can often do in SP systems. For example, a macroeconomic event, such as unemployment or interest rate change, as a market input can move the financial market in a certain way, but market participants cannot actively design and use such inputs to achieve desired market movement (output) unless they are central bankers or policymakers. Even policymakers are restricted in their power of managing input factors and cannot change many inputs, such as consumer confidence. In contrast, market participants in commercial product markets can actively manage many market inputs, such as a firm's marketing budget, product pricing, and research and development (R&D) investment, to achieve desired market outputs, such as sales targets. SP models are therefore poised to be useful in these types of product marketing research, e.g., to analyze and model market dynamics and optimize marketing strategies and investments. We present a market evolution analysis as an example to demonstrate the use of SP methods in marketing research.

In a commercial product market, when using a time-series model for market dynamics, first we would be able to observe the output of the market response [54]. For a company, among the most important observable market output time series is product sales. Evolving sales are always desirable, i.e., sales are a unit-root process, not a stationary process. However, many causes can lead to evolving sales, and managers need to know how to maintain sales performance. Meanwhile, market inputs, such as advertising, price cuts, promotions, R&D, and competitor investments, may affect market outputs, such as sales. Managers can indeed actively manage their marketing budgets and decisions. They search for market opportunities in which positive effects of a one-time investment can persist at least partly (i.e., have a hysteresis effect) and hope that the negative effect of a price war or product defect is short lived and has no hysteresis. Accurately understanding market input–output relationships is critical for managers to make the right investment decisions and take real-time actions to capture fleeting market opportunities.

For simplicity, we use the advertising-sales relationship as an example of the general marketing and business input–output dynamics. To clearly demonstrate basic concepts, first-order models are used whenever possible. Our first objective is to discriminate the different market dynamics, as shown in Figure 9.

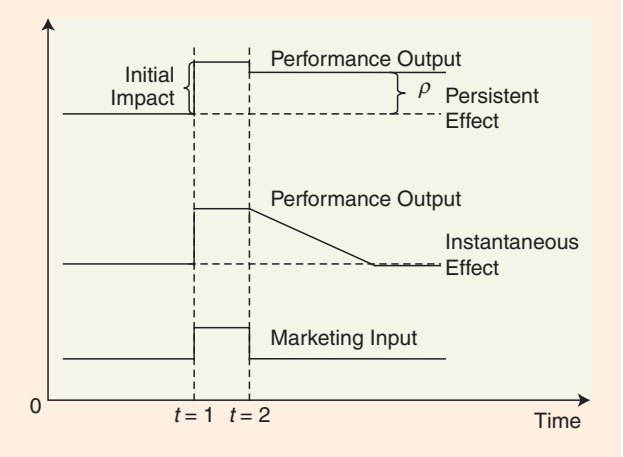


FIGURE 9. The different market input–output dynamics.

To identify market output characteristics, conventionally the following first-order time-series model is used to test whether the sales S_t is evolving (a unit-root test with null hypothesis $H_0: \phi = 1$):

$$S_t = \mu + \phi S_{t-1} + e_t.$$

When the sales S_t is evolving, managers need to know the role of the advertising investment A_t in creating and supporting sales evolution to budget advertising investment. The following market input–output model is established given known A_t:

$$S_t = c + \alpha S_{t-1} + \beta A_t + e_t.$$

Note that the sales S_t and advertising A_t are monetary value time series. To examine whether the sales evolution is intrinsic (i.e., a favorable market feature) or supported by advertising (i.e., a less favorable market feature with which continuous advertising needs to be budgeted), an intrinsic market evolution (IME) test is proposed [10] to test null hypothesis $H_0: \alpha = 1$. If H_0 is rejected, the market is intrinsically stationary; i.e., a short-term advertising investment has a short-term instantaneous effect on sales. Sales evolution needs to be supported by the persistent advertising spending. In SP language, the system function has no poles on or outside the unit circle. The unit root (pole) of the output $S(Z)$ is indeed generated by the unit root of input $A(Z)$. In reality, when advertising of a product increases, sales increase, but when advertising is withdrawn, sales revert to the original level. Marketing managers need to evaluate the investment return to maintain the most profitable sales level.

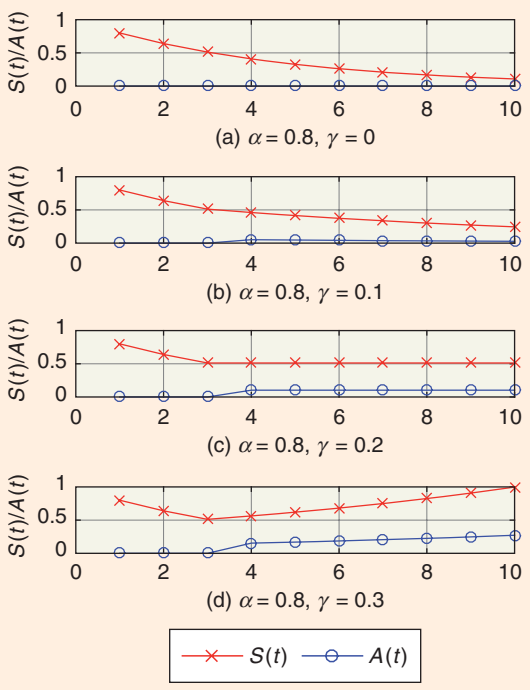


FIGURE 10. The different sales effects of percentage advertising budget practices with different γ_b given $\alpha = 0.8$, i.e., in an intrinsically stationary market. (a) The sales die out without maintaining advertising. (b) The sales still die out at a slower rate (along with the advertising) with inadequate advertising percentage budget. (c) The continuous advertising budget as a percentage of sales induces and supports the sales evolution. (d) The continuous advertising budget as a larger percentage of sales leads to increasing sales (and subsequent increasing advertising).

Conversely, if H_0 is not rejected with the IME test, the market is inherently evolving. A short-term advertising investment will have a long-term persistent effect on sales. In SP language, the system function itself has poles on or outside the unit circle. The unit root (pole) of the output $S(Z)$ is indeed generated by the intrinsic market dynamic. Such a phenomenon is not common in SP systems, and researchers may question why it is important to identify intrinsically evolving markets. In the economic and business world, an intrinsically evolving market indicates a more favorable business environment in which a short-term advertising campaign drives sales up; when the advertising is withdrawn, sales maintain their level or even continue to grow. Such a phenomenon could be caused by the intrinsically superior product characteristics that retain loyal customers and attract new customers by word of mouth, or a growing emerging market yearning for such products. Short-term advertising only acts as ignition. When an intrinsically evolving market is identified, managers can increase their marketing campaign budgets to capture such opportunities in time [12].

In addition to identifying the market nature, SP models build budgeting strategy. For example, a simple percentage budgeting model is

$$A(t) = f(S(t-1)) = \gamma_b S(t-1).$$

That is, the marketing budget is based on sales of the previous period. By this budgeting model, the optimal budgeting sequence can be quantified according to various constraints and costs to support the sales evolution in an intrinsically stationary market. Figure 10 shows an example of various advertising budget effects with different γ_b given $\alpha = 0.8$. More detailed methodology, analysis, and applications are available in [10]–[12].

Real-world marketing dynamics can be more sophisticated. With possible lag effects of market inputs and outputs and multiple variables, multivariate [12] and higher-order models [11] can be built based on these concepts. With more complexity in data and model analysis, SP model-based econometric analysis can certainly play an instrumental role in quantitative marketing and business research.

Time-varying risk models based on Kalman filtering and Gaussian processes

Time-varying systematic risk analysis based on Kalman filtering

In financial markets, the systematic risk is represented by the market risk beta in the CAPM or the multifactor betas in the FF

three-factor model, including market beta, size beta, and value beta. In practice, the ordinary least squares (OLS) time-series regression is used to estimate beta by assuming constant beta over the observation period. Prior research has shown evidence of time-varying beta [55]. The OLS time-series regression does not account for any significant economic event that could affect systematic risks of an asset.

The piecewise mean-reverting (PMR) model is based on the observations of the FF three-factor systematic risk behaviors in empirical tests. The betas tend to jump with relevant significant events and revert to their means with different rates depending on the type of events. The reverting rate indicates how quickly the systematic risk of a stock recovers from a sudden change. For the CAPM model, the PMR model consists of a system equation that represents the PMR dynamic of the hidden time-varying variable beta and an observation equation of the stock return given by the CAPM. The system equation is

$$\beta_t = (1-\phi_t)\bar{\beta} + \phi_t\beta_{t-1} + z_t u_t + \xi_t,$$

where $\bar{\beta}$ represents the average β over time and ϕ_t is the mean-reverting rate in the range of [0,1]. The larger is ϕ_t, the slower is the mean reverting. The jump process is based on the Bernoulli random variable z_t, and a zero-mean normally distributed variable, $u_t \sim N(0, \sigma_u^2)$, represents the amount of jump, i.e., z_t takes the value of 1 with a given probability p and the value of 0 with probability $1-p$; $\xi_t \sim N(0, \sigma_\xi^2)$ represents a random perturbation of β. The observation equation is simply the CAPM, i.e., the asset return time series

$$R_t - r_f = \alpha_t + \beta_t(R_{M,t} - r_f) + \varepsilon_t.$$

This model assumes that significant economic events can lead to the abnormal changes in beta. A modified Kalman filter can be used to estimate and track the PMR beta [56], [57]. In addition, the methodology can be extended to multifactor models, such as the FF three-factor model [57].

As previously discussed, model validation is always an open issue. To achieve validation, researchers can compare the model with alternative models and try to obtain the real-world data to examine whether betas change when major events occur. However, such case studies are confirmatory rather than conclusive.

Gaussian process regression stochastic volatility model

Volatility modeling is one of the most active research areas of financial time series. With the recent development of Bayesian nonparametric modeling in SP and machine-learning communities, flexible tools and modeling methods, such as the Gaussian process (GP) [58]–[60] and copula process [61], can be applied to model financial data volatility.

By combining the GP state-space modeling framework with the stochastic modeling concept [62], a GP regression stochastic volatility (GPRSV) model can be built to solve the problem of modeling and predicting time-varying variance of financial time-series data. In GPRSV models, a GP prior is placed over the state transition function, and the state transition function is a random function sample from the GP. It is therefore not limited to a fixed linear AR form, as in GARCH-type models and pure stochastic volatility (SV) models. For a zero-mean normally distributed asset return time series, Y_t, with time-varying variance σ_t^2, a GPRSV model is represented by the following set of equations:

$$\begin{aligned} a_t &= Y_t - \mu = \sigma_t \varepsilon_t, \\ v_t &= \log(\sigma_t^2) = f(v_{t-1}) + \tau\eta_t, \\ f &\sim \mathcal{GP}(m(x), k(x, x')), \\ \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} &\sim N(0, \Sigma), \\ \Sigma &= \begin{bmatrix} 1 & \rho\tau \\ \rho\tau & \tau^2 \end{bmatrix}, \end{aligned}$$

where μ is the mean of Y_t, a_t is the innovation of the return series, v_t is the logarithm of variance σ_t^2 at time t, and ε_t and η_t are independent and identically distributed zero-mean standard Gaussian distributed white noises. A well-known asymmetric effect is called *financial leverage*, representing a negative correlation between today's return and tomorrow's volatility [63], [64]. This asymmetric leverage effect is captured by the correlation ρ between ϵ_t and η_t. The unknown parameters τ and ρ are to be estimated. The hidden state transition function f is assumed to follow a GP, defined by the mean function $m(x)$ and covariance function $k(x, x')$. The parameters in $m(x)$ and $k(x, x')$ are called *hyperparameters*. For example, if the mean function is defined as $m(x) = cx$, then c is a hyperparameter. The mean function $m(x)$ can encode prior knowledge of system dynamics. Figure 11 shows the graphical model representation of a GPRSV model.

Particle-filter-based Markov chain Monte Carlo learning methods can be used to efficiently estimate the GPRSV model and make volatility inferences. The GPRSV model demonstrates

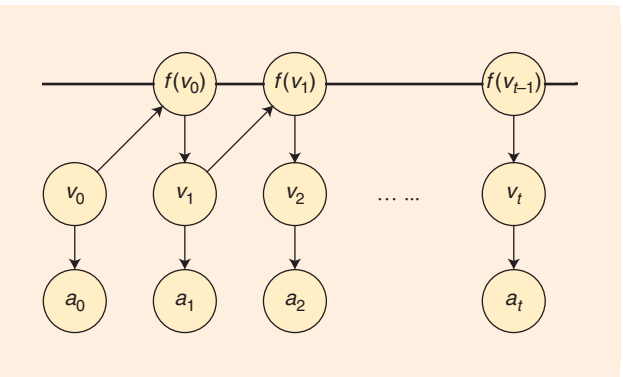


FIGURE 11. A graphical model representation of a GPRSV model of time-varying volatility/ a_t: the observation variable at time t, v_t: the hidden variable (logarithm of volatility), and f_t: the transition function sampled from a GP. The thick horizontal line represents fully connected nodes.

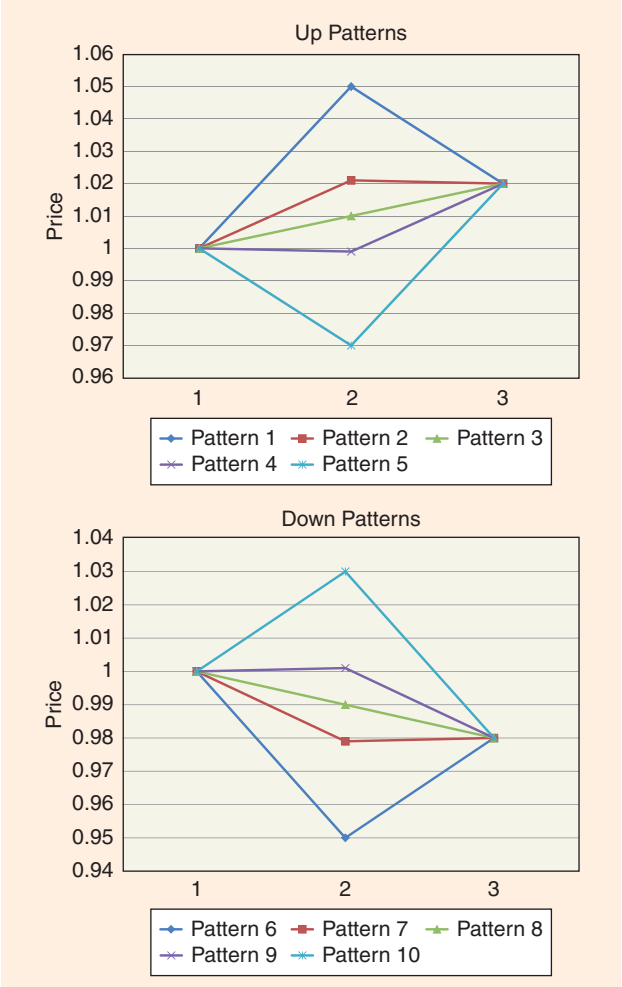


FIGURE 12. The ten preset three-point patterns: five with up direction with different paths and five with down direction with different paths.

Table 1. The statistical analysis of historical stock returns conditional on ten three-point patterns.

				KS Test		t-Test	
Pattern	**Matches**	**Mean**	**Std**	***H***	***p***	***H***	***p***
All	2,855,470	0.000	0.996	0	1	0	1.000
1	189,760	−0.016*	0.969	1	0	1	0.000
2	333,443	0.003	0.937	1	0	0	0.067
3	445,424	0.004*	0.916	1	0	1	0.021
4	362,097	0.010*	0.937	1	0	1	0.000
5	170,181	0.007*	0.938	1	0	1	0.002
6	168,005	−0.001	0.975	1	0	0	0.598
7	337,782	0.011*	1.051	1	0	1	0.000
8	435,628	0.012*	1.106	1	0	1	0.000
9	306,364	−0.032*	1.076	1	0	1	0.000
10	163,672	−0.034*	1.012	1	0	1	0.000

Null hypothesis H_0: the mean return is not significantly different from zero. The null hypothesis is rejected in eight of ten patterns (significant values are marked by an asterisk under $\alpha = 0.05$).

superior volatility prediction performance with both simulated and empirical financial data compared with alternative GARCH and SV models. The previous examples show that SP can contribute to investigating time-varying characteristics of finance and economic systems.

Big data analysis of financial data based on SP

Returning to stock returns, we present some empirical statistical facts to illustrate the potential information that can be sifted from financial big data.

Does the path matter in momentum patterns?

Stock price momentum is a well-known phenomenon in which the stock return continues its direction in the short run. The momentum factor is included in the Carhart four-factor model [25]. From an SP perspective, the momentum is a simple two-point pattern. A natural question is, would multiple point patterns (paths) predict the stock return behavior?

The data set is CRSP monthly price data from 1965 to 2012. The CRSP data set is the gold standard in historical stock return research. To compare stock returns of different stocks, all stock returns are normalized using the methodology in [6]. Ten preset three-point price patterns are constructed. Specifically, ten groups of three-point price patterns in the data set are created according to their correlation with each of the ten preset patterns, as shown in Figure 12. The correlation similarity threshold is 0.95. The subsequent one-month stock returns conditional on each of the ten three-point patterns are analyzed. The null hypothesis H_0 is that the mean conditional return of the subsequent month is zero, i.e., the same as unconditional returns. Both a nonparametric Kolmogorov–Smirnov (KS) test and parametric t-test are conducted. Table 1 summarizes the results.

There are more than 2.8 million three-point patterns. The means of all stock returns are normalized to zero. As can be seen, eight of ten conditional returns have statistically significant means with different directions. The p-values are reliable given the large number of samples. This statistical fact shows that in historical stock return data, 1) the three-point patterns contain information about future returns and 2) the path does matter in addition to the two-point momentum pattern.

That said, the results do not necessarily mean that one can profit from these patterns because potential constraints, such as liquidity and transaction costs, exist to prevent trading profit, and the results are just a summary of historical data. Rather, these results along with the results presented in [6] show that fine structures exist in stock returns and markets (more than traditional economic research indicates) and that SP can provide powerful analytical tools.

Can the past inform about the future?

All theories and models are based on historical data with the assumption that the past can represent the future. In SP, researchers do not often ask the question of whether the past can inform about the future because they are confident about natural physical laws. However, in financial markets, people

have always been skeptical about models and statistical results summarized from historical data, especially so after the 2008 financial crisis. Some people have even gone so far as to say that the past cannot inform about the future, and therefore models and statistics coming from historical data are not useful or trustworthy.

In the following, we show some interesting empirical results as a small step forward to answer this question. The stock daily returns of Russell 3000 component stocks from January 1995 to September 2014 are examined. The Russell 3000 is used because those stocks are relatively liquid and the price data are a good reflection of real market transactions. Note that the component stocks of the Russell 3000 are changing over time. The test period is set to about four years from January 2011 to September 2014. Starting from the first trading day of 2011, the return distribution of the historical data of the past 16 years is used to predict the return distribution of the next day. We want to examine overall how accurate it is to use the historical stock return distribution as a representation of the future distribution. The following procedure is used to examine the differences between historical distributions and future distributions.

All theories and models are based on historical data with the assumption that the past can represent the future.

At time t, we use all historical daily returns of the preceding 16 years to create a distribution f. We then define K equal probability quantile bins Q_k, $k = 1, \ldots, K$ with bin (stock return) boundaries $q_1, \ldots q_{K-1}$. Apparently, $P(R_\tau \in Q_k) = (1/K), \forall \tau \leq t$. If the future distribution is the same as the past distribution, we can expect that at time t, the returns of the 3000 stocks will fall into each bin uniformly, i.e., $P(R_t \in Q_k) = (1/K)$. The chi-square test can be conducted to test this hypothesis. We aggregate all bin counts and plot two scenarios in Figure 13 with 50 bins, i.e., $K = 50$. Figure 13(a) shows the performance of the past unconditional distribution. Figure 13(b) shows the performance of the past distribution conditional on the preceding ten-day patterns.

The bar charts should be flat if the past distribution and the future distribution are the same. Given the large number of samples, by chi-square tests, we can statistically reject the null hypothesis that the future distribution is the same as the past distribution for the testing periods. Note that both distributions underestimate tail risks. The statistical test results justify people's concern that the past data do not represent the future. However, as Figure 13 shows the past distribution does contain some information about the future, e.g., the mean values. In addition, the past distributions conditional on the preceding ten-day price pattern contain more information about the future than the unconditional past distributions.

These findings from the data are encouraging for SP because they indicate that sophisticated structures containing information in data need to be identified and understood. Meanwhile, these findings also pose challenges because hypotheses and systems in these economic data are different from those in typical SP applications. Related SP problems and solutions for such big data and modeling applications need to be carefully formulated.

Other related works in SP

We have mainly focused on the literature in economics and business research and provided the SP understanding on the literature. Note that there have been three special issues on SP for finance [1]–[3]: one in *IEEE Signal Processing Magazine* in 2011, and the other two in *IEEE Journal of Selected Topics in Signal Processing* in 2012 and 2016. Readers can find more SP examples there, especially on portfolio and risk analysis, HFT, and algorithmic trading. When going through SP technologies, readers can focus more on the economic problem formulation and evaluations of solutions with the concepts discussed in this tutorial. Also, an overview of business analytics related to SP can be found in [65] and [66], which provide perspectives on system modeling of a business.

Conclusions and thoughts

In this tutorial, we present an introduction to some fundamental economic theories that govern financial markets and people's economic decisions, including expected utility, RP, portfolio theory and asset-pricing models, EMH, prospect theory, and behavioral economics. We also introduce basic econometric tools and theories from an SP perspective.

We emphasize that when analyzing data and building economic models, researchers should keep in mind existing economic theories and hypotheses. The burden of proof is high when findings contradict or are inconsistent with existing theories. For example, any price anomaly found in financial markets needs to be carefully examined because it may contradict existing models and the EMH and could be a chance result. It is always difficult to predict prices because these are usually determined by market equilibrium no matter how

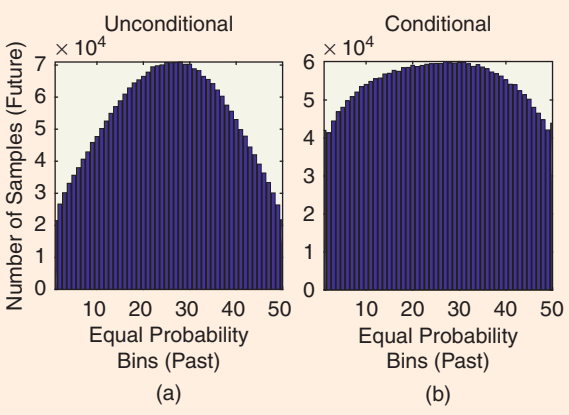


FIGURE 13. The bar charts should be flat if the past distribution and the future distribution are the same.

unreasonably they may seem. Excess economic benefits are associated with risks.

Time-series models, system models, and statistical models are certainly useful tools to analyze economic data. SP researchers can try to formulate the economic problems into filtering, array SP, impulse responses, finite-state machine, Markov models, and state-space models. Whenever a lagged transform seems to be useful, researchers can consider the Z transform, Laplace transforms, and spectral analysis, which may provide additional insights into economic systems and problem solutions.

Caution also needs to be exercised when formulating input–output relationships and objective functions for an economic system. In many cases, humans are only observers of the system and have little power to change the input, e.g., consumer price indices, or the output, e.g., sales or GDP. Meanwhile, humans are oftentimes also part of an economic system, and any human actions or predictions may change the system or system equilibrium. For example, a buying order for a stock may inevitably change the price of the stock. There are times when humans are indeed actively responsible for the input of the system, such as an investment budget decision on R&D or marketing for a company. The different nature of inputs and outputs shapes not only problems and models but also the methodology to the solution and interpretation of results.

In terms of mathematical formulations and modeling tools, many similarities exist between SP and econometrics. However, the weaknesses of SP methodology come from what is often taken for granted in SP, such as the knowledge of physical systems, ground-truth data, controlled experiments, or the confidence in simulated data. Mean-squared errors or variances are often good statistics in SP performance evaluation but are not sufficient in economic and business studies. Indeed, when making investment suggestions and decisions, i.e., putting one's money to bet on a model, researchers need to understand more of the probability and statistical significance of their predictions, estimation, data samples, model validity, etc. They need to understand causal relationships and examine all possible alternative models and hypotheses about the data because an accurate mechanism of a social or economic system is always unknown.

In the big data era, there is a wealth of data analysis and processing work for which SP tools are useful, e.g., denoising and data cleaning, feature extraction, pattern detection, tracking, and abnormality detection. Note that traditional economics are more about equilibrium, and SP can play a big role in understanding the process and transition patterns to reach equilibrium from data analysis.

Indeed, numerical data are at the core of SP. SP is promising in exploring economic big data to learn information better and faster and in finding sophisticated subtle structures within data. Researchers need to be careful not to be biased by the data at hand.

As SP research and applications are expanding, SP researchers can broaden their vision and take advantage of the methodologies and advancements in econometrics, economics, and statistical methods for SP problems. For example, game theory has been an active research tool in communication systems and resource allocation. More cross-pollination between SP and economics will produce fruitful results in research as well as practical applications affecting people's daily lives.

Acknowledgments

We thank EidoSearch Inc. (www.eidosearch.com) for providing the empirical test results in the section "Big Data Analysis of Financial Data Based on SP."

Authors

Xiao-Ping (Steven) Zhang (xzhang@ryerson.ca) received the B.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in electronic engineering and the M.B.A. degree in finance, economics, and entrepreneurship from the University of Chicago, Illinois. He is a professor of electrical and computer engineering and is cross appointed to the Finance Department at the Ted Rogers School of Management at Ryerson University, Toronto, Canada. His research interests include signal processing, electronic systems, machine learning, big data, finance, and marketing. He is the cofounder and chief executive officer for EidoSearch, an Ontario-based company offering a content-based search and analysis engine for financial big data.

Fang Wang (fwang@wlu.ca) received the M.B.A. degree in finance from the University of Texas, San Antonio, and the Ph.D. degree in management information systems from Michael G. DeGroote School of Business, McMaster University, Hamilton, Ontario, Canada. She is an associate professor of marketing at the Lazaridis School of Business and Economics, Wilfrid Laurier University, Waterloo, Ontario, Canada. Her research interests include market response models, marketing data mining/analysis, long-term marketing productivity, information systems management, big data, and social media applications in marketing and finance.

References

[1] *IEEE Signal Process. Mag. (Special Issue on Signal Processing for Financial Applications)*, vol. 28, no. 5, Sept. 2011.

[2] *IEEE J. Sel. Topics Signal Process. (Special Issue on Methods in Finance and Electronic Trading)*, vol. 6, no. 4, Aug. 2012.

[3] *IEEE J. Sel. Topics Signal Process. (Special Issue on Financial Signal Processing and Machine Learning for Electronic Trading)*, vol. 10, no. 6, Sept. 2016.

[4] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *J. Finance*, vol. 25, no. 2, pp. 383–417, May 1970.

[5] J. H. Cochrane, *Asset Pricing*, rev. ed. Princeton, NJ: Princeton Univ. Press, 2005.

[6] A. W. Lo, H. Mamaysky, and J. Wang, "Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation," *J. Finance*, vol. 55, no. 4, pp. 1705–1765, Aug. 2000.

[7] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–291, Mar. 1979.

[8] R. J. Shiller, "Do stock prices move too much to be justified by subsequent changes in dividends?" *Am. Econ. Rev.*, vol. 73, no. 1, pp. 236–237, Mar. 1983.

[9] E. F. Fama, "Market efficiency, long-term returns, and behavioral finance," *J. Finance Econ.*, vol. 49, no. 3, pp. 283–306, Sept. 1998.

[10] F. Wang and X. P. Zhang, "Reasons for market evolution and budgeting implications," *J. Marketing*, vol. 72, no. 5, pp. 15–30, Sept. 2008.

[11] F. Wang, X. P. S. Zhang, and M. Ouyang, "Does advertising create sustained firm value? The capitalization of brand intangible," *J. Acad. Marketing Sci.*, vol. 37, no. 2, pp. 130–143, 2009.

[12] D. M. Hanssens, F. Wang, and X. P. Zhang, "Performance growth and opportunistic marketing spending," *Int. J. Research Marketing*, vol. 33, no. 4, pp. 711–724, Dec. 2016.

[13] E. Budish, P. Cramton, and J. Shim, "The high-frequency trading arms race: Frequent batch auctions as a market design response," *Quarterly J. Economics*, vol. 130, no. 4, pp. 1547–1621, Nov. 2015.

[14] J. D. Hamilton, *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press, 1994.

[15] H. Levy and H. M. Markowtiz, "Approximating expected utility by a function of mean and variance," *Am. Econ. Rev.*, vol. 69, no. 3, pp. 308–317, 1979.

[16] Y. Kroll, H. Levy, and H. Markowitz, "Mean-variance versus direct utility maximization," *J. Finance*, vol. 39, no. 1, pp. 47–61, 1984.

[17] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, 1952.

[18] W. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *J. Finance*, vol. 19, no. 3, pp. 425–442, 1964.

[19] J. Lintner, "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets," *Rev. Economics Statistics*, vol. 47, no. 1, pp. 13–37, 1965.

[20] F. Black, "Capital market equilibrium with restricted borrowing," *J. Business*, vol. 45, no. 3, pp. 444–455, 1972.

[21] E. F. Fama and K. R. French, "The capital asset pricing model: Theory and evidence," *J. Economic Perspectives*, vol. 18, no. 3, pp. 25–46, Sept. 2004.

[22] J. V. Jacob Viner, "Cost curves and supply curves," *Zeitschrift fr Nationalkonomie/J. Economics*, vol. 3, no. 1, pp. 23–46, 1931.

[23] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *J. Financial Economics*, vol. 33, no. 1, pp. 3–56, Mar. 1993.

[24] E. F. Fama and K. R. French, "The cross-section of expected stock returns," *J. Finance*, vol. 47, no. 2, pp. 427–465, June 1992.

[25] M. M. Carhart, "On persistence in mutual fund performance," *J. Finance*, vol. 52, no. 1, pp. 57–82, 1997.

[26] E. F. Fama and K. R. French, "A five-factor asset pricing model," *J. Financial Economics*, vol. 116, no. 1, pp. 1–22, 2015.

[27] E. F. Fama, "Multifactor portfolio efficiency and multifactor asset pricing," *J. Financial Quantitative Analysis*, vol. 31, no. 4, pp. 441–465, 1996.

[28] R. C. Merton, "An intertemporal capital asset pricing model," *Econometrica*, vol. 41, no. 5, pp. 867–887, 1973.

[29] H. L. Van Trees, *Detection, Estimation, and Modulation Theory,* part 4, *Optimum Array Processing*. Hoboken, NJ: Wiley, 2002.

[30] G. Ganesan, "A subspace approach to portfolio analysis," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 49–60, Sept. 2011.

[31] M. U. Torun, A. N. Akansu, and M. Avellaneda, "Portfolio risk in multiple frequencies," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 61–71, Sept. 2011.

[32] A. N. Akansu and M. U. Torun, "Toeplitz approximation to empirical correlation matrix of asset returns: A signal processing perspective," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 4, pp. 319–326, Aug. 2012.

[33] F. Rubio, X. Mestre, and D. P. Palomar, "Performance analysis and optimal selection of large minimum variance portfolios under estimation risk," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 4, pp. 337–350, Aug. 2012.

[34] D. M. Malioutov, A. A. Corum, and M. Çetin, "Covariance matrix estimation for interest-rate risk modeling via smooth and monotone regularization," *IEEE J. Sel. Topics Signal Process., vol. 10*, no. 6, pp. 1006–1014, Sept. 2016.

[35] M. Senneret, Y. Malevergne, P. Abry, G. Perrin, and L. Jaffrs, "Covariance versus precision matrix estimation for efficient asset allocation," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 6, pp. 982–993, Sept. 2016.

[36] E. Jay, P. Duvaut, S. Darolles, and A. Chrétien, "Multifactor models," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 37–48, 2011.

[37] B. Malkiel, *A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing*, 10th ed. New York: Norton, 2011.

[38] M. S. Fischer Black, "The pricing of options and corporate liabilities," *J. Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.

[39] J. Y. Campbell and R. J. Shiller, "The dividend-price ratio and expectations of future dividends and discount factors," *Rev. Financial Studies*, vol. 1, no. 3, pp. 195–228, 1988.

[40] J. H. Cochrane, "Presidential address: Discount rates," *J. Finance*, vol. 66, no. 4, pp. 1047–1108, 2011.

[41] R. J. Shiller, *Irrational Exuberance*, 2nd ed. Princeton, NJ: Princeton University Press, 2005.

[42] J. H. Cochrane. Eugene Fama: Efficient markets, risk premiums, and the Nobel Prize. presented at the University of Chicago Syposium, Nov. 4, 2013. [Online]. Available: https://faculty.chicagobooth.edu/john.cochrane/research/papers/Fama_panel_nov_2013.pdf

[43] X. P. Zhang and D. Kedmey, "Techware: Financial data and analytic resources," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 138–141, Sept. 2011.

[44] R. Tsay, *Analysis of Financial Time Series*. Hoboken, NJ: Wiley, 2010.

[45] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *J. Amer. Statist. Assn.*, vol. 74, no. 366, pp. 427–431, 1979.

[46] C. Granger and P. Newbold, "Spurious regressions in econometrics," *J. Econometrics*, vol. 2, no. 2, pp. 111– 120, 1974.

[47] R. F. Engle and C. W. J. Granger, "Co-integration and error correction: Representation, estimation, and testing," *Econometrica*, vol. 55, no. 2, pp. 251–276, 1987.

[48] C. Granger, "Testing for causality: A personal viewpoint," *J. Economic Dynamics Control*, vol. 2, pp. 329–352, 1980.

[49] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969. [Online]. Available: http://www.jstor.org/stable/1912791

[50] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, vol. 31, pp. 307–327, 1986.

[51] R. Engle, "New frontiers for arch models," *J. Applied Econometrics*, vol. 17, no. 5, pp. 425–446, 2002.

[52] N. Wiener, "The theory of prediction," in *Modern Mathematics for the Engineer*, E. F. Beckenbach, Ed. New York: McGraw-Hill, 1994.

[53] D. Creal, "A survey of sequential Monte Carlo methods for economics and finance," *Econometric Revs.*, vol. 31, no. 3, pp. 245–296, 2012.

[54] D. Hanssens, L. Parsons, and R. Schultz, *Market Response Models: Econometric and Time Series Analysis* (International Series in Quantitative Marketing). New York: Springer-Verlag, 2012.

[55] S. Sunder, "Stationarity of market risk: Random coefficients tests for individual stocks," *J. Finance*, vol. 35, no. 4, pp. 883–896, 1980.

[56] T. Rajbhandary, X. P. Zhang, and F. Wang, "Piecewise constant modeling and Kalman filter tracking of systematic market risk," in *Proc. 2013 IEEE Global Conf. Signal Information Processing (GlobalSIP),* p. 1144.

[57] L. Vo, X.P. Zhang, and F. Wang, "Multifactor systematic risk analysis based on piecewise mean reverting model," in *Proc. 2013 IEEE Global Conf. Signal Information Processing (GlobalSIP),* p. 1142.

[58] J. Han, X. P. Zhang, and F. Wang, "Gaussian process regression stochastic volatility model for financial time series," *IEEE J. Sel. Topics Signal Process.* (*Special Issue on Financial Signal Processing and Machine Learning for Electronic Trading*), vol. 10, no. 6, Sept. 2016.

[59] R. Frigola, Y. Chen, and C. Rasmussen, "Variational Gaussian process state-space models," in *Proc. Advances Neural Information Processing Systems*, 2014, pp. 3680–3688.

[60] Y. Wu, J. M. Hernández-Lobato, and Z. Ghahramani, "Gaussian process volatility model," in *Proc. Advances Neural Information Processing Systems*, 2014, pp. 1044–1052.

[61] A. Wilson and Z. Ghahramani, "Copula processes," in *Proc. Advances Neural Information Processing Systems*, 2010, pp. 2460–2468.

[62] S. J. Taylor, "Modeling stochastic volatility: A review and comparative study," *Mathematical Finance*, vol. 4, no. 2, pp. 183–204, 1994.

[63] A. A. Christie, "The stochastic behavior of common stock variances: Value, leverage and interest rate effects," *J. Financial Economics*, vol. 10, no. 4, pp. 407–432, 1982.

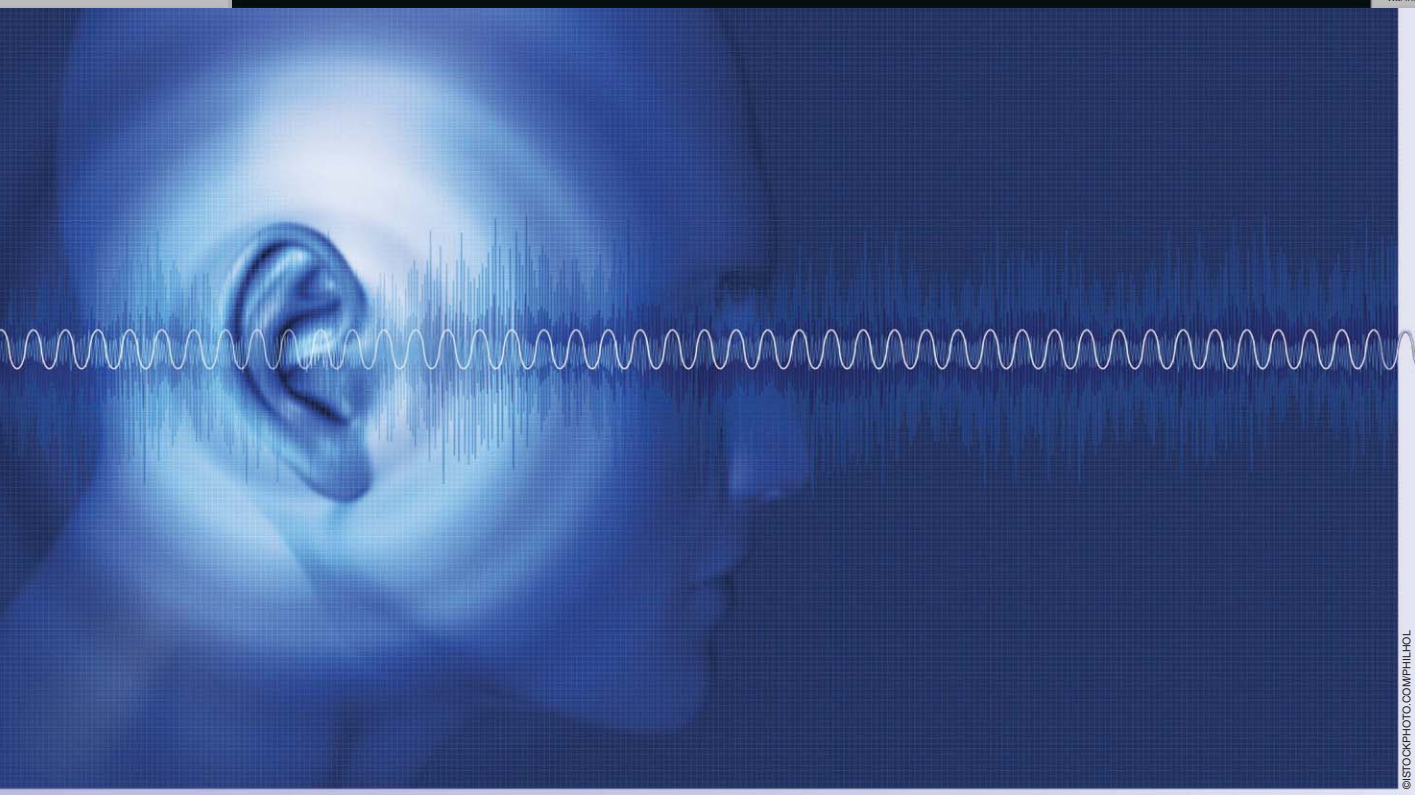
[64] G. Wu, "The determinants of asymmetric volatility," *Rev. Financial Studies*, vol. 14, no. 3, pp. 837–859, 2001.

[65] K. R. Varshney and A. Mojsilovi, "Business analytics based on financial time series," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 83–93, Sept. 2011.

[66] K. R. Varshney. (2012). Introduction to business analytics. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing tutorial. [Online]. Available: https://informationashvins.files.wordpress.com/2012/04/varshney_icassp2012.pdf

[67] J. H. Cochrane. (2008, Nov. 12). Is now the time to buy stocks? *Wall Street Journal*. [Online]. Available: https://www.wsj.com/articles/SB122645226692719401

SP



Perceptual Spatial Audio Recording, Simulation, and Rendering

Digital Object Identifier 10.1109/MSP.2017.2666081
Date of publication: 26 April 2017

Developments in immersive audio technologies have been evolving in two directions: physically motivated systems and perceptually motivated systems. Physically motivated techniques aim to reproduce a physically accurate approximation of desired sound fields by employing a very high equipment load and sophisticated, computationally intensive algorithms. Perceptually motivated techniques, however, aim to render only the perceptually relevant aspects of the sound scene by means of modest computational and equipment load. This article presents an overview of perceptually motivated techniques, with a focus on multichannel audio recording and reproduction, audio source and reflection culling, and artificial reverberators.

The spatial sound scene

Since Alan Blumlein introduced the original concept of stereophonic recording using a pair of figure-eight microphones, spatial sound technologies have steadily grown in sophistication, complexity, and capabilities. Delivering a convincing illusion of a desired sound field requires finding solutions to several problems at the intersection of physics, psychoacoustics, and engineering. First, the relevant sound-field information needs to be identified, and methods for its acquisition devised, which amounts to designing an array of microphones. Then, methods for rendering the identified spatial audio information in some optimal way need to be developed. This requires the design of a playback system, including a hardware configuration and the necessary signal processing algorithms. If the spatial sound field is virtual, i.e., not generated by an actual acoustic event, the required playback signals need to be synthesized rather than recorded. To that end, an accurate approximation of the desired sound field would ideally be computed and then recorded using a virtual microphone array, to be played via the corresponding actual loudspeaker array. This process is referred to as *auralization*. However, due to the very high numerical complexity of sound-field simulation methods, the auditory perspective is typically first rendered via level differences between pairs of loudspeakers and then overlaid by room effects. The

An overview of spatial-audio techniques based on psychoacoustics

Hüseyin Hacıhabiboğlu, Enzo De Sena, Zoran Cvetković, James Johnston, and Julius O. Smith III

past nine decades of spatial audio reproduction and synthesis have seen innovations and developments in all these directions.

Generating the experience of a spatial sound scene can be achieved in a number of ways. Comparing different methods, at one extreme there are binaural techniques [1], which provide a convincing experience over two channels by presenting stereophonic audio cues, i.e., interaural time, level, and spectral differences, which are known as *ear signals*. Binaural presentations work best over headphones. However, with crosstalk cancelation [2], they can also be successfully used with a pair of loudspeakers, although the effect is confined to a very narrow listening area. For a listener who is not static, the auditory illusion can be maintained via head-tracking mechanisms combined with the real-time adaptation of the binaural signals. The advent of virtual and augmented reality systems has recently revived interest in binaural systems. However, some inherent problems of binaural audio, such as individualization, remain [3], limiting the spatial quality of the auditory experience they provide.

At the other extreme, there are systems that aim to reconstruct an accurate physical approximation of a sound field. Notable examples include wave field synthesis (WFS) [4] and higher-order Ambisonics (HOA) [5]. WFS is based on the Huygens principle and Kirchhoff–Helmholtz integral, which together state that the sound field due to a primary source can be exactly synthesized by infinitely many secondary sources on the surface enclosing a reproduction volume. Such a system can achieve a spatially extensive listening area and can be used in large auditoriums, such as film theaters. Ambisonics is based on sound-field approximation using its spherical harmonics at the center of the listening area. HOA is capable of achieving results comparable to WFS close to the center of the reproduction rig. While both WFS and HOA provide elegant solutions to the spatial recording and reproduction problem, they have high equipment load requirements, which can reach several hundreds of carefully positioned loudspeakers. For this reason, their application domain has so far been confined to specialist high-end systems. WFS and HOA can also run on systems with a more practical equipment load by including perception-inspired corrections. Comprehensive reviews of WFS and HOA have recently been published [6], [7].

In between these two extremes are systems with five to ten channels that are suitable for use in small to medium-size listening rooms. Such systems do not possess a sufficient number of channels to physically reconstruct a sound field in a wide listening area, nor are they capable of accurately reconstructing the ear signals for listeners in multiple locations. Therefore, they must rely to a large degree on perceptual effects similar to those used for binaural systems to generate the illusion of a desired sound field within not overly confined areas.

As with recording and reproduction technologies, there are many techniques for sound-field simulation. At one extreme, there are physically motivated methods, which aim to calculate an approximate solution of the wave equation. For that purpose, several numerical methods have been developed that achieve a very high level of accuracy. However, they typically have prohibitively high computational costs. Examples include the finite-difference time domain, finite element method (FEM), and boundary element method (BEM) [8]. While these approaches lend themselves to parallelization, the associated computational cost is still too high for real-time operation at interactive rates and on low-cost devices.

Conversely, there are methods that try to render only some higher-level perceptual effects. These methods, called *artificial reverberators*, require only a fraction of the computational load associated with physically motivated room simulators and typically aim to mimic only certain characteristics of the tail of typical room impulse responses, such as modal density, echo density, and timbral quality [9]. They do not explicitly model a given physical space but, rather, are used to obtain a pleasing reverberant effect and have been widely used for artistic purposes in music production.

In between these two extremes are methods that aim to render a certain physical sound scene, but by modeling only its most perceptually relevant aspects. Full-blown room auralization systems typically aim to render each and every reflection and diffraction up to a given order for each source [10], [11]. More recent methods achieve remarkable computational savings by accurately rendering only first-order reflections, while replacing higher-order reflections with their progressively coarser approximations [12]. Further computational savings are possible by eliminating sources whenever they are inaudible, a

process referred to as *audio source culling*. This article is concerned with spatial audio systems and methodologies that substantially rely on psychoacoustics.

Spatial auditory perception

The psychophysics of spatial hearing has been an active research area for the past century. Most of the information given in this section has been thoroughly reviewed by Blauert in [13]. Interested readers are referred to this excellent volume for more information and an extensive set of further references. The primary mechanism humans use to localize sound sources in the horizontal plane is based on the differences between the signals received by the two ears. Due to the spatial separation between the ears, the sound wave generated by a sound source reaches the two ears with a different delay, called the *interaural time difference* (*ITD*). Moreover, the sound wave is scattered by the head, causing the level of the signal at the ear farther away from the source, the contralateral ear, to be reduced in comparison with the level of the signal at the ear closer to the source, the ipsilateral ear. This level difference is called the *interaural level difference* (*ILD*).

The interaural time delay for a typical human head can vary between ±750 μs in the acoustic free field. Humans can detect ITDs as low as 10–20 μs at the front direction, corresponding to about 1° in the horizontal plane. Similarly, the ILD is frequency dependent and can be as high as 21 dB at 10 kHz. Sensitivity to changes in the ILD is also frequency dependent. For instance, for pure tones, it varies between 0.5 and 2.5 dB. In contrast with the ITD, which is the primary localization cue at low frequencies, ILD cues are more important in sound source localization at higher frequencies. This is due to the low level of scattering at low frequencies when the wavelength is close to or larger than the size of the head. ITD and ILD cues also change with the distance of a sound source and the size of the head.

Note that ITD and ILD pairs do not uniquely specify the source direction. For the purpose of illustration, if we assume a spherical head, binaural cues will be identical for sound sources placed on cone-shaped surfaces at each side of the head. These surfaces are called *cones of confusion*. In the horizontal plane, sources on the conic section that is the intersection of the horizontal plane with the cone of confusion will have front–back ambiguity. Humans can typically resolve this ambiguity by small head movements.

The elevation of a sound source is perceived based primarily on the spectral shaping of its signal that occurs as a result of the scattering of the sound around the head. This spectral shaping depends on the elevation in a manner determined by the sizes and shapes of the pinnae, head, and torso. Consequently, the frequency content of the sound itself also affects the perception of the elevation of its source.

Subjective localization of sound sources involves a significant level of uncertainty. Localization blur is the smallest change in the direction of a source that will result in a change in its perceived direction. For sources in the horizontal plane, localization blur is generally less than around 10°. For sources in the median plane, localization blur on the order of 20° can be observed. A related concept is locatedness, which refers to the perception of the spatial extent of a sound source. This is an important attribute because the center of mass of a sound source can be localized accurately, yet the source can still be diffusely located. Two other measures of spatial resolution of hearing are the minimum audible angle (MAA) and minimum audible movement angle (MAMA). The MAA corresponds to the minimum change in the direction of a static source for a listener to discriminate it as being to the left or to the right of the original direction. The MAMA, on the other hand, is a measure of spatial resolution for moving sources; it quantifies the smallest arc that a moving sound source must travel to be discriminable from a stationary source [14].

The perception of the distance of a sound source is both less reliable and less well understood than the perception of the direction of a sound source. Several cues affect the perception of distance. Among these, intensity is the only cue inherently related to the sound source and is also the only absolute cue. The other distance cues are related either to the environment (the direct-to-reverberant energy ratio and lateral reflections), the physical properties of the listener (e.g., auditory parallax), or cognitive aspects (e.g., familiarity) [15].

An interesting property of distance perception is the overestimation and underestimation of distance at different ranges and for different sounds. Apparent distances of sources far away from a listener are underestimated, and those closer than around 1–2 m are overestimated [15]. Familiarity, which is a cognitive cue related to prior exposure to and knowledge of the characteristics of the sound source, also has a similar effect. For example, the distance of whispered speech is underestimated, while that of shouted speech is overestimated.

An important capability of the human auditory perception mechanism lies in its ability to localize sources in reverberant environments such as rooms and other enclosed spaces. This is made possible by suppressing reflections that come immediately after the direct sound. When a broadband impulse and a delayed copy of it are presented from different directions with a short delay of less than 1 ms in between, a single auditory event is perceived at a direction between the directions of the two sources, gradually shifting toward the leading source as the lag in the time of arrival increases. (An *auditory event* is defined as an event perceived by a listener typically, but not necessarily, in response to a sound event.) This effect is known as *summing localization*, and both sources contribute to the perceived direction of the auditory event. When the delay is between 1 and 5 ms, a single fused auditory event close to the leading source can be heard. Within this delay range, the presence of the lagging source is audible since it changes the timbre of the auditory event, but its direction cannot be easily discriminated. Above 5 ms, the broadband click and its echo are perceived as distinct sound events. The time delay above which two distinct events are heard is called the *echo threshold*. While the classic demonstration of these effects involves broadband click pairs, different signals will have different echo thresholds. For example, the echo threshold can be as high as 20 ms for speech and music signals.

The effect that the direction of the auditory event depends predominantly on the leading source is known as *localization*

dominance, whereas the effect that a single auditory event is perceived when there are two sound events is called *fusion*. The effect that the discrimination of the direction of the lagging sound source is suppressed is termed *lag discrimination suppression*. These three are collectively known as the *precedence effect* [16].

Another important binaural cue is interaural coherence (IC), which is a measure of the coherence of signals received by the two ears. IC is high for sounds coming directly from the source, where the two ear signals are highly correlated, and low in the diffuse sound field, where the correlation is low. Therefore, IC provides information about the level of reverberation and thus about the spaciousness of the environment.

The history of perceptually motivated spatial audio

Binaural audio and multichannel stereophony are two of the most common spatial audio technologies, predating more recent technologies such as WFS by more than half a century. Binaural audio has found extensive use and received renewed interest, especially for virtual reality (VR) applications, while multichannel systems have been the de facto standard for home entertainment and automotive audio systems. Simultaneously, due to the popularity and market dominance of two-channel audio formats, stereophony using two loudspeakers is still commonly employed.

Binaural audio

Binaural audio is based on a simple assumption: if the signals that would be received at the ears of a listener as a result of an acoustic event are provided to the listener with sufficient accuracy, the person will perceive an auditory event corresponding to the original acoustic event. These ear signals can either be recorded with microphones implanted in the ear canals of an artificial human head, such as Knowles Electronic Manikin for Acoustic Research or Neumann KU-100, or synthesized using signal processing methods. In both cases, the signals are usually presented over a pair of headphones.

The microphones used for recording binaural audio are also termed *dummy head* microphones and are manufactured to resemble a typical human head. The external ears of these microphones are typically molded in silicon and are modeled after the external ears of humans who have exceptional spatial hearing acuity. The recorded signals need to be played back using headphones equalized appropriately with free-field or diffuse-field equalization, depending on the environment in which the recording was made [1].

Binaural synthesis is based on the knowledge of the acoustic transfer paths between the source and the two ears. These paths are characterized by their impulse responses, referred to as the *head-related impulse response* (*HRIR*) and *head-related transfer function* in the frequency domain. For each source position, there will be two of them, one for the left ear and one for the right. When HRIRs are convolved with dry source signals, the resulting signals will incorporate the necessary binaural cues for the given source position. In the case of a sound field created by P sources in the far field, the right and left ear signals can be synthesized as

$$x_L(n) = \sum_{p=1}^{P} x_p(n) * h_{L,\theta_p,\phi_p}(n), \quad (1)$$

$$x_R(n) = \sum_{p=1}^{P} x_p(n) * h_{R,\theta_p,\phi_p}(n), \quad (2)$$

where $x_p(n)$ is the pressure signal due to source p; $h_{L,\theta_p,\phi_p}(n)$ and $h_{R,\theta_p,\phi_p}(n)$ represent the HRIRs for the left and the right ears for a source at a direction (θ,ϕ), where θ and ϕ are the azimuth and elevation angles, respectively; and * denotes convolution. This approach assumes that the acoustical system consisting of these sources and the listener is linear and time invariant and that the resulting left and right ear signals provide the necessary spatial hearing cues pertaining to the acoustic field that would be generated by these P sources.

In free field, HRIRs can be considered finite and are typically up to 12-ms long, corresponding to approximately 512 samples at the 44.1-kHz sampling rate. This does not present a significant computational cost for a single component. However, as the number of components increases, such as when a source and its reflections in a room are being rendered, the computational cost of convolution becomes an important bottleneck. To overcome this limitation, different filter design approaches have been proposed, e.g., [17]. These filters are designed to capture salient binaural cues while significantly reducing the computational cost.

Two essential requirements of binaural synthesis are 1) the availability of a set of HRIR measurements densely sampled on a spherical shell and 2) the match between these HRIRs and the actual HRIRs of the listener. Regarding the first requirement, interpolation methods such as kernel regression [18] can be used to increase the granularity of the available directions. The second requirement necessitates the measurement of individualized HRIRs, which is both time consuming and costly. For that reason, many existing research-grade and commercial solutions use generic HRIRs. This, however, is not an ideal solution, since there are significant differences between the spectra of the generic HRIRs and individual HRIRs of the listener, and these cues are essential for elevation perception [13]. Practical setups that allow quick measurement of HRIRs around a geodesic sphere surrounding the listener's head have recently been developed [19]. There also exist commercial products that allow tailoring a stored set of HRIRs based on head size (https://www.ossic.com). However, head size alone can improve only the ITD and ILD cues provided by the system, not the spectral cues used in the perception of source elevation.

Binaural synthesis also allows interactivity if the position and orientation of the listener's head can be tracked [20]. High-precision and high-accuracy magnetic trackers had been the de facto method for tracking a listener's head. Recent developments made it possible to track a user's head with inexpensive devices (http://www.3dsoundlabs.com). These developments make binaural synthesis an excellent solution for VR applications. For binaural synthesis, a side effect of system errors—such as a pair of improperly equalized headphones, an HRIR set that does not well match the HRIRs of the user, or inaccurate head tracking—is inside-the-head localization [21]. This

undesirable effect can be partly alleviated by adding simulated reflections and artificial reverberation.

Binaural audio can also be presented via a pair of loudspeakers; however, each ear then receives not only its intended signal but also the signal intended for the other ear, which impairs the coherence of binaural cues. This effect is known as *crosstalk* [22]. There are methods for crosstalk cancelation based on predicting the response of the crosstalk path and inverting it [23]. Such methods preprocess the left and right channels using a 2×2 crosstalk cancelation filter matrix, which is obtained as the inverse of the matrix containing the direct and crosstalk acoustic transfer paths. When a listener is sitting still at the position for which crosstalk cancelation is made, a single set of crosstalk cancelation filters can be very effective. However even small head movements require filter adaptation, which increases the computational overhead associated with crosstalk cancelation.

Crosstalk-canceled binaural audio, also known as *transaural audio*, has distinct benefits in comparison with two-channel stereophony and headphone-based binaural presentation: 1) it can simulate sources behind the listener, even when there is no corresponding physical source (i.e., a loudspeaker), and 2) it provides a better externalization of the simulated sources due to the presentation being made over loudspeakers [24].

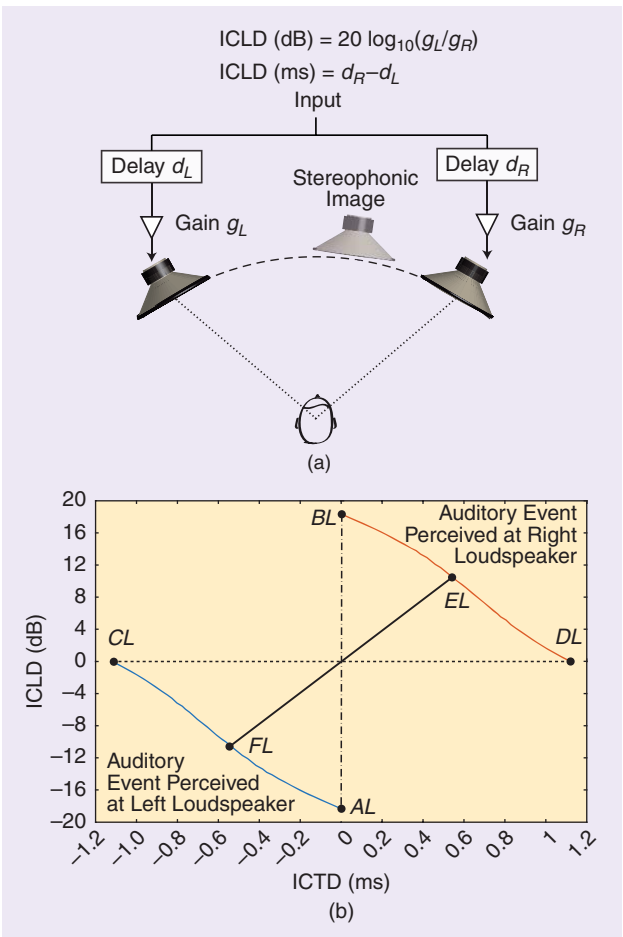


FIGURE 1. (a) The standard stereophonic setup according to Williams [27], where g_L and g_R are the left and right channel gains and d_L and d_R are the left and right channel delays, respectively, and (b) the associated psychoacoustical curves representing the ICTD and ICLD that are necessary to pan the direction of a virtual source from the direction of the left loudspeaker to the right one.

Two-channel stereophony

Two-channel stereophony is an alternative spatial audio technology that requires the minimal number of channels to produce the impression of spatial sound. In its usual implementation, two-channel stereophony uses two loudspeakers, each at the same distance from the listener, positioned 30° to either side of the front direction, providing a frontal auditory scene within a base angle of 60°. The ideal listening position, referred to as the *sweet spot*, thus forms an equilateral triangle with the loudspeakers. Two-channel stereophony creates the illusion of a sound source in a given direction within the base angle by means of the interchannel time differences (ICTDs) and interchannel level differences (ICLDs) of the two channels over which the source signal is presented. Figure 1(a) shows the standard stereophonic setup and illustrates how the gains and delays of each channel are linked to the ICTD and ICLD. Although it is intuitively clear that the direction of the virtual source is pulled toward the loudspeaker that produces the louder and earlier version of the signal, knowing the precise relationship between the perceived source direction and the presented (ICTD and ICLD) pairs requires extensive psychoacoustic measurements.

The first comprehensive study of the relationships between ICTD and ICLD, termed *stereophonic panning laws*, was conducted by Franssen [25]. Another study, by Williams [26], combined earlier studies on ICTD and ICLD. The panning curves presented in that study are now known as *Williams's curves*. Williams's psychoacoustic curves are illustrated in Figure 1(b), which shows curves of ICTD/ICLD pairs that create a virtual source in the direction of the left loudspeaker (the blue curve) and the right loudspeaker (the orange curve). ICTD/ICLD pairs that are below or above the two curves are also localized at the left and right loudspeaker, respectively. Virtual sources in directions between the loudspeakers are then created by means of ICTD/ICLD pairs that evolve along a line that connects two points on the psychoacoustic curves.

Note that there are many different ICTD/ICLD pairs that can create a virtual source in the same direction. Intensity stereophony is achieved when the ICTD is zero and only ICLDs are used for generating the stereophonic auditory perspective. In Figure 1(b), intensity panning curves are associated to the vertical axis, e.g., the dashed vertical line connecting points *AL* and *BL*. An ICLD of ±18 dB is sufficient to pan a virtual source exactly in the direction of the right or the left loudspeaker. Time-of-arrival stereophony is achieved when the ICLDs are zero and only ICTDs are used for generating the stereophonic auditory perspective. In Figure 1(b), time-of-arrival panning curves are associated to the horizontal axis, e.g., the dotted horizontal line connecting points *CL* and *DL*. An ICTD of ±1.2 ms is sufficient to create a virtual source in the direction of the right or the left loudspeaker.

Time–intensity stereophony is achieved when a combination of ICTDs and ICLDs is used. The solid line in the figure

connecting points *EL* and *FL* is an example of a time–intensity panning curve. Here, an ICTD of −0.5 ms combined with an ICLD of −12 dB will result in a virtual source aligned with the left loudspeaker. As the ICTD and ICLD are increased toward zero, the virtual source shifts from the left direction to the midline direction. Increasing the ICTD and ICLD further to 0.5 ms and 12 dB, respectively, shifts the virtual source to the direction of the right loudspeaker.

Based on whether interchannel time and level differences are obtained naturally while recording an acoustic scene or are introduced artificially, stereophony can be divided into two categories: recorded (true) stereophony and synthetic stereophony [27]. Recorded stereophony is constrained by the characteristics of the available physical microphones, primarily in terms of their directivity patterns. Microphones with first-order directivity patterns are typically used because of their affordability and availability. These microphones have directivity patterns of the following types:

$$\Gamma_L(\theta) = (1-\alpha_L) + \alpha_L \cos(\theta - \theta_L), \tag{3}$$

$$\Gamma_R(\theta) = (1-\alpha_R) + \alpha_R \cos(\theta - \theta_R), \tag{4}$$

where $\Gamma_L(\theta)$ and $\Gamma_R(\theta)$ are the directivity patterns that represent the directional sensitivity of the left and the right microphones, respectively; θ is the angle defined counterclockwise from the acoustic axis of the corresponding microphone; and θ_L and θ_R are the rotation angles of the left and the right microphones, respectively. Designing stereophonic microphone pairs then requires optimizing the ICTD and ICLD by careful selection of 1) α_L and α_R, 2) θ_L and θ_R, and 3) the distance D between the two microphones.

One of the first stereophonic recording microphone pairs was developed by Alan Blumlein and consisted of two coincident bidirectional microphones (i.e., $\alpha_L = \alpha_R = 1$) positioned at right angles with each other. Many different microphone configurations have been devised since then. These methods can be categorized roughly into three groups: coincident, near coincident, and spaced [28]. Coincident pairs have two colocated ($D = 0$) directional microphones, resulting in the recorded left and right channel signals that have only amplitude differences. Examples of coincident microphone pairs are the Blumlein pair, the XY stereo pair, and the M/S pair [28]. Spaced arrays, such as the AB pair [28], typically use omnidirectional microphones ($\alpha_L = \alpha_R = 0$), with a separation D that is many multiples of the desired wavelength. This makes the ICTD the main cue used to pan the sound source. Near-coincident recording techniques, on the other hand, use directional microphones separated by a small distance comparable to the size of a human head and record both the ICTD and ICLD. Two notable examples are the Nederlandse Omroep Stichting (NOS) and Office de Radiodiffusion Télévision Française (ORTF) pairs, which both use cardioid ($\alpha_L = \alpha_R = 0.5$) microphones and have separations of $D_{\text{ORTF}} = 17$ cm and $D_{\text{NOS}} = 30$ cm [28].

Synthetic stereophony has been predominantly based on intensity panning, since it is thought to provide the most stable virtual sound imaging. Indeed, the inclusion of ICTDs is sometimes considered to yield audible artifacts, such as tonal coloration due to comb filter effects. Another often-cited reason to avoid using ICTDs is the difficulty of controlling the direction of a virtual source by means of time delays. This view has been recently challenged [29] and will be discussed in the "Perceptual Sound Field Reconstruction" section.

The general form of an intensity panning law relates the gains g_L and g_R of the left and right loudspeakers, respectively, to a function of the source direction θ_s and the stereophonic base angle θ_B between the loudspeakers. More specifically, a panning law has the form

$$\frac{g_L(\theta_s) - g_R(\theta_s)}{g_L(\theta_s) + g_R(\theta_s)} = \frac{f(\theta_s)}{f(\theta_B)}. \tag{5}$$

The total power can be maintained via the constant power constraint $g_L(\theta_s)^2 + g_R(\theta_s)^2 = 1$. Two commonly used functions are $f(\theta) = \sin(\theta)$ and $f(\theta) = \tan(\theta)$, which give rise to the so-called sine panning law and tangent panning law, respectively.

The tangent panning law was derived based on perceptual considerations independent of known psychoacoustic curves [30]. In the context of Williams's psychoacoustic curves [see Figure 1(b)], the tangent panning law operates along the vertical axis, i.e., zero ICTD, and connects two points with $\pm\infty$ level differences. Thus, as opposed to panning laws described by Williams's curves, which specify the minimal level differences needed to create virtual sources in loudspeaker directions, the tangent law achieves the same effect by employing maximal level differences.

Multichannel stereophony

An early work by Steinberg and Snow [31] in 1934 suggested that a better auditory perspective is possible if at least three independent microphones are used to capture a frontal sound field and these signals are played back via three loudspeakers. Due to the hardware requirements and technical difficulties in the integration of a three-channel system in radio broadcasts, however, this finding has been obscured by the success and widespread adoption of two-channel stereophony.

The advent of quadrophony and cinematic sound spurred interest in multichannel systems. Traditionally, there are two different types of multichannel audio formats: discrete and matrix [32], [33]. In discrete multichannel audio, there is a one-to-one correspondence between channels and speakers. The storage and transmission of multichannel audio are made using the same number of channels. In matrix multichannel, the original channels are encoded to a smaller number of channels (e.g., two) for transmission or storage over common channels or media and then decoded back to the original channel multiplicity prior to playback. This requires appending auxiliary information to the encoded audio to be used at the decoding stage. More recently, object-based formats have appeared where content and context are encoded separately.

Surround sound is the more commonly known name for multichannel stereophony. There exist several reproduction setups, such as 5.1, 7.1, 10.2, and 22.2, which use five, seven, ten, and 22 main channels, respectively, and one or two low-frequency channels, as described in an International Telecommunication Union (ITU) report (ITU-R BS.2159-4). There are also commercial, object-based formats such as Dolby ATMOS (http://www.dolby.com/us/en/brands/dolby-atmos.html), DTS-X (http://dts.com/dtsx), and Auro-3D (http://www.auro-3d.com), which are very flexible and are likely to dominate the cinematic sound industry into the foreseeable future, considering the new International Organization for Standardization/International Electrotechnical Committee standards such as MPEG-D and MPEG-H.

Commercial microphone arrays for multichannel recording exist, but these arrays are based more on practice in the field than on a solid theory and understanding of the underlying acoustic processes. The microphone arrays used for recording 5.1 multichannel audio typically include cardioid, supercardioid, or hypercardioid microphones positioned on a tree structure [34], [35]. These arrays can in general be separated into two groups: 1) five-channel main microphone techniques and 2) front–rear separation techniques. The former uses five closely positioned microphones mapped directly to the five main channels of a 5.1 reproduction setup. The latter uses two separate arrays to record direct field and ambience separately. For some arrays (such as the INA-5 [36]), there is a one-to-one correspondence between the microphone and loudspeaker channels. For some other arrays (such as the Soundfield Microphone [37], Fukada Tree [38], or Hamasaki Tree [39]), the signals obtained from individual microphone channels need to be mixed.

Some well-known multichannel arrays used for recording multichannel audio are shown in Figure 2. It may be observed that a variety of microphone arrangements exist that try to address the common objective of obtaining an authentic auditory perspective and a high level of envelopment and immersion using existing first-order microphone directivity patterns. The microphone arrays for recording 10.2 multichannel stereophony are still rather experimental (see the ITU-R BS.2159-4 report). Similarly, recording for a 22.2 reproduction system will depend strongly on the venue and context. In fact, multichannel stereophonic systems with higher channel counts, by virtue of the degrees of design freedom they provide, allow for more flexibility but also make it more difficult to design recording setups with strict perceptual rationale.

Recommended reproduction setups for multichannel systems are either standardized (e.g., ITU-R BS.775-1) or in the process of standardization by different standardization bodies [40]. These setups rely mainly on the frontal channels for the presentation of audio content that accompanies visual content (usually films or games). The left and right front channels typically correspond exactly to the two-channel stereophonic setup for cross- and backward compatibility. The difference in these setups is mainly about how ambience is played back. Some of the standards, such as ITU-R BS.1116.1 and ITU-R BS.1534.1, define formal procedures for the subjective evaluation of these systems.

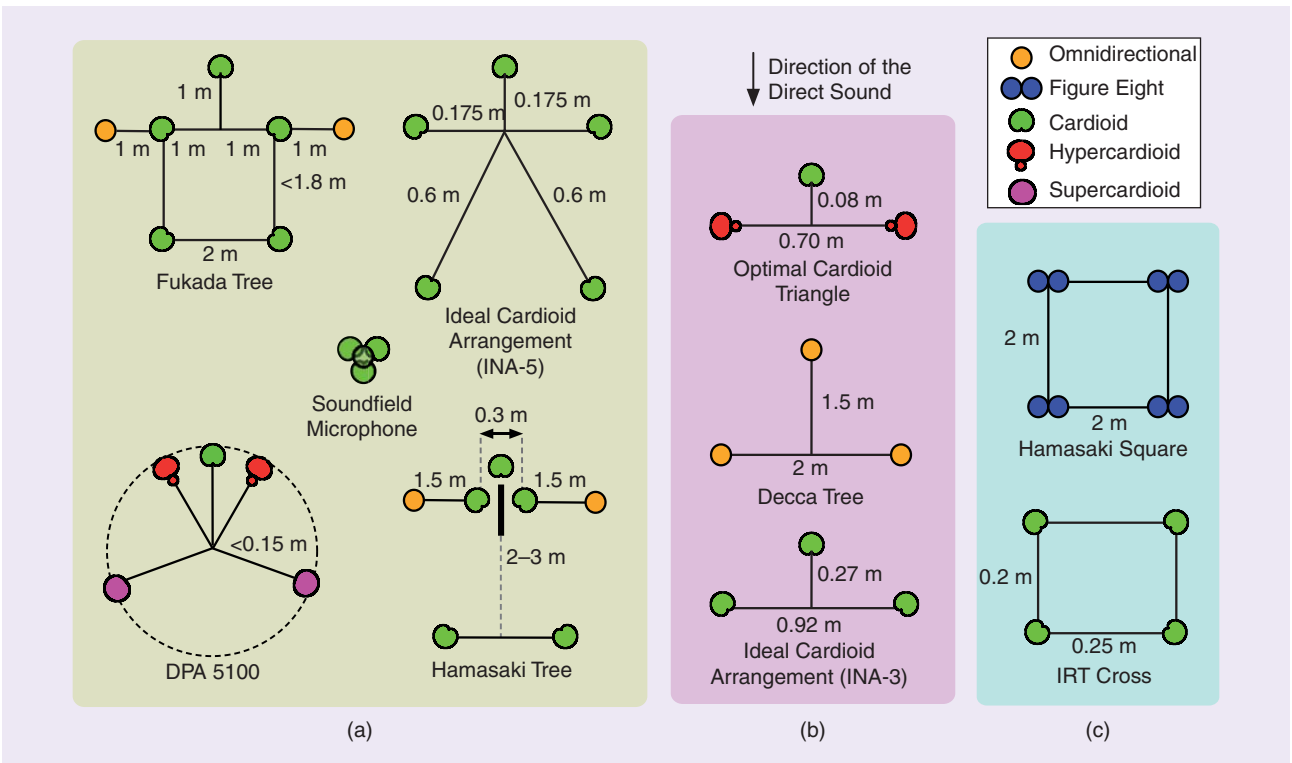


FIGURE 2. Some different microphone array configurations for recording 5.1 multichannel audio showing (a) complete arrays, (b) arrays used to record the frontal scene, and (c) rear-only arrays. The dimensions of the arrays are not drawn to scale. IRT: Institut für Rundfunktechnik.

Perceptually motivated multichannel recording and reproduction

There has been some recent work in the direction of developing systematic frameworks for the design of multichannel stereo systems, most notably vector-base amplitude panning (VBAP), directional audio coding (DirAC), and perceptual sound-field reconstruction (PSR).

Vector-base amplitude panning

It was shown as early as 1973 that tangent panning provides a stereophonic image that is more robust to head rotations than sine panning for the standard stereophonic loudspeaker setup [30]. Pulkki showed that tangent panning can be expressed using an equivalent, vector-based formulation in the horizontal plane and also proposed a three-dimensional (3-D) extension to two-channel intensity panning that allows rendering elevated virtual sources over flexible loudspeaker rigs [41]. This method is VBAP.

Originally, VBAP was designed for a loudspeaker array with elements placed on a geodesic dome's vertices that are situated at the acoustic far field of the listener. Figure 3 shows a section of such a sphere with three loudspeakers, with a listener positioned at the center of the array. The directions of the three loudspeakers are indicated as $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, and the corresponding gains as g_1, g_2, and g_3. A virtual source in a direction $\mathbf{v}_s$ between the loudspeakers can be generated by selecting the gains that satisfy $\mathbf{v}_s = \mathbf{Vg}$, where $\mathbf{V}$ is a matrix whose columns are the directions of the loudspeakers and $\mathbf{g} = [g_1 g_2 g_3]^T$. In addition, the calculated loudspeaker gains are normalized to keep the total power constant.

On the full geodesic sphere, active regions are selected based on the closest three points on the grid, and only those loudspeakers are used for source rendition. This is in contrast with physically based approaches such as Ambisonics, where even for a single source from a single direction, all loudspeakers are potentially active. A major assumption behind VBAP in three dimensions is that summing localization would occur not only with two sources but also with three. This assumption was subjectively tested for different setups and virtual source directions, and it was shown to result in a good subjective localization accuracy for elevated virtual sources [42], [43].

An issue resulting from the utilization of intensity panning in VBAP is the nonuniformity of the spatial spread of the panned source. More specifically, sources panned closer to the actual loudspeakers in the reproduction rig have a smaller spatial spread, while virtual sources panned to directions between loudspeakers have a larger spatial spread. The main cause of this issue is the use of a single loudspeaker when the virtual source direction coincides with the direction of that loudspeaker.

This issue was addressed by panning the virtual source to multiple directions by using three loudspeakers (instead of two) for all source directions in the horizontal plane, or four loudspeakers (instead of three) in the 3-D case. This approach was called *multiple-direction amplitude panning* (*MDAP*) [44]. In a study comparing VBAP with MDAP, it was shown that both provide good subjective localization accuracy, with MDAP being more accurate than VBAP [45]. In another, more recent evaluation carried out within the context of the MPEG-H standard, VBAP resulted in very good subjective localization accuracy, including not only the source azimuth but also its distance [46]. In yet another study, VBAP was shown to provide good localization performance also for sources in the median plane [47]. Note that VBAP is a technology for sound-field synthesis, and in the context of sound-field recording and reproduction it is used at the reproduction end of schemes such as DirAC.

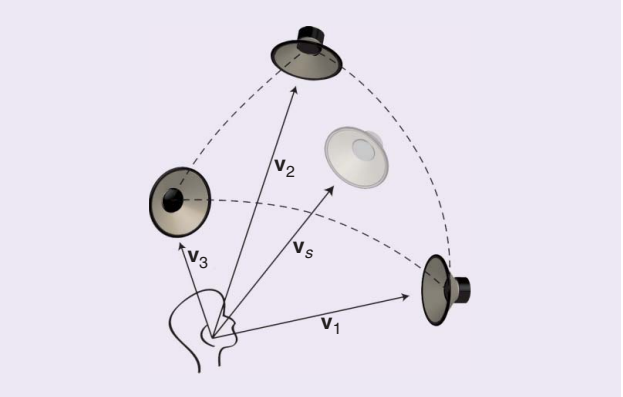


FIGURE 3. An arrangement of three loudspeakers and a phantom image panned using VBAP. The vectors used in the formulation of VBAP are also shown.

Spatial encoding methods

A class of multichannel audio methods involves dividing recorded signals into time or time–frequency bins and estimating certain spatial attributes within each bin. One of these methods is the spatial impulse response rendering (SIRR) method [48], [49]. At the recording stage, SIRR records the impulse response of a room using a B-format microphone, i.e., a microphone that provides the omnidirectional sound pressure component as well as the three axial pressure-gradient components of the sound field [28]. The impulse response is first transformed into a time–frequency representation and is then processed to obtain estimates of the acoustic intensity vectors at each time–frequency bin. It is assumed that each time–frequency bin corresponds to a single plane wave and thus that the direction of the acoustic intensity vector also represents the direction of that plane wave. A diffuseness estimate is obtained for each time–frequency bin using the ratio of the real part of the acoustic intensity to the total energy. These parameters, along with the sound pressure component obtained from the B-format recording, form the basis of the reproduction stage.

At the reproduction stage, direct and diffuse parts of the signal are treated differently. For the direct part, azimuth and elevation estimates in each time–frequency bin are used to pan portions of the B-format omnidirectional component, accordingly using VBAP. The diffuse part is reproduced by generating multiple decorrelated copies of the recorded sound played back from all the loudspeakers. The so-obtained channel impulse responses are then convolved with the desired anechoic sound sample. A similar method, called the *spatial decomposition method* (*SDM*), was recently proposed in [50].

Instead of using a time–frequency representation of the room impulse response, SDM simply divides it into time frames. As with SIRR, SDM assumes that within each time frame there is at most a single acoustic event (e.g., a reflection from the room walls), the direction of which is calculated using available direction-of-arrival estimation algorithms. Using this estimate, each time frame of the impulse response is panned between the loudspeakers using VBAP. The loudspeaker signals are then convolved with the desired anechoic sound sample.

Notice that, as opposed to SIRR, SDM does not explicitly differentiate direct and diffuse components. However, the later part of the room impulse response is still rendered as diffuse. This is due to the fact that, as time passes, a progressively larger number of echoes appears within each time frame, and, as a consequence, the direction-of-arrival algorithm tends to provide random estimates. In a formal listening experiment using synthesized room impulse responses, SDM was shown to outperform SIRR [50].

SIRR and SDM are not designed for continuous signals but for spatial room impulse responses, which are then convolved with an anechoic signal. In other words, they cannot be used for actual recordings of dynamic sound scenes. DirAC is a flexible spatial audio system for recording, coding, compression, transmission, and reproduction based on SIRR that overcomes this limitation [51]. As with SIRR, DirAC starts with an energy analysis of the recorded sound to assign a direction and a diffuseness level to each instant of the output channels of a filter bank that approximates the equivalent rectangular bandwidth scale. The direction predictions are then smoothed to imitate the temporal resolution of the auditory system. At the reproduction stage, these components are panned using VBAP. Figure 4 shows DirAC's recording, processing, and reproduction stages.

DirAC was evaluated and compared with Ambisonics (with different decoders) for reproduction quality using listening tests similar to multiple stimulus with hidden reference and anchor (also known by the acronym *MUSHRA*) [52]. The evaluation included different loudspeaker rigs (with four, five, eight, 12, and 16 loudspeakers), different audio material (music, speech, singing voice, and percussion), different simulated reverberation characteristics, and different listener positions. It was found that DirAC provides an excellent reproduction quality (better than an average of 80 or more out of 100) for the central listening position, and acceptable reproduction quality (better than an average of 60 or more out of 100) for the off-center positions. Ambisonics reproductions obtained using both decoders were rated consistently below DirAC. These results provide an instructive example of a perceptually motivated reproduction method achieving better subjective performance than a physically motivated approach.

Perceptual sound field reconstruction

PSR [29], [53] is a recently developed flexible multichannel recording, reproduction, and synthesis technology. As with DirAC, it provides a systematic framework for recording and reproduction of sound scenes. However, in contrast to DirAC, which performs panning of individual time–frequency components and renders the diffuse sound field via all channels relying on extensive processing of microphone array recordings to extract the necessary directional information and components, PSR relies on designing the underlying microphone arrays in a way that captures the required directional cues. When the recorded signals are played back with no additional processing, the directions of the wave fronts of all sound sources and all reflections are rendered accurately. A block diagram of a five-channel PSR system with uniform distribution of channels is shown in Figure 5. Another difference between DirAC and PSR is that, while DirAC uses only ICLDs for rendering auditory perspective, PSR employs both ICLDs and ICTDs and allows for trading one for the other while designing the directivity patterns of the microphones used in the arrays.

PSR uses near-coincident circular microphone arrays to capture the time differences between channels. The difference in the time of arrival of a sound wave propagating from a direction θ between microphones at angles ϕ_l and ϕ_{l+1} (see Figure 6) is

$$\tau_l(\theta) = \frac{2r_a}{c}\sin\left(\frac{\phi_{l+1}-\phi_l}{2}\right)\sin\left(\frac{\phi_{l+1}+\phi_l}{2}-\theta\right),$$

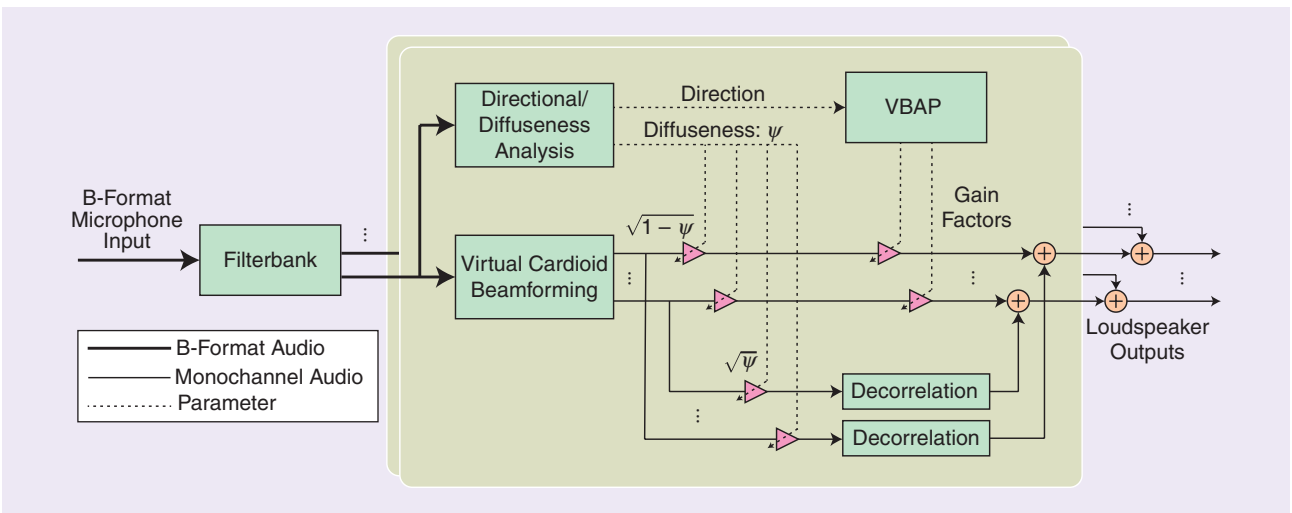


FIGURE 4. A DirAC flow diagram indicating different stages of analysis. (Image adapted from [51]. Figure used courtesy of the Audio Engineering Society.)

where r_a is the radius of the microphone array and c is the speed of sound. Microphone directivity patterns $\Gamma_l(\theta)$ are designed such that the level differences with which the sound wave is recorded are equal to the level difference that, in combination with the time differences as given in the above, creates a perception of the sound source in the direction θ. One way to achieve this is by designing $\Gamma_l(\theta)$ to satisfy the following relationship:

$$\frac{\Gamma_{l+1}(\theta)}{\Gamma_l(\theta)} = \frac{\sin(\theta - (\phi_l - \beta))}{\sin((\phi_{l+1} + \beta) - \theta)}, \quad (6)$$

where β is selected in such a way that for θ, which coincides with the direction of one of the microphones, the level difference is equal to the level difference needed to create the perception of the sound wave in the direction of the corresponding loudspeaker. That level difference is labeled *EL* (or *FL* with the reversed sign) in Figure 1(b) for the case where the maximal ICTD is 0.6 ms. Thus, as the direction of the sound source moves between the two microphones, the captured time and level differences traverse a curve connecting two end points, illustrated by the straight line between points *EL* and *FL* in Figure 1(b), which correspond to virtual sources in the directions of two corresponding loudspeakers. Microphones are additionally required to satisfy the constant power condition and $|\Gamma_l(\theta)|^2 + |\Gamma_{l+1}(\theta)|^2 = 1, \phi_l \le \theta \le \phi_{l+1}$ and sufficiently high attenuation outside the sector between the axes of the two adjacent microphones, so that every sound wave is effectively recorded and rendered only by the pair of the two closest channels.

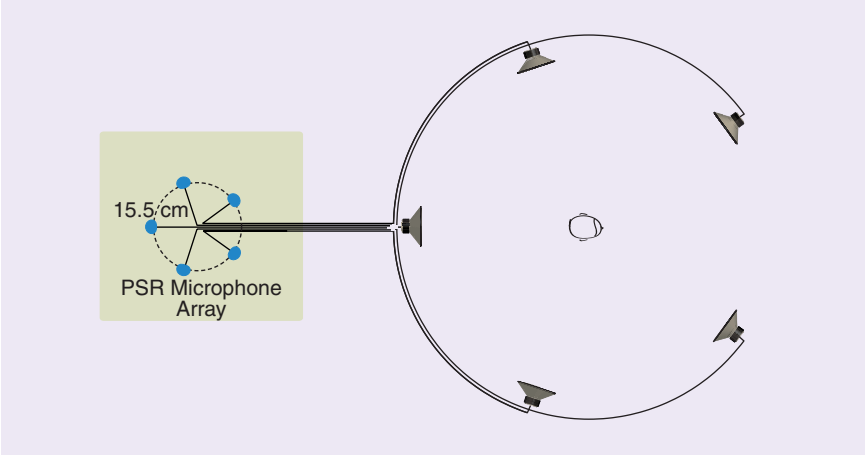


FIGURE 5. A block diagram of a PSR system indicating a one-to-one correspondence between the microphone and loudspeaker channels and no intermediate processing. (Figure used courtesy of [29].)

The degree to which time differences are present is controlled by the radius of the microphone array, which is present implicitly in the β factor in (6). In the special case where $\beta = 0$, the system implements intensity stereophony based on the tangent panning law. An example of a directivity pattern designed according to PSR principles for a five-channel, uniformly spaced system for an array of with $r_a = 15$ cm is shown in Figure 7, along with its second-order approximation and the polar pattern that corresponds to $\beta = 0$, which is the pattern designed for intensity stereophony according to the tangent law.

The five-channel PSR system design based on intensity and time–intensity principles as specified earlier in this section was subjectively evaluated and compared with second-order Ambisonics in terms of subjective localization accuracy [29]. Figure 8 shows the results of a localization test carried out using different recording/reproduction systems. The time–intensity PSR technology performed well, especially at off-center listening positions, while it performed worse for localization at lateral source directions, which is due to the fact that psychoacoustic curves for frontal presentation were used for all the pairs of loudspeakers. Another set of tests considered the localtedness of generated phantom sources and showed that time–intensity-based PSR provides better localtedness of phantom sources than techniques based on intensity alone (shown in the bar charts in Figure 8), which is attributed to the higher naturalness of the presented binaural cues [54].

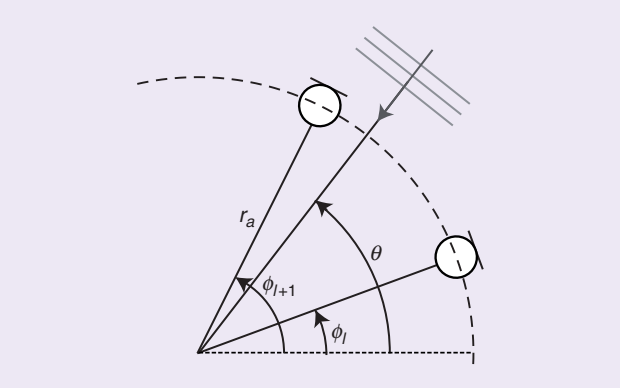


FIGURE 6. Two neighboring elements of the microphone array used in the PSR recording system. (Figure used courtesy of [29].)

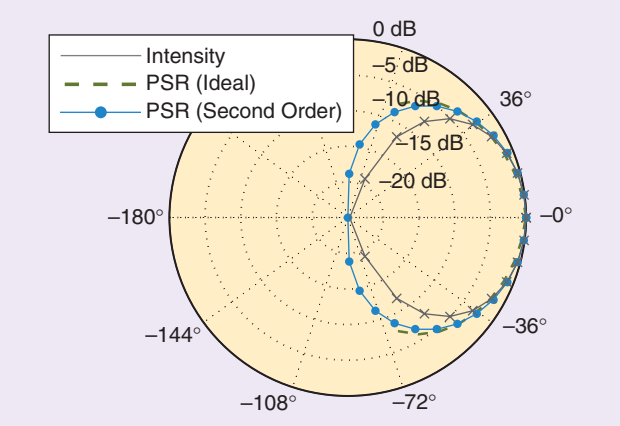


FIGURE 7. The directivity patterns of the PSR and intensity methods. Also shown is the directivity pattern of a second-order implementation of the ideal PSR directivity used in subjective evaluations. (Figure used courtesy of [29].)

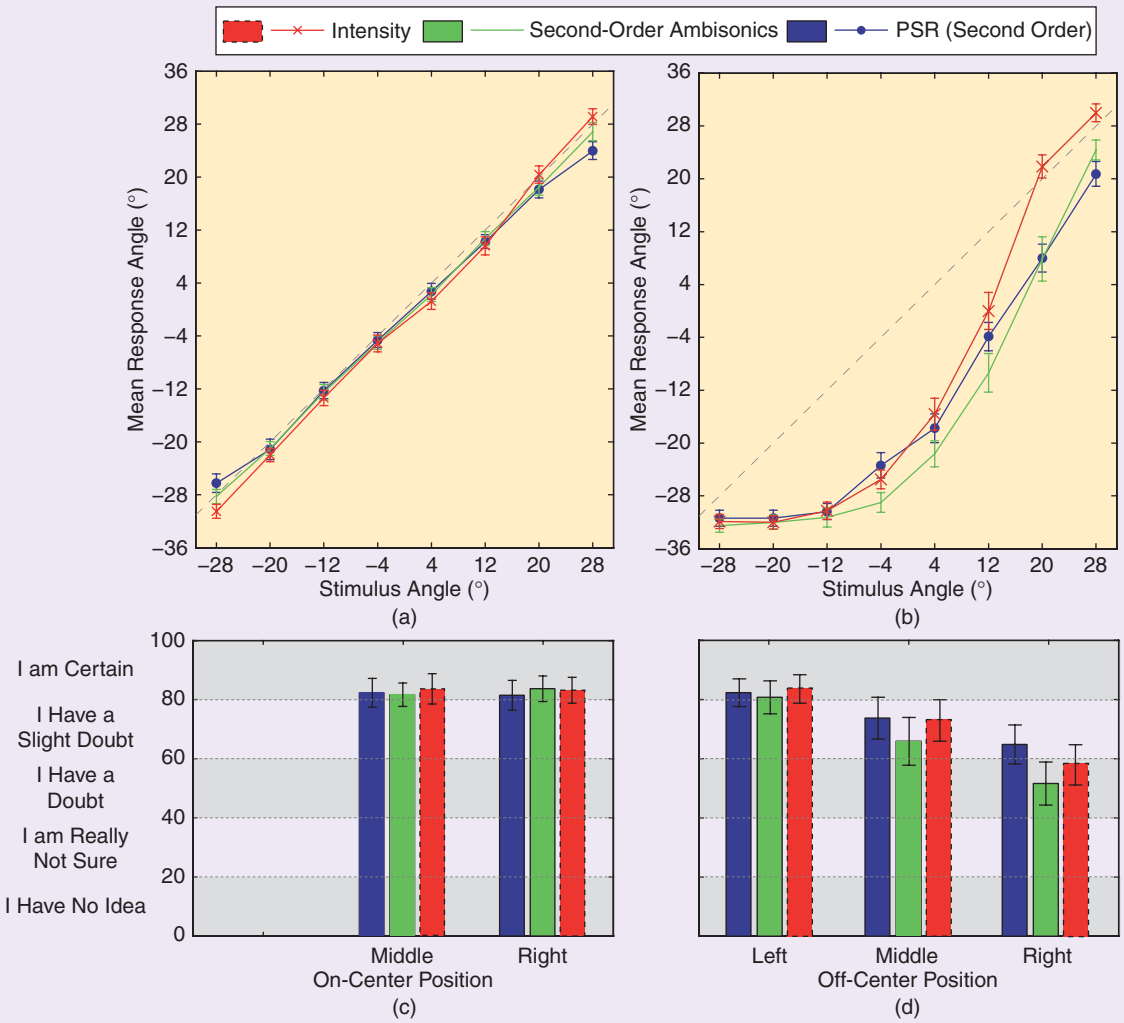


FIGURE 8. The results of a subjective listening experiment with the second-order Ambisonics, PSR, and intensity methods: (a) and (b) indicate localization performance, while (c) and (d) indicate locatedness performance. (a) and (c) are for a listener in the center of the listening area facing the midline between two loudspeakers (see Figure 5) and facing a loudspeaker, respectively. (b) and (d) are for an off-center listener—more precisely, 20 cm behind and to the left with respect to the on-center position. (Figure used courtesy of [29]).

The issue of intensity versus time–intensity techniques is a matter of debate among audio engineers and recording artists, and the widely held view is that although time–intensity stereophony provides more natural-sounding sources, intensity stereophony provides more stable imaging. This result provides a new insight into the issue and demonstrates that time–intensity techniques, if designed with a careful consideration of the underlying psychoacoustical requirements, are capable of actually providing a stable auditory perspective. The development of techniques for higher-order differential microphone arrays [55], [56] enabled the design of more sophisticated directivity patterns than those achievable by commonly used first-order microphones. This allowed the implementation of different panning laws and psychoacoustical panning functions in the multichannel microphone array design process.

Enlarging the optimal listening area

When a listener moves away from the center of the sweet spot, the auditory event shifts in the direction of the closest loudspeaker. This is due to the fact that the signal from the closest loudspeaker arrives earlier when compared with what is observed at the sweet spot. Position-independent stereo [57], [58] aims to alleviate this problem by designing loudspeaker directivity patterns in a manner that compensates for the incongruent time delay via appropriate intensity differences.

The design method proposed for this purpose by Rodenas et al. [59] consists of two separate optimization procedures. The first procedure involves finding a common directivity pattern for the left and right loudspeakers of a stereophonic setup to provide the level differences needed to compensate for the incongruent time differences over a desired listening area. Such directivity patterns can be obtained by beamforming using an array of

loudspeakers. The other optimization procedure involves finding the filter coefficients to be used for beamforming.

Figure 9 illustrates the problem of off-center listening and loudspeaker directivity patterns implemented to counteract this problem using a loudspeaker array with two drivers. It may be observed that when the listener moves to the left of the ideal listening position, the signal from the left loudspeaker will arrive earlier and at a higher level than the right loudspeaker, shifting the perceived direction of the virtual source toward the left. This problem can be compensated for by adjusting the right loudspeaker to have a higher level than the left loudspeaker at the corresponding direction, effectively shifting the virtual source back. The loudspeaker directivity patterns shown in the figure were designed to achieve this and thus to allow the listening area to be enlarged. Rodenas et al. [59] report results of an informal listening test with a system realized using loudspeaker arrays with two tweeter and two midrange drivers each and state that the proposed approach widens the sweet spot for standard stereophonic material.

While designing loudspeaker directivity patterns for robust stereophony is a promising idea, technical difficulties such as the equalization of drivers, compensation for diffraction from the edges of the loudspeaker cabinets, and the required number of loudspeakers used in the design may limit its practical use. These techniques, along with the generalization of the design approach to multichannel systems and combination with other technologies such as PSR, are interesting directions for future research.

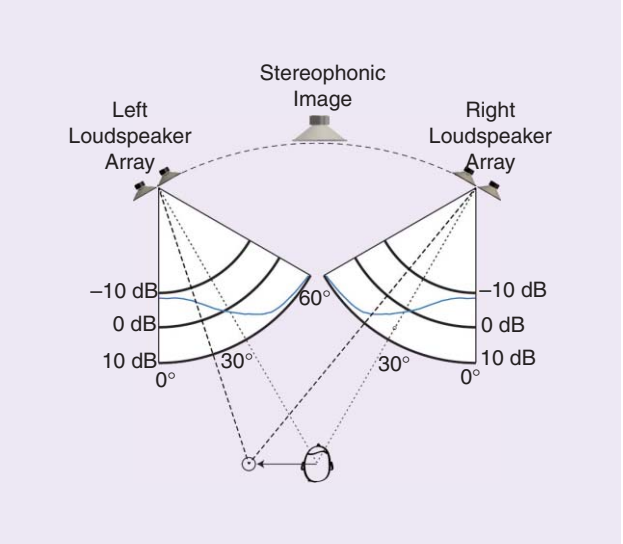


FIGURE 9. A conceptual depiction of position-independent stereo with optimal loudspeaker directivity patterns. When the listener moves toward the left or right of the stereophonic sweet spot, the directivity pattern of the loudspeaker array compensates for the resulting delay, and the direction of the virtual source remains unchanged. (Figure adapted from [59]. Reprinted with permission. Copyright 2003, Acoustical Society of America.)

Perceptually motivated room auralization

In cases where an acoustic scene actually exists, as in a live concert or a tennis match, the scene is recorded and reproduced by the techniques reviewed above. These techniques also capture the acoustics of the environment in which these recordings are made. However, there are applications where such scenes exist only virtually, as in computer games or VR. In such cases, the acoustics of the environment that contains the scene to be rendered need to be synthesized. Auralization is the process of making the acoustics of a real or virtual environment such as a room or a concert hall audible [8].

Rooms are multipath environments where the recording of a source by a microphone will include not only the direct path but also early reflections, reverberation tail, and diffraction components. Many different models have been proposed in the past 50 years to simulate room acoustics. A recent review article provides a summary of the research on room acoustics modeling [60] and divides algorithms into three classes: 1) convolutional algorithms, 2) delay networks, and 3) computational acoustics models. Convolutional algorithms involve measuring the impulse response of an actual room and convolving it with a desired input signal. Delay networks, which will be discussed in more detail in the "Perceptually Motivated Artificial Reverberation" section, are algorithms where the input is filtered and fed back along a number of delay paths designed according to desired reverberation characteristics. Computational acoustics models aim to simulate the propagation of sound waves in the modeled space.

Among computational acoustics models, there are geometric models, which use geometric arguments to calculate the room impulse response. These include the image-source method (ISM) [61], [62] and ray tracing [63] or beam tracing [64] and its variants [65]. Other computational acoustics models such as finite-difference methods [66], digital waveguide mesh [67], FEMs [68], and BEMs [69] are based on the time- and space-discretized solutions of the wave equation; hence, individual reflections are not rendered explicitly, but their effects are merged into the overall simulated wave fields. Computational acoustics models are capable of providing very accurate results (at least for certain frequency ranges) and are therefore used in architectural acoustics. However, their physical accuracy comes at a very high computational cost. While some computation can be carried out offline, auralization will typically require real-time operation at interactive rates—for instance, to allow a user to explore a virtual environment. The main computational bottleneck that this entails is associated with the different filtering operations involved in calculating and synthesizing reflections and edge diffraction components for each source.

Despite their high computational complexity, extremely accurate room auralization methods will always be in demand for applications such as architectural acoustics. However, they are not suitable for applications such as immersive games and VR, where a low computational cost is paramount. Such applications warrant the simplification of the model to the lowest possible number of components and sources to be rendered, which is typically achieved by removing perceptually irrelevant content.

Simplification of room acoustics models

The lack of a comprehensive mathematical model of the precedence effect, analogous to models of monaural masking, has,

for a long time, made it difficult to predict whether an individual reflection would be audible in the presence of the direct sound and other reflections. This is mainly due to the fact that the audibility of a reflection depends on many parameters.

One of the first models that aimed to parameterize the audibility of reflections, named *reflection masked threshold* (*RMT*), was proposed by Buchholz et al. [70]. The RMT is the lowest level at which a reflection will be audible, and it is a function of the directions of the reflection and of the corresponding direct sound, the time delay of the reflection with respect to the direct sound, the level of the direct sound, the difference of the frequency spectra between the direct sound and the reflection, the effect of other reflections and reverberation, and the signal content. The RMT can be used for simplifying room acoustic models via culling inaudible reflections. A simpler decision rule for culling inaudible early reflections was proposed by Begault et al. [71]–[73], based on the relative level of the reflection. In the absence of reverberation, the audibility threshold of a reflection is 21 dB below the level of direct sound for a delay of 3 ms. The presence of diffuse reverberation has the effect of increasing this threshold by 11 dB. This threshold is also known to decrease with the angle between the direct sound and the early reflection.

Some properties of binaural hearing, such as the precedence effect, may also make some reflections inaudible. The exclusion of those reflections from the audio rendering pipeline can further reduce the associated computational cost. To that end, a model of the precedence effect was proposed in [74], according to which perceived directions of acoustic events are modeled as normally distributed variables. If the direct path and a reflection are present, then the distribution of the perceived direction is a mixture of two Gaussians. The audibility of the reflection was then shown to be related to the number of modes in the mixture: if the mixture is unimodal, the reflection is masked, and if it is bimodal, it is audible. The derivation of the model parameters was made via subjective localization experiments. This model was applied for the culling of reflections in binaural room auralization [75]. More specifically, the ISM was used to obtain a number of secondary sources, and these were clustered according to their distance from the listener position and their azimuth angle. A single reflection masker was obtained for each cluster using the precedence effect model, and the rest of the secondary sources in the same cluster are excluded from the rendering pipeline, thereby reducing the computational cost. Subjective evaluations were carried out using different audio material, different room geometries, and different listening positions to compare the room auralizations using full-room response, level-based reflection selection, and perceptually motivated selection based on the precedence effect model. These experiments showed that reflection culling based on the precedence effect is capable of reducing the number of early reflections by over 60%, without any significant degradation in subjective localization, spaciousness, presence, and envelopment experience.

Another approach to perceptually motivated simplification of auralization based on the absolute threshold of hearing was recently proposed [76]. According to this model, the duration of ray tracing for calculating the room impulse responses for a given source depends on a temporal cutoff point determined by the last audible ray. It was shown that this approach resulted in noticeable improvements in the computation time of impulse responses without significantly degrading the auditory experience.

Perceptually motivated artificial reverberation

Room impulse responses can be divided in two parts: early reflections, where reflections are separated in time and have strong directional characteristics, and the reverberation tail, where higher-order reflections begin to overlap in time and the sound field becomes diffuse (i.e., omnidirectional). The human auditory system is sensitive to the direction of the direct wave front and the early reflections, while it cannot discern the directions of individual reflections within the reverberation tail [77]. The level and directions of lateral early reflections are related directly to the perception of the width of a sound source and the spatial impression of an enclosure [78].

As the density of reflections increases, statistical properties like reflection density and decay slope become more important than the fine temporal structure. In real enclosures, sound energy decays exponentially, and the point at which the total energy of the room impulse response drops 60 dB below its initial value is called the *reverberation time* [79]. The reverberation time has a strong influence on how spacious an enclosure is perceived to be [77]. Other quantities that have a strong influence on the perceived quality of reverberation include the density of the individual reflections in the late reverberation tail, called the *reflection density* [79]; the time-dependent profile of the reflection density, called the *echo density profile* [80]; and the number of damped resonant frequencies per Hertz, called the *mode density* [81]. The typical objective of perceptually motivated artificial reverberators is to accurately render the reverberation properties described previously.

Since the early part and the reverberation tail are perceived differently, a common approach is to model and render them separately in a typical room auralization algorithm. For the reverberation tail, a statistically compatible model is usually acceptable, due to the fact that the human auditory system is not sensitive to its fine structure. Figure 10 shows the diagram of a typical binaural auralization system. Here, one module simulates and renders binaurally the direct path and a number of early reflections, while an artificial reverberator unit renders the reverberation tail. In this context, we refer to an artificial reverberator as a room acoustic model (typically a delay network) that aims only at reconstructing important perceptual features of room reverberation with little regard to its physical accuracy. By targeting only the perceptual aspects of room reverberation, vast reductions in computational complexity are possible.

Various room auralization systems have been developed in the past 20 years [10], [82], [83]. The Digital Interactive Virtual Acoustics system [10], one of the first parametric interactive room auralization systems, simulates all the first- and second-order reflections and synthesizes them binaurally or for rendition over loudspeakers. It is capable of simulating the absorption characteristics of different wall materials, air absorption, and source directivity. Late reverberation is provided via an artificial

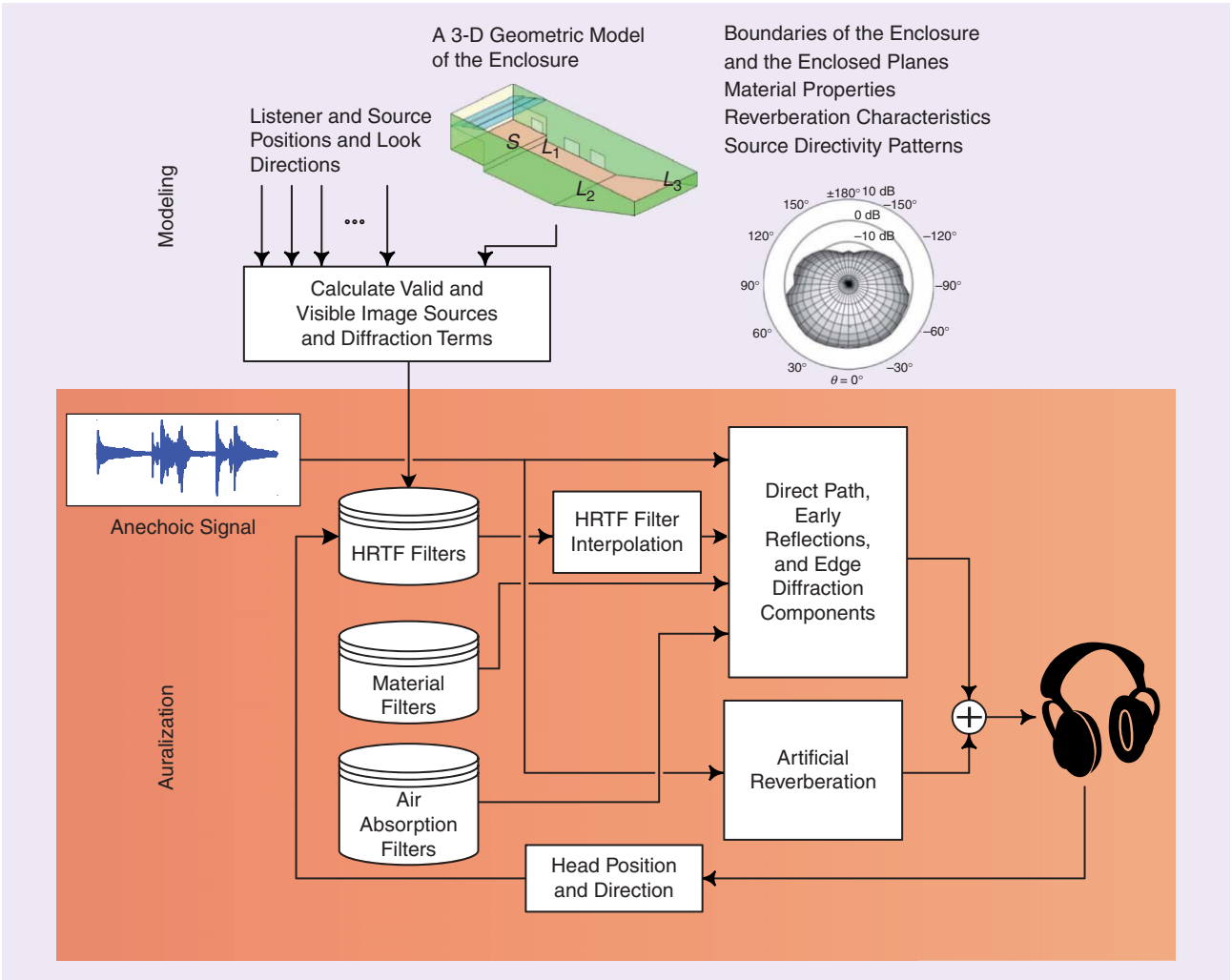


FIGURE 10. Typical processing stages in a binaural room auralization system using the image-source method to model the room acoustics.

reverberator consisting of a recursive structure using comb and all-pass filters.

The choice of artificial reverberation in a room auralization system is dictated not only by perceptual considerations but also by computational cost, and the holy grail in artificial reverberator design is an algorithm that can achieve good perceptual quality at a reasonable computational cost. The earliest digital artificial reverberators were proposed by Schroeder in the 1960s and consisted of comb filters connected in parallel to simulate the frequency modes of a room and all-pass filters to simulate a dense reverberation tail [84]. The original designs by Schroeder sometimes produced a metallic-sounding reverberation, and various improvements were subsequently proposed [9], [85]. These improvements, however, did not provide any means to explicitly or easily control the characteristics of the synthesized reverberation.

Feedback delay networks (FDNs) were developed as a multichannel extension of the Schroeder reverberator [87], [88]. The FDN is a recursive delay network that can generate reverberation for a number of input channels, such as individual audio channels of a four-channel (i.e., quadrophonic) system. Each of the input channels is delayed, fed back recursively through a feedback loop, attenuated, and mixed with the incoming channels. The delay lines are designed to have incommensurate lengths, and the feedback loop consists of multiplication with a unitary matrix.

Jot and Chaigne extended the FDN design and proposed a simple and structured procedure to design good-quality reverberators with a desired frequency-dependent reverberation time [86]. They also introduced the design principle that to avoid isolated ringing modes that tend to sound metallic, all the structure modes should decay at the same rate. A conceptual block diagram of Jot's reverberator is shown in Figure 11. Notice the absorption filters in the feedforward path that allow controlling the decay rate at different frequencies, and a tonal correction filter that is used to equalize the reverberator frequency response so that the generated reverberation sounds more natural. The original design uses a Householder matrix for the feedback path, but other unitary matrices can also be used [89]. These matrices can also be time varying, resulting in improved perceptual characteristics [90].

Equivalent to a wide class of FDNs are the digital waveguide networks (DWNs) [91], [98]. A DWN consists of a number of

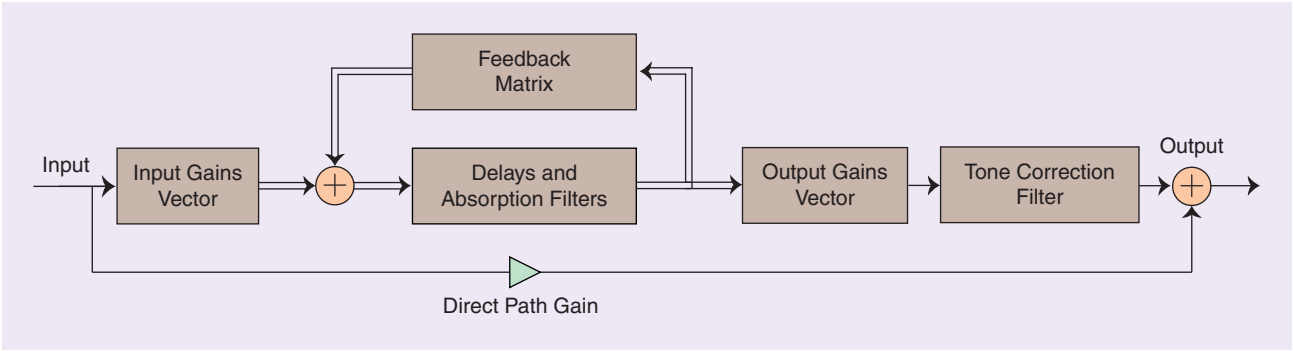


FIGURE 11. A conceptual block diagram of the modified FDN reverberator as proposed by Jot and Chaigne [86].

digital waveguides (the digital equivalent of analog propagation lines, formed of two opposite delay lines of equal length) connected at lossless scattering junctions (see Figure 12). Each scattering junction carries out a simple matrix multiplication to scatter the incoming signals on digital waveguides from each of its neighbors to generate outgoing signals to be distributed back to the same digital waveguides in the opposite direction. A signal reverberated using a DWN can be obtained by summing all the outgoing signals of one of the scattering nodes. DWNs have appealing stability properties and significant design flexibility owing to the different possible network graphs, types of lossless scattering, and lengths of the digital waveguides. While both Jot's reverberator and DWNs are capable of producing responses with a high perceptual quality, the parameters of these models are not explicitly linked to the physical characteristics of a particular room.

Artificial reverberators that are more tightly linked to room acoustics also exist. One of the earlier designs proposed by Kendall et al. [92] was based on recirculating delay elements whose lengths were determined by using an image-source model of a rectangular room. A similar approach was also used in [93]. Karjalainen et al. proposed a class of DWNs designed to simulate early reflections and axial modes of rectangular rooms [94]. A drawback of their algorithm is that many of the algorithm's internal parameters still require hand tuning to achieve a satisfactory reverberation.

An artificial reverberator that inherits all its parameters from the physical characteristics of the room it simulates was recently proposed [12]. This reverberator, termed the *scattering delay network* (*SDN*), is a modified DWN where the length of the digital waveguides and the topology of the network, as illustrated conceptually in Figure 13, are derived directly from the geometry of the simulated space. In particular, the SDN is a minimal network connecting as many scattering nodes as there are walls in the room, and where each scattering node is positioned at the point where first-order reflections impinge on the wall.

This design ensures that first-order reflections are rendered exactly, while second- and higher-order reflections are simulated

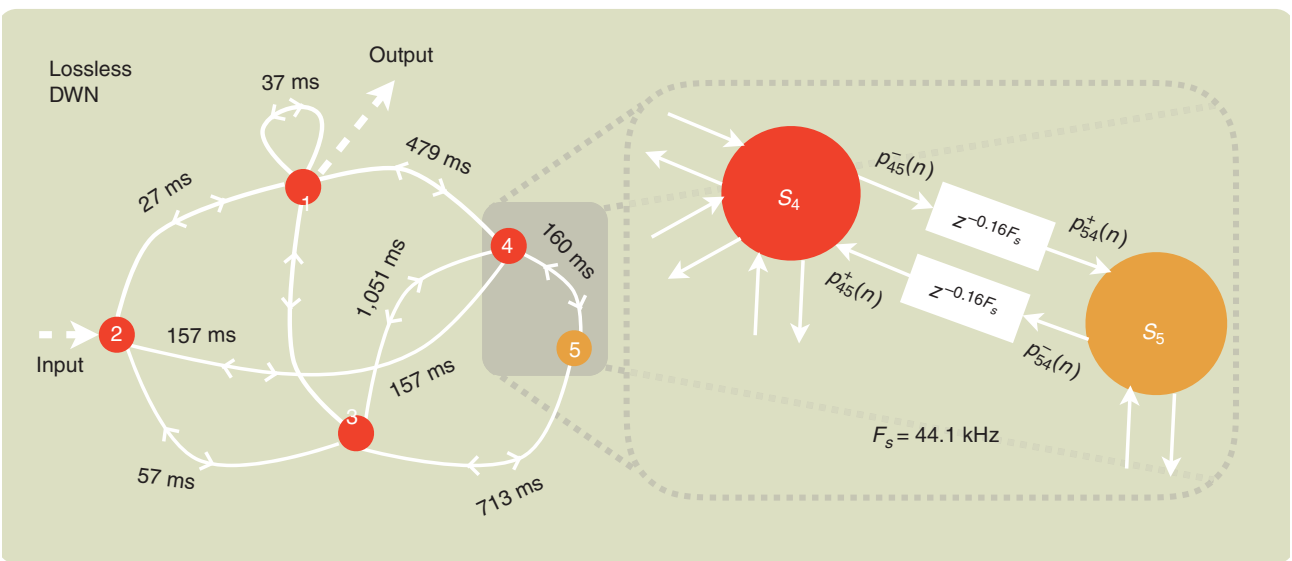


FIGURE 12. A conceptual depiction of a DWN. The figure shows a DWN with five nodes (indicated as circles) connected via bidirectional delay lines (curves with double arrows). To maximize the reflection density, the delay lengths are chosen to be coprime numbers. The input signal can be fed and the output can be obtained from any node. The inset shows the connection between two nodes where incoming and outgoing signals and the individual delay elements are clearly visible. (Figure adapted from [91].)

with a gradually diminishing accuracy. Since first-order early reflections are perceptually more important than the higher-order reflections, the resulting reverberation is perceptually realistic and statistically very similar to that of an actual room. A by-product of this design is that the SDN does not require separate modules for early reflections and late reverberation, while still allowing precise and explicit control of the room geometry, source and receiver directivity patterns, and wall absorption characteristics. Furthermore, it enables a straightforward implementation of virtual multichannel recordings and binaural auralization. A block diagram of the SDN reverberator is depicted in Figure 14, showing that it has a structure very similar to the FDN, while allowing direct and explicit control over all the physical characteristics of the space it simulates.

Audio source culling

Complex virtual environments typically include many sound sources, which makes synthesizing their acoustics a challenging task in terms of the associated computational cost. This difficulty is especially pronounced when rendering such audio content over devices with limited computational power, such as mobile phones. In a typical scenario involving many concurrent sources, it may be necessary to select and render only a few.

State-of-the-art game engines typically use volumetric culling of sound sources. Each sound source has an associated culling volume (cube, sphere, or cylinder), and when the listener is within this volume, the sound is rendered. This is a simple approach that does not incur any significant computational cost, apart from the relatively simple comparison operations between the bounding boxes of the listener and each of the sound sources. However, this approach does not take into account the relative levels of the sound sources. It also does not limit the number of sound sources that can be simultaneously active. This proximity makes the only determinant of whether or not a sound source will be rendered, completely disregarding its perceptual salience. Sound sources are dynamically activated in response to user-generated events in interactive applications such as games and VR. For scenes comprising multiple concurrent sound sources, many of these sources will be masked by the others. This makes it redundant to process these inaudible sources.

Tsingos et al. [95], [96] provide a perceptually based source-culling approach. The approach is based on ranking the sources

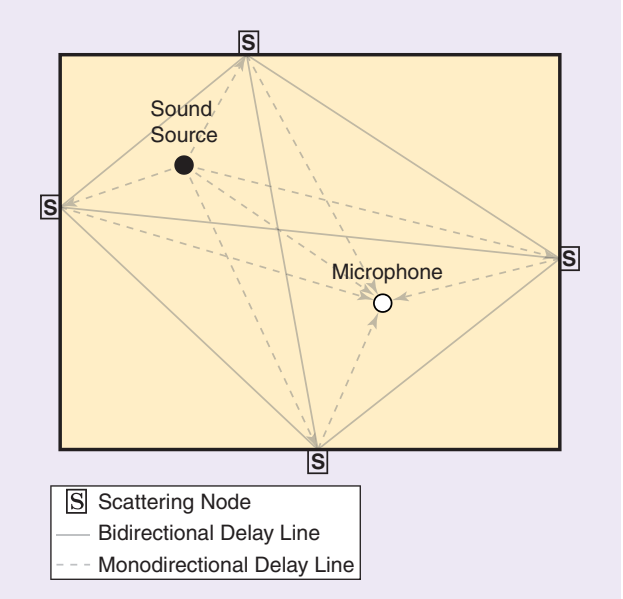


FIGURE 13. A conceptual depiction of the SDN reverberator. The figure shows a rectangular room as observed from above, with the associated delay lines interconnected at scattering junctions on the wall. Other delay lines associated to the floor and ceiling are also present but are not shown here for clarity. (Figure used courtesy of [12].)

in the scene using their binaural loudness at different frequency bands as a measure of perceptual salience. Loudness values are used to calculate a masking threshold from a time–frequency representation of the sound sources and stored for use during runtime. As a new sound event occurs, the decision to render the new sound source is made at the audio-frame level. Each frame is compared with the existing mix for evaluating whether the mix can mask it. If it can, the frame is culled. As a result of the frame-level temporal resolution, several frames from a single sound source can be culled while others are rendered. This results in each sound source being only partially culled. A similar algorithm was proposed by Metan and Hacıhabiboğlu [97]. The audibility calculation in this algorithm is slightly different from that of Tsingos et al. As a new sound event is generated, a look-ahead algorithm checks for the audibility of each frame of a sound source, given the current mix being played. The decision to render a sound source is based on the ratio of audible

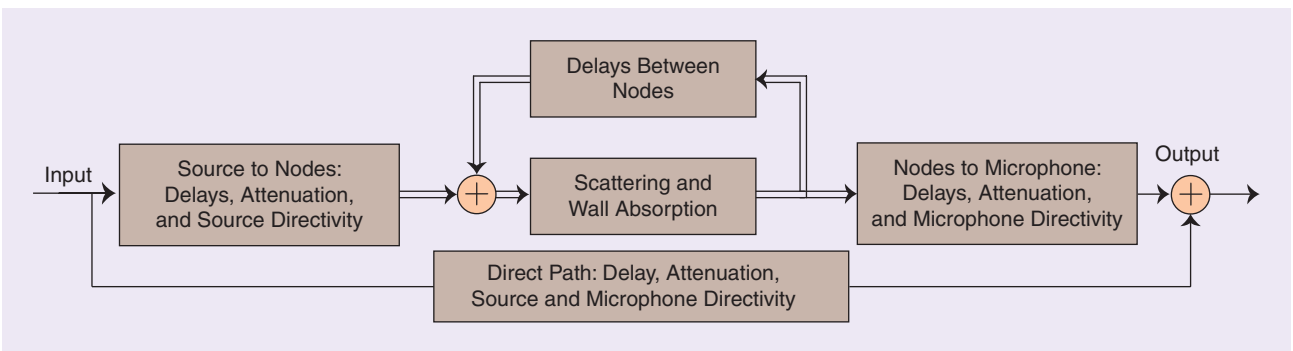


FIGURE 14. A conceptual block diagram of the SDN reverberator.

frames to the total number of frames in the audio signal to be rendered. This way, the whole source and not a portion of it is rendered or culled. The advantage of these methods is that they potentially allow the preprocessing stage of the source-culling process to be integrated with existing perceptual audio-coding algorithms such as MPEG-1 Layer I Audio.

Summary

The body of knowledge on spatial hearing and the mechanisms that govern it has been steadily growing. However, a comprehensive model that can account for all the different aspects of spatial hearing is yet to be developed. Still, the existing knowledge can be used to design audio systems and algorithms that have lower computational and hardware costs but can provide a subjective performance as good as more complicated physically motivated systems.

While developments in computer hardware could make it possible to overcome issues due to computational limitations, physical limitations such as the size of electroacoustic transducers or data bandwidth will remain. Similarly, the energy cost of carrying out simple operations such as multiplication or memory access is likely to diminish but will never vanish, and the power efficiency of mobile devices will also continue to be relevant. These issues will make it even more desirable to design simpler audio systems and algorithms. The importance of using our knowledge of auditory perception to that end will thus remain high.

Acknowledgments

This work was supported by the Turkish Scientific and Technological Research Council under research grant 113E513, "Spatial Audio Reproduction Using Analysis-Based Synthesis Methods," and by the European Commission under grant 316969 in the FP7-PEOPLE Marie Curie Initial Training Network, "Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS)."

Authors

Hüseyin Hacıhabiboğlu (hhuseyin@metu.edu.tr) received his B.Sc. (honors) degree from the Middle East Technical University (METU), Ankara, Turkey, in 2000 and his M.Sc. degree from the University of Bristol, United Kingdom, in 2001, both in electrical and electronics engineering, and his Ph.D. degree in computer science from Queen's University Belfast, United Kingdom, in 2004. He held research positions at the University of Surrey, Guildford, United Kingdom, (2004–2008) and King's College London, United Kingdom (2008–2011). Currently, he is an associate professor of signal processing and head of the Department of Modeling and Simulation in the Graduate School of Informatics, METU. He also coordinates the recently established multimedia informatics graduate program in the same department. His research interests include audio signal processing, room acoustics, multichannel audio systems, psychoacoustics of spatial hearing, microphone arrays, and game audio. He is a member of the IEEE Signal Processing Society, Audio Engineering Society, Turkish Acoustics Society, and European Acoustics Association. He is a Senior Member of the IEEE and an associate editor of *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.

Enzo De Sena (e.desena@surrey.ac.uk) received his B.Sc. degree in 2007 and his M.Sc. degree cum laude in 2009, both from the Universita degli Studi di Napoli Federico II in telecommunication engineering. In 2013, he received his Ph.D. degree in electronic engineering from King's College London, United Kingdom. Between 2013 and 2016, he was a postdoctoral research fellow at the Katholieke Universiteit Leuven, Belgium. Since September 2016, he has been a lecturer in audio at the Institute of Sound Recording of the University of Surrey, Guildford, United Kingdom. He held visiting positions at the Center for Computer Research in Music and Acoustics at Stanford University, California (2013), in the Signal and Information Processing section at Aalborg University, Denmark (2014–2015), and in the Speech and Audio Processing Group at Imperial College London, United Kingdom (2016). He is a former Marie Curie fellow, and his current research interests include room acoustics modeling, multichannel audio systems, microphone beamforming, and binaural modeling. He is a Member of the IEEE.

Zoran Cvetković (zoran.cvetkovic@kcl.ac.uk) is a professor of signal processing at King's College London, United Kingdom. He received his Dipl. Ing. and Mag. degrees from the University of Belgrade, Yugoslavia; his M.Phil. degree from Columbia University, New York; and his Ph.D. degree in electrical engineering from the University of California, Berkeley. He held research positions at EPFL (1996), Lausanne, Switzerland, and at Harvard University (2002–2004), Cambridge, Massachusetts. Between 1997 and 2002, he was a member of the technical staff of AT&T Shannon Laboratory. His research interests are in the broad area of signal processing, ranging from the theoretical aspects of signal analysis to applications in neuroscience and audio and speech technology. From 2005 to 2008, he served as an associate editor of *IEEE Transactions on Signal Processing*. He is a Senior Member of the IEEE.

James Johnston (j.d.johnston@ieee.org) received his B.S. and M.S. degrees in electrical engineering from Carnegie-Mellon University, Pittsburgh, Pennsylvania. He was the primary researcher and inventor of the MPEG-2 AAC audio coding algorithm, and a principal contributor to the MP3 algorithm. He was awarded the IEEE James L. Flanagan Signal Processing Field Award (2006), elected a fellow of the Audio Engineering Society (1997), received the AT&T Technology Medal and AT&T Standards Award (1998), and received a New Jersey Inventor of the Year Award (2001). He was a Distinguished Lecturer of the IEEE Signal Processing Society in 2011. His current research interests include acoustic scene modeling, loudspeaker design, loudspeaker pattern control, cochlear modeling, masking threshold models, stereo imaging models and stereo imaging sensitivity models, methods of reproducing sound fields, both actively steered and time-invariant microphone and sound-field capture techniques, and speech and audio coding methods in general. He is a Fellow of the IEEE.

Julius O. Smith III (jos@ccrma.stanford.edu) received his B.S. degree from Rice University, Houston, Texas, in 1975 in electrical engineering, control, circuits, and communication. He received his M.S. and Ph.D. degrees in electrical engineering from Stanford University, California, in 1978 and 1983, respectively. From 1982 to 1986, he was with the Adaptive Systems Department at Systems Control Technology, Palo Alto, California, where he worked in the areas of adaptive filtering and spectral estimation. From 1986 to 1991, he was employed at NeXT Computer Inc., responsible for sound, music, and signal processing software for the NeXT computer workstation. He then became an associate professor at the Center for Computer Research in Music and Acoustics at Stanford University, teaching courses and pursuing research related to signal processing techniques applied to music and audio systems. He is currently a professor of music and (by courtesy) electrical engineering at Stanford University. He is a Member of the IEEE.

References

[1] H. Møller, "Fundamentals of binaural technology," *Appl. Acoust.*, vol. 36, no. 3, pp. 171– 218, 1992.

[2] J. Köring and A. Schmitz, "Simplifying cancellation of cross-talk for playback of head-related recordings in a two-speaker system," *Acta Acustica United with Acustica*, vol. 79, no. 3, pp. 221–232, 1993.

[3] M. Maazaoui and O. Warusfel. (2016, May). Estimation of individualized HRTF in unsupervised conditions. presented at Audio Engineering Society Conv. 140. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18219

[4] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Amer.*, vol. 93, no. 5, pp. 2764–2778, 1993.

[5] J. Daniel and S. Moreau, "Further study of sound field coding with higher order ambisonics," in *Proc. 116th Conv. Audio Engineering Society*, Berlin, Germany, May 2004, Preprint no. 6017.

[6] J. Ahrens, *Analytic Methods of Sound Field Synthesis*. Berlin: Springer-Verlag, 2012.

[7] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter, "Spatial sound with loudspeakers and its perception: A review of the current state," *Proc. IEEE*, vol. 101, no. 9, pp. 1920–1938, 2013.

[8] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Berlin: Springer-Verlag, 2007.

[9] W. G. Gardner, "Reverberation algorithms," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. New York: Springer-Verlag, 2002, pp. 85–131.

[10] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc.*, vol. 47, no. 9, pp. 675–705, 1999.

[11] T. Lokki, P. Svensson, and L. Savioja, "An efficient auralization of edge diffraction," in *Proc. Audio Engineering Society Conf.: 21st Int. Conf.: Architectural Acoustics and Sound Reinforcement*, June 2002. Paper no. 84.

[12] E. De Sena, H. Hacıhabiboğlu, Z. Cvetković, and J. O. Smith, "Efficient synthesis of room acoustics via scattering delay networks," *IEEE/Association for Computing Machinery Trans. Audio, Speech, Language Process.*, vol. 23, no. 9, pp. 1478–1492, Sept. 2015.

[13] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press, 1997.

[14] B. C. Moore, *Hearing*. New York: Academic, 1995.

[15] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: A summary of past and present research," *Acta Acustica United with Acustica*, vol. 91, no. 3, pp. 409–420, 2005.

[16] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," *J. Acoust. Soc. Amer.*, vol. 106, no. 4, pp. 1633–1654, 1999.

[17] A. Kulkarni and H. S. Colburn, "Infinite-impulse-response models of the head-related transfer function," *J. Acoust. Soc. Amer.*, vol. 115, no. 4, pp. 1714–1728, 2004.

[18] Y. Luo, D. N. Zotkin, H. Daumé, and R. Duraiswami, "Kernel regression for head-related transfer function interpolation and spectral extrema extraction," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'13)*, May 2013, pp. 256–260.

[19] R. Duraiswami and N. A. Gumerov, "Method for measurement of head related transfer functions," U.S. Patent 7,720,229, May, 18 2010.

[20] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *J. Audio Eng. Soc.*, vol. 49, no. 10, pp. 904–916, Oct. 2001.

[21] N. I. Durlach, A. Rigopulos, X. D. Pang, W. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel, "On the externalization of auditory images," *Presence: Teleoperators and Virtual Environments*, vol. 1, no. 2, pp. 251–257, 1992.

[22] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *J. Audio Eng. Soc.*, vol. 37, no. 1/2, pp. 3–19, 1989.

[23] W. G. Gardner, *3-D Audio Using Loudspeakers*. New York: Springer-Verlag, 1998.

[24] W. G. Gardner, "Spatial audio reproduction: Towards individualized binaural sound," in *Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2004 NAE Symposium on Frontiers of Engineering*. Washington, D.C.: National Academies Press, 2005.

[25] N. V. Franssen, *Stereophony*. Eindhoven, The Netherlands: Philips Technical Library, 1964. [Database]

[26] M. Williams, "Unified theory of microphone systems for stereophonic sound recording," in *Proc. 82nd Conv. Audio Engineering Society*, London, Mar. 1987, Preprint no. 2466.

[27] C. Ceoen. (1972). Comparative stereophonic listening tests. *J. Audio Eng. Soc.* [Online]. 20(1), pp. 19–27. Available: http://www.aes.org/e-lib/browse.cfm?elib=2104

[28] F. Rumsey and T. McCormick, *Sound and Recording: An Introduction*. Boca Raton, FL: CRC, 2012.

[29] E. De Sena, H. Hacıhabiboğlu, and Z. Cvetković, "Analysis and design of multichannel systems for perceptual sound field reconstruction," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 21, no. 8, pp. 1653–1665, 2013.

[30] B. Bernfeld, "Attempts for better understanding of the directional stereophonic listening mechanism," in *Proc. 44th Conv. Audio Engineering Society*, Rotterdam, The Netherlands, 1973, pp. C–4.

[31] J. Steinberg and W. B. Snow, "Auditory perspective-physical factors," *Trans. Amer. Inst. Elect. Engineers*, vol. 53, no. 1, pp. 12–17, 1934.

[32] D. H. Cooper and T. Shiga, "Discrete-matrix multichannel stereo," *J. Audio Eng. Soc.*, vol. 20, no. 5, pp. 346–360, May 1972.

[33] M. A. Gerzon, "Surround-sound psychoacoustics," *Wireless World*, vol. 80, no. 1468, pp. 483–486, 1974.

[34] G. Theile, "Multichannel natural recording based on psychoacoustic principles," in *Proc. 108th Conv. Audio Engineering Society*, Paris, 2000, Preprint no. 5156.

[35] R. Kassier, H. K. Lee, T. Brookes, and F. Rumsey, "An informal comparison between surround-sound microphone techniques," in *Proc. 118th Conv. Audio Engineering Society*, Barcelona, Spain, 2005, Preprint no. 6429.

[36] U. Herrmann and V. Henkels, "Main microphone techniques for the 3/2-stereo-standard," *20th Tonmeister Tagung*, 1998.

[37] P. G. Craven and M. A. Gerzon, "Coincident microphone simulation covering three dimensional space and yielding various directional outputs," U.S. Patent 4,042,779, Aug. 16 1977.

[38] A. Fukada, "A challenge in multichannel music recording," in *Proc. 19th Conf. Audio Engineering Society: Surround Sound - Techniques, Technology, and Perception*, Elmau, Germany, June 2001, p. 1881.

[39] K. Hamasaki and K. Hiyama, "Reproducing spatial impression with multichannel audio," in *Proc. 24th Conf. Audio Engineering Society: Multichannel Audio, The New Reality*, Banff, Canada, June 2003, p. 19.

[40] K. Hamasaki, "22.2 multichannel audio format standardization activity," *Broadcast Technol.*, vol. 45, pp. 14–19, 2011.

[41] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, June 1997.

[42] V. Pulkki and M. Karjalainen, "Localization of amplitude-panned virtual sources: I—Stereophonic panning," *J. Audio Eng. Soc.*, vol. 49, no. 9, pp. 739–752, Sept. 2001.

[43] V. Pulkki, "Localization of amplitude-panned virtual sources: II—Two-and three-dimensional panning," *J. Audio Eng. Soc*, vol. 49, no. 9, pp. 753–767, Sept. 2001.

[44] V. Pulkki, "Uniform spreading of amplitude panned virtual sources," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'99)*, New Paltz, NY, Oct. 1999, pp. 187–190.

[45] M. Frank, "Localization using different amplitude-panning methods in the frontal horizontal plane," in *Proc. EAA Joint Symp. Auralization and Ambisonics*, Berlin, 2014, pp. 41–47.

[46] T. Sporer, J. Liebetrau, S. Werner, S. Kepplinger, T. Gabb, and T. Sieder. (2015, Mar.). Localization of audio objects in multichannel reproduction systems. *Proc. 57th Conf. Audio Engineering Society: The Future of Audio Entertainment*

Technology—Cinema, Television and the Internet. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17613

[47] R. Baumgartner, P. Majdak, and B. Laback, "Assessment of sagittal-plane sound localization performance in spatial-audio applications," in *The Technology of Binaural Listening*. J. Blauert, Ed. Berlin: Springer-Verlag, 2013, pp. 93–119.

[48] J. Merimaa and V. Pulkki, "Spatial impulse response rendering: I—Analysis and synthesis," *J. Audio Eng. Soc.*, vol. 53, no. 12, pp. 1115–1127, 2005.

[49] V. Pulkki and J. Merimaa, "Spatial impulse response rendering: II—Reproduction of diffuse sound and listening tests," *J. Audio Eng. Soc.*, vol. 54, no. 1/2, pp. 3–20, 2006.

[50] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc.*, vol. 61, no. 1/2, pp. 17–28, 2013.

[51] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516, June 2007.

[52] J. Vilkamo, T. Lokki, and V. Pulkki, "Directional audio coding: Virtual microphone-based synthesis and subjective evaluation," *J. Audio Eng. Soc.*, vol. 57, no. 9, pp. 709–724, Sept. 2009.

[53] J. D. Johnston and Y. H. V. Lam, "Perceptual soundfield reconstruction," in *Proc. 109th Conv. Audio Engineering Society*, Los Angeles, CA, Sept. 2000, Preprint no. 5202.

[54] E. D Sena and Z. Cvetković, "A computational model for the estimation of localisation uncertainty," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'13)*, Vancouver, BC, Canada, May 2013, pp. 388–392.

[55] E. De Sena, H. Hacıhabiboğlu, and Z. Cvetkovic, "On the design and implementation of higher order differential microphones," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 1, pp. 162–174, Jan. 2012.

[56] J. Benesty and C. Jingdong, *Study and Design of Differential Microphone Arrays*. Berlin: Springer-Verlag, 2013.

[57] R. M. Aarts, "Enlarging the sweet spot for stereophony by time/intensity trading," *in Proc. 94th Conv. Audio Engineering Society*, Berlin, Mar. 1993, Preprint no. 3473.

[58] J. A. Rodenas and R. M. Aarts, "Position independent stereophonic sound reproduction," *J. Acoust. Soc. Netherlands*, no. 153, pp. 37–47, Nov. 2000.

[59] J. A. Ródenas, R. M. Aarts, and A. Janssen, "Derivation of an optimal directivity pattern for sweet spot widening in stereo sound reproduction," *J. Acoust. Soc. Amer.*, vol. 113, no. 1, pp. 267–278, Jan. 2003.

[60] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 5, pp. 1421–1448, May 2012.

[61] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[62] J. Borish, "Extension of the image model to arbitrary polyhedra," *J. Acoust. Soc. Amer.*, vol. 75, no. 6, pp. 1827–1836, June 1984.

[63] A. Krokstad, S. Strom, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *J. Sound and Vibration*, vol. 8, no. 1, pp. 118–125, Aug. 1968.

[64] A. Kulowski, "Algorithmic representation of the ray tracing technique," *Appl. Acoust.*, vol. 18, no. 6, pp. 449–469, June 1985.

[65] A. Farina, "RAMSETE: A new pyramid tracer for medium and large scale acoustic problems," in *Proc. 2nd European Conf. Noise Control (Euronoise'95)*, Lyon, France, Mar. 1995, p. 55.

[66] K. Kowalczyk and M. van Walstijn, "Room acoustics simulation using 3-D compact explicit FDTD schemes," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 1, pp. 34–46, 2011.

[67] S. A. Van Duyne and J. O. Smith, "The tetrahedral digital waveguide mesh," in *Proc. IEEE Workshop on Application Signal Processing Audio and Acoustics (WASPAA'95)*, Oct. 1995, pp. 234–237.

[68] T. Shuku and K. Ishihara, "The analysis of the acoustic field in irregularly shaped rooms by the finite element method," *J. Sound and Vibration*, vol. 29, no. 1, pp. 67–76, Jan. 1973.

[69] S. M. Kirkup, *The Boundary Element Method in Acoustics*. Integrated Sound Software Heptonstall, 2007.

[70] J. M. Buchholz, J. Mourjopoulos, and J. Blauert, "Room masking: Understanding and modelling the masking of room reflections," in *Proc. 110th Conv. Audio Engineering Society*, Amsterdam, The Netherlands, May 2001, Preprint no. 5312.

[71] D. R. Begault, "Audible and inaudible early reflections: Thresholds for auralization system design," in *Proc. 100th Conv. Audio Engineering Society*, Copenhagen, Denmark, May 1996, Preprint no. 4244.

[72] D. R. Begault, B. U. McClain, and M. R. Anderson, "Early reflection thresholds for virtual sound sources," in *Proc. Int. Workshop on Spatial Media*, 2001.

[73] D. R. Begault, B. U. McClain, and M. R. Anderson, "Early reflection thresholds for anechoic and reverberant stimuli within a 3-D sound display," in *Proc. 18th Int. Congr. Acoustics (ICA'04)*, Kyoto, Japan, 2004.

[74] H. Hacihabiboğlu and F. Murtagh, "An observational study of the precedence effect," *Acta Acustica United with Acustica*, vol. 92, no. 3, pp. 440–456, Mar. 2006.

[75] H. Hacihabiboğlu and F. Murtagh, "Perceptual simplification for model-based binaural room auralisation," *Appl. Acoust.*, vol. 69, no. 8, pp. 715–727, Aug. 2008.

[76] C. Schissler and D. Manocha, "Adaptive impulse response modeling for interactive sound propagation," in *Proc. 20th Association for Computing Machinery Special Interest Group on Computer Graphics and Interactive Techniques Symp. Interactive 3D Graphics and Games*, Redmond, WA, Feb. 2016, pp. 71–78.

[77] F. Rumsey, *Spatial Audio*. Burlington, MA: Focal Press, 2001.

[78] M. Barron and A. H. Marshall, "Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure," *J. Sound and Vibration*, vol. 77, no. 2, pp. 211–232, 1981.

[79] H. Kuttruff, *Room Acoustics*. Boca Raton, FL: CRC, 2009.

[80] J. S. Abel and P. Huang, "A simple, robust measure of reverberation echo density," in *Proc. 121st Conv. Audio Engineering Society*, San Francisco, CA, Oct. 2006, Preprint no. 6985.

[81] M. Karjalainen and H. Jarvelainen, "More about this reverberation science: Perceptually good late reverberation," in *Proc. 111th Conv. Audio Engineering Society*, New York, Nov. 2001, Preprint no. 5415.

[82] S. Favrot and J. M. Buchholz, "LoRA: A loudspeaker-based room auralization system," *Acta Acustica United with Acustica*, vol. 96, no. 2, pp. 364–375, Feb. 2010.

[83] T. Okamoto, B. F. Katz, M. Noisternig, Y. Iwaya, and Y. Suzuki, "Implementation of real-time room auralization using a surrounding 157 loudspeaker array," in *Proc. Int. Workshop Principles and Applications of Spatial Hearing (IWPASH 2009)*, Sendai, Japan, Nov. 2009, pp. 373–382.

[84] M. R. Schroeder, "Natural sounding artificial reverberation," *J. Audio Eng. Soc.*, vol. 10, no. 3, pp. 219–223, Mar. 1962.

[85] J. A. Moorer, "About this reverberation business," *Comp. Music J.*, vol. 3, no. 2, pp. 13–28, 1979.

[86] J.M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *Proc. 90th Conv. Audio Engineering Society*, Paris, Feb. 1991, Preprint no. 3030.

[87] J. Stautner and M. Puckette, "Designing multi-channel reverberators," *Comp. Music J.*, vol. 6, no. 1, pp. 52–65, 1982.

[88] D. Rocchesso, "Maximally diffusive yet efficient feedback delay networks for artificial reverberation," *IEEE Signal Process. Lett.*, vol. 4, no. 9, pp. 252–255, Sept. 1997.

[89] F. Menzer and C. Faller, "Unitary matrix design for diffuse Jot reverberators," in *Proc. 128th Conv. Audio Enginering Society*, London, June 2010, Preprint no. 7984.

[90] S. J. Schlecht and E. A. Habets, "Time-varying feedback matrices in feedback delay networks and their application in artificial reverberation," *J. Acoust. Soc. Amer.*, vol. 138, no. 3, pp. 1389–1398, Mar. 2015.

[91] J. Smith, "Digital signal processing using closed waveguide networks," U.S. Patent 5,448,010, Sept. 5, 1995.

[92] G. Kendall, W. Martens, D. Freed, D. Ludwig, and R. Karstens, "Image model reverberation from recirculating delays," in *Proc. 81st Conv. Audio Engineering Society*, Los Angeles, CA, Nov. 1986, Preprint no. 2408.

[93] T. Wendt, S. Van De Par, and S. D. Ewert, "A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation," *J. Audio Eng. Soc.*, vol. 62, no. 11, pp. 748–766, 2014.

[94] P. Huang, M. Karjalainen, and J. O. Smith, "Digital waveguide networks for room response modeling and synthesis," in *Proc. 118th Conv. Audio Engineering Society*, Barcelona, Spain, May 2005, Preprint no. 6394.

[95] N. Tsingos, E. Gallo, and G. Drettakis, "Perceptual audio rendering of complex virtual environments," *ACM Trans. Graphics*, vol. 23, no. 3, pp. 249–258, Aug. 2004.

[96] N. Tsingos, "Scalable perceptual mixing and filtering of audio signals using an augmented spectral representation," in *Proc. 8th Int. Conf. Digital Audio Effects (DAFx-05)*, 2005, pp. 277–282.

[97] A. C. Metan and H. Hacihabiboğlu, "Perceptual audio source culling for virtual environments," in *Proc. 19th Int. Conf. Digital Audio Effects (DAFx-16)*, 2016, pp. 201–208.

[98] J. O. Smith, "A new approach to digital signal processing using closed waveguide networks," in *Proc. Int. Computer Music Conf.*, Burnaby, B.C., Canada, 1985, pp. 47–53.

SP

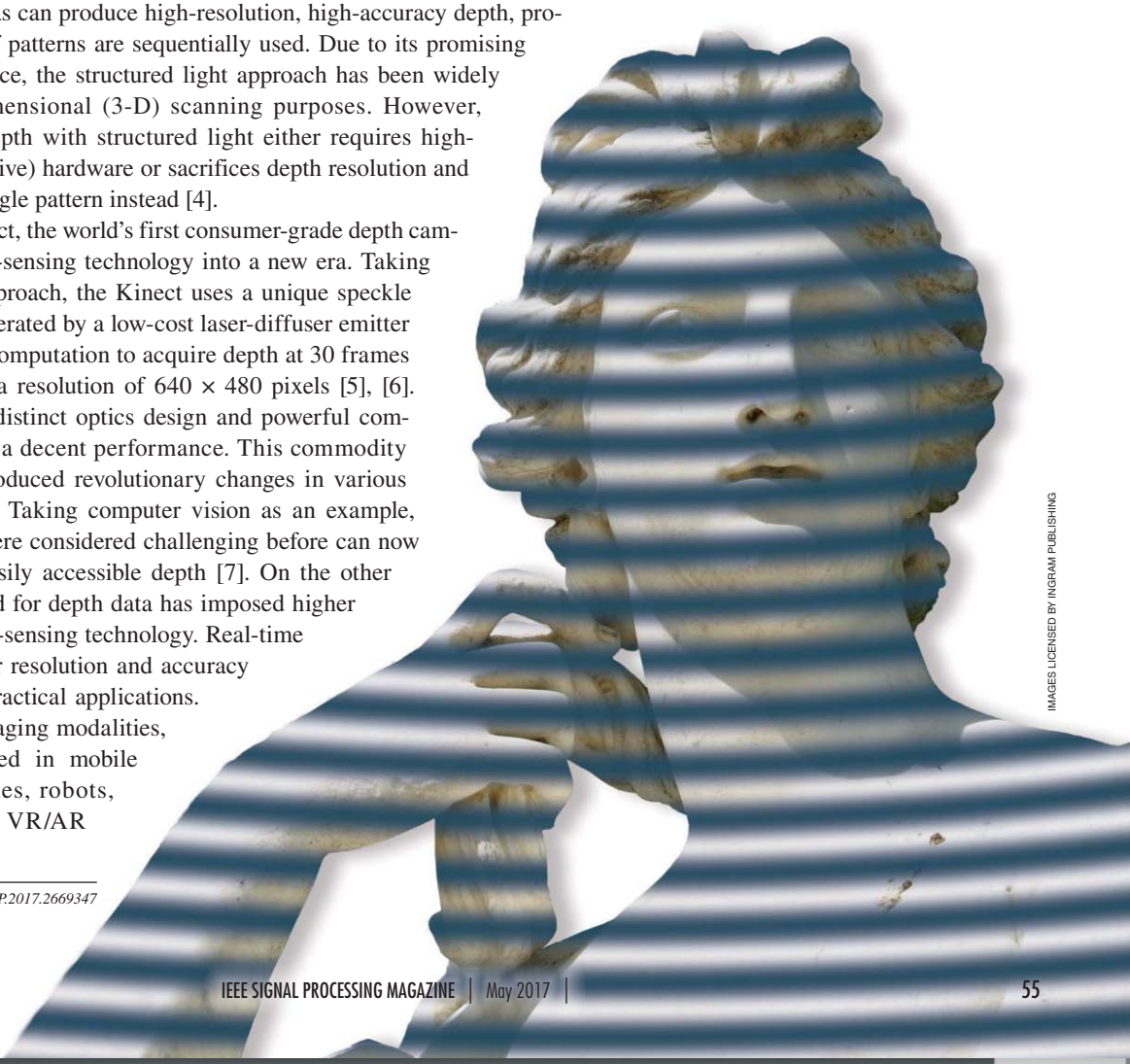
Zhiwei Xiong, Yueyi Zhang, Feng Wu, and Wenjun Zeng

Computational Depth Sensing

Toward high-performance commodity depth cameras

Depth information plays an important role in a variety of applications, including manufacturing, medical imaging, computer vision, graphics, and virtual/augmented reality (VR/AR). Depth sensing has thus attracted sustained attention from both academia and industry communities for decades. Mainstream depth cameras can be divided into three categories: stereo, time of flight (ToF), and structured light. Stereo cameras require no active illumination and can be used outdoors, but they are fragile for homogeneous surfaces [1]. Recently, off-the-shelf light field cameras have demonstrated improved depth estimation capability with a multiview stereo configuration [2]. ToF cameras operate at a high frame rate and fit time-critical scenarios well, but they are susceptible to noise and limited to low resolution [3]. Structured light cameras can produce high-resolution, high-accuracy depth, provided that a number of patterns are sequentially used. Due to its promising and reliable performance, the structured light approach has been widely adopted for three-dimensional (3-D) scanning purposes. However, achieving real-time depth with structured light either requires high-speed (and thus expensive) hardware or sacrifices depth resolution and accuracy by using a single pattern instead [4].

The Microsoft Kinect, the world's first consumer-grade depth camera, has brought depth-sensing technology into a new era. Taking the structured light approach, the Kinect uses a unique speckle pattern that can be generated by a low-cost laser-diffuser emitter and relies on parallel computation to acquire depth at 30 frames per second (fps) with a resolution of 640 × 480 pixels [5], [6]. The Kinect leverages distinct optics design and powerful computing capabilities for a decent performance. This commodity depth camera has introduced revolutionary changes in various related research areas. Taking computer vision as an example, many problems that were considered challenging before can now be solved with this easily accessible depth [7]. On the other hand, the large demand for depth data has imposed higher requirements for depth-sensing technology. Real-time depth with even higher resolution and accuracy is desirable in many practical applications. In the future, more imaging modalities, such as those deployed in mobile devices, game consoles, robots, drones, and emerging VR/AR



IMAGES LICENSED BY INGRAM PUBLISHING

Digital Object Identifier 10.1109/MSP.2017.2669347
Date of publication: 26 April 2017

devices, will largely rely on high-performance commodity depth cameras. The recently released Microsoft HoloLens [8] and Google Tango [9] are two good examples.

Inspired by the success and philosophy of the Kinect, much research has been done in the past few years targeting real-time, high-performance depth sensing. These works share a common insight that the codesign of image sensor systems and signal processing algorithms is essential to achieve a superior performance. Such codesign is called *computational depth sensing*. In this article, we introduce this important concept and provide an overview of the latest representative techniques. By bringing together and analyzing interdisciplinary research from signal processing, computer vision, and optics communities, our goal is to shed light on the development of future commodity depth cameras, which is a potential great interest to a broad audience. Specifically, this article will focus mainly on the structured light approach, which provides a large degree of freedom for the design of depth cameras. A recent review of ToF cameras is given in [10], and another on light-field cameras is given in [11]. Also, a comprehensive review on traditional structured light cameras can be found in [4]. This article will be complementary to these reviews.

Due to its promising and reliable performance, the structured light approach has been widely adopted for three-dimensional scanning purposes.

Computational depth sensing

For ordinary digital cameras, what you see is what you get under appropriate ambient illumination. By contrast, depth cameras do not sense depth information directly, but, rather, through either the space deformation or the time delay of light signals. In this sense, computation is actually indispensable for all depth cameras. In this article, however, computational depth sensing is defined as the redesign of image sensor systems or elements followed by advanced signal processing algorithms. Computational depth sensing can improve the capabilities of traditional depth cameras, introduce features that were not possible with traditional depth cameras, or reduce the cost or size of camera elements.

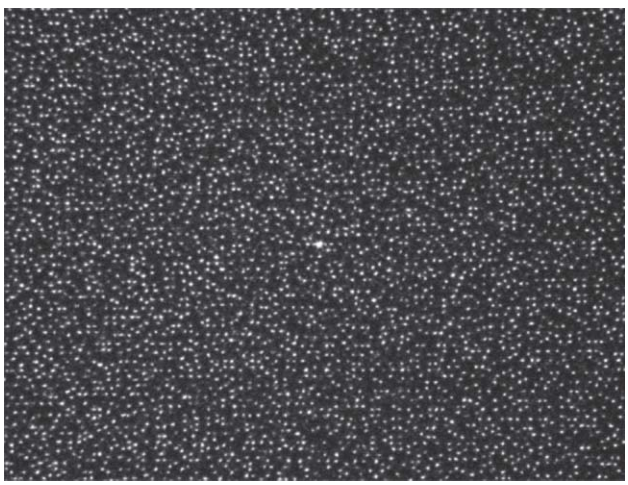


FIGURE 1. An infrared camera image of part of the speckle pattern projected by the Kinect.

For the completeness of this review, we first briefly introduce traditional structured light techniques, through which it is easier to understand the merits of computational techniques. Traditional structured light techniques can be roughly divided into two categories according to the pattern design strategy: time coding and space coding. Time-coding methods project multiple patterns sequentially to identify each point in the scene with a time series, e.g., binary code [12], Gray code [13], N-ary code [14], and phase-shifted fringe [15]. Due to the pixel-independent encoding, time-coding methods can provide high-accuracy depth with simple decoding algorithms. However, time coding cannot deal with dynamic scenes unless expensive high-speed hardware is used [16]. In contrast, space-coding methods using a single pattern are suitable for 3-D capture of dynamic scenes. Generally, this pattern needs to be designed in a way that each point is uniquely encoded by its neighborhood, e.g., De Bruijn sequences [17], M-arrays [18], and color-coded [19] and symbol-coded [20] patterns. Nevertheless, space-coding methods rely on an assumption of local depth smoothness that will be violated by abrupt depth changes. Thus, the accuracy of depth obtained from space coding is limited.

We still take the Kinect as an example to further explain the concept of computational depth sensing. The pattern used in the Kinect belongs to space coding for structured light. In traditional space-coding methods, color or grayscale information is generally used. These color or grayscale patterns need to be emitted by projectors. While projectors are often expensive, bulky, low-energy, and offer a small depth of field, the emitted light is also susceptible to ambient illumination and scene albedo. All of these issues hinder the practical application of space-coding depth cameras. The revolutionary change in the Kinect is to use a tiny laser-diffuser emitter to replace the projector. The diffuser bears a binary, pseudorandom speckle pattern, and the laser emits infrared light. Therefore, while being invisible to human eyes, an infrared speckle pattern is generated and projected onto the scene, as shown in Figure 1. Due to the well-designed pseudorandom distribution of the bright dots, this speckle pattern has high distinguishability when observed in a local window. Compared with the projector, the laser-diffuser emitter has obvious advantages: low cost, compact size, high energy, and large depth of field. Moreover, compared with the projector-generated color or grayscale patterns, this binary, infrared speckle pattern is robust to ambient illumination and scene albedo. All of these issues contribute to the reliable performance of the Kinect depth camera.

Besides the unique optics design, the Kinect also adopts an efficient signal processing algorithm, that is, how to compute depth once the deformed speckle pattern is captured. The triangulation principle is used here. A reference image

of the speckle pattern is stored in the memory of the Kinect. The reference image is acquired only once, by projecting the speckle pattern onto a planar object at a proper distance. The captured image is then processed with the reference image to find a relative shift of the speckle pattern. This is implemented using a correlation-based image-matching algorithm. A small sliding window scans the captured image and correlates it with the reference image, and the correlation peak indicates the relative shift that can be mapped to depth through triangulation. Since the laser-diffuser emitter is strictly calibrated with the infrared camera in the Kinect, the shift of the speckle pattern will only occur in the horizontal direction, which significantly reduces the search range for the image matching. The winner-takes-all strategy is applied as postprocessing to refine the results. The whole implementation is accelerated by parallel computing to produce real-time depth.

The Kinect serves as a typical example of computational depth sensing, where a distinct optics design in combination with powerful computing capabilities carves out a way toward high-performance commodity depth cameras. For the first time, for a few hundred U.S. dollars, everyone can have a decent depth camera with millimeter-level accuracy at a distance of approximately 1 m. (The accuracy decreases as the distance increases.) Nevertheless, the accuracy of the Kinect depth is still limited by the space-coding approach. There is a large demand for real-time commodity depth cameras with improved accuracy. Driven by the increasing demand, different computational techniques either inheriting the merits of the Kinect or adopting new principles have been developed recently. Before we go into the details about these techniques, we first categorize them into four groups from the perspective of methodology: phase period coding, space-time coding, binary phase shifting, and sensor fusion. Phase period coding advances the classical phase-shifting profilometry to significantly reduce the number of patterns required for accurate depth acquisition. Space-time coding adaptively exploits the space-time-coded information toward a scalable depth-sensing paradigm. Binary phase shifting uses defocused, dithered, or filtered binary patterns in pursuit of super-high-speed, yet low-cost, depth cameras. Sensor fusion jointly takes advantage of two different depth sensors for an improved performance while maintaining backward compatibility.

Driven by the increasing demand, different computational techniques either inheriting the merits of the Kinect or adopting new principles have been developed recently.

Toward high-performance commodity depth cameras

Phase period coding

Phase shifting is a well-known structured light technique for high-accuracy 3-D measurement [15]. It requires a set of sinusoidal fringe patterns with an incrementally shifted phase to be projected and captured. Taking three-step phase shifting as an example, each of the three fringe patterns is shifted by a phase of $2\pi/3$ from the previous one. Assume the fringe is horizontal, as shown in Figure 2(a), the nth $(n = 1,2,3)$ fringe pattern can be readily generated as

$$P_n(x,y) = C + A\cos\left(2\pi\, fy + \frac{2\pi}{3}n\right), \tag{1}$$

where A, C, and f represent the amplitude, the dc offset, and the frequency of fringe, respectively. The corresponding captured image can be described as

$$I_n(u,v) = I_C(u,v) + I_A(u,v)\cos\left[\phi(u,v) + \frac{2\pi}{3}n\right], \tag{2}$$

where

$$I_A(u,v) = \alpha(u,v)\,A,\ I_C(u,v) = \alpha(u,v)\,[C + \beta(u,v)]. \tag{3}$$

One captured image for a real scene is shown in Figure 2(b). To differentiate from (x,y) denoting the projector coordinates, we use (u,v) to denote the camera coordinates. $\alpha(u,v)$ and $\beta(u,v)$ represent the albedo and the ambient illumination at each scene point. There are three unknowns in (2), the sinusoidal amplitude $I_A(u,v)$, the background intensity $I_C(u,v)$,

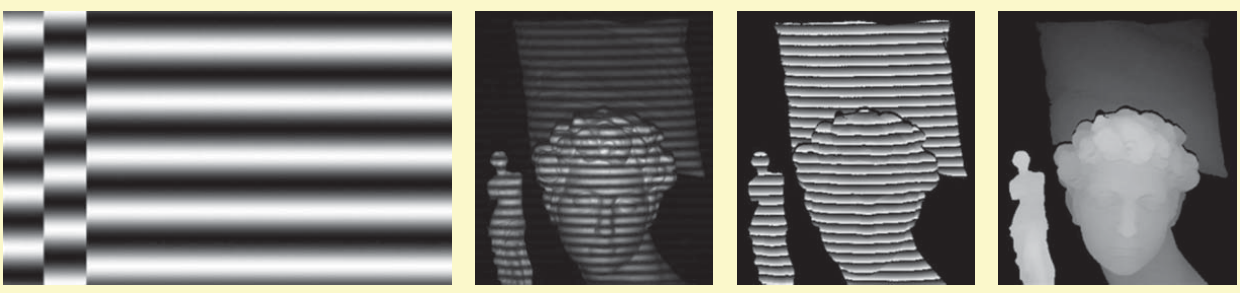


FIGURE 2. An example of phase shifting. These images depict (a) three projected fringe patterns of the same size, overlapped to demonstrate the phase difference; (b) one captured image; (c) a wrapped phase map; and (d) an absolute depth map.

and the phase $\phi(u,v)$, so the theoretical minimum number of fringe patterns required is three. Using the three captured images together, the phase can be calculated as

$$\phi(u,v) = \arctan\left[\sqrt{3}\frac{I_2(u,v) - I_1(u,v)}{2I_3(u,v) - I_2(u,v) - I_1(u,v)}\right]. \quad (4)$$

The phase obtained by (4) ranges from $-\pi$ to π, which results in discontinuities in the phase map as shown in Figure 2(c). Thus, a phase unwrapping procedure is needed to convert the wrapped phase to the absolute phase as

$$\phi_{\text{abs}}(u,v) = \phi(u,v) + 2d(u,v)\pi, \quad (5)$$

where $d(u,v)$ is an integer disparity within $[0, M-1]$ and M is the number of periods of the fringe pattern. Once the absolute phase is obtained, the absolute depth at each scene point can be calculated through triangulation between the projector and the camera. An exemplar depth map is shown in Figure 2(d).

Phase shifting belongs to time coding for structured light, as it uses multiple patterns. By concentrating the usable codeword within a narrow fringe period and eliminating the influence of ambient illumination and scene albedo, phase shifting can achieve higher accuracy compared with other time-coding methods, especially when high-frequency fringes are used [21]. Nevertheless, due to the periodical nature of the fringe signal, a major problem of this technique is the phase ambiguity. That is, only depth within a range equivalent to one fringe period can be measured directly, and depth exceeding this range will be wrapped. While high-frequency fringes are preferred for high-accuracy measurement, they also lead to severe ambiguity. To recover the absolute depth, a prevalent solution is temporal phase unwrapping using a large number of additional patterns, e.g., Gray code or multifrequency fringes [22], which greatly limits the application of phase shifting in time-critical scenarios.

Different depth sensors have complementary advantages, and a combined device could outperform each single component.

Recently, the advancement of projector technology has enabled fast and accurate 3-D measurement with phase shifting. One pioneering phase-shifting-based system is described by Zhang and Huang in [23]. To achieve real-time depth sensing, they modify a color digital light processing (DLP) projector and use the red, green, and blue (RGB) channels to project three fringe patterns. Moreover, to avoid the time-consuming arctangent function when calculating the phase, they propose a new phase-shifting method using trapezoidal fringes instead of sinusoidal fringes. The 3-D acquisition and reconstruction speed of their system is up to 40 fps at 532 × 500 pixels. In this work, spatial phase unwrapping [24] is applied to recover a continuous phase map from a wrapped one, yet it cannot solve the ambiguity when multiple isolated objects or abrupt depth changes are present. Although restricted in the scope of application, this work demonstrates the feasibility of high-resolution, real-time depth sensing with phase shifting.

To maintain a minimum number of required images for absolute depth recovery, several phase-shifting methods using period coding have been proposed. The common idea is to encode a period-specific signal into the fringe, and the phase ambiguity is solved by decoding the embedded signal. In [25], Liu et al. embed a unit-frequency sinusoidal fringe into a high-frequency one. The high-frequency fringe plays the role of phase shifting, while the unit-frequency fringe serves as period codes. At least five images are needed to derive a phase pair. The unit-frequency phase has no ambiguity, as there is only a single period, so it can be used to disambiguate the wrapped high-frequency phase. To achieve a real-time performance, they also propose a lossless table-lookup method for phase calculation and 3-D point cloud generation. The 3-D reconstruction speed of their system can reach 228 fps at 640 × 480 pixels, though the actual 3-D acquisition speed depends on the projector and the camera being used.

In [26], Wissmann et al. propose another phase-shifting method with period coding. Specifically, they embed a one-dimensional (1-D) binary De Bruijn sequence into a four-step phase-shifting pattern set, where the embedded signal will not affect the phase calculation. They also design a corresponding decoding strategy to extract the period index from the De Bruijn sequence. One highlight of this work is the cost-effective hardware design. Instead of using an off-the-shelf DLP projector, they develop a high-speed spatial light modulator (SLM) for pattern generation. The four patterns are arranged around the rotational center of the SLM. The SLM is illuminated by a light source and imaged through a conventional camera lens. Pattern switching is achieved via synchronized timing of camera exposure intervals, integrating the image of the rotating SLM during the rotation of a pattern segment. With this dedicated system, the 3-D acquisition speed is claimed to be 50 fps, while the 3-D reconstruction speed is 11 fps at 640 × 480 pixels.

The aforementioned two methods solve the phase ambiguity using a relatively small number of images. However, one drawback is that the amplitude of fringe needs to be reduced to accommodate the embedded period codes, which sacrifices the signal-to-noise ratio (SNR) and thus the accuracy of phase shifting. To retain the SNR of phase shifting, Wang et al. propose a period-coding strategy without reducing the amplitude of the fringe in [27]. Their key observation is that the spatial intensity efficiency of phase-shifted fringes is less than 100%, leaving a margin of the available intensity dynamic range for adding a period cue signal. Given the step of phase-shifting N and the number of periods of fringe M, the period codes can be designed by following several basic properties. It should be noted that a smaller N and a larger M will decrease the margin of intensity dynamic range and increase the difficulty of code design. If using four patterns, 16 periods can be supported. In an extreme case of three patterns, at most four periods can be used for reliable decoding. The computational cost of this method is low, and up to

120 fps can be achieved for 3-D acquisition and reconstruction at 640 × 480 pixels. Still, the limited range of frequency of fringe, especially for three patterns, limits the measurement accuracy.

To minimize the number of required images while retaining both the full amplitude and the high frequency of fringe, Zhang et al. propose to use three speckle-embedded fringe patterns in which the pseudorandom speckle signal solves the phase ambiguity [28]. This work is inspired by the speckle pattern used in the Kinect, yet here the speckle signal is adapted for phase unwrapping. The nth$(n = 1,2,3)$ speckle-embedded fringe pattern can be described as

$$P_n^E(x,y) = C + B^E(x,y)Z(x,y) + A\cos\left(2\pi\ fy + \frac{2\pi}{3}n\right). \quad (6)$$

Compared to (1), the newly added second term consists of two parts: $Z(x,y)$ denotes the distribution of a pseudorandom speckle signal similar to Figure 1, and $B^E(x,y)$ controls the intensity of the speckle. Similar to previous period-coding methods, this embedded speckle signal will not affect the phase calculation.

Suppose the intensity dynamic range of the projected patterns is L-bit grayscale. Generally, the amplitude and the dc offset of the fringe signal are set as $A = C = 2^{L-1}$. To embed the speckle signal without reducing the amplitude of fringe, a straightforward solution is to assign $B^E(x,y)$ the largest available intensity as

$$B^E(x,y) = 2^L - \max_{n=1,2,3}\{P_n(x,y)\}, \quad (7)$$

where $P_n(n = 1,2,3)$ denotes the original sinusoidal fringe patterns. However, little speckle signal can be embedded when $P_n(x,y)$ equals or gets very close to 2^L. Actually, as period codes, the intensity of the speckle can be either positive or negative. Therefore, a better strategy is to assign $B^E(x,y)$ as

$$B^E(x,y) = \begin{cases} 2^L - B^E_{\max} & 2^L - B^E_{\max} \geq B^E_{\min} \\ -B^E_{\min} & \text{otherwise} \end{cases}, \quad (8)$$

where

$$B^E_{\max} = \max_{n=1,2,3}\{P_n(x,y)\},\ B^E_{\min} = \min_{n=1,2,3}\{P_n(x,y)\}. \quad (9)$$

The 1-D illustration for speckle intensity decision is given in Figure 3(a). As can be seen, the margin of intensity dynamic range in three phase-shifted fringes is fully exploited in this way. The three speckle-embedded fringe patterns are shown in Figure 3(b). The positive speckle intensity corresponds to the brighter dots, and the negative speckle intensity corresponds to the darker dots on the fringe background.

The phase-unwrapping process follows the idea of patch-based image matching used in the Kinect. Since the disparity in (5) is an integer within $[0, M-1]$, at most M possible disparities need to be checked for each wrapped phase. To improve the robustness of phase unwrapping when high-frequency fringes are employed in pursuit of high accuracy,

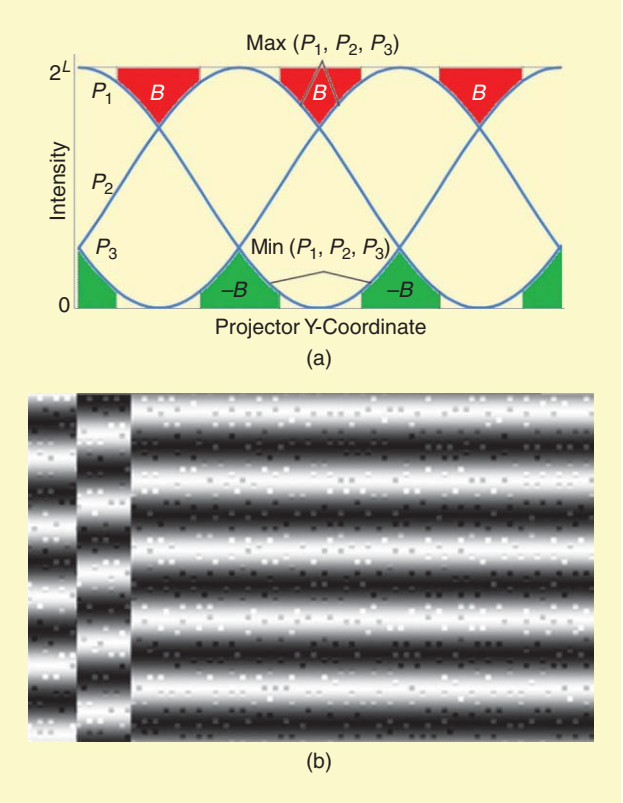


FIGURE 3. (a) The speckle intensity decision; red regions indicate positive values and green regions indicate negative values. (b) Three speckle-embedded fringe patterns.

an efficient voting strategy is further proposed to recover the absolute phase region by region instead of pixel by pixel. Figure 4 shows the online 3-D reconstruction results of a dynamic scene using this method, which demonstrates the accuracy of phase shifting that can be achieved by using the theoretical minimum number of images. Quantitatively, the root-mean-square reconstruction error is submillimeter at a distance of approximately 1 m. In other words, the measurement accuracy is an order of magnitude higher compared with the Kinect. A potential drawback of this method is that phase map segmentation is required, which may cause reliability issues in case of complex scenes.

Table 1 lists the main features of the previously discussed phase period coding methods. Compared with traditional phase shifting using a large number of patterns, systems based on these methods can all achieve real-time 3-D acquisition and reconstruction, making high-accuracy depth sensing possible in time-critical scenarios. However, even with the minimum of three patterns, motion effects still pose challenges for developing commodity depth cameras based on phase shifting. For example, if a 30-fps acquisition speed is expected (which should be enough for general-purpose usage), the projector and the camera need to be synchronized at 90 fps. Such high-speed hardware requirements would greatly increase the system cost. It is worth mentioning that since the phase relationship between any two consecutive patterns is

FIGURE 4. Online 3-D reconstruction results of a dynamic scene (facial expressions with black eyeglasses) using three speckle-embedded fringe patterns.

Table 1. A comparison of different phase period coding methods.

Period Codes	Image Number	3-D Reconstruction Speed	Computing Unit	Remarks
None [23]	3	40 fps at 532 × 500 pixels	2.8 GHz CPU	Single smooth surface
Unit-frequency fringe [25]	5	228 fps at 640 × 480 pixels	3.0 GHz CPU	Fringe amplitude reduced
De Bruijn sequence [26]	4	11 fps at 640 × 480 pixels	2.0 GHz CPU	Fringe amplitude reduced
Customized [27]	4	120 fps at 640 × 480 pixels	3.0 GHz CPU	Fringe frequency limited
Speckle [28]	3	15 fps at 640 × 480 pixels	2.83 GHz CPU	Phase segmentation required

the same, any newly captured image can be combined with its preceding two images to produce a depth map. Therefore, one can have a pseudo frame rate of 30 fps even when the projector and the camera work at the same speed, by using a sliding window strategy in the time dimension. However, motion occurring during the acquisition time of three images (100 ms in this example) is likely to cause measurement errors. To address the motion problem, there are two main directions. One is pattern design to further reduce the time dependency, and the other is pattern generation to further reduce the hardware cost. In both cases, the accuracy of depth measurement should be retained as much as possible. The efforts along these two directions are detailed next.

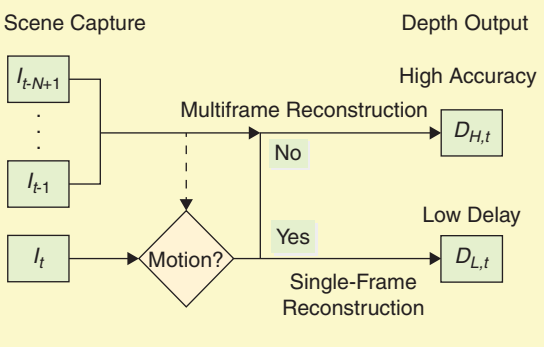


FIGURE 5. A general framework of scalable depth sensing with space-time coding.

Space-time coding

Human eyes will focus on the object details in a static scene. Once the scene becomes dynamic, human attention will be largely attracted by the object motion and thus be less sensitive to the object details. Therefore, it is possible to take advantage of this property to integrate space coding and time coding within a scalable depth-sensing paradigm. The basic idea is to design a uniform space-time coding through which a high-accuracy depth map can be reconstructed from multiple consecutive frames for a temporally static scene, while a low-delay depth map can be reconstructed from a single frame once the scene becomes dynamic. The transition between these two modes should be immediate and seamless.

Figure 5 shows a general framework for scalable depth sensing with space-time coding. Suppose there is a synchronized projector-camera system working with structured light illumination, and the projected patterns have a temporal period of N. The currently captured frame I_t is first compared with the prior $N-1$ frames for motion detection. If the scene remains static within N frames, the image set $(I_{t-N+1},\ldots,I_{t-1},I_t)$ is processed to produce a high-accuracy depth map $D_{H,t}$. Otherwise, if the scene becomes dynamic during the N frames, only I_t is processed to produce a low-delay depth map $D_{L,t}$. The scalability can be further extended. On one hand, if the scene remains static within $n\,(1 \le n \le N)$ frames, the image set $(I_{t-n+1},\ldots,I_{t-1},I_t)$ can be processed to produce a depth map $D_{n,t}$ with progressively improved accuracy as the value of n increases. On the other hand, a regionwise optimal depth map can be produced through region-wise motion detection.

An early work of space-time coding is described by Zhang et al. in [29]. They first design a color stripe pattern by realizing a De Bruijn sequence as shown in Figure 6(a), which supports single-frame depth reconstruction for high-speed scans of moving objects. By smoothing the color-stripe pattern with a Gaussian filter and projecting its shifted copies (in the direction perpendicular to the stripe) over time, multiframe depth reconstruction is then supported for high-accuracy scans of static scenes. For single-frame reconstruction, they develop a multipass dynamic programming algorithm that eliminates global smoothness assumptions and strict ordering constraints present in previous formulations. For multiframe reconstruction, space-time analysis is conducted at each sensor pixel to obtain interframe depth localization. Using a short sequence of seven time-shifted stripe patterns, the accuracy of the multiframe reconstructed depth map is significantly improved for static scenes compared with the approach that first obtains seven single-frame reconstructed depth maps independently and then combines them into a high-resolution one, which demonstrates the effectiveness of space-time analysis. Still, this work is one step away from scalable depth sensing, as it does not support the transition between the two types of reconstruction.

Once the scene becomes dynamic, human attention will be largely attracted by the object motion and thus be less sensitive to the object details.

The first embodiment of scalable depth sensing is realized by Ishii et al. in [30]. As shown in Figure 6(b), the base pattern is periodical in one direction (vertical for example) with a period of N, while each column represents a 1-D Gray code pattern. By shifting the base pattern $N-1$ times in the vertical direction, one can get a set of patterns that are both spatially and temporally periodical based on the Gray code, which thus enables space-time coding with a flexible number of frames. This work takes $N=8$ as an example, so there are eight decoding types: the first type uses eight codewords along the time dimension in eight different frames, the second type adds one more codeword in space from the latest frame and removes the dependency on the earliest frame, and so on. Finally, the eighth type uses all eight codewords in space from the latest frame and gets rid of the dependency on previous frames. Obviously, the first decoding type is effective for measuring static objects, while the eighth type is effective for measuring moving objects. The adaptive selection of decoding types $n(1 \le n \le N)$ is determined based on the frame differencing features that detect the decoding errors e caused by motion. Specifically, n is proportional to e. In other words, larger decoding types that are robust to motion are selected when e increases, and smaller decoding types that enable accurate measurement are selected when e decreases. In this way, scalable depth sensing is elegantly realized with a smooth transition between single-frame and multiframe depth reconstruction.

In [31], Taguchi et al. propose another scalable depth-sensing method, where they also use a De Bruijn color stripe as the base pattern similar to [29] and obtain a temporally periodical pattern set by shifting the base pattern regularly over time. The whole pattern set contains eight patterns and allows for decoding with a flexible number of frames (e.g., one, two, four, or eight) at every pixel. In other words, motion-aware spatiotemporal window selection is supported during decoding. The spatial window size decreases as more temporal frames are used. If eight frames are used, the spatial window size shrinks to only one pixel. One highlight of this work is that pixel-wise optimal depth reconstruction is supported based on the plane-sweeping algorithm. That is, for each pixel, the optimal depth value should give the maximum matching score in terms of the normalized cross-correlation. The matching score is computed between the captured images and the projected patterns, using all possible spatiotemporal window sizes at a given depth layer. This space-time coding can handle dynamic scenes with different motions in different parts of the scene and improves the accuracy of depth measurement for static or slowly moving parts.

As mentioned previously, phase-shifted fringes provide higher accuracy depth measurement than other time-coding methods, while the pseudorandom speckle pattern provides more robust depth measurement than other space-coding methods. Combining these two distinct components, Zhang et al. introduce a set of hybrid patterns for scalable depth sensing in [32], where the nth$(n=1,2,3)$ pattern can be described as

$$P_n^H(x,y) = B^H(x,y)Z(x,y) + \lambda\left[C + A\cos\left(2\pi\, fy + \frac{2\pi}{3}n\right)\right]. \tag{10}$$

Compared to (6), the main difference here is that the amplitude of fringe needs to be reduced to support single-frame reconstruction from the speckle, and $\lambda \in [0,1]$ denotes the percentage of intensity dynamic range that the fringe component occupies. It has been experimentally investigated that $\lambda \in [0.2, 0.4]$ can provide competitive accuracy of depth measurement in both

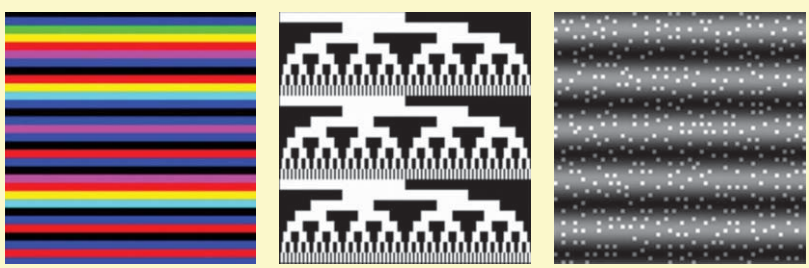


FIGURE 6. Base patterns of space-time coding: (a) a De Bruijn color stripe, (b) a periodical Gray code, and (c) a hybrid speckle and fringe. (Images (a) and (b) are from [29] and [30], respectively.)

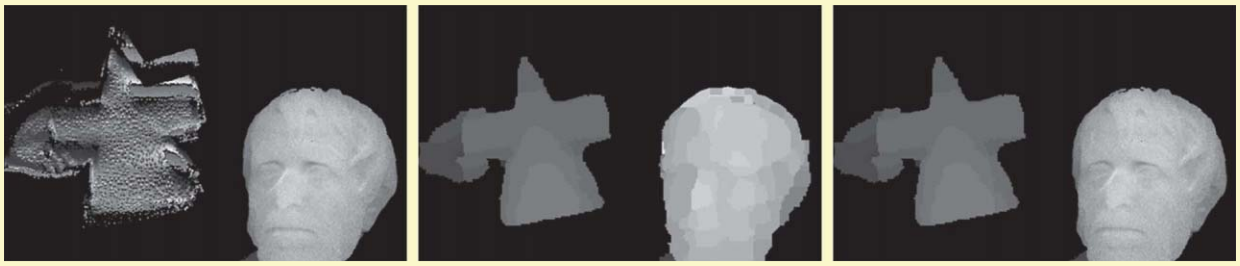


FIGURE 7. Depth reconstruction results for a scene with one moving object and one static object: (a) a multiframe reconstruction, (b) a single-frame reconstruction, and (c) a scalable depth sensing.

single-frame and multiframe reconstruction. Accordingly, the intensity of the speckle is set as

$$B^H(x,y) = 2^L - \lambda \max_{n=1,2,3}\{P_n(x,y)\}. \tag{11}$$

One of the hybrid patterns is shown in Figure 6(c). In this work, two sets of hybrid patterns with different frequencies of fringes are used for phase unwrapping when performing multiframe reconstruction, and patch-based image matching is adopted for single-frame reconstruction from the speckle as in the Kinect. To enable region-wise mode transition between single-frame and multiframe reconstruction, a motion mask that separates multiple objects with different motions is estimated. Figure 7 shows an example where there are two objects in the scene: one is moving and one remains static. As can be seen, this scalable depth-sensing method adaptively produces low-delay depth for the moving object and high-accuracy depth for the static object. In general, single-frame depth reconstruction is computationally intensive due to the correlation-based image matching. For example, it is reported in [31] that the correlation computation for generating one depth map takes up to seven seconds on a standard personal computer with an Intel Core i7-950 processor. Although the acquisition can be performed at a speed as high as 500 fps in [30], the reconstruction needs to be conducted offline. Using the central processing unit (CPU) for multiframe reconstruction and the graphics processing unit (GPU) for single-frame reconstruction, this work realizes real-time scalable depth sensing for the first time. The depth reconstruction speed is 20 fps at 640×480 pixels.

Since scalable depth sensing well coincides with the characteristics of the human visual system and can be easily realized with low-cost hardware, it provides a practical solution for many real-world applications when both accuracy and speed of depth sensing are concerned. In fact, a compromise always has to be made between accuracy and speed in developing commodity depth cameras according to the application scenarios, and scalable depth sensing achieves an efficient balance between these two requirements.

Binary phase shifting

Phase shifting can achieve high-accuracy 3-D measurement, but it is vulnerable to motion. To alleviate the motion effect, high-speed hardware is required, but the high-speed DLP projector is expensive. In practice, there is a speed limit to generate 8-b grayscale patterns using a DLP projector. However, it is much faster to generate 1-b binary patterns, since only switchings between two states are needed. Moreover, as demonstrated by the Kinect, it is possible to generate binary patterns using a low-cost laser-diffuser emitter. Therefore, it will greatly enhance the speed or reduce the cost of phase shifting if binary patterns can be utilized to produce sinusoidal fringes.

An early work along this line is described by Lei and Zhang in [33]. The basic idea is to use projector defocusing. The defocused projector is modeled by a point spread function that can be approximated as a Gaussian smoothing filter. They find that a sinusoidal fringe pattern can be produced by applying the Gaussian filter to a square stripe pattern, as shown in Figure 8(a). Therefore, they simulate three-step phase shifting using the defocused square stripes, where the phase shift is realized by shifting the stripes spatially. This method turns out to work quite well, and the phase error is negligible, given a proper degree of projector defocusing and a proper width of the stripes. Using a DLP projector

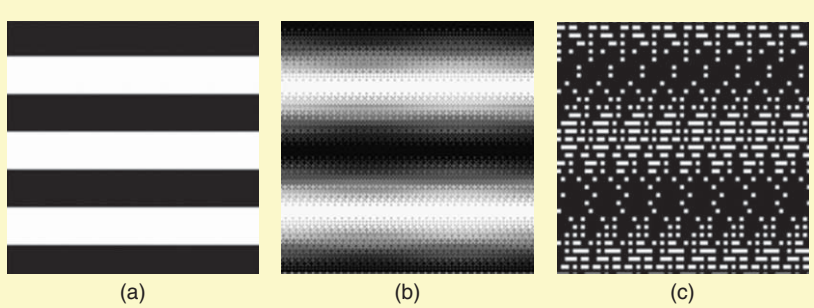


FIGURE 8. Base patterns of binary phase shifting: (a) a square binary pattern, (b) a dithered binary pattern, and (c) a density-modulated binary pattern.

and the square binary patterns, a superfast 3-D measurement system has been developed with a speed of 667 fps at 480 × 480 pixels [34]. Under such a high frame rate, the motion effect could be largely alleviated. Nevertheless, the square binary pattern has two main drawbacks. First, the depth measurement range needs to be small to guarantee a proper degree of projector defocusing, since the stripes will not become sinusoidal when the degree of defocusing is not proper. Second, when the stripes are wide, it requires a large degree of defocusing to reshape the square stripes to the sinusoidal fringes, which may exceed the capability of the projector lens.

To overcome the shortcomings of the square binary patterns, in [35] Wang and Zhang propose another kind of binary pattern by borrowing the idea of dithering, also known as *half toning*, which has been extensively used in digital image processing. The dithered binary pattern they use is shown in Figure 8(b). Specifically, the desired sinusoidal fringes are produced using the Bayer-ordered dithering technique or the error-diffusion dithering technique. Different from the square binary pattern, the dithered pattern appears sinusoidal even before processing. Therefore, even when the fringes are very wide, or the projector is nearly focused, the dithered binary pattern can produce ideal sinusoidal fringes. In other words, the dithered binary pattern is less dependent on the degree of projector defocusing and the width of the fringes. The former robustness extends the depth measurement range, while the latter robustness enables multifrequency phase shifting for absolute depth measurement. Using the dithered binary patterns, a superfast two-frequency phase-shifting method is developed for absolute 3-D measurement of live rabbit hearts at 800 fps with a resolution of 576 × 576 pixels [36].

The aforementioned defocusing techniques generally require much calibration effort, since most of the existing calibration techniques assume the projector to be in focus. Alternatively, Yang et al. propose density-modulated binary patterns to carry the phase information [37]. Different from defocusing, the phase is implicitly represented by the 1-D density variation of bright dots in the pattern. Figure 8(c) shows an exemplar pattern with density variation in the vertical direction. Specifically, when denoting the pattern as $P^B(x,y)$, the number of bright dots in different rows of $P^B(x,y)$ is defined as a sinusoidal function

$$k(y) = \text{Round}\left\{\left[\sin\left(2\pi\frac{y}{T}+\theta\right)+1\right]\times\gamma+1\right\}, \qquad (12)$$

where Round() is a function to round a floating number to an integer, T is the number of rows in a sinusoidal period, and γ is a scaling factor controlling $k(y)$ as an integer from 1 to K that determines the number of different densities in a period. The three patterns for three-step phase shifting are generated by setting θ to $-2\pi/3$, 0, and $2\pi/3$, respectively. A low-pass filter is then applied to the captured images to extract the phase, which is accomplished by calculating the average energy in a small sliding window. It is verified that the energy images well approximate the phase-shifted sinusoidal fringes and thus support high-accuracy depth measurement. Furthermore, an error correction method is proposed to reduce the quantization errors introduced by approximating sinusoidal fringes with a limited number of densities.

Figure 9 shows the depth reconstruction results of some indoor scenes in comparison with original phase shifting and the Kinect. It can be seen that depth from the density-modulated patterns is consistently better than that from the Kinect, in terms of surface details and object boundaries. Compared with original phase shifting, some details are lost as binary patterns cannot represent phase as perfectly as grayscale patterns. However, a distinct property of the density-modulated binary pattern is that it still preserves the locally unique distribution of bright dots in the vertical direction. Therefore, using three density-modulated binary patterns, the advantages of scalable depth sensing in [32] and unambiguous 3-D measurement in

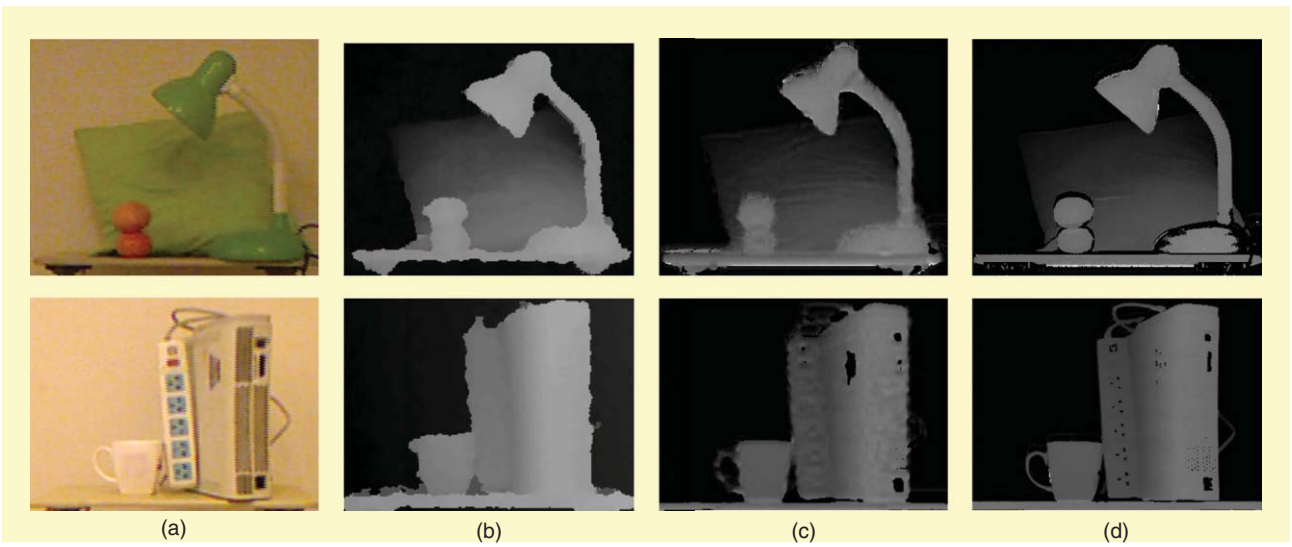


FIGURE 9. Depth reconstruction results for two indoor scenes: (a) the color image, (b) the Kinect results, (c) the results of density-modulated binary patterns, and (d) the results of original phase shifting.

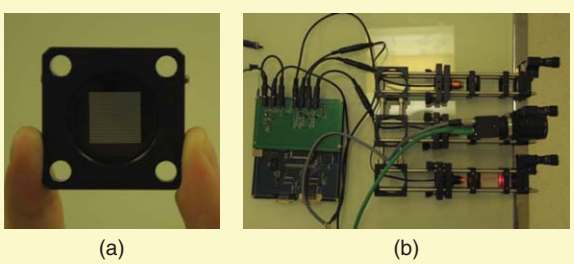
[28] can be inherited. While single-frame depth reconstruction is supported simultaneously, the phase ambiguity can be eliminated through patch-based image matching. These advantages are beyond the capabilities of original phase shifting using grayscale patterns.

Moreover, the density-modulated binary patterns can be generated using low-cost laser-diffuser emitters as is done in the Kinect, which gets rid of expensive DLP projectors. This property is extremely important for developing commodity depth cameras. Figure 10(a) shows the coded optical element customized for the usage of the diffuser. It is done on an opaque base with a 25-mm diameter. The pattern is in the center of the base with a 15-mm length in each dimension. The size of a bright dot is $20\ \mu m \times 20\ \mu m$. Figure 10(b) shows the depth camera prototype on the fly. Since the binary pattern is fixed on the diffuser in front of the laser, three lasers are needed to be aligned with three properly shifted patterns. A timing circuit then controls the three lasers to be on and off sequentially, together with a synchronized camera. Depth can be reconstructed from either a single frame or three consecutive frames, depending on the motion detected in the scene. Note that the physical size of the laser-diffuser emitter can be much more compact in actual production. Besides its low cost and compact size, the laser emitter also offers high energy and large depth of field that far surpass DLP projectors.

A common insight behind most of the techniques discussed previously is to improve the efficiency of space-time multiplexing in terms of the light signal usage.

Sensor fusion

A depth camera is ideally fast, accurate, and robust. Unfortunately, no existing depth camera is perfect on its own. A feasible way to enhance the performance of depth cameras is sensor fusion, as different depth sensors have complementary advantages, and a combined device could outperform each single component. Due to its passive properties, stereo is easy to combine with active illumination sensors. For example, structured light illumination has been used to improve the accuracy of stereo sensors, since structured light can provide distinguishable features for stereo matching on textureless surfaces [38].



(a) (b)

FIGURE 10. (a) The coded optical element customized for the diffuser usage. (b) The depth camera prototype using laser-diffuser emitters to generate density-modulated binary patterns.

The combination of stereo and structured light has other forms. To achieve high-accuracy, low-delay depth sensing, Weise et al. built a novel depth camera based on three-step phase shifting, where the phase ambiguity is solved by stereo [39]. Their system consists of a DLP projector, two high-speed monochrome cameras, and a color camera. The two monochrome cameras are synchronized and record the three images of phase-shifted fringes. As mentioned previously, the phase ambiguity is an integer disparity within $[0, M-1]$, where M is the number of periods of the fringe pattern. While stereo matching can be performed between the two monochrome cameras to solve the ambiguity, the number of possible disparity is limited to M. Therefore, no dense stereo matching is needed, which allows for a fast implementation. To increase the robustness to noise, specularities, and occlusions, an optimization algorithm using loopy belief propagation followed by a consistency check is proposed to reduce the phase-unwrapping error. Meanwhile, a motion compensation method is proposed to reduce the motion error. The resulting system can give accurate depth measurement of complex dynamic scenes at 17 fps with the assistance of a GPU. This stereo-assisted phase-unwrapping idea is further extended in [40] by enforcing viewpoint and temporal consistencies.

Stereo has also been used to improve the accuracy of ToF sensors, since stereo can make use of high-resolution commodity cameras, while ToF sensors are limited in resolution. On the other hand, ToF sensors perform better than stereo on textureless surfaces, while stereo is more reliable on surfaces with rich textures that poses difficulties for ToF sensors due to the large variation of scene albedo. In [41], Zhu et al. describe a multisensor system with two color cameras and a SwissRanger SR3000 ToF sensor. The probability distribution functions of depth estimates from each sensor modality are fused using a Markov random field model, and belief propagation is then applied on the combined data from ToF and stereo to produce enhanced depth estimates through global regularization. Note that the ToF and stereo sensors need to be calibrated into a common Euclidean coordinate system in advance.

Generally, the combination of passive stereo and active illumination sensors performs better than either alone. But there is an inherent limitation for stereo: failure on textureless surfaces. In this case, only active illumination sensors can work. In [42], Zhang et al. propose a novel fusion framework to combine ToF and phase shifting, which is the first attempt to combine two active illumination sensors. The basic idea is to use the coarse, low-resolution depth from ToF to disambiguate the wrapped, high-resolution depth from phase shifting. The proposed system is shown in Figure 11, which consists of a second-generation ToF Kinect (the Kinect Gen2), an off-the-shelf DLP projector, and a monochrome camera. Specifically, two key technical issues are addressed in this work. First, both ToF and phase shifting emit light signals, so they will inevitably

interfere with each other unless precise synchronization is conducted. To this end, an interference-free synchronization strategy is designed with minimal restrictions on the system, after analyzing the properties of light signals emitted by the ToF and phase-shifting sensors. Second, calibration is essential for establishing an accurate mapping between the depth measurements from ToF and phase shifting. Thanks to the high-resolution RGB camera that is well calibrated with the ToF sensor in the Kinect Gen2, a cross-modal calibration strategy is designed to reliably relate the depth measurements from ToF and phase shifting. Based on the calibration and synchronization strategies, a 30-fps real-time depth-sensing system has been developed with the assistance of a GPU. Figure 12 shows some 3-D reconstruction results of fusion in comparison with ToF and phase shifting alone. As can be seen, this fusion method produces accurate and robust depth with high resolution and low delay, which successfully integrates the advantages of ToF and phase shifting. Note these two scenes are both with textureless surfaces, which poses difficulty for the previous fusion methods using passive stereo.

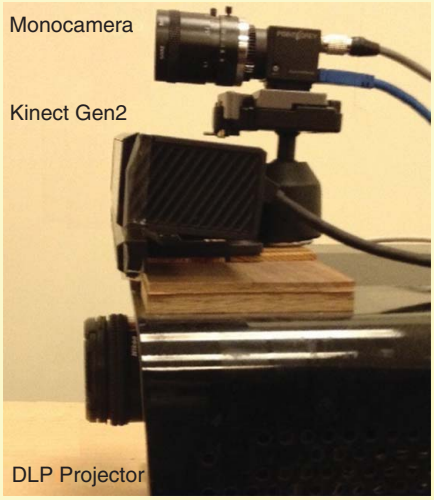


FIGURE 11. The system of ToF and phase-shifting fusion.

Summary

Now that we have reviewed the representative computational depth-sensing techniques, let us summarize by taking a deeper look into the common essence of these techniques and how these techniques will facilitate the future development of high-performance commodity depth cameras.

A common insight behind most of the techniques discussed previously is to improve the efficiency of space-time multiplexing in terms of the light signal usage, which is different from traditional structured light cameras that explicitly separate space coding and time coding for specific application scenarios (dynamic capture or static scanning). Actually, to achieve the goal of fast and accurate depth sensing, space-time multiplexing is indispensable. From this perspective, phase period coding exploits the space redundancy in time-multiplexing signals (i.e., phase-shifted fringes) to reduce the time dependency of high-accuracy phase shifting, space-time coding exploits the time redundancy in space-multiplexing signals (e.g., De Bruijn sequence, Gray code, and speckle pattern) for improved accuracy of motion-insensitive regions in the scene, and sensor fusion exploits the cross-modal redundancy to pursue even higher efficiency of space-time multiplexing through precise calibration and synchronization. On the other hand, cost is an important consideration. For structured light cameras, DLP projectors, especially high-speed ones, are usually the most expensive components. Therefore, improving the efficiency

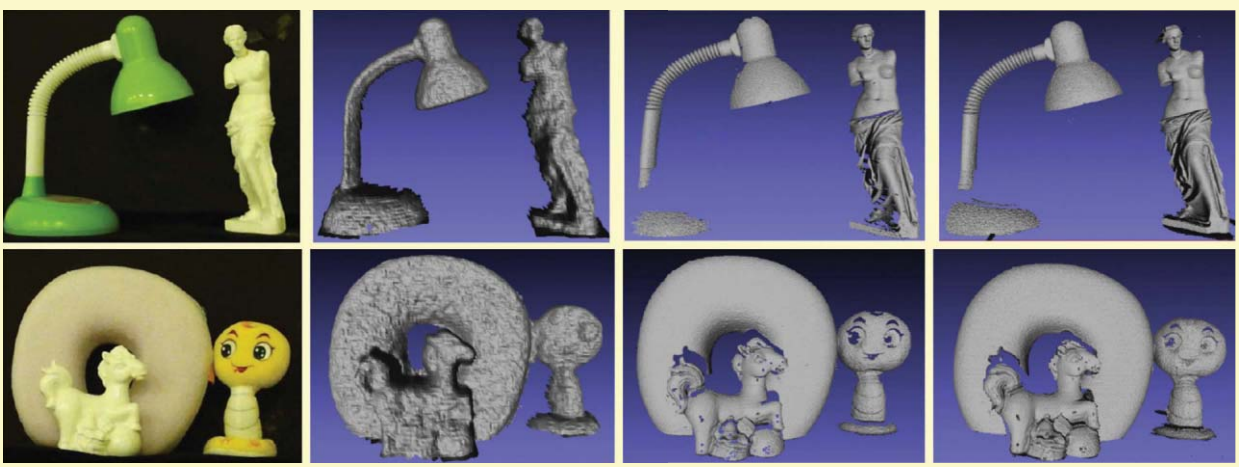


FIGURE 12. Three-dimensional reconstruction results of two scenes with textureless surfaces: (a) the color image, (b) the ToF results, (c) the results of three-frequency phase shifting (nine patterns), and (d) the results of ToF and phase-shifting fusion (three patterns).

of space-time multiplexing promotes the usage of commodity hardware while maintaining the measurement accuracy. Specifically, it is worth mentioning that binary phase shifting is a promising approach to enable the design of high-performance commodity depth cameras, as binary patterns can be easily generated using relatively inexpensive DLP projectors or laser-diffuser emitters that cost much less than DLP projectors.

Another common insight behind these techniques is that interdisciplinary design of system principles and components could break the performance limit of traditional depth sensing. From this perspective, phase period coding implicitly uses space coding to solve the phase ambiguity in phase shifting (a kind of time coding) without using a large number of patterns, space-time coding explicitly integrates advantages of space coding and time coding within a scalable depth-sensing paradigm, and sensor fusion directly combines different types of depth sensors for an improved performance while maintaining backward compatibility. Furthermore, binary phase shifting borrows the idea of dithering from digital image processing to achieve high-accuracy depth measurement with elegantly simulated sinusoidal fringes from binary patterns.

Essentially, the aforementioned insights can be summarized into two aspects: advanced signal processing algorithms and redesign of image sensor systems or elements, which together mark the era of computational depth sensing. Along with the ever-increasing computing power and the advancement of light source technology, we believe computational techniques will play a more and more important role in the development of future depth cameras.

Advanced issues

Open problems

Despite the rapid progress made in the field of computational depth sensing, there are still some major challenges that are considered important open problems.

Outdoor and global illumination

One issue that restricts the application of active depth cameras is undesired illumination. A typical scenario is outdoors, where the presence of strong sunlight severely interferes with active illumination. Attempts to overcome this difficulty usually have side effects of degraded measurement accuracy or speed [43]. Another kind of undesired illumination, called *global illumination*, refers to interreflections, surface scattering, and other effects that are not directly from the light source. Global illumination effects frequently pose obstacles for active depth cameras, and avoiding them has been investigated in the literature [44]. Nevertheless, it remains challenging to maintain measurement accuracy, speed, and reliability at the same time.

Hardware-friendly algorithm design

For real-time, high-resolution depth sensing, parallel computing is indispensable. Both the signal processing algorithm and the hardware equipped in depth cameras should support highly parallel computing. While the processing for traditional techniques such as space coding and time coding are parallelizable in principle, more sophisticated algorithms may be required for an improved performance in computational depth sensing, which are not necessarily easy to be parallelized. On the other hand, previous works generally demonstrate parallel implementation on CPU/GPU, yet seldom on resource-constrained hardware such as field-programmable gate arrays [45]. The latter, however, is closer to the real case for the deployment of commodity depth cameras. Therefore, the design of parallel signal processing algorithms adapted to resource-constrained hardware should receive more attention.

In this mobile-first world, high-accuracy depth information would enable a great deal of applications if available on mobile phones.

Depth cameras on mobile phones

In this mobile-first world, high-accuracy depth information would enable a great deal of applications if available on mobile phones. However, it is difficult to integrate active depth cameras into mobile phones, mainly due to the physical size of the light source. Although the laser-diffuser emitter used in the Kinect could meet the rigorous requirements, safety is another issue to be considered, as lasers may not be preferred for near body usage. Fortunately, the advancement of projector technology could relieve this problem. The miniaturized version of DLP projectors would soon be available on mobile phones [46]. The Lenovo Phab 2 Pro [47], the first commodity Tango-powered mobile phone with a built-in ToF depth camera, is already on the market. Still, practical issues such as energy efficiency and measurement range need to be taken into consideration for the development of depth cameras on mobile phones.

Promising applications

With improved performance and reduced cost, commodity depth cameras have found their applications in a number of emerging scenarios. Here we discuss several application scenarios that are expected to have a large impact in the future.

VR/AR

VR and AR are becoming extremely popular. However, the richness of the content is one bottleneck for this industry. As 3-D information is necessary for generating an immersive user experience, depth cameras play an indispensable role in this scenario. Equipped with a high-performance depth camera, VR/AR devices can easily obtain the 3-D information of the environment in real time and then render the reality as one wishes. That is the working principle behind the Microsoft HoloLens [8] and Google Tango [9]. To be competitive in the market, depth cameras on VR/AR devices should have low cost, compact size, and large measurement range. As can be expected, the growth of the VR/AR industry will greatly promote the development of commodity depth cameras.

Artificial intelligence/robots

Artificial intelligence has seen its renaissance recently, and intelligent robots are again attracting mass attention. While RGB-depth images are widely used in computer vision tasks and show significant advantages, a pair of RGB and depth cameras would become the default configuration for intelligent visual analysis. Future robots that are able to segment, detect, recognize, and track different objects in an environment will heavily rely on the accurate depth information. Still, depth cameras need to be customized to meet the specific requirements under different environments. For example, on drones, where depth information is essential for navigation, depth acquisition should be highly robust to ambient illumination; whereas when human beings are involved for interaction, it is better to use infrared light instead of visible light for the light source.

Future robots that are able to segment, detect, recognize, and track different objects in an environment will heavily rely on the accurate depth information.

Human–computer interaction

Game consoles have been demonstrated as a killer application of depth cameras, where human–computer interaction can be conducted in an easy and natural way. However, the potential of high-performance commodity depth cameras is far greater. Eventually, scenes from science fiction films will become real. We will be able to easily control the equipment around us with our bodies, just as we control the characters in the Kinect games. That will greatly facilitate every aspect of our daily lives—entertainment, education, surgery, and disability assistance, to name just a few.

Conclusions

By adding one missing dimension back to digital imaging, depth cameras provide indispensable information for interpreting the environment. Because of the rapid advancement of computing power and light sources, commodity depth cameras have been developed with improved accuracy and speed through the redesign of image sensor systems and advanced signal processing algorithms. In this article, we provided an overview of representative techniques of computational depth sensing, analyzed their pros and cons, and summarized the common insights behind these techniques. Looking forward, we also discussed the open problems for computational depth sensing as well as promising applications that can be explored with the availability of high-performance commodity depth cameras.

Besides depth, there are other dimensions missing in traditional digital imaging. Taking the spectrum dimension as an example, multi/hyperspectral cameras can capture color details beyond the capability of ordinary RGB cameras. Similar to depth cameras, computational techniques have been extensively used to improve the performance of multi/hyperspectral imaging when both accuracy and speed are concerned [48]. The future trend of computational imaging will integrate more dimensions, for example, depth and spectrum together [49], [50], toward the ultimate goal of resolving the plenoptic function in a single shot. We hope the principles of computational depth sensing revealed in this article help lift the roadblocks ahead and inspire brand-new imaging modalities.

Acknowledgments

We acknowledge funding from the Natural Science Foundation of China grants 61671419 and 61425026 and the Chinese Academy of Sciences Pioneer Hundred Talents Program.

Authors

Zhiwei Xiong (zwxiong@ustc.edu.cn) received his B.S. and Ph.D. degrees in electrical engineering from the University of Science and Technology of China, Hefei, in 2006 and 2011, respectively, where he has worked as a research professor since 2016. He was previously a researcher with Microsoft Research Asia. His research interests include computational imaging, inverse problems in computer vision, and visual data compression.

Yueyi Zhang (zhyue@microsoft.com) received his B.S. degree in applied mathematics and his Ph.D. degree in electrical engineering from the University of Science and Technology of China, Hefei, in 2010 and 2015, respectively. From 2011 to 2014, he was a research intern with Microsoft Research Asia, where his work concentrated on depth sensing and three-dimensional imaging. He is currently a software development engineer with Microsoft in Vancouver, British Columbia, Canada.

Feng Wu (fengwu@ustc.edu.cn) received his B.S. degree in electrical engineering from Xidian University in Xi'an, China, in 1992, and his M.S. and Ph.D. degrees in computer science from the Harbin Institute of Technology, China, in 1996 and 1999, respectively. He has been a full professor at the University of Science and Technology of China, Hefei, since 2014. He previously served as a principal researcher and research manager with Microsoft Research Asia. His research interests include computational imaging, image and video compression, media communication, and media analysis and synthesis. He is a Fellow of the IEEE.

Wenjun Zeng (wezeng@microsoft.com) received his B.E. degree from Tsinghua University, Beijing, China; his M.S. degree from the University of Notre Dame, Indiana; and his Ph.D. degree from Princeton University, New Jersey. He was with the Computer Science Department of the University of Missouri from 2003 to 2016, most recently as a full professor. He has worked for PacketVideo Corporation, Sharp Labs of America, Bell Labs, and Panasonic Technology; he is currently a principal research manager at Microsoft Research Asia. His research interests include mobile-cloud video acquisition and analysis, social network/media analysis, multimedia communications and networking, and content/network security. He is a Fellow of the IEEE.

References

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vision*, vol. 47, pp. 7–42, Aug. 2002.

[2] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 972–986, May 2012.

[3] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (tof) cameras: A survey," *IEEE Sensors J.*, vol. 11, no. 9, pp. 1917–1926, Sept. 2011.

[4] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," *Pattern Recogn.*, vol. 43, no. 8, pp. 2666–2680, Aug. 2010.

[5] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE Multimedia Mag.*, vol. 19, no. 2, pp. 4–10, Feb 2012.

[6] J. Smisek, M. Jancosek, and T. Pajdla, "3D with Kinect," in *Proc. IEEE Int. Conf. Computer Vision Workshops*, 2011, pp. 1154–1160.

[7] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with Microsoft Kinect sensor: A review," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, June 2013.

[8] Microsoft HoloLens. [Online]. Available: https://www.microsoft.com/microsoft-hololens/

[9] Google Tango. [Online]. Available: https://www.google.com/atap/project-tango/

[10] A. Bhandari and R. Raskar, "Signal processing for time-of-flight imaging sensors: An introduction to inverse problems in computational 3-D imaging," *IEEE Signal Process. Mag.*, vol. 33, no. 5, pp. 45–58, Sept. 2016.

[11] I. Ihrke, J. Restrepo, and L. Mignard-Debise, "Principles of light field imaging: Briefly revisiting 25 years of research," *IEEE Signal Process. Mag.*, vol. 33, no. 5, pp. 59–69, Sept. 2016.

[12] J. Posdamer and M. Altschuler, "Surface measurement by space-encoded projected beam systems," *Computer Graphics Image Process.*, vol. 18, no. 1, pp. 1–17, Jan. 1982.

[13] S. Inokuchi, K. Sato, and F. Matsuda, "Range imaging system for 3-D object recognition," in *Proc. Int. Conf. Pattern Recognition*, Jan. 1984, pp. 806–808.

[14] D. Caspi, N. Kiryati, and J. Shamir, "Range imaging with adaptive color structured light," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 5, pp. 470–480, May 1998.

[15] V. Srinivasan, H. C. Liu, and M. Halioua, "Automated phase-measuring profilometry of 3-D diffuse objects," *Appl. Optics*, vol. 23, no. 18, pp. 3105–3108, 1984.

[16] S. G. Narasimhan, S. J. Koppal, and S. Yamazaki, "Temporal dithering of illumination for fast active vision," in *Proc. European Conf. Computer Vision*, Oct. 2008, pp. 830–844.

[17] P. Vuylsteke and A. Oosterlinck, "Range image acquisition with a single binary-encoded light pattern," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 2, pp. 148–164, Feb. 1990.

[18] E. M. Petriu, T. Bieseman, N. Trif, W. S. McMath, and S. K. Yeung, "Visual object recognition using pseudorandom grid encoding," in *Proc. Int. Conf. Intelligent Robots Systems*, July 1992, 1617–1624.

[19] S. Y. Chen, Y. F. Li, and J. Zhang, "Vision processing for realtime 3-D data acquisition based on coded structured light," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 167–176, Feb. 2008.

[20] C. Albitar, P. Graebling, and C. Doignon, "Robust structured light coding for 3D reconstruction," in *Proc. IEEE Int. Conf. Computer Vision*, Oct. 2007, p. 1–6.

[21] S. Zhang, "Recent progresses on real-time 3D shape measurement using digital fringe projection techniques," *Opt. Laser Eng.*, vol. 48, no. 2, pp. 149–158, Feb. 2010.

[22] T. Pribanic, S. Mrvos, and J. Salvi, "Efficient multiple phase shift patterns for dense 3D acquisition in structured light scanning," *Image Vision Comput.*, vol. 28, no. 8, pp. 1255–1266, Aug. 2010.

[23] S. Zhang and P. S. Huang, "High-resolution, real-time three-dimensional shape measurement," *Opt. Eng.*, vol. 45, no. 12, pp. 123601:1–8, Dec. 2006.

[24] S. Zhang, X. Li, and S. T. Yau, "Multilevel quality-guided phase unwrapping algorithm for real-time three-dimensional shape reconstruction," *Appl. Optics*, vol. 46, no. 1, pp. 50–57, Jan. 2007.

[25] K. Liu, Y. Wang, D. L. Lau, Q. Hao, and L. G. Hassebrook, "Dual-frequency pattern scheme for high-speed 3-d shape measurement," *Opt. Express*, vol. 18, no. 5, pp. 5229–5244, 2010.

[26] P. Wissmann, R. Schmitt, and F. Forster, "Fast and accurate 3D scanning using coded phase shifting and high speed pattern projection," in *Proc. Int. Conf. 3D Imaging Modeling Processing Visualization Transmission*, May 2011, pp. 108–115.

[27] Y. Wang, K. Liu, Q. Hao, D. L. Lau, and L. G. Hassebrook, "Period coded phase shifting strategy for real-time 3-D structured light illumination," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3001–3013, Nov. 2011.

[28] Y. Zhang, Z. Xiong, and F. Wu, "Unambiguous 3d measurement from speckle-embedded fringe," *Appl Optics*, vol. 52, no. 32, pp. 7797–7805, Nov. 2013.

[29] L. Zhang, B. Curless, and S. M. Seitz, "Rapid shape acquisition using color structured light and multi-pass dynamic programming," in *Proc. 3D Data Processing Visualization Transmission*, 2002 pp. 24–36

[30] I. Ishii, K. Yamamoto, K. Doi, and T. Tsuji, "High-speed 3D image acquisition using coded structured light projection," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, Oct./Nov. 2007, pp. 925–930.

[31] Y. Taguchi, A. Agrawal, and O. Tuzel, "Motion-aware structured light using spatio-temporal decodable patterns," in *Proc. European Conf. Computer Vision*, Oct. 2012, pp. 832–845.

[32] Y. Zhang, Z. Xiong, Z. Yang, and F. Wu, "Real-time scalable depth sensing with hybrid structured light illumination," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 97–109, Jan. 2014.

[33] S. Lei, and S. Zhang, "Flexible 3-D shape measurement using projector defocusing," *Opt. Lett.*, vol. 34, no. 20, pp. 3080–3082, 2009.

[34] S. Zhang, D. V. D. Weide, and J. Oliver, "Superfast phase-shifting method for 3-d shape measurement," *Opt. Express*, vol. 18, no. 9, pp. 9684–9689, Apr. 2010.

[35] Y. Wang and S. Zhang, "Three-dimensional shape measurement with binary dithered patterns," *Appl. Opt.*, vol. 51, no. 27, pp. 6631–6636, Sept. 2012.

[36] Y. Wang, J. I. Laughner, I. R. Efimov, and S. Zhang, "3D absolute shape measurement of live rabbit hearts with a superfast two-frequency phase-shifting technique," *Opt. Express*, vol. 21, no. 5, pp. 5822–5832, Mar. 2013.

[37] Z. Yang, Z. Xiong, Y. Zhang, J. Wang, and F. Wu, "Depth acquisition from density modulated binary patterns," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2013, pp. 25–32.

[38] J. Davis, D. Nehab, R. Ramamoorthi, and S. Rusinkiewicz, "Spacetime stereo: a unifying framework for depth from triangulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 296–302, Feb. 2005.

[39] T. Weise, B. Leibe, and L. Van Gool, "Fast 3D scanning with automatic motion compensation," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, June 2007, pp. 1–8.

[40] R. Garcia and A. Zakhor, "Consistent stereo-assisted absolute phase unwrapping methods for structured light systems," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 5, pp. 411–424, Sept. 2012.

[41] J. Zhu, L. Wang, R. Yang, J. Davis, and Z. Pan, "Reliability fusion of time-of-flight depth and stereo geometry for high quality depth maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1400–1414, July 2011.

[42] Y. Zhang, Z. Xiong, and F. Wu, "Fusion of time-of-flight and phase shifting for high-resolution and low-latency depth sensing," in *Proc. IEEE Int. Conf. Multimedia Expo*, June/July 2015, pp. 1–6.

[43] M. Gupta, Q. Yin, and S. K. Nayar, "Structured light in sunlight," in *Proc. IEEE Int. Conf. Computer Vision*, Dec. 2013.

[44] M. Gupta and S. K. Nayar, "Micro phase shifting," in *Proc. Computer Vision and Pattern Recognition*, June 2012, pp. 813–820.

[45] J. Wang, Z. Xiong, Z. Wang, Y. Zhang, and F. Wu, "FPGA design and implementation of kinect-like depth sensing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 6, pp. 1175–1186, June 2016.

[46] Texas Instruments. "Products that use TI DLP technology." [Online]. Available: http://www.ti.com/lsds/ti/dlp-technology/products/dlp-pico-mobile-projectors/smartphone-projector.page

[47] Lenovo. Phab 2 Pro Smartphone. [Online]. Available: http://shop.lenovo.com/us/en/smartphones/phab-series/phab2-pro/

[48] X. Cao, T. Yue, X. Lin, S. Lin, X. Yuan, Q. Dai, L. Carin, and D. J. Brady, "Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world," *IEEE Signal Process. Mag.*, vol. 33, no. 5, pp. 95–108, Sept. 2016.

[49] J. Wu, B. Xiong, X. Lin, J. He, J. Suo, and Q. Dai, "Snapshot hyperspectral volumetric microscopy," *Scientific Reports*, vol. 6, pp. 24624:1–10, Apr. 2016.

[50] L. Wang, Z. Xiong, G. Shi, W. Zeng, and F. Wu, "Simultaneous depth and spectral imaging with a cross-modal stereo system," *IEEE Trans. Circuits Syst. Video Technol.,* to be published. doi:10.1109/TCSVT.2016.2616374

SP

Mónica F. Bugallo and Angela M. Kelly

Engineering Outreach: Yesterday, Today, and Tomorrow

This article discusses the current landscape of outreach efforts in the United States to engage K–12 students in engineering. It then provides an overview of two programs run by the College of Engineering and Applied Sciences, and the Institute for Science, Technology, Engineering, and Mathematics Education at Stony Brook University (SBU) to promote student participation and interest in engineering. These efforts are aligned with the recently released Next Generation Science Standards (NGSS), which emphasize incorporating engineering design principles in K–12 science education. We describe two models, one in the form of an on-campus summer camp and the other as a series of after-school activities with both on- and off-campus offerings. These experiences are rarely available in K–12 schools and have the added benefit of exposing students to engineering faculty and researchers. The programs are focused on electrical and computer engineering with emphasis on signal and information processing and analysis and have hosted more than 200 students for the past six years.

We argue that offering continuing education opportunities to teachers and counselors at schools will have a considerably higher impact, and we describe two innovative programs targeting those populations as well as a new format of student experiences based on one-day campus visits.

Overview

A major challenge in our increasingly technological society is the need to build awareness and excitement for engineering careers to help attract the engineers of the future. Unfortunately, students often view engineering as an unattractive and inaccessible subject and career option [1]. Contributing to this view is the traditional lack of engineering instruction in elementary, middle, and high schools (known collectively as the K–12 schools), compounded by a limited awareness of engineering knowledge and careers among teachers and school counselors [2], [3]. However, the recent adoption of the NGSS [4] by 16 states has shown tremendous promise for widespread proliferation of engineering in K–12 education.

The NGSS explicitly integrates science content knowledge, engineering practices, and cross-cutting concepts so students may identify, explain, and solve everyday problems through engineering design. The NGSS complement the American Society for Engineering Education (ASEE) K–12 Science, Technology, Engineering, and Mathematics (STEM) Guidelines for All Americans [5], which emphasize the scope of engineering practice, understanding engineering design, and applying STEM concepts to technological challenges. These standards and guidelines present an opportunity for universities to impact precollege engineering education by sharing resources and expertise. This article explores the current status of engineering education and careers in the United States, successful engineering outreach programs and curricula, and the efforts at SBU in advocating for a broader participation in engineering through university–community partnerships.

Recent reports have documented the chronic shortage of engineering talent in the United States [6]–[8]. The U.S. Bureau of Labor Statistics projected an increase of 365,000 engineering job openings, due to replacing current engineers between 2010 and 2020, and an additional need for 160,300 engineers due to new job growth [9]. It is questionable whether colleges and universities will be able to maintain the pace with engineering employment demands. There were 106,658 bachelor's degrees awarded in all engineering disciplines in 2014–2015 in the United States [5], yet there are indications that retention and diversity in undergraduate engineering programs are persistent concerns [10]. Undergraduate engineering enrollment in the United States was 541,705 in 2014, including 104,033 women, 27,163 African Americans, and 60,017 Hispanic students [11]. The graduation rate of engineering majors in the United States was 60% over six years [12], with both academic preparation and nonacademic factors contributing to attrition [12], [13]. The traditional disparities in undergraduate engineering education are reflected in

Digital Object Identifier 10.1109/MSP.2017.2673018
Date of publication: 26 April 2017

the demographics of the current workforce. In 2014, there were 1,680,854 engineers in the United States, of whom 12.9% were women, 3.6% were African American, and 8.3% were Hispanic [11]. Disparities in engineering interest develop before college, with 14.5% high school men and just 2.5% of high school women intending to major in engineering [14]. Students entering higher education to study engineering need to be prepared differently, and it is essential that more diverse populations are attracted to and retained in the field [15].

Issues in K–12 STEM education

Intrinsic to the shortage of engineering talent is a lack of tradition of engineering education in K–12 schools [16], [17]. This has been manifested in current policies and existing science and mathematics curricula, inadequate teacher preparation, and limited resources for providing appropriate student guidance. There has been much discussion about the inclusion of engineering in K–12 classrooms and disjointed state efforts to do so [18], [19], yet research into STEM integration has not kept pace with changes in policy [20]. New York State, for example, has no distinct K–12 teacher certification in engineering and does not allow engineering coursework to meet licensing requirements. New York State is not prepared to meet the incorporation of engineering content and practices necessitated when the NGSS are fully adopted [21]. These standards originated in the U.S. National Research Council's Framework for K–12 Science Education [22], which recommended curricula that incorporate cross-cutting concepts, scientific practices, and disciplinary core ideas. With most states moving forward on the adoption of the NGSS, school districts will be required to provide engineering experiences embedded within traditional science, mathematics, and technology curricula.

K–12 STEM teachers and the need for professional development

Current science and mathematics teachers will require significant professional development to incorporate engineering knowledge and design principles in their classroom teaching. Most science teachers are unfamiliar with engineering practices, lack confidence in teaching design principles, and have an inaccurate understanding of the skills and training required for engineering careers [23], [24]. Limited resources are currently available to help them overcome these restrictions [20]. Research-based principles provide a compelling model for effective professional development supporting effective NGSS implementation, which complements the goals of ASEE [25], [26]. In-service teacher professional development is necessary to facilitate meaningful integration of science content and engineering design.

Purzer et al. proposed that teacher professional development should emphasize evidence-based decision making through collaborations with science and engineering education researchers. New curricula should be developed that integrate science content and engineering practices; teachers should encourage critical thinking and the development of literacy skills with the aid of formative assessments [25]. Stereotypical contexts should be avoided (e.g., building fast cars) to reduce inequitable practices in engineering education and encourage diverse participation in engineering careers [27]. A related STEM knowledge integration model suggests that students should learn diverse ideas about science and engineering, develop evaluative criteria, and test their ideas by collecting evidence [28].

In Bell and Gilbert's model of effective professional development [29], successful teachers express a desire to improve their practice, reflect critically on their pedagogy, integrate new ideas, and become empowered to implement new strategies and inspire others through collaboration. Frequent opportunities for interactions with colleagues and mentors contribute to curricular reform efforts [30]. Theoretical support is further evidenced by research suggesting professional development should be sustained over time [31], [32]. Effective training programs typically require 50–80 hours of instruction in authentic settings before significant treatment effects are evident [33], [34].

School counselors and the need for STEM training

The need for school guidance personnel trained in appropriate precollege academic preparation for engineering study is acute. Many high school students depend on the advice of counselors in choosing elective coursework and deciding where to send college applications. Counselors often have final decision-making authority on which courses a student will take. Engineering is a discipline where gateway precollege STEM coursework determines access to and success in the college major [35], [36]. Counselors and science teachers have been highly influential in encouraging students to pursue STEM-related careers, particularly those whose parents cannot advise on necessary choices [37], [38]. Access to high-quality STEM counseling is typically limited for underresourced schools [39], where students may be dissuaded from pursuing advanced science and mathematics in ill-conceived efforts to prioritize graduation rates [40]. Also, engineering is often viewed by counselors and teachers as a course of study for the academic elite, which further diminishes encouragement [24].

Underresourced student guidance has had a dramatic impact on the preparation of high needs students to pursue engineering. Just 4% of underrepresented students have taken the mathematics and science courses required for admission to the majority of engineering schools [41]. Each successive level of mathematics and science course-taking has been associated with an 8.2% increase in the likelihood of declaring a STEM major [42]. Coursework in physics and calculus is particularly significant [43]. Research has suggested that providing support and training school counselors regarding the value of STEM coursework will have a positive impact on students' STEM performance, course choices, and awareness of STEM careers [44], [45]. Counselors are well positioned to manage the alignment between students' career expectations and curricular decisions [46]. More work needs to be done with school professionals in the position to educate high school students on the challenges and rewards of engineering study.

Theoretical support for precollege engineering education

Student-related engineering program designs are supported by a sociopsychological theoretical framework that synthesizes elements of the expectancy-value model and the theory of planned behavior. Career choice has often been associated with outcome expectations or the anticipation of probable results from chosen actions [47]. This construct is partially explained by the expectancy-value model, which suggests that behaviors are based upon two considerations: the anticipation of actual outcomes and the importance or value attached to that choice [48]. Students generally do not choose careers in which they do not feel competent, and they do not see their relevance and social value early in the academic pipeline; this is particularly true for traditionally underrepresented students in math-intensive fields such as engineering [49]. The choice of engineering majors and persistence in the field has been linked to whether students possess an engineering identity that is consistent with their sense of self [50].

The theory of planned behavior [51] built upon the expectancy-value model by suggesting that one's controllability of career choice is predicated by self-efficacy. The theory states that human behavior is guided by likely consequences, the normative expectations of others, and beliefs about inhibiting factors. Engineering may be viewed as an achievable career choice if students have the confidence that they can overcome potential obstacles along the way. For example, undergraduate engineering majors have often experienced declines in self-efficacy early in their academic majors, and social supports are necessary for overcoming their self-doubts [52]–[54]. Our educator-related project designs also incorporate expectancy value and the theory of planned behavior in professional development; we believe teachers must be aware of these constructs in appropriately advising students about engineering careers.

Current outreach efforts in the United States

Research has shown that early exposure to engineering activities can significantly increase student awareness of engineering as a rewarding career path. Effective engineering programs in K–12 education have tended to incorporate inductive teaching approaches, which are referred to as problem-based or discovery learning. Collaborative knowledge construction is another strategy for modeling engineering practices [55], [56]. When working with diverse groups of students, engineering pedagogy that is interactive and student centered helps students recognize their cultural capital and improves their overall engagement [57]. Engineering education based upon NGSS and ASEE guidelines can improve engineering knowledge and skills as well as the scientific literacy necessary to understand and solve real-world problems [58]. These pedagogical principles have guided many engineering education projects. We provide some examples of these existing programs to situate our own work in building upon successful models.

The core objective of many outreach efforts is to align activities and workshops consistent with ASEE's goal that all Americans will be able to apply concepts of science, technology, and mathematics to engineering processes and problems [5]. Previous work in the field has generated engineering curricula for students and research on their impacts has been mostly positive [58], [59], [60]. Some curricula are for full-year courses specifically in engineering. For example, Project Lead the Way (PLTW) [61] developed curricula for one-year high school courses in introductory engineering, aerospace engineering, civil engineering, digital electronics, and other engineering-related focus areas. They provide professional development, resources, and ongoing training for teachers to implement PLTW curricula effectively. A review of PLTW research revealed that participating students performed as well or slightly better than non-PLTW peers, while teachers reported increasing their STEM integration over time [62]. Engineer Your World, a high school curriculum developed by UTeach at University of Texas Austin [63], is a one-year engineering design course based on socially relevant issues. Students learn about engineering design and habits of mind while also exploring the breadth of engineering professions.

Other engineering education innovations were designed for teachers to incorporate engineering in their existing science, technology, or mathematics curricula. The Infinity Project [64] provides two-day professional development for teachers to create and implement individual design projects in their middle and high school classrooms. Engineering Is Elementary was created for elementary and middle school teachers to include engineering activities related to real-world problems [65]. Out-of-school time (OST) programs, such as In the Middle of Engineering (IME) [66], provide informal exposure to engineering activities that parallel their school-based science and physics curricula. IME is targeted toward girls in middle and high school and involves women engineers as teachers and role models. These programs and others have resulted in increased STEM interest among participating students [17], [67], an internal construct that often leads to further STEM persistence.

Current outreach efforts at SBU

Our current outreach efforts focus on OST programs targeting high school students. Participation in OST programs has been shown to improve students' interest in STEM study and careers [68], so we have developed these programs to increase the number of students who intend to major in engineering. We describe two of our outreach initiatives here to illustrate how research and best practices informed our project designs. The first one is an engineering summer camp for high school sophomores and juniors. The second program comprises school and SBU-based OST engineering programs for freshman, sophomore, and junior students. For both programs, special emphasis has been placed on recruitment of underrepresented and high needs students and financial support has been obtained to promote their inclusion.

The goal of these programs is to expose students to the challenge, passion,

and opportunity of engineering through an ample menu of hands-on activities in engineering with particular focus on the field of electrical and computer engineering. Whenever possible, tasks related to signal and information processing and data analysis are included as part of the activities. The general learning objectives include: 1) understanding and gaining appreciation for what engineers do, and, in particular, what electrical and computer engineers do; 2) learning basic theoretical and practical concepts related to the electrical and computer engineering fields; and 3) learning how to analyze an engineering challenge both qualitatively and quantitatively, how to design a solution for a problem by breaking it into smaller pieces, and how to evaluate and test the proposed solution.

The activities have been created and initiated by SBU faculty in engineering, physics, and science education with the assistance of staff and graduate students and with the support of both internal and external funding. Industry experts guide and advise on topics of interest for the activities and STEM teachers affiliated with SBU provide pedagogical and curriculum insights. The activities have a class size of 20–24 students.

There are different ranges of difficulty for the activities depending on students grade levels and backgrounds. In all activities, students are assessed on their knowledge, practical application of engineering skills, justification of designs based on data, and their ability to engage effectively in the peer-review process. Activities involve different engineering disciplines in general, but as mentioned previously, the greatest focus to date has been on electrical and computer engineering as well as computer science, leveraging the expertise of the College of Engineering and Applied Sciences faculty. As we shall see, many projects incorporate sensing or signal/data analysis, whereby students are introduced to elementary forms of signal processing techniques and basic concepts. The pedagogical design of each activity is currently aligned with the NGSS with the following guiding principles:

- Each performance expectation must combine a relevant practice of science or engineering with a core disciplinary idea and cross-cutting concept [4]. The activities combine science concepts with engineering design; for example, students learn about basic electromagnetism principles when building metal detectors.
- Students engage collaboratively in argumentation from evidence [22]. They advocate for their chosen designs by explaining their reasoning and associated evidence for their claims. For example, when building a pilotable helium balloon, they present their prototype in a peer-review process and debate various design components. They respond to diverse perspectives and assess the merits of counter arguments [4].
- When developing models, students have the opportunity to revise the designs based on evidence to optimize performance [4]. Students consider the relationships among the components of their system when making modifications.

In addition to academic activities, the programs include presentations by engineers from local industry and the Office of Admissions and Career Center at SBU to discuss career opportunities and requirements for engineering programs. This is consistent with research that suggested students career expectations are important when choosing pathways to specific postsecondary careers [47], [69].

Engineering Summer Camp

The Engineering Summer Camp was developed for high school students in their sophomore and juniors years [70]. This residential two-week university-based program was offered from 2009 to 2015 and is currently being redesigned for broader implementation. A total of 93 students have attended the camp (23 female), 16 of whom were totally or partially awarded scholarships to attend the camp based on their socioeconomic status.

The menu of activities has changed over the years and has been modernized and adjusted according to students' and instructors' feedback as well input from a board of advisors comprising teachers and social science experts. Figure 1 shows some students participating in the 2012 camp. Here we briefly describe four activities that have been offered over the years, although more than 20 different ones have been developed and instructed. Most of them have a duration of one camp day (about six hours of instruction) although there are some exceptions that require up to two days.

Understanding sonar, radar, and GPS

This activity consists of a series of experiments to highlight the simplicity of measurement of the speed of sound (which is the key to sonar operation) and object localization [which is the fundamental principle used in sophisticated applications like radar and global positioning system (GPS)]. We briefly describe two experiments related to the activity.

Experiment 1—Measuring the speed of sound

The speed of sound is measured using an experimental setup consisting of a speaker and two microphones. The speaker generates a pulse waveform, which is recorded on each of the microphones. The time delay between the subsequent arrivals of the waveform at the microphones is measured using a PC-based software oscilloscope and the sound card. Based on the known distance between the microphones and the measured delay, the students can calculate the speed of sound.

Experiment 2—Object localization

Students learn the concepts of trilateration and multilateration. These methods allow determination of an object location in a sonar or radar system using time of flight or time difference of arrival, respectively. Multilateration is also used to determine location in GPS receivers. The effect of measurement errors is also discussed, along with some techniques to optimize the solution in that case. The students calculate the algebraic solution for the location of an object using trilateration in a noise-free case. The data for this case can either come from oscilloscope measurements performed by the students (in which case there is some small error), or the students can be given synthesized data.

(a) (b)

FIGURE 1. Different activities from the 2012 Engineering Summer Camp for high school students: (a) understanding frequency with speech and music and (b) line-following robot.

The students also develop a method for solving the multilateration problem with noisy measurements using the computer. This can be accomplished either in Excel or in a programming language such as MATLAB, depending on the students' background.

Understanding frequency with speech and music

This is a series of experiments in which students are led to an understanding of the importance of the concept of frequency in everyday signals, mainly in speech and music. The experiments are performed in real time on dedicated digital signal processing (DSP) chips, using a visual programming environment. Audio clips and the students' own voices are taken as inputs via microphones, and loudspeakers are used as the main outputs. In addition, preset oscilloscopes are used to obtain a real-time visual concept of the outputs. The experiments enable the students to create sound effects on their own, in addition to performing assigned tasks. Figure 1(a) shows students in the DSP laboratory working on various experiments.

Experiment 1—Effects of suppressing and removing frequencies from a signal

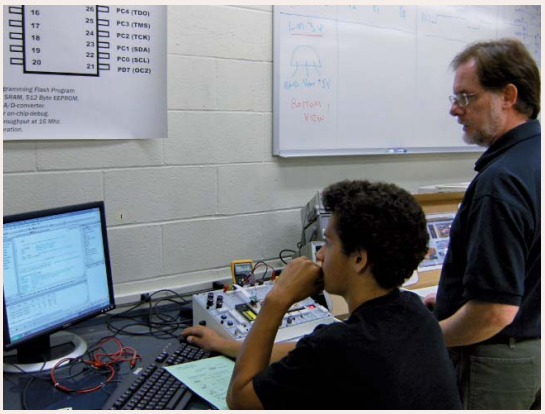
Students first synthesize and play sinusoids of various frequencies and change frequencies while listening to the outputs. They then synthesize and play sums of sinusoids, both harmonically related and nonharmonic, and change relative amplitudes while listening to the outputs. By inputting an audio signal (music or the student's voice) to the DSP board and listening to the effects of preprogrammed filters (high pass, bandpass, and low pass) on the signal, they observe how these effects change as the cut-off frequencies are altered. The same exercise is repeated by inputting a recorded audio signal corrupted by noise and using a low-pass filter to lower the noise audibility. Finally, some initial concepts related to Fourier manipulation of signals is introduced, and students synthesize or input a square-wave and listen to the output. Then they pass the square-wave through a narrow-band, bandpass filter, and vary the center frequency to identify the sinusoidal components. The outputs are observed both audibly and on the oscilloscope.

Experiment 2—Effects of shifting and scrambling frequencies

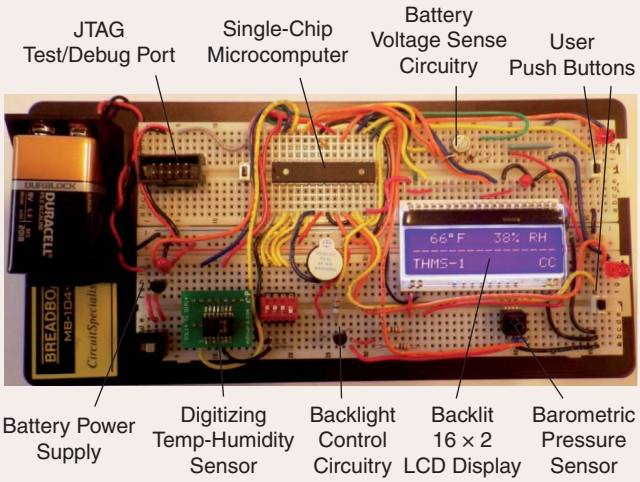
A preprogrammed frequency shifter is used to shift the frequencies of voice inputs in both directions (up and down) by up to an octave to demonstrate the effects of pitch changes. The frequency shifter demonstrates the limitations of sampling by continually raising the output frequency until aliasing converts it into a low frequency. Also, the frequency shifter is set to half of the sampling frequency, which results in spectral inversion (high frequencies are converted to low and vice versa). The result is a simple voice scrambler, which is tested on the students' voices. Finally, a more complex voice scrambler, based on multiband spectral shifting and inversion, is demonstrated and again tested with the students' voices.

Line-following robot

Students learn concepts related to a line-following robot, a mobile machine that automatically follows a specified path without the need for human steering [Figure 1(b)]. Such a machine has various applications in areas such as industrial automation, warehousing, and automatic guided vehicles on roads of the future. A line-following robot has three main components: a sensing system, a drive system, and a microcontroller. The sensing system is responsible for determining the position of the robot with respect to the line it has to follow; the drive system generates the motion of the robot; and the microcontroller runs the control algorithm that controls the speed and direction of the robot along the specified line. Students build, program, and test a line-following robot. The sensing system consists of six reflective optical sensors. These sensors have a light-emitting



(a)



(b)

FIGURE 2. The creating prototypes activity at the 2012 Engineering Summer Camp: (a) the lab session and (b) the final device, a portable battery-powered temperature, humidity, and barometric pressure-monitoring system. JTAG: Joint Test Action Group.

diode (LED) and a phototransistor. The LED emits a light toward a surface, and the phototransistor enables measurement of this light reflected from the surface. The line is specified with a black color on a white/light colored surface. When an optical sensor is above the white surface, a large amount of light is reflected back to the phototransistor. If, on the other hand, the optical sensor is directly above the black line, very little light is reflected back to the phototransistor. Thus, by using a sensor array on the bottom of the robot, students determine the position of the robot with respect to the line by measuring the outputs of all the sensors. The drive system for our robot consists of two small dc motors. The shafts of the motors are coupled to rubber wheels attached to axles connected to the main body of the robot. The torque generated by the motors is transferred to the wheels to give motion to the robot.

- Students use an Atmel ATmega8 microcontroller. A control algorithm is implemented that controls the speed and direction of the robot. The microcontroller has six analog-to-digital converter channels that are connected to the outputs of the six sensors. This allows the microcontroller to determine the position of the robot with respect to the line and control the speed and the direction of the robot accordingly. Pulsewidth modulation (PWM) varies the speed of the dc motors and allows for varying speed of the motor by changing the width of successive pulses sent to the motor. These pulses are fed to the motor through a simple drive circuit consisting of a logic-level metal–oxide–semiconductor field-effect transistor and a diode. The higher the width of the pulses, the faster the motor rotation and robot speed and vice versa. The direction of the robot is controlled by a differential mechanism whereby the speed of one of the motors is increased or decreased with respect to the other to turn the robot in a particular direction.
- The complexity of the path and the speed with which the robot can follow it depends upon the control algorithm implemented in the microcontroller. Students explore three types of algorithms; in other words, bang-bang control, proportional control, and proportional-derivative-integral control. Based on the observed results of the line following, they tune the control parameters of these algorithms to achieve better performance.

Creating prototypes

This topic is essentially "Microcomputers 101." Students learn and utilize fundamental microcomputer system design techniques, resulting in the construction of a fully working design prototype. The design is a simple ambient temperature monitoring system. This two-day activity has a lecture/laboratory format. A full design overview is provided, and by the end of the second day, each student has constructed, fully tested, and optimized the system prototype.

- Each day students spend approximately two hours in lecture, and the remaining time in the Embedded Systems Design Laboratory. The lectures present important theoretical descriptions of the hardware and software utilized for the implementation of the system. The lab periods are spent constructing, testing, troubleshooting, and verifying proper system operation of their prototype [Figure 2(a) shows a lab session of the activity].
- Theoretical concepts during the initial day include discussions about system block diagram (high level), the system schematic, and fundamental operations, as well as basic breadboarding concepts. The lab session is used to give an overview of the

breadboarding system and wiring techniques, explanations on the solderless breadboard and system parts layout, a description of interconnect and wiring techniques, and system wiring. For the second day, the lecture revolves around concepts of data processing and collection, application program high-level flow chart, introduction to application modules and coding, and system troubleshooting and operation. The practice consists of continuation of system wiring, applying power (the strobe test), troubleshooting and verifying basic system operation, and adding one or more system extensions as time permits. Figure 2(b) shows the final device with the different components.

In addition to the strictly academic activities, students are also exposed to campus life and a variety of extracurricular events, for example, meeting with engineering admissions staff and having lunch with engineers working in university laboratories and industry. On the last day of the camp, there is a showcase of built devices and experimental results from the different activities (see Figure 3), and a panel of judges decide on different awards that are given during a closing ceremony.

Data were collected from 38 students over a two-year period (2012–2013); this was done after the researchers improved the previous instruments to collect more nuanced data on the students' engineering knowledge and attitudes. Students significantly improved their knowledge of electrical and computer engineering principles and processes as measured by pre/postassessments; this outcome was observed for both female and male participants. Students' confidence in performing engineering tasks also significantly improved as a result of their participation, although motivation for engineering careers did not change (likely due to self-selection for the camp) [70]. Qualitative data revealed students felt empowered by making connections between engineering principles and their personal experiences and interests, as well as optimizing and improving functionality of their designs. Survey responses indicated students particularly enjoyed meeting with university researchers and industrial engineers. These interactions helped students strengthen their engineering self-identity and envision themselves in engineering occupations in the future [71]. The camp activities were modified over several years and serve as the test bed for our other initiatives with expanded outreach to students and science teachers.

After-school engineering offerings

The after-school engineering program was developed with local school districts to inspire high school students in grades 9–11 with the opportunities and rewards of participating in engineering. It was piloted in the fall of 2015 and is mostly an off-campus program hosted by local schools with at least a one-time campus visit per student group. It consists of several out-of-school offerings (of about two and a half hours each) spread throughout the academic year for at least 24 hours of exposure to engineering disciplines as well as computer science, with an emphasis on the processing of signals and data related to different technological problems. The activities combine exploration of theoretical concepts with hands-on practice. Approximately 72 students have benefited from these preliminary offerings with nearly half of the attendees identified as female students. Moreover, in these first offerings, all students attended schools in high needs districts. Qualitative and quantitative data were collected to measure student impacts. Our preliminary research showed that students were enthusiastic about learning about engineering and programming to design solutions, and they were more motivated to pursue engineering after participating in the program. However, they did note that they were generally dissatisfied with school counseling on engineering study and careers—a finding that confirms our recent efforts [72], [73]. In the future, we will train K–12 science teachers to incorporate the activities in their curricula.

The activities are continuously reviewed and adapted according to state-standardized curricula and feedback from teachers and students. Figure 4 displays students engaged during the 2015 offerings. We briefly describe three activities as examples of our efforts: persistence of vision clock, discovering the radio, and a night-light. We note that, due to the time constraints at each school visit, one activity is usually spread out across more than one day.

Persistence of vision clock

In this activity, students learn how our vision is somewhat deceptive, and many types of visual displays take advantage

FIGURE 3. The 2013 Engineering Summer Camp for high school students at SBU. Students participate in the engineering exhibition and competition at the closing ceremony.

of these optical illusions. The offering is motivated with real-world examples, and students learn that our perception of a rapidly flickering light source being constantly illuminated is called the *persistence of vision*. We take advantage of this property to display the time and other text using a single row of LEDs.

- The foundation of the project is a small microcontroller that we custom program in assembly language. This microcontroller is capable of executing millions of instructions each second and is responsible for flashing the LEDs at the required speed. The LEDs are moved across our field of vision leaving a trail of flashes that appear as text floating in space.
- The project is based on a custom printed circuit board and requires soldering skills in its assembly. Most of the computer code is prewritten, and the student makes changes to customize the unit to display the desired text.

Experiment 1—Understanding persistence of vision

Students use function generators and LEDs to demonstrate the phenomenon of persistence. We detect the lowest flashing rate that appears constant to each student, we move the LED and observe the "trail" that the flickering LED leaves, and we observe the effect of duty cycle on apparent brightness.

Experiment 2—Building the project and coding

Students solder to assemble the project and test the board. They learn enough assembly language programming to make simple changes to the microcontroller program. This enables the unit to display an arbitrary string of text that the student chooses [see Figure 4(a)]. In addition, the unit is capable of displaying the time of day.

Experiment 3—Strobe effects

The project has a mode that flashes the LEDs at an adjustable rate. Students use this feature to observe rotating objects and measure their corresponding rotational speeds. They also demonstrate effects related to sampling at speeds greater than the Nyquist frequency.

Discovering the radio

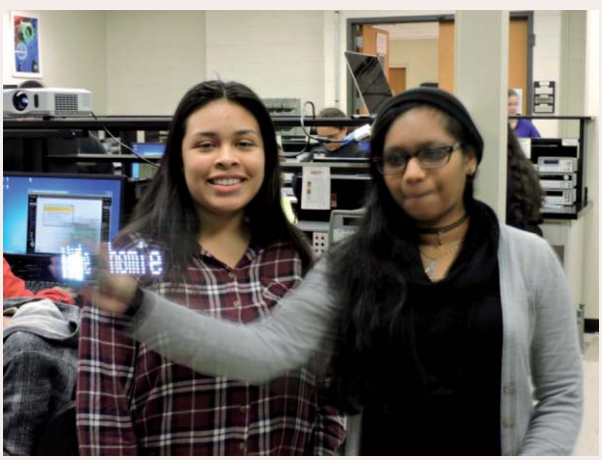
In this laboratory exercise, students learn the basic theory of amplitude modulation and detection as used in the transmission and reception of AM radio signals. They build a tuned-radio-frequency (TRF) one-chip AM radio from a dedicated kit.

- In the process of building the radio kit, students become familiar with circuit components such as variable capacitors, air-wound inductors, electrolytic capacitors, resistors, and, of course, the single integrated circuit chip used for detection. They also learn about transistor audio amplifier stages and become acquainted with the notion that the job of engineers is to design and build properly functioning circuits. Students learn the processes of AM tuning, detection, and audio amplification as they complete the various stages of the kit.
- Familiarity with small hand tools is useful but not required, as the skill can be rapidly acquired in this experiment. Soldering is required; however, students quickly learn the necessary techniques even with limited prior experience [see Figure 4(b)].

A nightlight

This is a simple project that allows novice engineers to apply basic electrical engineering concepts to daily life. The students are given materials to create an optical switch-activated LED module or, in layman's terms, a nightlight. Concepts related to voltage dividers, photo-resistors, transistor functioning, and the handling of a prototype breadboard are introduced. With the completion of this project, the students have the introductory skills necessary to design their own electrical engineering projects.

- This project uses a straightforward dc-analog design. A single 9-V battery powers the circuit. The voltage



(a)

(b)

FIGURE 4. Activities from the 2015 After-School Engineering Program for students in grades 9–11: (a) persistence of vision clock and (b) discovering the radio.

divider uses a photoresistor, which can vary from 27,000 Ω to 200,000 Ω, a sufficient range to function as a switch. Following the voltage divider, a simple positive-negative-positive bipolar junction transistor is implemented in common collector mode to increase current flow. Finally, the output consists of an array of LEDs connected in a series to serve as a light source.

- For those students who wish to work on more advanced designs, additional modules are available.

Experiment 1—On/Off toggle switch

An additional toggle switch can be added to break the connection from the battery source to the rest of the circuit. This can help tremendously in saving battery life.

Experiment 2—Fine-tuning with a potentiometer

A potentiometer is used to replace the 100,000 resistor in the circuit. Varying the impedance on this potentiometer adjusts the sensitivity of the light switch. Students get a schematic, the materials for the project, and a short lesson on how the electronic circuit works. Figure 5(a) displays the students working in the lab to build the night-light and (b) shows the schematic of the project.

On-going endeavors

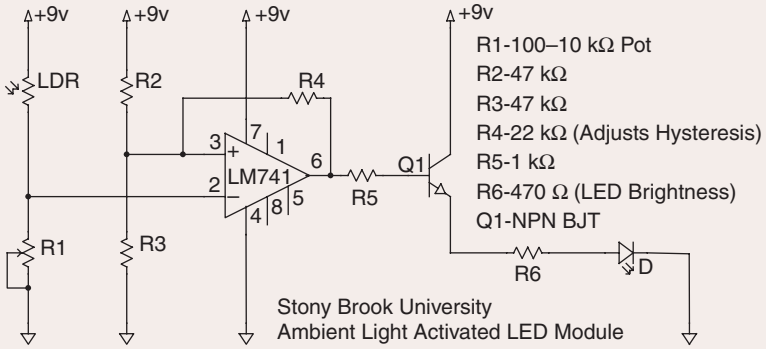
Our ongoing work builds upon what we have learned from past outreach efforts as well as the research base in engineering education. These projects are to be implemented in the coming year with a pilot design phase and accompanying research components to measure impacts empirically. This allows us to make formative changes and maximize programmatic effectiveness for scaling the following year. Our projects are designed to attract, retain, and support precollege students in engineering. We plan to educate school counselors and science teachers on the diversity of engineering career pathways as well as engineering disciplinary knowledge and process skills. By targeting these two groups, we will build capacity and competence for studying engineering, a profession that contributes to global technological advancement. The overarching goals are twofold. Engineering should not be viewed as a separate discipline but rather an essential component of students scientific literacy, complementing traditional science content with structured opportunities to design solutions to scientific problems. Furthermore, the field of engineering will be diversified with students from an untapped talent pool to contribute to the global competitiveness of the United States.

Counselors and teachers

Our work with science teachers and school counselors began in 2017, with our previous work developing engineering activities for students as the starting point for professional development. We will train science teachers to incorporate engineering design in their instruction, and we will work with school counselors on their efforts to advise students on pre-engineering coursework and the diversity of engineering careers. The professional development workshops will be modeled upon previous similar offerings at SBU. With external support, summer STEM education workshops have been offered for elementary teachers, middle school mathematics and science teachers, and high school chemistry and physics teachers [74]. Our theoretical model reinforces our emphasis on professional development in authentic settings for both science teachers and school counselors. In addition, the teacher workshops will incorporate ASEE's Standards for Preparation and Development for Teachers of Engineering [75], which include literacies in engineering design, engineering careers, and engineering and society.

The science teacher workshops will educate teachers to incorporate engineering aspects of the NGSS in their New York State-standardized science curricula. Twenty-four participants will attend each four-part workshop, with each teacher impacting approximately 150 students per academic year. We expect to recruit four cohorts during the first year with expansion in subsequent years. Participants will build their engineering skills by applying design principles while teaching science content and process. Each cohort will be immersed in a program of mutually reinforcing

(a)



(b)

FIGURE 5. The night-light activity at the 2015 After-School Engineering Program: (a) students soldering the circuit and (b) the schematic of the project.

components: 1) introductory work in engineering related to their curricula in living environment, physics, chemistry, and earth science; 2) classroom-based action research that builds teachers' ability to use data as a formative assessment to inform instruction; and 3) collaborations with engineers and STEM researchers to learn about engineering pathways and careers. Teachers will learn activities that we have previously piloted and have the flexibility to modify them for their students. Each activity will include a detailed explanation of science content and how it relates to state standards and the NGSS, followed by instruction in engineering pedagogical content knowledge. The teachers will learn a variety of assessment strategies for informing their instruction, for example, rapid, response systems, performance tasks, and questioning techniques. They will be encouraged to participate in professional learning communities to share their knowledge with other teachers in their districts and strengthen their commitment to engineering integration.

School counselors will participate in workshops to build their knowledge base in advising students about appropriate precollege engineering coursework and engineering career pathways. Once again, our prior work with students provided data to inform the content and structure of this professional development. The counselors will be immersed in a one-day training including diversity training; introductory work in engineering related to supporting competencies in science and mathematics curricula; and informative talks with engineers, STEM researchers, and university staff to learn about the diversity of engineering employment opportunities. The professional development workshops will be held at different off-campus sites and led by engineering and science education faculty and university staff. Discussions about science content and how it relates to New York State Standards and the NGSS will also be part of the training. The workshops will involve industrial engineers and staff from the Admissions Office and Career Center at SBU, so counselors will learn about qualifications for schools of engineering and specific disciplinary skill sets. The counselors will be recruited from the 125 school districts in the region, and the broader impact will be considerable since they interact with 175–300 students per academic year.

Engineering teaching laboratories

This outreach component has been modeled upon existing teaching labs in biotechnology and chemistry that have been offered at SBU since 1992, where students in grades 6–12 participate in inquiry-based experiences not readily available in their schools. More than 5,000 students have participated each year, and data have shown immediate increased student motivation to pursue STEM [76]. However, the initial offerings were a one-day-only experience for students and long-term impacts were not measured. This initiative expands and builds upon the models success, with the ultimate goal that teachers will adapt these OST engineering teaching laboratories into their classroom science instruction. In doing so, the project may be scaled to impact more students. The evaluation of prior pilot activities supports the age-appropriateness of this and other proposed activities for students in grades 9–11.

Students will come to campus during the school day and spend six hours working on a proposed hands-on activity that is aligned with the NGSS. Here we describe two activities that are in the pilot stage with full implementation scheduled for the coming year.

Linking fiber optics

The goal of this activity is to teach basic engineering concepts related to communications with an emphasis on fiber optics. The activity involves engineering theory related to transmitters and receivers; physics content knowledge related to Snell's law, refractive indices, Ohm's law, and electrical components of a circuit board (aligned with New York State's Physical Setting Standards [77]); and engineering skills such as soldering, testing functionality, debugging systems by detecting and isolating malfunctions, and minimizing signal distortion. Required materials include basic electronic components that will be purchased so students can build their own prototypes. Students discuss and debate the advantages and limitations of fiber-optic communication, optimal designs based on their own evidence, societal impacts of this technology, and potential future developments in communication.

Competitions are also part of the activity. For example, students receive an arbitrary length of fiber link, and they test the maximum distance for which reliable communication is maintained. They then increase this distance by using their knowledge and creativity. Solutions involve increasing the input power of the LED or the amplifier gain by using a different resistor.

Learning images through apps

Students learn about images through app programming. The activity involves the introduction to computer science-related concepts such as pixels, digital images, and movies; science content knowledge related to optics and communication (aligned with New York State's Physical Setting Standards [77]); and programming skills such coding, debugging, and code optimization. Required materials include a laptop with the appropriate software (we will use the open-source web application App Inventor for Android) and an Android tablet to download and test the product (we had most of the devices in place as part of previous outreach offerings, and we will renew existing materials as the project progresses). Most regional school districts indicated that these materials are available in their schools. Students will first learn image-related topics (pixel, RGB color model, or intensity) and programming concepts (for example, control flow instructions) using the open-source computing environment Octave. They then will learn how to develop mobile apps using App Inventor. Differences between Octave and App Inventor will be discussed, especially on issues related to their capabilities when dealing with images. Students will be instructed to create an app step by step, to troubleshoot and download the apps to Android devices and, finally, test them. Later, engineering teams will have an app competition.

Some prototype apps will be provided, and each group will decide to either add innovative elements to the existing prototypes or create a completely new one. At the end of the activity, each group will give a brief presentation.

Concluding remarks

Engineering education is at a crossroads with recent efforts to create inspiring engineering experiences for K–12 students. There is a persistent need to attract and retain students in engineering postsecondary study and careers, and educators and policy makers have responded with widespread adoption of the NGSS to incorporate engineering knowledge and skills in science instruction. We aim to advance engineering education by creating and refining programs that improve STEM teaching and learning by building passion, preparation, and confidence for engineering study among secondary students.

Our programs involve several stakeholders—students, teachers, and counselors—in a multifaceted effort to address weaknesses in precollege engineering accessibility. More students will be exposed to engineering as a means to solve problems by applying scientific knowledge, and their teachers and counselors will have the skills to communicate these processes and advise students on academic trajectories that lead to engineering careers. We will continue leveraging the expertise of engineering and science education faculty to designing innovative experiences that ultimately diversify the engineering talent pool. Although our previous and current student offerings focus on electrical and computer engineering with signal and information processing, we plan to expand to other engineering disciplines to offer students a broader vision of engineering careers. In doing so, we hope to contribute to the knowledge base in engineering education so effective outreach strategies might be incorporated in classroom teaching and advisement.

Acknowledgments

We would like to thank the National Science Foundation for its support through awards EEC-1647405 and CCF-1617986. We would also like to thank Dr. Keith Sheppard, director of the Institute for STEM education, for his support of our initiatives.

Authors

Mónica F. Bugallo (monica.bugallo@stonybrook.edu) is an associate professor of electrical and computer engineering and the faculty director of the Women in Science and Engineering program at Stony Brook University, New York. She received her B.S., M.S., and Ph.D. degrees in computer science and engineering from the University of A Coruña, Spain. Her research interests are in the field of statistical signal processing, with emphasis on the theory of Monte Carlo methods and its application to different disciplines including biomedicine, sensor networks, and finance. She is also committed to STEM education and has initiated several successful programs with the purpose of engaging students at all academic stages in the excitement of engineering and research, with a particular focus on underrepresented groups. She has authored and coauthored two book chapters and more than 150 journal papers and refereed conference articles.

Angela M. Kelly (angela.kelly@stonybrook.edu) is an associate professor of physics and astronomy and the associate director of the Science Education Program at Stony Brook University, New York. She attended La Salle University, Philadelphia, Pennsylvania, where she received her B.A. degree in chemistry, and completed her M.A. and Ph.D. degrees in science education (2000 and 2006, respectively) and her Ed.M. degree in curriculum and teaching (2007) at Teachers College, Columbia University. She is the recipient of the SUNY Chancellor's Award for Excellence in Teaching (2016); the Provost's Faculty Recognition Award for Excellence in Scholarship and Research from Lehman College, City University of New York (2010); and the Outstanding Teaching Award from Teachers College, Columbia University (2006). Her research has been rooted in a commitment to equity in precollege science and engineering.

References

[1] P. R. Aschbacher, E. Li, and E. J. Roth, "Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine," *J. Res. Sci. Teaching*, vol. 47, no. 5, pp. 564–582, 2010.

[2] R. H. McCuen and J. Greenberg, "Educating guidance counselors on engineering as a career and academic choice," *J. Prof. Issues Eng. Educat. Practice*, vol. 135, no. 3, pp. 91–94, 2009.

[3] M. Robinson, M. Sami Fadali, G. Ochs, and R. J. Quinn, "How do high school mathematics and science teachers cover engineering and technology?" *Frontier Educat.*, 2002, vol. 1, pp. T1CT–18.

[4] Achieve Inc. (2016). Next Generation Science Standards [Online]. Available: http://www.achieve.org/next-generation-science-standards

[5] American Society for Engineering Education. (2008). The Corporate Member Council: K–12 STEM guidelines for all Americans [Online]. Available: https://www.asee.org/member-resources/councils-and-chapters/corporate-member-council/special-interest-group/cmc-k12-stem-guidelines-for-all-americans.pdf

[6] National Academy of Sciences. "Rising above the gathering storm: Energizing and employing America for a brighter economic future," Tech. Rep., National Washington, DC, Academies Press, 2007.

[7] J. P. Holdren, M. Cora, and S. Suresh, "Federal STEM education, 5-year strategic plan: A report from the Committee on STEM Education," Tech. Rep., National Science Technology Council, Office of the President of the United States, Washington, DC, 2013.

[8] National Research Council. "Changing the conversation: Messages for improving public understanding of engineering," Tech. Rep., Washington, DC, National Academies Press, 2008.

[9] J. F. Sargent, "The U.S. science and engineering workforce: Recent, current, and projected employment, wages, and unemployment," Tech. Rep., Washington, DC, Congressional Research Service, 2013.

[10] J. F. Sargent, "Disparities in STEM employment by sex, race, and hispanic origin," Tech. Rep., American Community Survey Reports, U.S. Department of Commerce, U.S. Census Bureau, Washington, DC, 2013.

[11] National Science Board. (2016). Science and engineering indicators [Online]. Available: https://www.nsf.gov/statistics/2016/nsb20161

[12] R. M. Mara, K. A. Rodgers, D. Shen, and B. Bogue, "Leaving engineering: A multi-year single institution study," *J. Eng. Educat.*, vol. 101, no. 1, pp. 6–27, 2012.

[13] A. L. Griffith, "Persistence of women and minorities in STEM field majors: Is it the school that matters?" *Econ. Educat. Rev.*, vol. 29, no. 6, pp. 911–922, 2010.

[14] C. Hill, C. Corbett, and A. S. Rose, "Why so few? Women in science, technology, engineering, and mathematics," Tech. Rep., Washington, DC, American Association of University Women, 2010.

[15] H. Kimmel, J. Carpinelli, and R. Rockland, "Bringing engineering into K–12 schools: A problem looking for solutions?" in *Proc. 2006 American Society of Engineering Education Annu. Conf.*, 2007.

[16] National Academy of Engineering. "K–12 education," Tech. Rep. 3, Washington, DC, The Bridge, pp. 5–10, 2009.

[17] T. J. Moore, K. M. Tank, A. W. Glancy, and J. A. Kersten, "NGSS and the landscape of engineering in K–12 state science standards," *J. Res. Sci. Teaching*, vol. 52, no. 3, pp. 296–318, 2015.

[18] X. S. Apedoe, B. Reynolds, M. R. Ellefson, and C. D. Schunn, "Bringing engineering design into high school science classrooms: The heating/cooling unit," *J. Sci. Educat. Technol.*, vol. 17, no. 5, pp. 454–465, 2008.

[19] R. L. Carr, L. D. Bennett, and J. Strobel, "Engineering in the K–12 STEM standards of the 50 U.S. states: An analysis of presence and extent," *J. Eng. Educat.*, vol. 101, no. 3, pp. 539–564, 2012.

[20] G. H. Roehrig, T. J. Moore, H. H. Wang, and M. S. Park, "Is adding the E enough? Investigating the impact of K–12 engineering standards on the implementation of stem integration," *School Sci. Math.*, vol. 112, no. 1, pp. 31–44, 2012.

[21] A. M. Waldron, K. Sheppard, D. Spencer, and C. A. Batt, "Too small to see: Educating the next generation in nanoscale science and engineering," in *Nanofabrication Towards Biomedical Applications: Materials and Methods*, C. Kumar, J. Hormes, and C. Leuschner, Eds., Hoboken, NJ: Wiley, 2005, pp. 373–389.

[22] H. Schweingruber, T. Keller, and H. Quinn, "A framework for K–12 science education: Practices, crosscutting concepts, and core ideas," Tech. Rep., Washington, DC, National Academies Press, 2012.

[23] S. Yasar, D. Baker, S. Robinson-Kurpius, S. Krause, and C. Roberts, "Development of a survey to assess K–12 teachers perceptions of engineers and familiarity with teaching design, engineering, and technology," *J. Eng. Educat.*, vol. 95, no. 3, pp. 205–216, 2006.

[24] M. J. Nathan, N. A. Tran, A. K. Atwood, A. Prevost, and L. A. Phelps, "Beliefs and expectations about engineering preparation exhibited by high school STEM teachers," *J. Eng. Educat.*, vol. 99, no. 4, pp. 409–426, 2010.

[25] S. Purzer, T. J. Moore, D. Baker, and L. Berland. (2014). Supporting the implementation of the Next Generation Science Standards (NGSS) through research: Engineering [Online]. Available: http://narst.org/ngsspapers/engineering.cfm

[26] S. S. Guzey, K. Tank, H.-H. Wang, and G. Roehrig, "A high-quality professional development for teachers of grades 3–6 for implementing engineering into classrooms," *School Sci. Math.*, vol. 114, no. 3, pp. 139–149, 2014.

[27] C. Schnittka and R. Bell, "Engineering design and conceptual change in science: Addressing thermal energy and heat transfer in eighth grade," *Int. J. Sci. Educat.*, vol. 33, no. 13, pp. 1861–1887, 2010.

[28] H.-S. Lee, M. C. Lin, K. Varma, and O. L. Liu, "How do technology-enhanced inquiry science units impact classroom learning?" *J. Res. Sci. Teaching*, vol. 47, no. 1, pp. 71–90, 2010.

[29] B. Bell and J. K. Gilbert, *Teacher Development: A Model from Science Education*. New York: Psychology Press, 1996.

[30] K. S. Davis, "Change is hard: What science teachers are telling us about reform and teacher learning of innovative practices," *Sci. Educat.*, vol. 87, no. 1, pp. 3–30, 2003.

[31] E. Banilower, "Results of the 2001-2002 study of the impact of the local systemic change intiative on student achievement in science," Tech. Rep., Washington, DC, National Science Foundation, 2002.

[32] H. Borko, "Professional development and teacher learning: Mapping the terrain," *Educational Researcher*, vol. 33, no. 8, pp. 3–15, 2004.

[33] J. A. Supovitz and H. M. Turner, "The effects of professional development on science teaching practices and classroom culture," *J. Res. Sci. Teaching*, vol. 37, no. 9, pp. 963–980, 2000.

[34] K. S. Yoon, T. Duncan, S. Lee, B. Scarloss, and K. L. Shapley, "Reviewing the evidence on how teacher professional development affects student achievement," Tech. Rep., Department of Education, Washington, DC, 2007.

[35] G. Crisp, A. Nora, and A. Taggart, "Student characteristics, pre-college, college, and environmental factors as predictors of majoring in and earning a STEM degree: An analysis of students attending a hispanic serving institution," *Am. Educat. Res. J.*, vol. 46, no. 4, pp. 924–942, 2009.

[36] W. Tyson, R. Lee, K. M. Borman, and M. A. Hanson, "Science, technology, engineering, and mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment," *J. Educat. Students Placed at Risk*, vol. 12, no. 3, pp. 243–270, 2007.

[37] C. West-Olatunki, L. Shure, R. Pringle, T. Adams, D. Lewis, and B. Cholewa, "Exploring how school counselors position low-income African American girls as mathematics and science learners," *Professional School Counseling J.*, vol. 13, no. 3, pp. 184–195, 2010.

[38] J. L. Moore, "A qualitative investigation of African American males' career trajectory in engineering: Implications for teachers, school counselors, and parents," *Teachers College Rec.*, vol. 108, no. 2, pp. 246–266, 2006.

[39] Z. B. Corwin, K. M. Venegas, P. M. Oliverez, and J. E. Colyar, "How appropriate guidance affects educational equity," *Urban Educat.*, vol. 39, no. 4, pp. 442–457, 2004.

[40] V. E. Lee and R. B. Ekstrom, "Student access to guidance in high school," *Am. Educat. Res. J.*, vol. 24, no. 2, pp. 287–310, 1987.

[41] 2008 Graduating Scholars Survey. National Action Council for Minorities in Engineering Inc. (2008) [Online]. Available: http://www.nacme.org/research-publications

[42] M. E. Engberg and G. C. Wolniak, "College student pathways to the STEM disciplines," *Teachers College Rec.*, vol. 115, no. 1, pp. 010304, 2013.

[43] J. Trusty, "Effects of high school course-taking and other variables on choice of science and mathematics college majors," *J. Counseling Dev.*, vol. 80, no. 4, pp. 464–474, 2002.

[44] N. A. Fouad, "Career linking: An intervention to promote math and science career awareness," *J. Counseling Dev.*, vol. 73, no. 5, pp. 527–534, 1995.

[45] N. A. Fouad, G. Hackett, P. L. Smith, N. Kantamneni, M. Fitzpatrick, S. Haag, and D. Spencer, "Barriers and supports for continuing in mathematics and science: Gender and education level differences," *J. Vocational Behav.*, vol. 77, no. 3, pp. 361–373, 2010.

[46] D. T. Sciarra, "Predictive factors in intensive math course-taking in high school," *Professional School Counseling J.*, vol. 13, no. 3, pp. 196–207, 2010.

[47] R. W. Lent, S. D. Brown, and G. Hackett, "Toward a unifying social cognitive theory of career and academic interest, choice, and performance," *J. Vocational Behav.*, vol. 45, no. 1, pp. 79–122, 1994.

[48] J. Eccles and A. Wigfield, "Motivational beliefs, values, and goals," *Annu. Rev. Psychol.*, vol. 53, no. 1, pp. 109–132, 2002.

[49] S. J. Ceci and W. M. Williams, "Understanding current causes of women's underrepresentation in science," *Proc. Natl. Acad. Sci. USA*, vol. 108, no. 8, pp. 3157, 2011.

[50] H. M. Matusovich, R. A. Streveler, and R. L. Miller, "Why do students choose engineering? A qualitative, longitudinal investigation of students' motivational values," *J. Eng. Educat.*, vol. 99, no. 4, pp. 289–303, 2010.

[51] I. Ajzen, "Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior," *J. Appl. Social Psychol.*, vol. 32, no. 4, pp. 665–683, 2002.

[52] A. Kahveci, S. A. Southerland, and P. J. Gilmer, "Retaining undergraduate women in science, mathematics, and engineering," *J. College Sci. Teaching*, vol. 36, no. 3, pp. 34–38, 2006.

[53] L. S. Richman and M. Van Dellen, "How women cope: Being a numerical minority in a male-dominated profession," *J. Social Issues*, vol. 67, no. 3, pp. 492–509, 2011.

[54] R. W. Lent, M. J. Miller, P. E. Smith, B. A. Watford, R. H. Lim, and K. Hui, "Social cognitive predictors of academic persistence and performance in engineering: Applicability across gender and race/ethnicity," *J. Vocational Behav.*, vol. 94, pp. 79–88, June 2016.

[55] J. M. Kittleson and S. A. Southerland, "The role of discourse in group knowledge construction: A case study of engineering students," *J. Res. Sci. Teaching*, vol. 41, no. 3, pp. 267–293, 2004.

[56] T. A. Koszalka, Y. Wu, and B. Davidson, "Instructional design issues in a cross-institutional collaboration within a distributed engineering educational environment," in *Proc. World Conf. E-Learning in Corporate, Government, Healthcare, and Higher Education*, 2007, pp. 1650–1657.

[57] E. Rodriguez-Falcon, A. Hodzic, and A. Symington, "Learning from each other: Engaging engineering students through their cultural capital," *Eng. Educat.*, vol. 6, no. 2, pp. 29–38, 2011.

[58] S. Brophy, S. Klein, M. Portsmore, and C. Rogers, "Advancing engineering in P-12 classrooms," *J. Eng. Educat.*, vol. 97, no. 3, pp. 369–387, 2008.

[59] N. A. Tran and M. J. Nathan, "Pre-college engineering studies: An investigation of the relationship between pre-college engineering studies and student achievement in science and mathematics," *J. Eng. Educat.*, vol. 99, no. 2, pp. 143–157, 2010.

[60] S. S. Guzey, T. J. Moore, M. Harwell, and M. Moreno, "STEM integration in middle school life science: Student learning and attitudes," *J. Sci. Educat. Technol.*, vol. 25, no. 4, pp. 550–560, 2016.

[61] (2014). Project Lead the Way (PLTW) [Online]. Available: https://www.pltw.org/pltw-engineering

[62] R. H. Tai. (2012). An examination of the research literature on Project Lead the Way [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.361.548

[63] University of Texas Austin UTEACH. (2015). Engineer your world [Online]. Available: http://engineeryourworld.org

[64] Southern Methodist University. (1999). The infinity project [Online]. Available: https://www.smu.edu/Lyle/Institutes/CaruthInstitute/K–12Programs/InfinityProject

[65] Museum of Science Boston. (2016). Engineering is elementary [Online]. Available: http://www.eie.org

[66] National Girls Collaborative Project. (2016). In the middle of engineering [Online]. Available: https://ngcproject.org/mini-grant/middle-engineeringu-l

[67] C. Cunningham, "Elementary teacher professional development in engineering: Lessons learned from engineering is elementary," in *Proc. American Society for Engineering Education National Conf.*, 2008.

[68] K. P. Dabney, R. H. Tai, J. T. Almarode, J. L. Miller-Friedmann, G. Sonnert, P. M. Sadler, and Z. Hazari, "Out-of-school time science activities and their association with career interest in STEM," *Int. J. Sci. Educat. B*, vol. 2, no. 1, pp. 63–79, 2012.

[69] X. Wang, "Modeling student choice of STEM fields of study: Testing a conceptual framework of motivation, high school learning, and postsecondary context of support," Tech. Rep., Wisconsin Center for the Advancement of Postsecondary Education, Madison, WI, 2012.

[70] M. F. Bugallo, A. M. Kelly, and M. Ha, "Research on impacts of a university-based electrical engineering

(continued on page 100)

C.-C. Jay Kuo

The CNN as a Guided Multilayer RECOS Transform

There is a resurging interest in developing a neural-network-based solution to the supervised machine-learning problem. The convolutional neural network (CNN) will be studied in this lecture note. We introduce a rectified-correlations on a sphere (RECOS) [1] transform as a basic building block of CNNs. It consists of two main concepts: 1) data clustering on a sphere and 2) rectification. We then interpret a CNN as a network that implements the guided multilayer RECOS transform with two highlights. First, we compare the traditional single-layer and modern multilayer signal-analysis approaches, point out key areas that enable the multilayer approach, and provide a full explanation to the operating principle of CNNs. Second, we discuss how guidance is provided by labels through backpropagation (BP) in the training.

Relevance

CNNs are widely used in the computer vision field today. They offer state-of-the-art solutions to many challenging vision and image processing problems such as object detection, scene classification, room layout estimation, semantic segmentation, image superresolution, image restoration, and object tracking, to name a few. They are the mainstream machine-learning tool for big visual data analytics. A great amount of effort has been devoted to the interpretability of CNNs based on various disciplines and tools, such as approximation theory, optimization theory, and visualization techniques. We explain the CNN operating principle using data clustering, rectification, and transform, which are familiar to researchers and engineers in the signal processing and pattern recognition community. As compared with other studies, this approach appears to be more direct and insightful. It is expected to contribute to further research advancement on CNNs.

Digital Object Identifier 10.1109/MSP.2017.2671158
Date of publication: 26 April 2017

Prerequisites

The prerequisites consist of basic calculus, probability, and linear algebra. Statistics and approximation techniques could also be useful but are not necessary.

Problem statement

We will study the following three problems in this note.

1) *Neural networks architecture evolution.* We provide a survey on the architecture evolution of neural networks, including computational neurons, multilayer perceptrons (MLPs), and CNNs.
2) *Signal analysis via multilayer RECOS transform.* We point out the differences between the single- and multilayer signal-analysis approaches and explain the working principle of the multilayer RECOS transform.
3) *Network initialization and guided anchor vector update.* We carefully examine the CNN initialization scheme since it can be viewed as an unsupervised clustering and the CNN self-organization property can be explained. The supervised learning is achieved by BP using data labels in the training stage. It will be interpreted as guided anchor vector update.

Solution

CNN architecture evolution

We divide the architectural evolution of CNNs into three stages and provide a brief survey.

Computational neuron

A computational neuron (or simply neuron) is the basic operational unit in neural networks. It was first proposed by McCulloch and Pitts in [2] to model the "all-or-none" character of nervous activities. It conducts two operations in cascade: an affine transform of input vector $\mathbf{x}$ followed by a nonlinear activation function. Mathematically, we can express it as

$$\begin{aligned} y &= f(b), \\ b &= T_{\mathbf{a}}(\mathbf{x}) \\ &= \sum_{n=1}^{N} a_n x_n + a_0 \mu \\ &= \mathbf{a}^T \mathbf{x} + a_0 \mu = \mathbf{a}'^T \mathbf{x}', \end{aligned} \tag{1}$$

where y, $\mathbf{x} = (x_1, \ldots, x_N)^T \in R^N$, and $\mathbf{a} = (a_1, \ldots, a_N)^T \in R^N$ are the scalar output, the N-dimensional input and model parameter vectors, respectively; $a_0 \mu$ is a bias term with $\mu = (1/N)\sum_{n=1}^{N} x_n$ (i.e., the mean of all input elements), $f(.)$

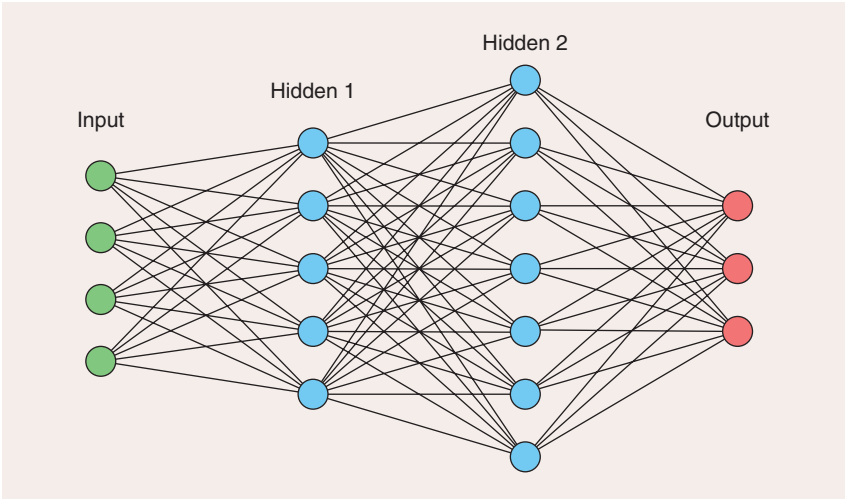


FIGURE 1. An exemplary MLP with one input layer, two hidden layers, and one output layer.

denotes a nonlinear activation function, and b is the intermediate result between the two operations.

The function, $f(.)$, was chosen to be a delayed step function in form of $f(b) = u(b - \phi)$ in [2], where $f(b) = 1$ if $b \geq \phi$ and 0 if $b < \phi$. A neuron is on (in the one state) if the stimulus, b, is larger than threshold ϕ. Otherwise, it is off (in the zero state). Multiple neurons can be flexibly connected into logic networks as models in theoretical neurophysiology.

For the vision problem, input $\mathbf{x}$ denotes an image (or image patch). The neuron should not generate a response for a flat patch since it does not carry any visual pattern information. Thus, we set $b = 0$ if all of its elements are equal to a nonzero constant. It is then straightforward to derive $\sum_{n=1}^{N} a_n + a_0 = 0$ or $a_0 = -\sum_{n=1}^{N} a_n$ is a dependent variable. We can form augmented vectors $\mathbf{x}' = (\mu, x_1, \ldots, x_N)^T \in R^{N+1}$ and $\mathbf{a}' = (a_0, a_1, \ldots, a_N)^T \in R^{N+1}$ for $\mathbf{x}$ and $\mathbf{a}$, respectively. Without loss of generality, we assume $\mu = 0$ in the following discussion. If $\mu \neq 0$, we can either consider the augmented vector space of $\mathbf{x}'$ or normalize input $\mathbf{x}$ to be a zero-mean vector before the processing and add the mean back after the processing.

Multilayer perceptrons

The perceptron was introduced by Rosenblatt in [3]. One can stack multiple perceptrons side by side to form a perceptron layer and cascade multiple perceptron layers into one network. It is called the *MLP* or the *feedforward neural network*. An exemplary MLP is shown in Figure 1. In general, it consists of a layer of input nodes (the input layer), several layers of intermediate nodes (the hidden layers), and a layer of output nodes (the output layer). These layers are indexed from $l = 0$ to L, where the input and output layers are indexed with 0 and L, and the hidden layers are indexed from $l = 1 \cdots L - 1$, respectively. Suppose that there are N_l nodes at the lth layer. Each node at the lth layer takes all nodes in the $(i-1)$th layer as its input. For this reason, it is called the *fully connected layer*. Clearly, the MLP is end-to-end fully connected. A modern CNN often contains an MLP as its building module.

MLPs were studied intensively in the 1980s and 1990s as decision networks for pattern recognition applications. The input and output nodes represent selected features and classification types, respectively. There are two major advances from simple neuron-based logic networks to MLPs. First, there was no training mechanism in the former since they were not designed for the machine-learning purpose. The BP technique was introduced in MLPs as a training mechanism for supervised learning. Since differentiation is needed in the BP yet the step function is not differentiable, other nonlinear activation functions are adopted in MLPs. Examples include the sigmoid function, the rectified linear unit (ReLU) and the parameterized ReLU (PReLU). Second, MLPs have a modularized structure (i.e., perceptron layers) suitable for parallel processing.

As compared with traditional pattern recognition techniques based on simple linear analysis (e.g., linear discriminant analysis, principal component analysis, etc.), MLPs provide a more flexible mapping from the feature space to the decision space, where the distribution of feature points of one class can be nonconvex and irregular. It is built upon a solid theoretical foundation proved by Cybenko [4] and Hornik et al. [5]. That is, a network with only one hidden layer can be a universal approximator if there are "enough" neurons.

Convolutional neural networks

Fukushima's neocognitron [6] can be viewed as an early form of a CNN. The architecture introduced by LeCun et al. in [7] serves as the basis of modern CNNs. The main difference between MLPs and CNNs lies in their input space—the former are features while the latter are source data such as image, video, speech, etc. This is not a trivial difference. Let us use the LeNet-5 shown in Figure 2 as an example, whose input is an image of size 32×32. Each pixel is an input node. It would be very challenging for an MLP to handle this input since the dimension of the input vector is $32 \times 32 = 1{,}024$. The diversity of possible visual patterns is huge. As explained later, the nodes in the first hidden layer should provide a good representation for the input signal. Thus, it implies a large number of nodes in hidden layers. The number of links (or filter weights) between the input and the first hidden layers is $N_0 \times N_1$ due to full connection. This number can easily go to the order of millions. If the image dimension is in the order of millions such as those captured by today's smartphones, the solution is clearly unrealistic.

Instead of considering interactions of all pixels in one step as done in the MLP, the CNN decomposes an input image into smaller patches, known as *receptive fields,* for nodes at certain layers. It gradually enlarges the receptive field to cover a larger portion of the image. For example, the filter size of the first two convolutional layers of LeNet-5 is 5×5. The first convolutional layer considers

interactions of pixels in the short range. Since the patch size is small, the diversity is less. One can use six filters to provide a good approximation to the 5×5 source patches, and all source patches share the same six filters regardless of their spatial location. After subsampling, the second convolutional layer examines interaction of pixels in the midrange. After another subsampling, the whole spatial domain shrinks to a size 5×5 so that it can take global interaction into account using full connection. Typically, the interaction contains not only spatial but also spectral elements (e.g., the RGB three channels and multiple filter responses at the same spatial location) and all interactions are modeled by computational neurons as given in (1).

It is typical to decompose a CNN into two subnetworks: the feature extraction (FE) subnet and the decision-making (DM) subnet. The FE subnet consists of multiple convolutional layers while the DM subnet is composed of a couple of fully connected layers. Roughly speaking, the FE subnet conducts clustering aiming at a new representation through a sequence of RECOS transforms. The DM subnet links data representations to decision labels, which is similar to the classification role of MLPs. The exact boundary between the FE subnet and the DM subnet is actually blurred in the LeNet-5. It can be either S4 or C5. If we view S4 as the boundary, then C5 and F6 are two hidden layers of the DM subnet. On the other hand, if we choose C5 as the boundary, then there is only one hidden layer (i.e., F6) in the DM subnet. Actually, since these two subnets are connected side by side, the transition from the representation to the classification happens gradually and smoothly.

One main advantage of CNNs over the support vector machine and the random forest classifiers is that the FE task is automatically done through the BP from the last layer to the first layer. Generally speaking, discriminant features are difficult to find for traditional classifiers such as the MLP, support vector machine, and random forest. Such tasks, called *feature engineering*, demand the domain knowledge. Furthermore, it is difficult to argue that ad hoc features found empirically are optimal in any sense. This explains why the traditional computer vision field is fragmented by different applications. After the emergence of CNNs, the domain knowledge is no longer important in FE, yet it plays a critical role in data labeling (known as *label engineering*). To give an example, anaconda, vipers, titanoboa, cobras, rattlesnake, etc. are finer classifications of snakes. It requires expert knowledge to collect and label their images. The CNN provides a powerful tool in data-driven supervised learning, where the emphasis is shifted from "extracting features from the source data" to "constructing data sets by pairing carefully selected data and their labels."

Single-layer RECOS transform

Our discussion applies to $\mathbf{x}$ and $\mathbf{a}$ if $\mu = 0$ or augmented vectors $\mathbf{x}'$ and $\mathbf{a}'$ if $\mu \neq 0$. For convenience, we only consider the case with $\mu = 0$. The generalization to $\mu \neq 0$ is straightforward.

Clustering on sphere's surface

There is a modern interpretation to the function of a single perceptron layer based on the clustering notion. Since data clustering is a well-understood discipline, one can understand the operation of CNNs better if a connection between the operation of a perceptron layer and data clustering can be established. This link was built in [1]. It will be repeated below. Let

$$S = \{\mathbf{x} \mid \|\mathbf{x}\| = 1\}.$$

be an N-dimensional unit hypersphere (or simply sphere). We consider clustering of points in S using the geodesic distance. For an arbitrary vector $\mathbf{x}$ to be a member in S, we need to normalize it by its magnitude $g = \|\mathbf{x}\|$. If $\mathbf{x}$ is an image patch, the magnitude normalization after its mean removal has a physical meaning: contrast adjustment. When g is smaller than a threshold, the patch is nearly flat. A flat patch carries little visual information yet its normalization does amplify noise. In this case, it is better to treat it as a zero vector. When g is larger than the threshold, vector $\mathbf{x}$ does represent a visual pattern, and humans perceive little difference between the original and normalized patches since the contrast has little effect on visual patterns. Although this normalization procedure is not implemented in today's CNNs, the following mathematical analysis can be significantly simplified while the essence of CNNs can still be well captured.

The geodesic distance of two points, $\mathbf{x}_i$ and $\mathbf{x}_j$ in S, is proportional to the

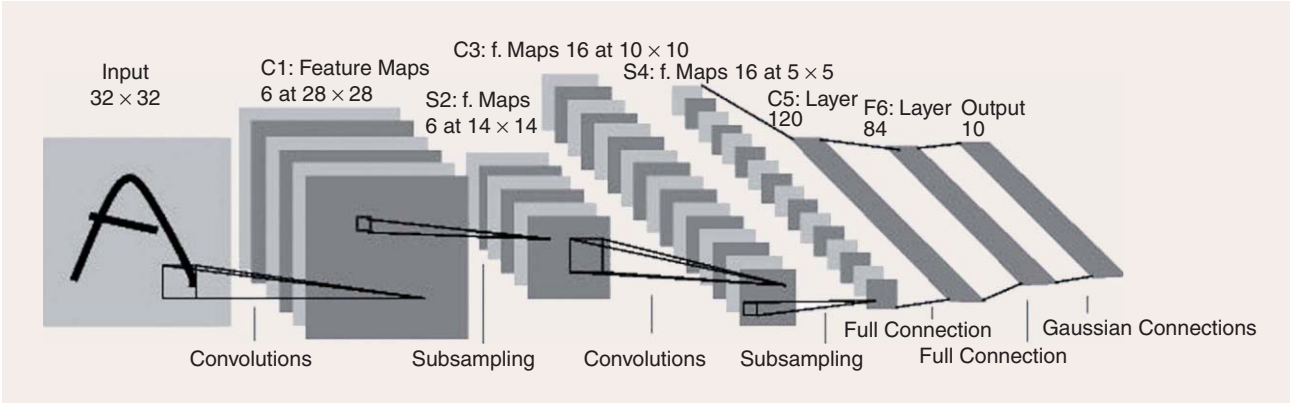


FIGURE 2. The LeNet-5 architecture [7] as an exemplary CNN.

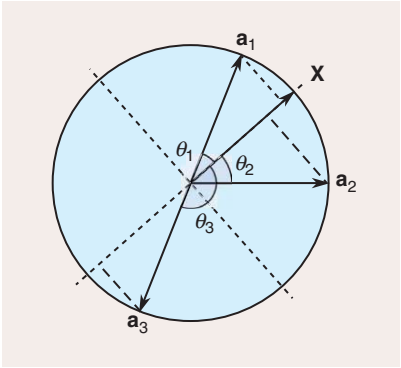


FIGURE 3. Illustrating the need of rectification [1].

magnitude of their angle, which can be computed by

$$\theta(\mathbf{x}_i, \mathbf{x}_j) = \cos^{-1}(\mathbf{x}_i^T \mathbf{x}_j).$$

Since $\cos\theta$ is a monotonically decreasing function for $0^\circ \le |\theta| \le 180^\circ$, we can use the correlation, $0 \le \mathbf{x}_i^T \mathbf{x}_j = \cos\theta \le 1$, as another distance measure between them, and cluster vectors in S accordingly. Note that, when $90^\circ \le |\theta| \le 180^\circ$, the correlation, $\mathbf{x}_i^T \mathbf{x}_j = \cos\theta$, is a negative value.

For MLPs and CNNs, a set of neurons is used to operate on a set of input nodes. For example, nodes in each hidden layer and the output layer in Figure 1 take the weighted sum of values of nodes in the preceding layer as their outputs. These outputs are treated as one inseparable unit that becomes the input to the next layer. Each neuron has a filter weight vector denoted by $\mathbf{a}_k, k = 1, \ldots, K$. In signal processing terminology, the set of neurons forms a filter bank. The RECOS model [1] describes the relationship between nodes of the $(l-1)$th and lth layers, $l = 1, 2, \ldots$, where the input layer is the 0th layer. There are three RECOS units in cascade for the MLP in Figure 1. One corresponds to a filter bank. The filter weight vector is called an *anchor vector* since it serves as a reference pattern associated with a neuron unit.

Need of rectification

A neuron computes the correlation between an input vector and its anchor vector to measure their similarity. There are K neurons in one RECOS unit. The projection of $\mathbf{x}$ onto all anchor vectors, $\mathbf{a}_k$, can be written in form of

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad \mathbf{A}^T = [\mathbf{a}_1 \cdots \mathbf{a}_k \cdots \mathbf{a}_K],$$

where $\mathbf{y} = (y_1, \ldots, y_k, \ldots y_K)^T \in R^K$, $y_k = \mathbf{a}_k^T \mathbf{x}$, and $\mathbf{A} \in R^{K \times N}$. For input vectors $\mathbf{x}_i$ and $\mathbf{x}_j$, their corresponding outputs are $\mathbf{y}_i$ and $\mathbf{y}_j$. If the geodesic distance of $\mathbf{x}_i$ and $\mathbf{x}_j$ in S is close, we expect the distance of $\mathbf{y}_i$ and $\mathbf{y}_j$ in the K-dimensional output space to be close as well.

To show the necessity of rectification, a two-dimensional (2-D) example is illustrated in Figure 3, where $\mathbf{x}$ and $\mathbf{a}_k$ $(k = 1, 2, 3)$ denote an input and three anchor vectors on the unit circle, respectively, and θ_i is their respective angle. Since θ_1 and θ_2 are lower than 90°, $\mathbf{a}_1^T \mathbf{x}$ and $\mathbf{a}_2^T \mathbf{x}$ are positive. The angle, θ_3, is larger than 90° and correlation $\mathbf{a}_3^T \mathbf{x}$ is negative. The two vectors, $\mathbf{x}$ and $\mathbf{a}_3$, are far apart in terms of the geodesic distance. Since $\cos\theta$ is monotonically decreasing for $0^\circ \le |\theta| \le 180^\circ$, it can be used to reflect the order of the geodesic distance in one layer.

However, when two RECOS units are in cascade, the filter weights of the second RECOS unit can take positive or negative values. If the response of the first RECOS unit is negative, the product of a negative response and a negative filter weight will produce a positive value. On the other hand, the product of a positive response and a positive filter weight will also produce a positive value. If the nonlinear activation unit did not exist, the cascaded system would not be able to differentiate them. For example, the geodesic distance of $\mathbf{x}$ and $-\mathbf{x}$ should be farthest. However, they yield the same result, and their original patches become indistinguishable under this scenario. Similarly, a system without rectification cannot differentiate the following two cases: 1) a positive response at the first layer followed by a negative filter weight at the second layer and 2) a negative response at the first layer followed by a positive filter weight at the second layer.

Rectifier design

Since a nonlinear activation unit is used to rectify correlations, it is called a rectifier here. To avoid the above-mentioned confusion cases, we impose the following two requirements on a rectifier.

1) The output $\mathbf{a}_k^T \mathbf{x}$ should be rectified to be a nonnegative value.
2) The rectification function should be monotonically increasing so as to preserve the order of the geodesic distance.

Three rectifiers are often used in MLPs and CNNs. They are the sigmoid function, the ReLU, and the PReLU as shown in Figure 4(a)–(c). The PReLU is also known as the *leaky ReLU*. Both the sigmoid and ReLU satisfy the aforementioned two requirements. Although the PReLU does not strictly satisfy the first requirement, it does not have a severe negative impact on spherical surface clustering. This is because a negative

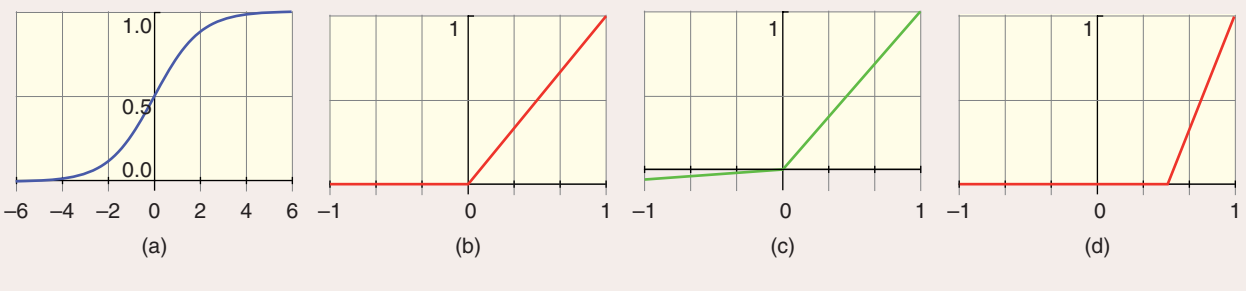


FIGURE 4. An illustration of four rectifiers: (a) the sigmoid function, (b) the ReLU (middle), (c) the leaky ReLU, and (d) the TReLU.

correlation is rectified to a significantly smaller negative value.

Based on the two requirements, one can design different rectifiers. One example is shown in Figure 4(d). It is called the *threshold ReLU* (*TReLU*). The rectification function can be defined as $\text{TReLU}(x) = 0$, if $x < \phi$ and $\text{TReLU}(x) = (x - \phi)/(1 - \phi)$ if $x \geq \phi$. When $\phi = 0$, TReLU is reduced to ReLU. For the LeNet-5 applied to the MNIST data set, we observe better performance as ϕ increases from 0 to 0.5 and then decreases. One advantage of $\text{TReLU}(\phi)$ with $\phi > 0$ is that we can block the influence of more anchor vectors. When $\phi = 0$, we block the influence of anchor vectors that have an angle larger than 90° with respect to the input vector. When $\phi = 0.5$, we block the influence of anchor vectors that have an angle larger than 60°. The design of an optimal rectifier for target applications remains to be an open problem.

Multilayer RECOS transform

Single-layer signal analysis via representation

Signal modeling and representation is commonly used in the signal processing field for signal analysis. Typically, we have a linear model in form of

$$\mathbf{x} = \mathbf{A}\mathbf{c}, \tag{2}$$

where $\mathbf{x} \in R^N$ denotes the signal of interest, $\mathbf{A} \in R^{N \times M}$ is a representation matrix, and $\mathbf{c} \in R^M$ is the coefficient vector. If $M = N$ and the column vectors of $\mathbf{A}$ form a set of basis functions, (2) defines a transform from one basis to another. The task is in selecting powerful basis functions to represent signals of interest. Fourier and wavelet transforms are well-known examples. Then, a subset of coefficient vector $\mathbf{c}$ can be used as the feature vector. If $M > N$, there exist infinitely many solutions in $\mathbf{c}$. We can impose constraints on $\mathbf{c}$, leading to the linear least-squares solution, sparse coding, among others. For the sparse representation, the task is in finding a good dictionary, $\mathbf{A}$, to represent the underlying signal effectively. Again, a subset of coefficient vector $\mathbf{c}$ can be chosen as features.

Multilayer signal analysis via cascaded transforms

The CNN approach provides a brand-new framework for signal analysis. Instead of finding a representation for signal analysis, it relies on a sequence of cascaded transforms that builds a link between the input signal space and the output decision space. The operation at each layer is to conduct a spherical surface's clustering of input samples with a rectified output (i.e., the RECOS transform).

For MLPs, each network corresponds to a simple cascade of multiple RECOS transforms. Mathematically, we have

$$\mathbf{d} = \mathbf{B}_L \cdots \mathbf{B}_l \cdots \mathbf{B}_1 x, \tag{3}$$

where $\mathbf{x}$ is an input signal, $\mathbf{d} = (d_1, \ldots, d_c, \ldots, d_C)$ is an output vector in the decision space indicating the likelihood in class c with $c = 1, \ldots, C$, and $\mathbf{B}_l$ is the lth layer RECOS transform matrix with $l = 1, \ldots, L$. The input and output to the lth layer RECOS transform $\mathbf{B}_l$ are denoted by $\mathbf{x}_{l-1}$ and $\mathbf{x}_l$, respectively. Thus, we get

$$\mathbf{x}_l = \mathbf{B}_l \mathbf{x}_{l-1}, \quad \text{where} \quad \mathbf{B}_l = \mathbf{R} \circ \mathbf{A}_l, \tag{4}$$

and where $\mathbf{R}$ is the element-wise rectification function operating on the output of $\mathbf{A}_l \mathbf{x}_{l-1}$. Clearly, we have $\mathbf{x}_0 = \mathbf{x}$ and $\mathbf{x}_L = \mathbf{d}$.

The ground truth $\mathbf{d}$ is that $d_i = 1$ if i is the target class while $d_j = 0$ if j is not the target class. It is called the *one-hot vector*. The training samples have both input $\mathbf{x}$ and its label $\mathbf{d}$. The testing samples have only input $\mathbf{x}$, and we need to predict its output $\mathbf{d}$ and convert it to its nearest one-hot vector. The task is in finding good $\mathbf{B}_l$, $l = 1, \ldots, L$, so as to minimize the classification error.

For CNNs, we have two types of $\mathbf{B}_l$ in the form of

$$\mathbf{B}_l^C = \mathbf{P} \bigcup_{s \in \Omega} \mathbf{R} \circ \mathbf{A}_{l,s}, \text{ and } \mathbf{B}_l^F = \mathbf{R} \circ \mathbf{A}_l, \tag{5}$$

where $\mathbf{A}_{l,s}$ denotes a convolutional filter at layer l with spatial index s, $\bigcup_s$ is the union of outputs from a neighborhood, Ω, and $\mathbf{P}$ denotes a pooling operation. The union of outputs from a set of parallel convolutional filters serve as the input to the filter at the next layer. The two RECOS transforms, $\mathbf{B}_l^C$ and $\mathbf{B}_l^F$, are called the *convolutional layer* and the *fully connected layer*, respectively, in the modern CNN literature. Clearly, $\mathbf{B}_l = \mathbf{B}_l^F$ in (4).

It is inspiring to compare the two signal-analysis approaches as given in (2) and (3). The one in (2) is a single-layer approach where no rectification is needed. The one in (3) is a multilayer approach and rectification is essential. The single-layer approach seeks for a better signal representation. For example, a multiple-scale signal representation was developed using the wavelet transform. A sparse signal representation was proposed using a trained dictionary. The objective is to find an "optimal" representation to separate critical components in desired signals from others.

In contrast, the CNN approach does not intend to decompose underlying signals. Instead, it adopts a sequence of RECOS transforms to cluster input data based on their similarity layer by layer until the output layer is reached. The output layer predicts the likelihood of all possible decisions (e.g., object classes). The training samples provide a relationship between an image and its decision label. The CNN can predict results even without any supervision, although the prediction accuracy would be low. The training samples guide the CNN to form more suitable anchor vectors (thus, better clusters) and connect clustered data with decision labels. To summarize, we can express the multilayer RECOS transform as

$$\text{MLP:} \quad \mathbf{x} = \mathbf{x}_0 \xrightarrow{\mathbf{B}_1^F} \mathbf{x}_1 \xrightarrow{\mathbf{B}_2^F} \cdots \xrightarrow{\mathbf{B}_{L-1}^F} \mathbf{x}_{L-1} \xrightarrow{\mathbf{B}_L^F} \mathbf{x}_L = \mathbf{d},$$

$$\text{CNN:} \quad \mathbf{x} = \mathbf{x}_0 \xrightarrow{\mathbf{B}_1^C} \mathbf{x}_1 \xrightarrow{\mathbf{B}_2^C} \cdots \xrightarrow{\mathbf{B}_m^C} \mathbf{x}_m \xrightarrow{\mathbf{B}_{m+1}^F} \mathbf{x}_{m+1} \xrightarrow{\mathbf{B}_{m+2}^F} \cdots \xrightarrow{\mathbf{B}_{L-1}^F} \mathbf{x}_{L-1} \xrightarrow{\mathbf{B}_L^F} \mathbf{x}_L = \mathbf{d}.$$

The output from the lth layer, $\mathbf{x}_l$, serves as the input to the $(l+1)$th layer. It is called the intermediate representation at the lth layer or the lth intermediate representation, in short.

It is important to have a deeper understanding on the compound effect of two RECOS transforms in cascade. This was

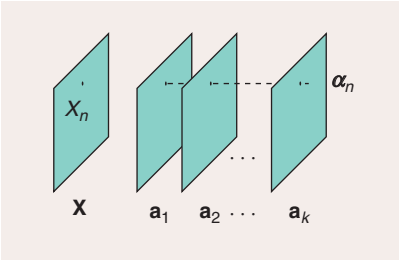


FIGURE 5. The visualization of the anchor-position vector α_n [1].

thoroughly studied in [1], and the main result is summarized below. A representative 2-D input and its corresponding anchor vectors are shown in Figure 5. Let α_n be a K-dimensional vector formed by the same position (or element) of $\mathbf{a}_k$. It is called the *anchor-position vector* since it captures the position information of anchor vectors. Although anchor vectors $\mathbf{a}_k$ capture global representative patterns of $\mathbf{x}$, they are weak in capturing position sensitive information. This shortcoming can be compensated for by modulating outputs with elements of the anchor-position vector α_n in the next layer.

Let us use layers S4, C5, and F6 in LeNet-5 as an example. There are 120 anchor vectors of dimension 400 from S4 to C5. We collect 400 anchor-position vectors of dimension 120, multiply the output at C5 by them to form a set of modulated outputs, and then compute 84 anchor vectors of dimension 120 from C5 to F6. Note that the output at C5 contains primarily the spectral information but not the position information. If a position in the input vectors has less consistent information, the variance of its associated anchor position vector will be larger and the modulated output will be more random. As a result, its impact on the formation of the 84 anchor vectors is reduced. For more details, we refer to the discussion in [1].

New clustering representation

We have a one-to-one association between a data sample and its cluster in traditional clustering schemes. However, this is not the case in the RECOS transform. A new clustering representation is adopted by MLPs and CNNs. That is, for an input vector $\mathbf{x}$, the RECOS transform generates a set of K nonnegative correlation values as the output vector of dimension K. This representation enables repetitive clustering layer by layer as given in (4). For an input, one can determine the significance of clusters according to the magnitude of the rectified output value. If its magnitude for a cluster is zero, $\mathbf{x}$ is not associated with that cluster. A cluster is called a relevant or irrelevant one depending on whether it has an association with $\mathbf{x}$. Among all relevant ones, we call cluster i the *primary* cluster for input $\mathbf{x}$ if

$$i = \arg\max_k \mathbf{a}_k^T \mathbf{x}.$$

The remaining relevant ones are *auxiliary* clusters.

The FE subnet uses anchor vectors to capture local, midrange, and long-range spatial patterns. It is difficult to predict the clustering structure since new information is introduced at a new layer. The DM subnet attempts to reduce the dimension of intermediate representations until it reaches the dimension of the decision space. We observe that the clustering structure becomes more obvious as the layer of the DM subnet goes deeper. That is, the output value from the primary cluster is closer to unity while the number of auxiliary clusters is fewer and their output values become smaller. When this happens, an anchor vector provides a good approximation to the centroid for the corresponding cluster.

The choice of anchor vector numbers, K_l, at the lth layer is an important problem in the network design. If input data $\mathbf{x}_{l-1}$ has a clear clustering structure (say, with h clusters), we can set $K_l = h$. However, this is often not the case. If K_l is set to a value too small, we are not able to capture the clustering structure of $\mathbf{x}_{l-1}$ well, and it will demand more layers to split them. If K_l is set to a value too large, there are more anchor vectors than needed, and a stronger overlap between rectified output vectors will be observed. As a result, we still need more layers to separate them. Another way to control the clustering process is the choice of the threshold value, ϕ, of the TReLU. A higher threshold value can reduce the negative impact of a larger K_l value. The tradeoff between ϕ and K_l is an interesting future research topic.

Network initialization and guided anchor vector update

Data clustering plays a critical role in the understanding of the underlying structure of data. The k-means algorithm, which is probably the most well-known clustering method, has been widely used in pattern recognition and supervised/unsupervised learning. As discussed previously, each CNN layer conducts data clustering on the surface of a high-dimensional sphere based on a rectified geodesic distance. Here, we would like to understand the effect of multiple layers in cascade from the input data source to the output decision label. For unsupervised learning such as image segmentation, several challenges exist in data clustering [8]. Questions such as "What is a cluster?" "How many clusters are present in the data?" and "Are the discovered clusters and partition valid?" remain open. These questions expose the limit of unsupervised data clustering methods.

In the context of supervised learning, traditional feature-based methods extract features from data, conduct clustering in the feature space, and, finally, build a connection between clusters and decision labels. Although it is relatively easy to build a connection between the data and labels through features, it is challenging to find effective features. In this setting, the dimension of the feature space is usually significantly smaller than that of the data space. As a consequence, it is unavoidable to sacrifice rich diversity of input data. Furthermore, the feature selection process is guided by humans based on their domain knowledge (i.e., the most discriminant properties of different objects). This process is heuristic. It can become overfit easily. Human efforts are needed in both data labeling and feature design.

CNNs offer an effective supervised learning solution, where supervision is conducted by a training process using data labels. This supervision closes the semantic gap between low-level representations (e.g., the pixel representation) and high-level semantics. Furthermore, the CNN self-organization capability was well discussed in the 1980s and 1990s, e.g., [6]. By self-organization, the network can learn with little supervision. To put the above two together, we expect

that CNNs can provide a wide range of learning paradigms—from the unsupervised, weakly supervised, to heavily supervised learning. This intuition can be justified by considering a proper anchor vector initialization scheme (for self-organization) and providing proper "guidance" to the proposed multilayer RECOS transform in anchor vector update (for supervised learning).

Network initialization

The CNN conducts a sequence of representation transforms using cascaded RECOS units. The dimension of transformed representations gradually decreases until it reaches the number of output classes. Since labels of output classes are provided by humans with a semantic meaning, the whole end-to-end process is called the *guided* (or *supervised*) transform.

Before examining the effect of label guidance, we first compare two network initialization schemes: 1) the random initialization and 2) the *k*-means initialization. For the latter, we perform *k*-means at each layer based on its corresponding input data samples (with zero-mean and unit-length normalization), and we repeat this process from the input to the output layer after layer. Today, random initialization is commonly adopted. Based on the previous discussion, we expect the *k*-means initialization to be a better choice. This is verified by our experiments in the LeNet-5 applied to the MNIST data set.

Once the network is initialized, we can feed the test data to the network and observe the output, which corresponds to unsupervised learning. The comparison of unsupervised classification results with the random and *k*-means initializations is given in Figure 6, where we show images that are closest to the anchor vectors (or centroids) of the ten output nodes. We see that the *k*-means initialization provides ten anchor vectors pointing to ten different digits while the random initialization cannot do the same. Different random initialization schemes will lead to different results, yet the one given in Figure 6 is representative. That is, multiple anchor vectors will point to the same digit.

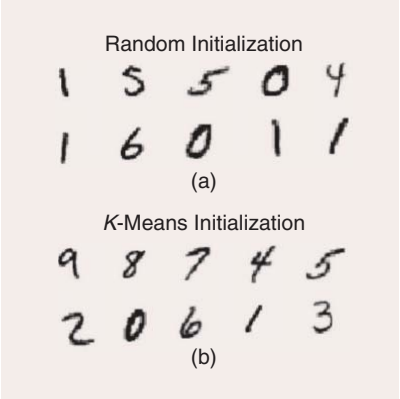


FIGURE 6. The comparison of MNIST unsupervised classification results of the LeNet-5 architecture with the (a) random and (b) *k*-means initializations, where the images that are closest to centroids of ten output nodes are shown.

Guided anchor vector update

We apply the BP for network training with a varying number of training samples. For a fixed number of training samples, we train the network until its performance converges and plot the correct classification rate in Figure 7. The two points along the y-axis indicate the correct classification rates without any labeled training sample. The rates are around 32 and 14% for the *k*-means and random initializations, respectively. Note that the 14% is slightly better than the random guess on the outcome, which is 10%. Then, both performance curves increase as the number of labeled training samples grows. The *k*-means can reach a correct classification rate of 90% when the number of labeled training samples is around 250, which is only 0.41% of the entire MNIST training data set (i.e., 60,000 samples). This shows the power of the LeNet-5 even under extremely low supervision.

To further understand the role played by label guidance, we examine the impact of the BP on the orientation of anchor vectors in various layers. We show in Table 1 the averaged orientation changes of anchor vectors in terms of radian (or

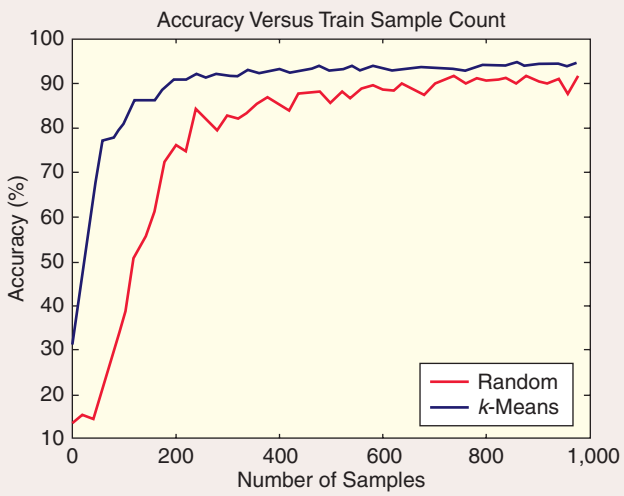


FIGURE 7. The comparison of MNIST weakly supervised classification results of the LeNet-5 architecture with the random and *k*-means initializations, where the correct classification rate is plotted as a function of training sample numbers.

Table 1. The averaged orientation changes of anchor vectors in terms of the radian (or degree) for the *k*-means and the random initialization schemes.

In/Out Layers	*k*-Means	Random
Input/S2	0.155 (or 8.881°)	1.715 (or 98.262°)
S2/S4	0.169 (or 9.683°)	1.589 (or 91.043°)
S4/C5	0.204 (or 11.688°)	1.567 (or 89.783°)
C5/F6	0.099 (or 5.672°)	1.579 (or 90.470°)
F6/output	0.300 (or 17.189°)	1.591 (or 91.158°)

FIGURE 8. Two subclasses obtained from the horned rattlesnake class using unsupervised clustering.

degree) for the two cases in Figure 6. They are obtained after the convergence of the network with all 60,000 MNIST training samples. This orientation change is the result due to label guidance through the BP. It is clear from the table that a good network initialization (corresponding to unsupervised learning) leads to a faster convergence rate in supervised learning.

Classes and subclasses

We use another example to gain further insights to the guided clustering process. We can zoom in on the horned rattlesnake class obtained by the AlexNet and conduct the unsupervised *k*-means on feature vectors in the last layer associated with this class to further split it into multiple subclasses. Images of two subclasses are shown in Figure 8. Images in the same subclasses are visually similar. However, they are not alike across subclasses. The two subclasses are grouped together under the horned rattlesnake class because they share the same class label (despite strong visual dissimilarity). That shows the power of label guidance. However, the feature distance is shorter for images in the same subclass and longer for images in different subclasses. This is due to the inherent clustering capability of CNNs.

Discussion and open issues

Discussion

A CNN was viewed as a guided multilayer RECOS transform in this article. The following known facts can also be explained using this interpretation.

- **Robustness to wrong labels.** Humans do clustering first, and then the CNN mimics humans based on the statistics of all labeled samples. It can tolerate small percentages of erroneous labels since these wrongly labeled data do not have a major impact on clustering results.
- **Overfitting.** Overfitting occurs when a statistic model describes noise instead of the underlying input/output relationship. For a given number of observations, this could happen for an excessively complex model that has too many model parameters. Such a model has poor prediction performance since it overreacts to minor fluctuations in the training data. Although a CNN has a large number of parameters (specifically, filter weights), it does not suffer much from overfitting for the following reason. When there are only input and output layers without any hidden layers in between, the CNN is degenerated to a linear system that solves a linear least-squared regression problem (where no rectification is needed.) It is well known the linear regression is robust to noisy data. When there are hidden layers, the filter weight determination is a cascaded optimization problem, which has to be solved iteratively. In the BP process, we update the filter weights layer by layer in a backward direction. Fundamentally, it still attempts to solve a regression problem at each layer. Although a rectifier conducts rectification on the output, it does not change the regression nature of MLPs and CNNs.
- **Data augmentation.** A low-cost way to generate more samples is data augmentation. This is feasible since minor perturbations in the image pixel domain do not change their class types.
- **Data set bias.** A CNN can be biased due to the inherent bias in the low level representation existing in training samples. Thus, the performance of a CNN can degrade significantly from one data set to the other in the same application domain due to this reason.

Open issues

There are many interesting open problems remaining for further exploration.

- **Network Architecture Design.** It is interesting to be able to specify the layer number and the filter number per layer for given applications automatically.
- **Decoder network analysis.** The classification network maps an image to a label. There are image processing networks that accept an image as the input and another image as the output. Examples include superresolution networks, semantic segmentation networks, etc. These networks can be decomposed into an encoder-decoder architecture. The analysis in this lecture note focuses on the encoder part. It is interesting to generalize the analysis to the decoder part as well.
- **Localization and attention.** Region proposals have been used in object detection to handle the object localization problem. Learning the object location and human visual attention from the network automatically without the use of proposals is desirable. The design and analysis of networks to achieve this goal is interesting.
- **Transfer learning.** It is often possible to fine-tune a CNN for a new application based on an existing CNN model trained by another data set in another application. This is because the low-level image representation corresponding to the beginning CNN layers can be very flexible and equally powerful.
- **Weakly supervised learning.** Unsupervised and heavily supervised

learning are two extremes. Weakly supervised learning occurs most frequently in our daily applications. The design and analysis of a weakly supervised learning mechanism based on CNNs is interesting and practical.

The interpretation of a CNN as a guided multilayer RECOS transform should be valuable to the investigation of these topics.

Summary

The operating principle of CNNs was explained as a guided multilayer RECOS transform in this note. A couple of illustrative examples were provided to support this claim. Several known facts were interpreted accordingly, and some open issues were pointed out at the end.

Acknowledgments

I would like to thank Andrew Szot, Shangwen Li, Zhehang Ding, and Gloria Budiman for their help in running experiments and drawing figures for this work. I am also grateful for the valuable feedback from friends, including Bart Kosko, Kyoung Mu Lee, Sun-Yuan Kung, and Jenq-Neng Hwang. This material is based on research sponsored by the Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory (AFRL) under agreement number FA8750-16-2-0173.

The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA and AFRL or the U.S. Government.

Author

C.-C. Jay Kuo (cckuo@ee.usc.edu) is a professor of electrical engineering at the University of Southern California, Los Angeles.

References

[1] C.-C. J. Kuo, "Understanding convolutional neural networks with a mathematical model," *J. Vis. Commun. Image Represent.*, vol. 41, pp. 406–413, Nov. 2016.

[2] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, 1943.

[3] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1957.

[4] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control Signals Syst.*, vol. 2, no. 4, pp. 303–314, 1989.

[5] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.

[6] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, pp. 193–202, 1980.

[7] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[8] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recogn. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.

Rodrigo Capobianco Guido

Effectively Interpreting Discrete Wavelet Transformed Signals

Following two decades of research focusing on the discrete wavelet transform (DWT) and driven by students' high level of questioning, I decided to write this essay on one of the most significant tools for time-frequency signal analysis. As it is widely applicable in a variety of fields, I invite readers to follow this lecture note, which is specially dedicated to show a practical strategy for the interpretation of DWT-based transformed signals while extracting useful information from them. The particular focus resides on the procedure used to find the time support of frequencies and how it is influenced by the wavelet family and the support size of corresponding filters.

Digital Object Identifier 10.1109/MSP.2017.2672759
Date of publication: 26 April 2017

Relevance

Frequently, DWT computation is much faster than that of the discrete Fourier transform (DFT) and even the fast Fourier transform (FFT), encouraging its usage. Furthermore and opposite to the DFT, to the FFT, and even to the short-time Fourier transform, the DWT reveals the time support of frequencies efficiently, as described in [1]. Thus, its study is of paramount importance.

Prerequisites

On one hand, essential knowledge of digital filters and wavelet families is only desirable. On the other hand, the basic aspects involving the procedures used to calculate DWTs from discrete-time signals are imperative for the full comprehension of this lecture note. If readers are not comfortable with the topic, I encourage you to consult [1]–[3], written in a friendly manner, before proceeding any further.

Problem statement and solution

Problem statement

Given the pair of signals $s[\cdot]$ and $y[\cdot]$, both of size M equal to a power of two and indexed from 0 to $M-1$, being the former and the latter the discrete-time input and its DWT, respectively, the problem is to extract and interpret

relevant information the latter contains about the former, particularly allowing the time support of frequencies to be accurately found.

Solution

Differently from the DFT, which converts a signal from the time to the frequency domain, the DWT *filters* it. Thus, significant amplitudes in $y[\cdot]$ come from frequencies that originally exist in $s[\cdot]$ and were not removed by the filters. Based on this process, our solution consists of a DWT-based filtering applied to $s[\cdot]$ followed by an inspection of the dominant amplitudes contained in the transformed signal $y[\cdot]$.

The practical procedure used to obtain $y[\cdot]$ from $s[\cdot]$ has already been described in detail in [2]; thus, it is not repeated at this time. Nevertheless, there are two main issues to be addressed to solve the stated problem: the ideal level of transformation, i.e., j, and the appropriate analysis filter pair, i.e., the low-pass filter $h[\cdot]$ and the high-pass mirrored finite impulse response half-band filter $g[\cdot]$ [4]. Once j is defined, the generalized and well-known DWT decomposition tree [4] will be more or less accurate depending on the order $N-1$ of those filters, N being their length, and on the family they belong to, such as Haar, Daubechies, Coiflets, Symmlets, and so on [4].

Notably, spectral description and temporal localization hold distinct requirements for ideality. On one hand, the former offers the finest resolution, whenever j and N are maximized and the associated filters exhibit linear-phase and maximally flat responses at their pass and stop bands. On the other hand, the latter is foolproof whenever j and N are minimized because each sample of $y[\cdot]$ refers just to a small part of $s[\cdot]$. Furthermore, according to Heisenberg's uncertainty principle [4], time and frequency information are antagonistic. The better that is, the worse this is, and vice versa. Considering the limits imposed by the real conditions, we may either balance time and frequency information or prioritize one of them based on our specific needs.

Disregarding temporal information, the finest spectral resolution is guaranteed whenever the deepest decomposition level is adopted, i.e., $j = \log_2(M)$. Additionally, the maximum possible length for N, for which there is no upper bound, is required. In practice, it usually implies filters for which $N \geq 40$. Maximally flat responses at their pass and stop bands, such as those provided by Daubechies' filters, may be used to avoid improper floatations in $y[\cdot]$ due to a reduced or an excessive gain at some subbands [4], causing inaccurate frequency magnitudes. Furthermore, distinct delays for different frequency bands are also present, whenever the phase responses of the filters are not linear; thus, Symmlets and Coiflets [4], which exhibit almost linear responses, are also interesting choices. Contrary to this, the best temporal resolution, which causes the poorest spectral description, requires the lowest decomposition level, i.e., $j = 1$, and the smallest support size of filters, i.e., $N = 2$, directly implying in the adoption of Haar's filters [4].

Opposed to an extreme time or frequency resolution, equilibrium is reached whenever the intermediary decomposition level is chosen, i.e., the situation in which j is the mean between the minimum, one, and the maximum, $\log_2(M)$. This results in $j = \lfloor (1 + \log_2(M))/2 \rfloor$, where $\lfloor \cdot \rfloor$ is the floor operator, and implies that the corresponding frequency resolution is $r = T/2^j = (T/2^{\lfloor (1+\log_2(M))/2 \rfloor})$ Hz, where $2T$ is the sampling rate at which $s[\cdot]$ was digitalized and, according to Nyquist's theorem [4], T is its maximum frequency content. Complementarily, a balanced time and frequency accuracy also requires an intermediary value for N, for which two is the lower bound and there exists no upper bound. To circumvent the missing bound, a careful inspection of the procedure used to calculate DWTs, as described in [2], is useful. It allows the statement that at the jth-level transformation based on a filter of support size N, information from $(N-1)(2^j-1)+1$ subsequent samples of $s[\cdot]$ is grouped together, directly influencing the accurate time support of frequencies.

The final issue to be solved, before formulating a solution to the stated problem, involves the specific map to be adopted: the regular DWT map or the DWT-packet map. On one hand, the former only provides the finest resolution for the lowest frequencies, which are usually those that carry most of the useful information. On the other hand, the latter provides a uniform and equally distributed time-frequency resolution for all subbands, once j, N, and the wavelet family are defined. At each decomposition level of the former map, the transformation produces two half-size and half-band signals—trend and fluctuation, respectively—containing the low and high frequencies of the input signal that are subsequently concatenated to establish the transformed signal. Contrastingly, the latter map contains 2^j subband signals of size $(M/2^j)$ at the jth decomposition level, as detailed in [5]. Figure 1 helps to recall these schemes.

Based both on the previous notes and Mallat's algorithm [2], the use of a regular DWT map, as exemplified in Figure 1(a), requires strategy STG_A-(i) and (ii) to solve our problem, assuming an ideal situation:

i) The energy of the ith sample of the jth-level trend contains the amplitude of frequencies between 0 and $(T/2^j)$ Hz, which are located within the range $\{s_{i2^j}, s_{(i+1)2^j-1}\}$, for $i = \{0,1,2,...,(M/2^j)-1\}$.
ii) The energy of the ith sample of the jth-level fluctuation contains the amplitude of frequencies between $(T/2^j)$ and $(T/2^{j-1})$ Hz, which are located within the range $\{s_{i2^j}, s_{(i+1)2^j-1}\}$.

Accordingly, the DWT-packet map built considering the natural frequency ordering (NFO) [5, p. 111], which is exemplified in Figure 1(b), is associated with the strategy STG_B:

The energy of the ith sample of the bth subband at the jth-level contains the amplitude of frequencies between $(bT/2^j)$ and $((b+1)T/2^j)$ Hz, which are located within the range $\{s_{i2^j}, s_{(i+1)2^j-1}\}$, for $b = \{0,1,2,...,2^j-1\}$ and $i = \{0,1,2,...,(M/2^j)-1\}$.

To lessen the imprecise temporal localization caused by the influence of N, correction COR is required right after calculating the DWT but prior to using those strategies:

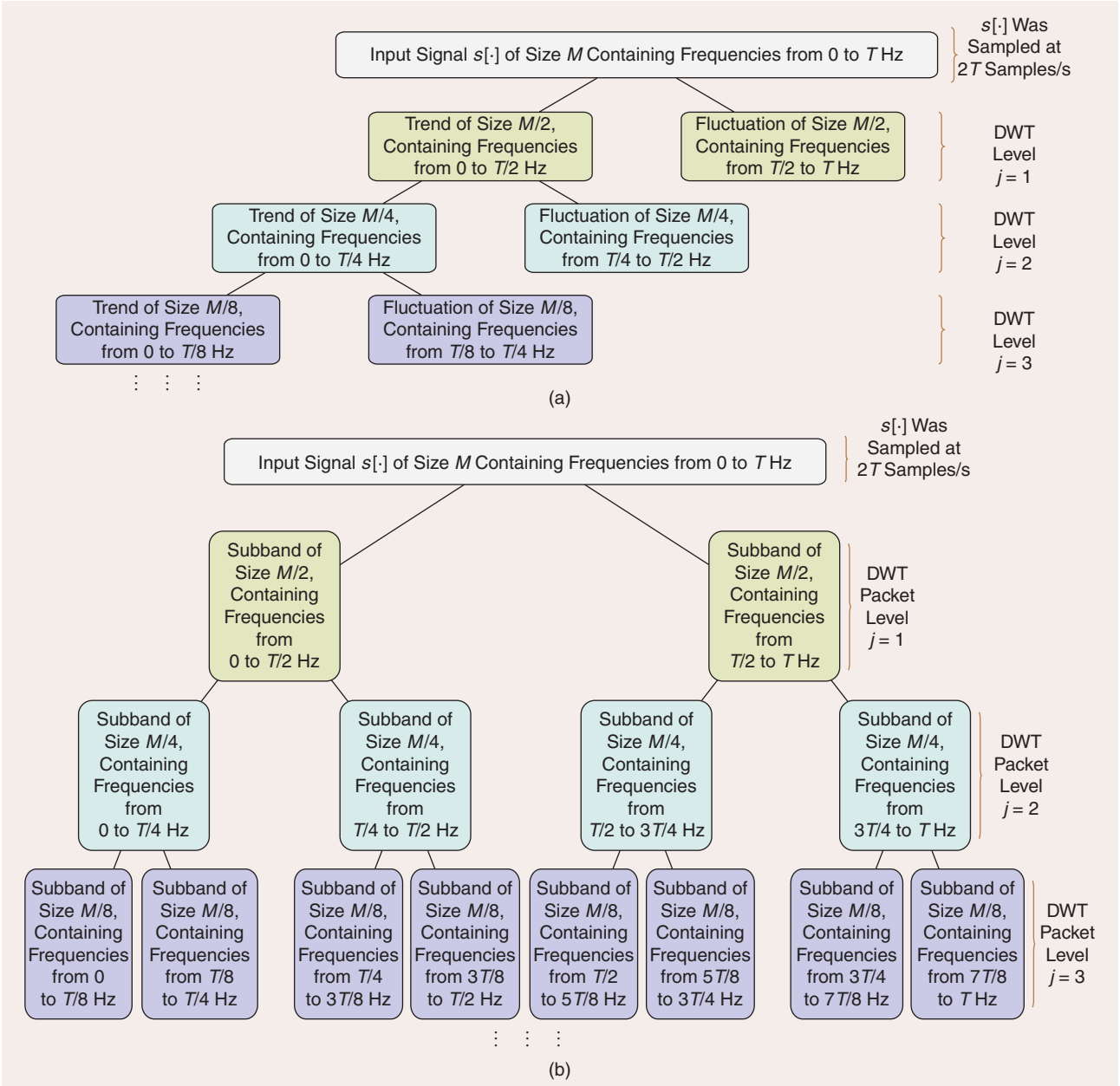


FIGURE 1. (a) The traditional DWT decomposition tree, exemplified for $j = 3$, where the $(M/8) + (M/8) + (M/4) + (M/2) = M$-sample long signal $y[\cdot]$ corresponds to the concatenation of all the leafs, considering a left-to-right and bottom-up procedure. (b) The corresponding DWT-packet tree, built based on the NFO.

Apply individually, for each leaf subband contained in the wavelet decomposition tree at the level j, regardless of coming from the regular DWT map or the DWT-packet map, a circular right-sided shift of $(N-2)/2$ samples.

COR comes from the fact that the convolution between $s[\cdot]$ and a wavelet filter produces a signal for which the length corresponds to $M+N-1$, primarily. Then, once the downsampling by two is applied [2], this length is reduced to $(M+N-1)/2 = (M/2) + (N-1)/2$. Finally, the wraparound procedure [2] vanishes the effect of the term $(N-1)/2$, remaining only a $(M/2)$-sample-long resulting signal. Thus, even though N has no influence over the length of the transformed signal, it does cause shifts. Particularly, due to the fact that N is usually even and half-shifts are not defined in discrete wavelet systems, $(N-1)/2$ can be treated as being $(N-2)/2$, implying that all the filters but Haar's cause a shifting. Therefore, COR consists of a circular right-sided shift of $(N-2)/2$ samples, regardless of j. Notably, COR does not correct delays originated from nonlinear phase responses of the filters, which can be attenuated with the use of specific wavelet families, as aforementioned.

Finally, a refined time-frequency description based on the previously defined strategies requires the use of

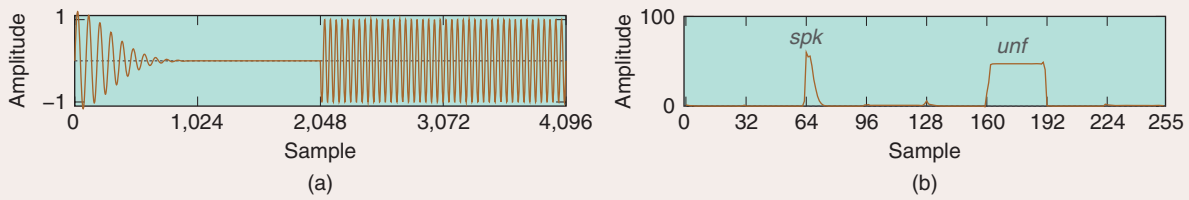


FIGURE 2. The original and transformed data for the example in the section "Numerical Example." (a) The original signal under analysis ($s[\cdot]$) and (b) the first four subbands of the corresponding DWT-packet map, modified and converted based on COR and TEG, respectively ($\tilde{y}[\cdot]$).

Teager's energy operator [6], instead of the ordinary energy [7], to convert all the samples of each leaf subband into their instantaneous potentials, according to the TEG procedure:

For the subband $b = 0, 1, ..., 2^j - 1$ of the deepest level j, do:

- $\tilde{y}_{0,b} \leftarrow (y_{0,b})^2$
- $\tilde{y}_{i,b} \leftarrow (y_{i,b})^2 - (y_{i-1,b}) \cdot (y_{i+1,b})$, for $(1 \le i \le (M/2^j) - 2)$, whenever the length of b is at least four
- $\tilde{y}_{(M/2^j)-1,b} \leftarrow (y_{(M/2^j)-1,b})^2$, whenever the length of b is at least two

where the left arrows mean attribution. Summarizing, the ideal time-frequency analysis of $s[\cdot]$ based on the DWT requires four steps:

- **Step 1**: Select the most appropriate wavelet family and the support size of filters, i.e., N. Then, calculate the M-sample-long transformed signal $y[\cdot]$ from $s[\cdot]$ based either on the regular DWT map or the DWT-packet map.
- **Step 2**: Apply COR, shifting $y[\cdot]$ accordingly.
- **Step 3**: Apply TEG, creating $\tilde{y}[\cdot]$ from shifted $y[\cdot]$.
- **Step 4**: Interpret $\tilde{y}[\cdot]$ based either on STG_A or STG_B, depending on which map was chosen to carry out Step 1.

Numerical example

Problem statement

Let $s_i = 1.2 \sin((2\pi \cdot 44i)/4095)) e^{-(\frac{i}{409.6})^2}$ for $\{0 \le i < 1{,}024\}$, $s_i = 0$, for $\{1{,}024 \le i < 2{,}048\}$, and $s_i = \sin((2\pi \cdot 85i/4{,}095)$ for $\{2{,}048 \le i < 4{,}096\}$ to be the 1-s long signal under analysis, sampled at $2T = 4{,}096$ samples/s, as shown in Figure 2(a). The problem is to find out the time support of its frequencies with a balance between temporal and spectral accuracies.

Solution

A balanced time-frequency analysis of $s[\cdot]$, which contains frequencies up to $T = 2{,}048$ Hz, requires $j = \lfloor (1 + \log_2(M))/2 \rfloor = \lfloor (1 + \log_2(4{,}096))/2 \rfloor = 6$. Correspondingly, $r = (T/2^j) = (2{,}048/2^6) = 32$ Hz is the frequency resolution for this decomposition, which is based on the DWT-packet map to allow a uniform analysis, i.e., to keep r intact for all the subbands. Symmlets of size $N = 16$ were chosen to perform the analysis because of their almost linear phase responses, practically avoiding frequency components to become misaligned.

Proceeding with the calculations and applying COR and TEG, we get $\tilde{y}[\cdot]$, for which only the first four subbands appear in Figure 2(b) to allow magnification. Its inspection reveals two main elements: *spk* and *unf*. The former is a spike decaying from samples 64 to 79, and the latter consists of a uniform step from samples 160 to 191. Observing that each one of the $2^j = 2^6 = 64$ subbands contains $(M/2^j) = (4096/2^6) = 64$ samples, indexed from 0 to 63, we conclude that *spk* occurs within the interval $\{i = 0, i = 15\}$ of the second subband, i.e., $b = 1$, and *unf* occurs within the interval $\{i = 32, i = 63\}$ of the third subband, i.e., $b = 2$. Thus, according to STG_B:

- For *spk*, the energy of the sample $i = 0$ of the subband $b = 1$ at the level $j = 6$ contains the amplitude of frequencies between $(1 \cdot 2{,}048/2^6) = 32$ and $((1 + 1) \cdot 2{,}048)/2^6 = 64$ Hz, located within the range $\{s_{0 \cdot 2^6}, s_{(0+1)2^6-1}\}$, i.e., $\{s_0, s_{63}\}$. Accordingly, the energy of the sample $i = 15$ of that subband contains the amplitude of those frequencies, which are located within the range $\{s_{15 \cdot 2^6}, s_{(15+1)2^6-1}\}$, i.e., $\{s_{960}, s_{1023}\}$. It means that a frequency between 32 and 64 Hz exists in $s[\cdot]$, starting and ending somewhere between $\{s_0, s_{63}\}$ and $\{s_{960}, s_{1023}\}$, respectively. In fact, the frequency of 44 Hz was defined in $s[\cdot]$ within the interval $\{s_0, s_{1023}\}$. Furthermore, the decay starting from sample 64 of Figure 2(b) shows the corresponding power reduction clearly seen within the corresponding interval of Figure 2(a).
- For *unf*, the energy of the sample $i = 32$ of the subband $b = 2$ at the level $j = 6$ contains the amplitude of frequencies between $2 \cdot 2048/2^6 = 64$ and $((2 + 1) \cdot 2048)/2^6 = 96$ Hz, located within the range $\{s_{32 \cdot 2^6}, s_{(32+1)2^6-1}\}$, i.e., $\{s_{2048}, s_{2111}\}$. Complementarily, the energy of the sample $i = 63$ of that subband contains the amplitude of those frequencies, which were placed within the range $\{s_{63 \cdot 2^6}, s_{(63+1)2^6-1}\}$, i.e., $\{s_{4032}, s_{4095}\}$. It means that a frequency between 64 and 96 Hz exists in $s[\cdot]$, starting and ending somewhere between $\{s_{2048}, s_{2111}\}$ and $\{s_{4032}, s_{4095}\}$, respectively. In fact, the frequency of 85 Hz was uniformly defined in $s[\cdot]$ within the interval $\{s_{2048}, s_{4095}\}$.

What we have learned

Based on this article, readers could have learned how to efficiently interpret discrete wavelet-transformed signals, while extracting from them the time support of frequencies contained in the original signal under analysis. Supplementary to this, the criteria used

(continued on page 100)

Call for Papers and Sponsors

ICASSP2018

The 43rd IEEE International Conference on Acoustics, Speech and Signal Processing

April 22 - 27, 2018, Seoul, Korea

http://2018.ieeeicassp.org

Signal Processing and Artificial Intelligence: Changing the World

Submission of Papers

Authors are invited to submit papers of not more than four pages of technical content including figures and references, with an optional fifth page containing only references. Submission instructions, paper format templates, and other important information will be made available on the ICASSP 2018 website, http://2018.ieeeicassp.org.

Conference Topics

The conference will feature world-class international speakers, tutorials, exhibits, lectures and poster sessions from around the world. Topics include but are not limited to:

- Audio and acoustic signal processing
- Sensor array & multichannel signal processing
- Bio-imaging and biomedical signal processing
- Signal processing education
- Design & implementation of signal processing systems
- Signal processing for communications & networking
- Image, video & multidimensional signal processing
- Signal processing theory & methods
- Industry technology tracks
- Signal processing for big data
- Information forensics and security
- The Internet of Things & RFID
- Machine learning for signal processing
- Speech processing
- Spoken language processing
- Multimedia signal processing
- Remote sensing and signal processing
- Signal processing for brain machine interface
- Signal processing for smart systems
- Signal processing for cyber security
- Computational imaging

Call for Tutorials

Tutorials at ICASSP form an important part of the program, giving attendees the opportunity to learn about current research areas that are of growing interest to the signal processing community. Those who are interested in presenting a tutorial may want to contact one of the tutorial chairs before preparing a formal proposal. It is important to keep in mind, for any tutorial, that it should be tutorial in nature, and within the grasp of a wide audience.

Call for Special Sessions

The program for ICASSP 2018 will include Special Sessions that complement the traditional program with new and emerging topics of significant interest to the signal-processing community, particularly those that are in line with the theme of the conference. Please refer to the conference webpage for information about Special Session proposals.

Call for Exhibitors and Sponsors

ICASSP 2018 offers exhibitors and sponsors an opportunity to showcase their company's products and innovative solutions at the Signal Processing Society's flagship conference that will be held for the first time in the Korean Peninsula. Please refer to the conference webpage for information about signing up to become an exhibitor or sponsor at ICASSP.

Signal Processing Letters

Authors of IEEE Signal Processing Letters (SPL) papers will be given the opportunity to present their work at ICASSP 2018, subject to space availability and approval by the Technical Program Chairs. SPL papers published between January 1, 2017 and December 31, 2017 are eligible for presentation at ICASSP 2018. Because they are already peer-reviewed and published, SPL papers presented at ICASSP 2018 will neither be reviewed nor included in the proceedings.

Important Dates

August 4, 2017
Special Session Proposals Due

August 11, 2017
Tutorial Proposals Due

September 8, 2017
Notification of Special Session Acceptance

September 15, 2017
Notification of Tutorial Acceptance

October 27, 2017
Paper Submissions Due

January 12, 2018
Signal Processing Letters Due

January 26, 2018
Notification of Paper Acceptance

February 9, 2018
Revised Paper Upload Deadline

February 16, 2018
Author Registration Deadline

General Chairs

Monson Hayes
Hanseok Ko

Technical Program Chairs

Dan Schonfeld
Pascale Fung
Nam Ik Cho

Sponsored by

IEEE

IEEE Signal Processing Society

Digital Object Identifier 10.1109/MSP.2017.2681139

Jørgen Arendt Jensen, Carlos Armando Villagómez Hoyos,
Simon Holbek, and Kristoffer Lindskov Hansen

Velocity Estimation in Medical Ultrasound

This article describes the application of signal processing in medical ultrasound velocity estimation. Special emphasis is on the relation among acquisition methods, signal processing, and estimators employed. The description spans from current clinical systems for one- and two-dimensional (1-D and 2-D) velocity estimation to the experimental systems for three-dimensional (3-D) estimation and advanced imaging sequences, which can yield thousands of images or volumes per second with fully quantitative flow estimates. Here, spherical and plane wave emissions are employed to insonify the whole region of interest, and full images are reconstructed after each pulse emission for use in velocity estimation.

Ultrasound velocity estimation is widely used for diagnosing circulatory and cardiac problems in the human body. The investigations are conducted in real time and can directly reveal quantitative data about blood velocity, turbulence, volume flow, resistive index, and other haemodynamic quantities. Ultrasound velocity estimation can provide instantaneous images of the spatial velocity distribution as well as single site measurements of velocity distribution and its evolution over time. Conventional ultrasound velocity estimation is widely used due to the widespread ability, its inexpensive nature, no preparation of the patients, and safety, since nonionizing radiation is used. The underlying estimation schemes involve a combination of ultrasound physics, image acquisition, and velocity estimation. This article touches on all three aspects and spans from the currently available commercial systems using sequential data acquisition to the latest research techniques using parallel acquisition yielding thousand of images per second. They are capable of showing complex pulsating flow with vortices in two and three dimensions. The signal processing is tightly integrated with the underlying physics.

Digital Object Identifier 10.1109/MSP.2017.2665719
Date of publication: 26 April 2017

Physics of velocity estimation

Conventional ultrasound systems are often called *Doppler systems*, indicating that the instantaneous frequency shift of the emitted spectrum is estimated. This would indicate that a single pulse-echo measurement is sufficient to find the velocity. However, this is not done in practice as scattering and attenuation contribute to significant shifts in the received signal's mean frequency. Unless the unknown factors from attenuation and scattering are compensated for, a frequency estimation would be highly inaccurate. Thus, a more precise understanding is needed.

In basic blood ultrasound imaging, a single blood scatterer traverses the ultrasound beam. The position is denoted as $\vec{r}_1$ at the first measurement and $\vec{r}_2$ for the second measurement acquired T_{prf} s after. The interpulse motion and, thereby positional shift, gives rise to a delay t_s in the second received signal corresponding to

$$t_s = \frac{2\,|\,\vec{r}_1 - \vec{r}_2\,|\cos\theta}{c} = \frac{2\,|\,\vec{v}\,|\,T_{prf}\cos\theta}{c} = \frac{2v_z}{c}T_{prf}, \quad (1)$$

where θ is the angle between the ultrasound beam and the blood motion direction, and c is the speed of sound, usually 1,540 m/s. The axial velocity component is $v_z = |\,\vec{v}\,|\cos\theta$. Motion away from the transducer gives rise to a signal arriving with an increasing delay, while a scatterer moving toward the transducer will decrease the time from pulse emission to reception.

The emitted pulse $p(t)$ in ultrasound is sinusoidal, and for velocity estimation often consists of four to eight cycles. The received signal can be written as

$$x_i(t) = ba\sin\left(2\pi f_0\left(t - \frac{2d}{c} - i\cdot t_s\right)\right), \quad (2)$$

where a is the pulse amplitude, f_0 is the emitted frequency, d is the depth of the scatterer, and b is its amplitude. Making a measurement at one fixed time instance $t_d = n_d/f_s$ for sample index n_d, corresponding to a fixed depth, then gives a sampled signal as

$$x_s(n_d, i) = -ba\sin\left(2\pi\frac{2v_z}{c} \times f_0 T_{prf}\, i - \Theta_d\right),$$
$$\Theta_d = 2\pi f_0\left(n_d/f_s - \frac{2d}{c}\right), \quad (3)$$

where f_s is the sampling frequency and i is the pulse emission number. This data

acquisition over several pulse emissions gives a signal sample with a rate of $1/T_{prf} = f_{prf}$, which has a frequency of

$$f_p = \frac{2v_z}{c} f_0 = \frac{2v_z}{\lambda}, \tag{4}$$

directly proportional to the axial blood velocity. Here $\lambda = c/f_0$ is the wavelength of the transmitted pulse. The frequency content in the transmitted pulse is scaled by the factor $2v_z/c$ and also reduces the influence of scattering and attenuation by this factor. Velocity estimation, thus, relies on acquiring the flow signal over several pulse emissions and determining the motion between pulse emissions for estimating the axial velocity.

Axial velocity estimation

Several methods for finding the axial velocity have been introduced [1]. Spectral systems use short-time Fourier transform to determine the frequency content of $x_s(n_d,i)$ for the sample depth n_d, and these systems show the velocity distribution over time for a single position in the blood vessel as a spectral display [2]. An example of such a display is shown in the left part of Figure 1(a). The gray tone B-mode image displays the anatomy, and the broken yellow lines indicate the place for finding the velocity distribution. In this spectrogram, the brightness indicates the relative number of scatterers moving at a given speed. The yellow "wings" indicate the beam-to-flow angle, which has to be compensated for to yield the correct velocity magnitude. It is therefore preferred to keep this angle below 60° if possible to avoid large errors in the compensated velocities. This display gives a quantitative measure of velocity and is widely used in the clinic to evaluate peak velocities, volume flow, and resistive index for indicating resistance to flow in the vessel.

The velocity can also be directly estimated by cross-correlating two consecutive signals and then determining the time shift [3]:

$$\begin{aligned} R_{12}(n_d,k) &= \sum_n x_s(n_d+n,i)x_s \\ &\quad \times (n_d+n+k,i+1) \\ &= \sum_n x_s(n_d+n,i)x_s \\ &\quad \times (n_d+n-n_s+k,i) \\ &= R_{11}(n_d,k-n_s), \\ n_s(n_d) &= \frac{2v_z(n_d)}{c} T_{prf} f_s, \end{aligned} \tag{5}$$

where n is the sample index. Here $R_{12}(n_d,k)$ is the cross-correlation function and R_{11} is the autocorrelation of the received signal, which has a unique peak found at the lag $k = n_s$. The velocity is then

$$v_z(n_d) = \frac{n_s(n_d)}{2f_s T_{prf}} c. \tag{6}$$

For roughly monochromatic signals (long pulse emissions with four to eight cycles), the phase shift between the two signals can be determined using the lag one autocorrelation of the signals [4]. Here it is needed to perform a Hilbert transform, $y_s(n,i) = \mathcal{H}\{x_s(n,i)\}$, on the received signal to find the direction of the motion. This phase shift is directly proportional to the velocity and is found by

$$\begin{aligned} v_z(n_d) &= -\frac{cf_{prf}}{4\pi f_0} \\ &\times \arctan\left(\frac{\sum_{i=1}^{N_c-1} y_s(n_d,i+1)x_s(n_d,i) - x_s(n_d,i+1)y_s(n_d,i)}{\sum_{i=1}^{N_c-1} x_s(n_d,i+1)x_s(n_d,i) + y_s(n_d,i+1)y_s(n_d,i)} \right), \end{aligned}$$

where N_c is the number of emissions to average over.

The velocity can then be found as a function of depth, when using eight to

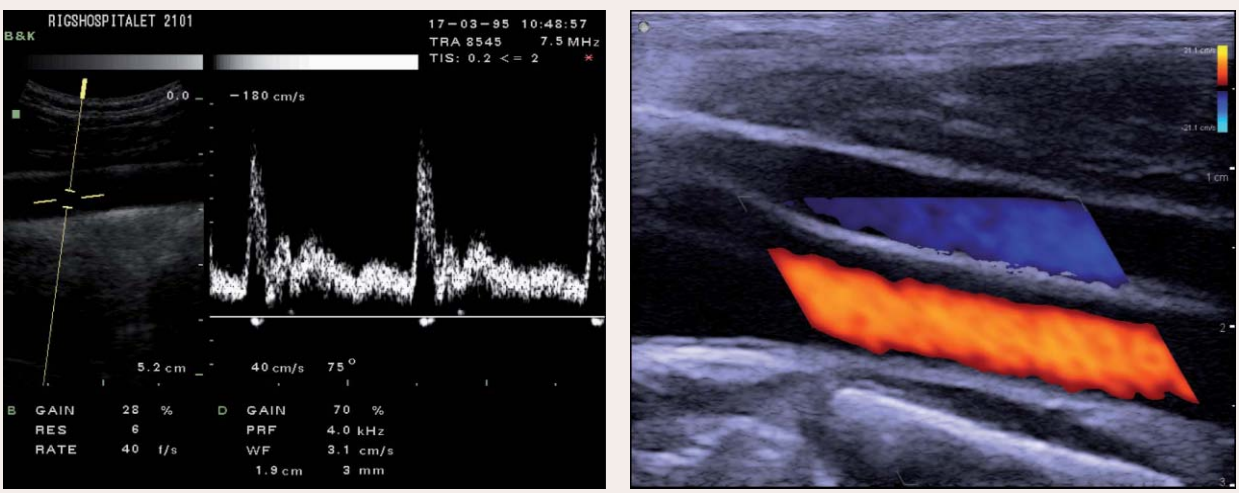


FIGURE 1. (a) A duplex ultrasound scan shows, at the left, the anatomic B-mode image. The yellow broken line indicates the position for velocity estimation and, at the right, shows the spectral display of the velocity distribution as a function of time. The brightness indicates the relative number of scatterers moving at a certain speed. (b) A color flow map where the gray tones indicate the anatomy and the red color indicates velocity toward the transducer and blue away from it. The images were acquired from the carotid artery and jugular vein in the neck of a human volunteer.

16 pulse emissions in the same direction. This is performed in a number of directions, and an image of the axial velocity distribution is acquired and displayed as shown in Figure 1(b). The image is acquired from the neck and shows the carotid artery supplying blood to the brain and the returning blood via the jugular vein with flow in the opposite direction. The underlying gray-level image shows the anatomy. The red color indicates velocities toward the probe, which is placed at the top of the image, and blue represents velocities away. Such color flow map (CFM) images can be acquired at a rate of 10–60 Hz, depending on the depth and the number of directions for performing velocity imaging. The images give a real-time visualization of flow in the arteries and an indication of the axial velocity direction and magnitude. CFM images are not quantitative due to the angle dependence of the estimates, and the standard deviation is usually high due to the few emissions employed in the velocity estimation. They are, however, very helpful in diagnosing many haemodynamic problems like stenosis and cardiac function, and widely used for detection of vascularization in, e.g., infection, inflammation, and cancer.

Challenges

The scattering from blood is weak, and the vessels in an ultrasound image appear black. It is, thus, vital to make a processing for reducing the noise influence. Matched filters are employed on the data, and averaging is performed in the velocity estimators over both a number of pulse emissions N_c and along the depth direction. This improves the variance of the velocity estimates proportionally to the number of independent lines and samples. Often eight to 16 emissions in the same direction are averaged at the penalty of a reduction in the frame rate. For triplex images that show the anatomic B-mode image, the spectrum, and a CFM image of the velocity, a frame rate down to 5 Hz is often attained. This is not sufficient to visualize the dynamics of the heart, where a frame rate above 20 Hz up to 60 Hz is preferred.

A further complication is in separation of tissue and blood. The tissue signal amplitude is often 20–40 dB larger than that of the blood and the tissue signal has to be removed to avoid bias in the blood velocity estimation. It is assumed that the tissue is roughly stationary, and a filtration along the emissions is made by, e.g., subtracting the mean signal for all emissions from the signals. Other more advanced methods are found, and this filtration is often the most challenging part of velocity estimation, as there is a limited amount of data available, and the filtration inevitable introduces noise especially for low velocities.

Estimating the velocity vector

The estimation schemes in the "Axial Velocity Estimation" section only find the axial velocity component, but most vessels run parallel to the skin surface, so the least important component is the axial velocity. This is often sought, compensated for by tilting the ultrasound beam, but it can be difficult to maintain a beam-to-flow angle below 60° needed for achieving an acceptable 10% precision of the velocity estimate. Also, flow in the human circulation is pulsating and complex with disturbed and turbulent flow patterns, and transitory vortices, which appear in short time periods, are often found. A single angle correction, thus, cannot be performed, and the angle should be estimated for all positions in each individual image.

Several methods for finding the velocity vector have been developed and investigated. The first methods were based on dual beam systems, where the velocities were estimated for two beams at two different angles [5]. The axial and lateral velocity components could then be found from the sum and difference of the two estimates. Other methods include tracking the speckle pattern in the ultrasound image [6], and various forms of beamforming transverse to the ultrasound beam or along the flow direction.

A commercially introduced approach is transverse oscillation (TO) [7] for which advanced beamforming is used to introduce an oscillation in the ultrasound field transverse to the propagation direction. The feature that enables axial velocity estimation is the sinusoidal oscillation of the emitted pulse. This makes it possible to find the frequency, time, or phase shift. Introducing a lateral oscillation enables the estimation of the lateral velocity component.

In the focal region of an ultrasound probe, there is a Fourier relation between the amplitude weighting (apodization) of the individual elements and the lateral beam pattern. Generating a sinusoidal oscillation is, thus, possible by having two separated peaks in the apodization function. Shaping the peaks with a window function limits the width of the oscillation. Commonly, two peaks with a von Hann or Gauss window shape are used during processing the receive signals, and the lateral oscillation period or wavelength λ_x is given by [7]

$$\lambda_x = \frac{2\lambda d}{P_d} = \frac{2\lambda d}{N_d P_i}, \tag{7}$$

where d is the depth and P_d is the distance between the two peaks in the apodization function. The transducer pitch is P_i, and the number of elements between the peaks is N_d. A fairly broad transmit field with a focus beneath the region of interest is used, and the apodization is applied during the receive processing making it possible to dynamically adapt the lateral wavelength. Two beams separated by a lateral distance of $\lambda_x/4$ are focused in parallel to make it possible to find the sign of the velocity. The beams are roughly 90° phase shifted compared to each other, and a complex signal is attained like for the axial velocity estimator. Each pulse emission, thus, gives four samples: two for each beam from the axial Hilbert transform. The measured signals are given by

$$\begin{aligned} r_{sq}(n,i) &= x_l(n,i) + j y_r(n,i) \\ r_{sqh}(n,i) &= \mathcal{H}\{x_l(n,i)\} + j\mathcal{H}\{y_r(n,i)\}, \end{aligned} \tag{8}$$

where $x_l(n,i)$ is the left beam signal and $y_r(n,i)$ the right. The received signals from the transducer are then Hilbert transformed in the temporal direction n to yield $r_{sqh}(n,k,i)$. Two new signals are then formed from

$$\begin{aligned} r_1(n,k,i) &= r_{sq}(n,k,i) + j r_{sqh}(n,k,i), \\ r_2(n,k,i) &= r_{sq}(n,k,i) - j r_{sqh}(n,k,i), \end{aligned} \tag{9}$$

for reducing the influence of the axial oscillation on the lateral oscillation. A fourth-order estimator has been derived to separately estimate the axial and lateral velocity components as

$$v_x = \frac{\lambda_x}{2\pi 2 T_{prf}} \times \arctan\left(\frac{\Im\{R_1(1)\}\Re\{R_2(1)\} + \Im\{R_2(1)\}\Re\{R_1(1)\}}{\Re\{R_1(1)\}\Re\{R_2(1)\} - \Im\{R_1(1)\}\Im\{R_2(1)\}} \right), \quad (10)$$

and similarly for v_z, where $R_1(1)$ is the complex lag-one autocorrelation value for $r_1(i)$ and $R_2(1)$ is the complex lag-one autocorrelation value for $r_2(i)$. $\Im$ denotes the imaginary part and $\Re$ denotes the real part of the argument.

This approach has been implemented on scanners from BK Ultrasound (Herlev, Denmark), and an example from the aorta (main artery from the heart) in a short axis view is shown in Figure 2. During the contraction of the heart, a rotational motion of the blood is often found in the aorta in addition to the main flow direction downward in the body. This rotation is shown in the figure, where the arrows indicate direction and magnitude. The color coding also indicates direction, and a full 360° rotation is found at the late systolic phase in the cardiac cycle. This demonstrates the spatial variation in angle throughout the image, precluding the use of a single angle correction factor.

Vector flow challenges

The predominate direction of the flow in Figure 2 however, is, out of the imaging plane, and a full visualization of the flow in the human circulation necessitates a complete estimation of the 3-D velocity vector. The TO approach has been extended to full 3-D imaging by employing a 2-D matrix probe with $N \times N$ elements. The ultrasound beam can then be controlled in both directions. A broad beam is again emitted, and two apodization patterns are applied to generate four beams in parallel with two apodization peaks in both the lateral and orthogonal elevation direction. From these four beams and eight samples per emission the axial, lateral, and elevation velocity components can be determined. This measurement makes the velocity determination completely independent of the angle between the ultrasound beam and the flow direction. Quantification of peak velocities, volume flow, and other parameters potentially becomes more reliable and the system also becomes operator independent.

The 3-MHz probe employed has 32 × 32 = 1,024 elements each generating around 40 MB of data/s for a combined data flow of 40 GB/s, which must be processed in real time. Connection to the probe through 1,024 coaxial cables is also a challenge, as such a cable has a diameter of 3 cm, which makes working with it cumbersome for clinicians.

Advanced imaging and continuous data

What currently limits the accuracy and frame rate of velocity imaging is the sequential data acquisition. The pulses are emitted in one direction at a time, and eight to 16 emissions are needed for a sufficient velocity accuracy. Imaging down to 15 cm takes 200 μs/pulse, and making flow estimation in 32 directions with eight emissions, combined with an anatomic image with 128 lines results in a frame rate of 13 Hz, barely acceptable for cardiac imaging. There is, thus, a link between imaging depth, estimation accuracy, and frame rate, which is currently a major obstacle for fast and quantitative flow imaging.

This complex problem can be solved by employing fast imaging schemes using spherical [8] or plane wave emissions [9], [10]. The full imaging region is insonified using a spherical or plane wave, which covers the full imaging region as shown in Figure 3. The scattered signal is then received by all the elements and dynamically focused during receive processing. Such imaging has no transmit focusing, but the receive focusing can be synthesized by sending out spherical waves from other parts of the probe and combine all the measurements. This is called *synthetic transmit aperture* (*STA*) imaging or *ultrafast imaging for plane waves* [11], [12]. The emitted energy for STA is low, but this can be compensated for by combining a number of elements and using long chirp excitation. The scheme gives dynamic focusing throughout the image with a frame rate of thousands of frames per second, if emission sequences with few emissions are used. Continuous data everywhere in the image are,

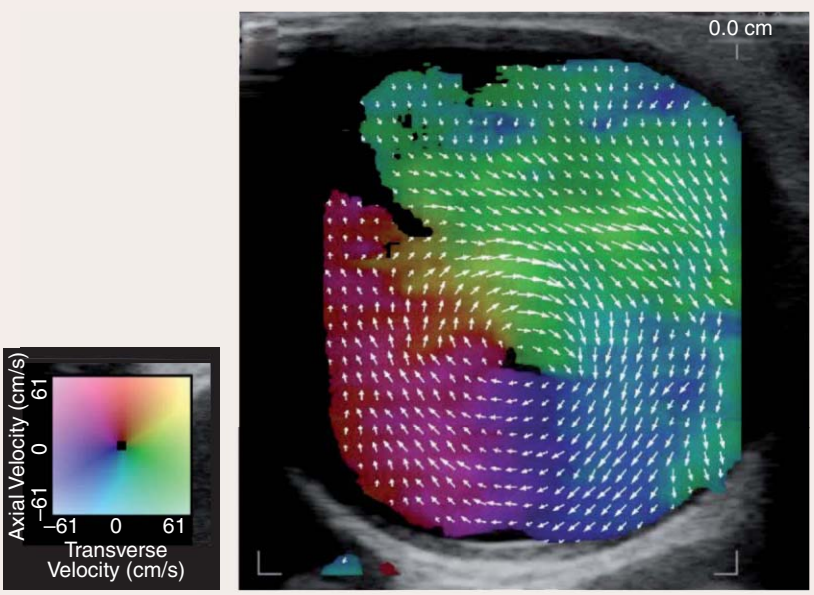


FIGURE 2. (a) The color wheel used for indicating direction and magnitude. (b) The vector flow image displays the velocities estimated at the late systolic phase in the ascending aorta in a short axis view.

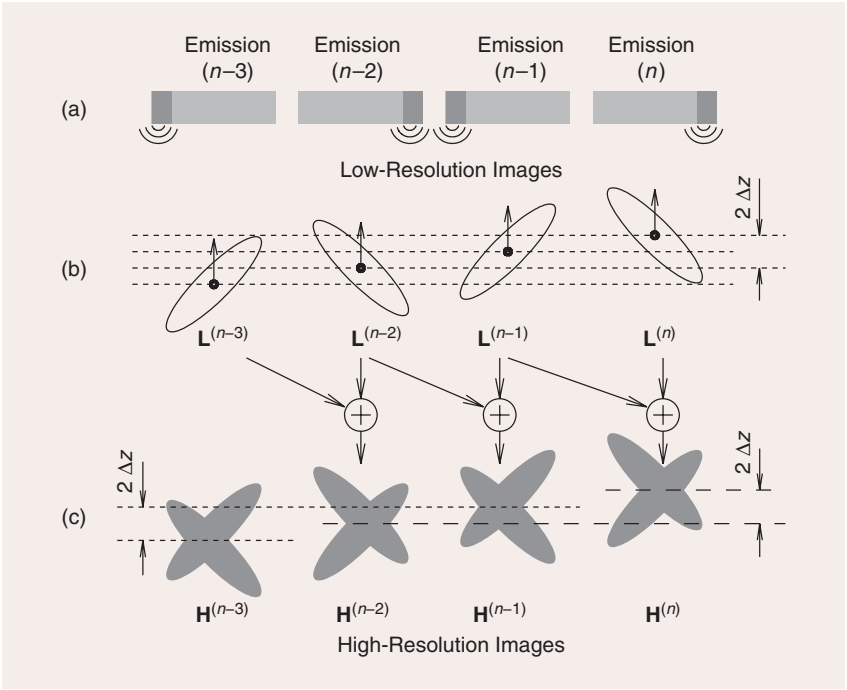


FIGURE 3. An STA imaging sequence with two emissions that are repeated: (a) the probe and from where the spherical emission emanates, (b) the PSF for each emission is indicated, and (c) the combination from several emissions is shown. (Figure used courtesy of [8].)

thus, available, and the data for velocity estimation can be averaged over as many emissions as the blood velocity is roughly stationary [11].

There is, however, one complication to this imaging scheme, as the scatterers move between emissions. This is illustrated in Figure 3. Figure 3(a) shows the probe and from where the spherical emission emanates. The sequence only consists of two emissions from the first and last elements of the probe. In (b), the elliptically shaped point spread function (PSF) beamformed from one emission is shown, and (c) shows the summation of several emissions. For each emission a low-resolution image (not focused in transmit) can be generated for the whole region. This is combined with the other emissions in (c) to yield a high-resolution image (HRI), but these are not fully in phase due to the motion. The PSF for different HRIs $H^{(n)}$ are slightly different. The HRI PSFs are, however, the same, if the emissions sequence is the same, so the pair $H^{(n-3)}$ and $H^{(n-1)}$, and the pair $H^{(n-2)}$ and $H^{(n)}$ have the same PSF, apart from a translation due to the motion. There is, thus, full correlation between the pairs $H^{(n-3)}$ and $H^{(n-1)}$ and $H^{(n-2)}$ and $H^{(n)}$. This might seem like a small detail, but it has a widespread impact on velocity imaging. The correlation of HRI data for time-shift estimation using (5) will yield the same correlation functions as long as the blood velocity is constant, and these correlation functions can be averaged to improve the velocity estimate. The continuous data, therefore, makes it possible to correlate the data over the time period the blood velocity can be considered constant, so it is the acceleration that limits observation time. The imaging can be conducted using either spherical [11] or plane waves [12].

The continuous data also decouples the number of imaging directions and accuracy, as any arbitrary number of directions can be beamformed from the acquired data. The standard deviation on the velocity estimates can be kept low by the averaging of a long period of time. The received data can be beamformed everywhere, and frame rates of hundreds to thousands of images per second with fully quantitative flow can be attained. The availability of data everywhere also makes it possible to introduce new advanced focusing methods. The TO method can be employed, and it is also possible to focus along the flow direction. This makes it possible to track the blood scatterers along their motion path and avoid decorrelation from motion in or out of the ultrasound beam to further increase the velocity accuracy. Another important aspect from continuously available data is the ability to use any echo canceling filter without being affected by initialization effects. This makes the suppression of the strong tissue signal much more efficient. These factors reduce the standard deviation (SD) by roughly a factor of ten to yield very accurate estimates with a relative SD lower than 1%.

An example of STA flow imaging is seen in Figure 4 and is taken at the carotid bifurcation in the neck. The sequence generates 2,500 images/s with fully quantitative magnitudes and directions. Experiments in a flow rig indicated an angle determination precision lower than 1.4°, and the relative SD for the magnitude was between 1.9 and 4.7% for this imaging method. The image is taken at the peak systole (maximum contraction of the heart). Flow with a magnitude around 1 m/s is seen in both vessel branches and the lower branch corresponding to the internal carotid artery also contains a vortex. It is formed upstream in the beginning of the heart contraction near the bifurcation and moves downstream along the outer vessel wall during the systolic phase. Such a vortex is found in nearly all healthy people, and it appears for roughly 100 ms, whereafter the flow again usually becomes laminar in both branches. During the diastolic phase, when the heart relaxes, the velocity decreases to below 0.2–0.1 m/s depending on the spatial position. The synthetic aperture (SA) flow imaging method can capture the high as well as low velocities due to both the fast image acquisition rate and the continuous available data. Fast flow can be tracked over a short period of time and low velocity flow, which is difficult to detect, can be observed over a longer period of time making removal of the

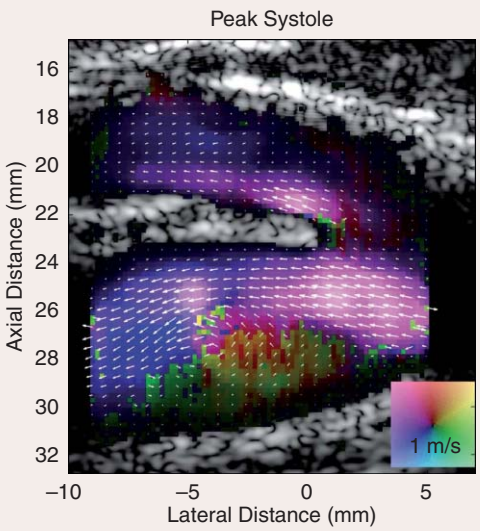


FIGURE 4. The velocity images of flow in the carotid bifurcation in the neck of a 52-year-old healthy individual. The arrows indicate direction and magnitude and the color indicates direction with intensity proportional to velocity magnitude. The frame is taken at the peak systolic phase in the cardiac cycle.

tissue signal easier. This has especially been demonstrated for plane wave flow imaging [12].

Discussion and conclusions

Velocity imaging in ultrasound has, since the 1980s, been used for easily accessible diagnosis of the human circulation for both peripheral and deep vessels as well as the heart. There has been a steady increase in the accuracy of the investigations. Two-dimensional vector flow imaging has been commercially introduced recently. The development is spurred by advances in signal processing combined with novel and precise digital beamforming methods. The field is very broad and spans more than 50 years of development reported in more than 1,000 articles. Space constraints in this article do not allow us to include more references, but consulting recent reviews can give a broader introduction to the literature.

Entirely new acquisition methods are being developed by researchers for clinical implementations. The fast SA spherical and plane wave images can easily reveal vortices and fast transitory events in the human circulation without preparation of the patient or injection of contrast agents [9]. The quantitative results also make it possible to derive diagnostic measures from the data. Peak velocities and velocity ratios before and after a constriction or stenosis can be used for grading the severity.

The continuous data make it possible to have very long echo-canceling filters to remove the signal from the tissue. Essentially, any filter can be used, and the effects of filter initialization are avoided. In this case, many emissions can be combined to reduce noise, and the SA and plane wave flow images are more sensitive to low velocity flow. This was used to map out the vasculature of the rat brain for, e.g., detecting the brain activity, when stimulating a single whisker, mapping out an epileptic seizure, and revealing the influence of odors on the brain [12].

There are still many challenges in ultrasound flow estimation, but the integrated approach by combining signal processing, estimator development with advanced acoustics, and new digital beamforming seems promising to yield full 3-D volumetric vector flow imaging in real time for both low and high velocity flows at many hundred volumes per second.

Authors

Jørgen Arendt Jensen (jaj@elektro.dtu.dk) received his M.S. and Ph.D. degrees in 1985 and 1989, respectively, from the Technical University of Denmark and the Dr.Techn. degree from the same university in 1996. Since 1993, he has been a full professor of biomedical signal processing in the Department of Electrical Engineering at the Technical University of Denmark, and the head of the Center for Fast Ultrasound Imaging since its inauguration in 1998. He has published more than 400 journal and conference papers on signal processing and medical ultrasound. His research focus is on the simulation of ultrasound imaging, synthetic aperture imaging, vector blood flow estimation, and construction of ultrasound research systems. He is a Fellow of the IEEE.

Carlos A. Villagómez Hoyos (cavh@elektro.dtu.dk) received his B.S. degree in electronic engineering in 2008 and his M.S. degree in digital signal processing in January 2013, both from the National Autonomous University of Mexico. He spent six months at the Ultrasound Laboratory at the Federal University of Rio de Janiero in 2012. He obtained his Ph.D. degree in biomedical engineering from the Center for Fast Ultrasound Imaging at the Technical University of Denmark in 2016, where he is currently a postdoctoral researcher.

Simon Holbek (sholbek@elektro.dtu.dk) received his M.S. degree in physics in 2013 from the Niels Bohr Institute, University of Copenhagen, Denmark, and his Ph.D. degree from the Technical University of Denmark in 2017. He is currently a postdoctoral researcher with the Center for Fast Ultrasound Imaging in the Department of Electrical Engineering, Technical University of Denmark. His current research topic is within three-dimensional vector flow imaging.

Kristoffer Lindskov Hansen (lindskov@gmail.com) received his M.D. and Ph.D. degrees from Copenhagen University, Denmark, in 2003 and 2010, respectively, and became a medical specialist in diagnostic radiology in 2014. He is currently with the Department of Radiology, Copenhagen University Hospital, Denmark, and is an associate professor at Copenhagen University. His research interest focuses on advanced ultrasound techniques with

the main emphasis on cardiac vector velocity estimation.

References

[1] D. H. Evans and W. N. McDicken, *Doppler Ultrasound, Physics, Instrumentation, and Signal Processing*. New York: Wiley, 2000.

[2] D. W. Baker, "Pulsed ultrasonic Doppler blood-flow sensing," *IEEE Trans. Son. Ultrason.*, vol. SU-17, no. 3, pp. 170–185, 1970.

[3] O. Bonnefous, P. Pesqué, and X. Bernard, "A new velocity estimator for color flow mapping," in *Proc. IEEE Ultrasonics Symp.*, 1986, pp. 855–860.

[4] C. Kasai, K. Namekawa, A. Koyano, and R. Omoto, "Real-time two-dimensional blood flow imaging using an autocorrelation technique," *IEEE Trans. Son. Ultrason.*, vol. 32, no. 3, pp. 458–463, 1985.

[5] B. Dunmire and K. W. Beach, "A brief history of vector Doppler," in *Proc. SPIE Medical Imaging*, Feb. 2001, vol. 4325, pp. 200–214.

[6] G. E. Trahey, J. W. Allison, and O. T. von Ramm, "Angle independent ultrasonic detection of blood flow," *IEEE Trans. Biomed. Eng.*, vol. 34, no. 12, pp. 965–967, 1987.

[7] J. A. Jensen and P. Munk, "A new method for estimation of velocity vectors," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 45, no. 3, pp. 837–851, 1998.

[8] S. I. Nikolov and J. A. Jensen, "In-vivo synthetic aperture flow imaging in medical ultrasound," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 50, no. 7, pp. 848–856, 2003.

[9] J. Udesen, F. Gran, K. L. Hansen, J. A. Jensen, C. Thomsen, and M. B. Nielsen, "High frame-rate blood vector velocity imaging using plane waves: Simulations and preliminary experiments," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 55, no. 8, pp. 1729–1743, 2008.

[10] J. Bercoff, G. Montaldo, T. Loupas, D. Savery, F. Meziere, M. Fink, and M. Tanter, "Ultrafast compound Doppler imaging: Providing full blood flow characterization," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 58, no. 1, pp. 134–147, Jan. 2011.

[11] J. A. Jensen, S. Nikolov, K. L. Gammelmark, and M. H. Pedersen, "Synthetic aperture ultrasound imaging," *Ultrasonics*, vol. 44, pp. e5–e15, 2006.

[12] M. Tanter and M. Fink, "Ultrafast imaging in biomedical ultrasound," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 61, no. 1, pp. 102–119, Jan. 2014.

SP

SP EDUCATION

(continued from page 80)

summer program for high school students," *Int. J. Eng. Educat.*, vol. 31, no. 5, pp. 1419–1427, 2015.

[71] M. Bugallo and A. M. Kelly, "A pre-college recruitment strategy for electrical and computer engineering study," in *Proc. Integrated STEM Education Conf. (ISEC)*, 2014, pp. 1–4.

[72] M. F. Bugallo and A. M. Kelly, "An outreach after-school program to introduce high school students to electrical engineering," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 2015, pp. 5540–5544.

[73] R. Gearns, A. M. Kelly, and M. F. Bugallo, "Shifts in students' views towards engineering in an out-of-school-time program," in *Proc. Annu. Conf. National Association of Research in Science Teaching (NARST)*, 2017.

[74] L. Padwa and K. Sheppard, "Real world chemistry: A model for secondary chemistry teacher professional development," in *Proc. 22nd Biennial Conf. Chemistry Education*, 2012.

[75] American Society for Engineering Education. (2014). Standards for preparation and professional development for teachers of engineering [Online]. Available: https://www.asee.org/conferences-and-events/outreach/egfi-program/k12-teacher-professional-development

[76] J. Kiely, A. M. Kelly, K. La Magna, D. J. Moloney, and R. D. Bynum, "Research on impacts of university-based biotechnology teaching laboratories on teacher professional development and student outcomes," in *Proc. National Association of Research in Science Teaching Conf.*, 2014.

[77] New York State Education Department. (2014). Physical setting/physics core curriculum [Online]. Available: http://www.p12.nysed.gov/ciai/mst/pub/phycoresci.pdf

SP

LECTURE NOTES

(continued from page 92)

to choose the wavelet family, the transformation level, and the corresponding support size of the filters were brought to light.

Author

Rodrigo Capobianco Guido (guido@ieee.org) is an associate professor at São Paulo State University (UNESP), São José do Rio Preto, Brazil. His home page at UNESP is available at www.sjrp.unesp.br/~guido/.

References

[1] R. C. Guido, P. Addison, and J. Walker, "Introducing wavelets and time-frequency analysis," *IEEE Eng. Biol. Med. Mag.*, vol. 28, no. 5, pp. 13, 2009.

[2] R. C. Guido, "Practical and useful tips on discrete wavelet transforms," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 162–166, 2015.

[3] R. C. Guido, "A note on a practical relationship between filters coefficients and the scaling and wavelet functions of the discrete wavelet transform," *Appl. Math. Lett.*, vol. 24, no. 7, pp. 1257–1259, 2011.

[4] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley, MA: Wellesley-Cambridge Press, 1997.

[5] A. Jensen and A. L. Cour-Harbo, *Ripples in Mathematics: The Discrete Wavelet Transform*. New York: Springer, 2000.

[6] J. F. Kaiser, "On a simple algorithm to calculate the energy of a signal," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Albuquerque, NM, 1990, pp. 381–384.

[7] R. C. Guido, "A tutorial on signal energy and its applications," *Neurocomputing*, vol. 179, pp. 264–282, Feb, 2016.

SP

Lifan Zhao, Xiumei Li, Lu Wang, and Guoan Bi

Autocalibrated Sampling Rate Conversion in the Frequency Domain

Frequency-domain sampling rate conversion (SRC) can be conveniently implemented by manipulating the discrete Fourier transform (DFT) of the input signal. This method has achieved the advantages of using less computation to obtain more accurate converted output. Conversion errors are mainly produced from the formulation process of the DFT of the output signal. This article presents a sparsity-based scheme to appropriately and automatically calibrate the conversion errors to make further improvement on the conversion accuracy at the cost of more computational complexity. The experimental results demonstrate that the proposed scheme can significantly decrease the mean-square errors (MSEs) and is particularly effective on minimizing the MSEs of phase spectrum.

Introduction

SRC has been an indispensable operation in various digital signal processing applications to achieve acceptable performance at affordable processing costs [1]. In the literature, various approaches have been reported to carry out the SRC either in the time or frequency domain, as illustrated in Figure 1. The time-domain SRC requires processing operations including up-sampling, low-pass filtering, and down-sampling as shown in Figure 1(a). Although various methods have been reported to implement the SRC, the success largely depends on carefully rearranging the up-/down-samplers and subfiltering functions to collectively perform the desired conversion process [1]. In contrast, the frequency-domain SRC directly manipulates the DFT of the input signal, as shown in Figure 1(b) [2], [3]. In this method, the output of the SRC is obtained by the inverse DFT (IDFT) of the manipulated DFT of the input signal [2], [4]. Although useful rules of manipulating the DFT have been given for achieving better conversion accuracy [2], [3], the optimal selection criteria for minimizing the conversion errors remains an open problem.

In spite of careful design efforts on various methods of implementing the SRC, there always exist conversion errors that are defined as the differences between the obtained and the ideal SRC outputs. In various applications including digital audio, communication, and multimedia systems, SRC with high accuracy is often desired. Therefore, it is of significant importance to investigate the possibilities of further increasing the conversion accuracy on the outputs of the SRC. Because these errors are relatively small and randomly distributed, we may consider them as perturbations. Inspired by the recent success of the sparse representation (SR) technique, let us develop a simple sparsity-based method in this article to autocalibrate the perturbations produced by the frequency-domain SRC method [3] to improve the conversion accuracy. This autocalibration method can be used as an additional measure for accuracy improvement after all of the possible wisdoms have been exhausted for implementing the frequency-domain SRC. Our formulation is different from the autocalibration problem in other SR problems [5] because sparsity assumption of the DFT coefficients is generally not valid in SRC applications. To properly facilitate error estimation in the frequency domain, sparsity in time-domain errors, which is a unique characteristic of frequency-domain SRC, is properly exploited. Simulated experimental results demonstrate that the conversion

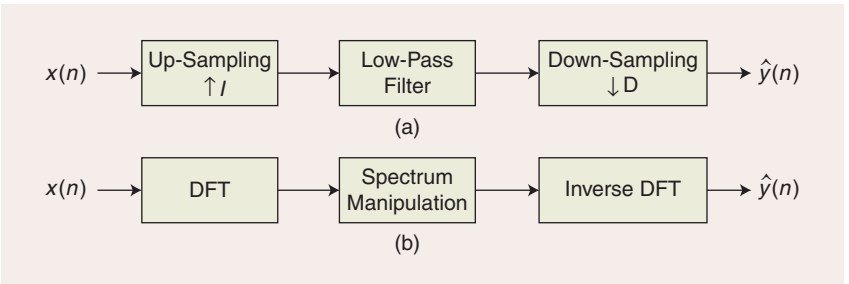


FIGURE 1. The steps of SRC: (a) the time-domain method and (b) the frequency-domain method.

Digital Object Identifier 10.1109/MSP.2017.2671412
Date of publication: 26 April 2017

accuracy can be significantly increased by the proposed calibration method based on sparsity constrained optimization.

Notations

All of the small letters represent time-domain signal, and all of the capital letters represent Fourier domain coefficients. Vectors and matrices are in bold case. The set of natural numbers is denoted by $\mathbb{N}$.

Preliminary of frequency-domain SRC

Let us denote an N_x-point input sequence as $x(n), n = 0, \ldots, N_x - 1$ that is obtained by sampling the given analog signal $x(t)$ at a sampling frequency F_x, and its DFT as $X(k), k = 0, \ldots, N_x - 1$. Similarly, we also define the ideal N_y-point output sequence as $y(m)$, $m = 0, \ldots, N_y - 1$ that can be obtained by sampling the analog signal $x(t)$ at a sampling frequency F_y, and its DFT as $Y(k), k = 0, \ldots, N_y - 1$. In general, the sequence $y(m)$ is considered as the ideal SRC output signal that is used as a reference to evaluate the accuracy of the output sequence produced by the SRC method. The factor of the sampling frequency conversion is defined as F_y/F_x, which can be easily proved to be equal to N_y/N_x and I/D, respectively, where I and D are the factors of up-sampling and down-sampling, respectively, as shown in Figure 1. Without loss of generality, it is assumed that the ratio, I/D, is irreducible. In practice, the SRC produces an output sequence, $\hat{y}(n)$, i.e., the approximation of $y(m)$ for $m = 0, \ldots, N_y - 1$. With the frequency-domain method in [4], its corresponding DFT, $\hat{Y}(k), k = 0, \ldots, N_y - 1$, is obtained by manipulating the DFT, $X(k)$, of the input sequence $x(n)$ according to [2] and [3].

- Decimation $(D > I)$. In the case of $F_y < F_x$,

$$\hat{Y}(k) = \begin{cases} \frac{I}{D}X(k), & 0 \le k < \frac{N_y}{2} \\ C_1, & k = \frac{N_y}{2} \\ \frac{I}{D}X(k + N_x - N_y), & \frac{N_y}{2} < k < N_y \end{cases} \quad (1)$$

- Interpolation $(D < I)$. In the case of $F_y > F_x$,

$$\hat{Y}(k) = \begin{cases} \frac{I}{D}X(k), & 0 \le k < \frac{N_x}{2} \\ C_2, & \frac{N_x}{2} \le k \le N_y - \frac{N_x}{2} \\ \frac{I}{D}X(k + N_x - N_y), & N_y - \frac{N_x}{2} < k < N_y. \end{cases} \quad (2)$$

In both (1) and (2), some values of $\hat{Y}(k)$ are replaced by the values of $X(k)$ scaled by I/D. For decimation, according to (1), the values of $X(k)$ for $k = N_y/2$ to $N_x - N_y/2$ are ignored in realizing $\hat{Y}(k)$. The value of $X(N_y/2)$ is set to be C_1, and $X(k)$ for $k = N_x - N_y/2 + 1$ to $N_x - 1$ are relocated to become $\hat{Y}(k)$ for $k = N_y/2 + 1$ to $N_y - 1$. For interpolation, (2) shows that a small value of C_2 is used to be the values of $\hat{Y}(k)$ for $k = N_x/2$ to $N_y - N_x/2$, and the values of scaled $X(k)$ for $k = N_x/2 + 1$ to $N_x - 1$ are used to approximate the values of $\hat{Y}(k)$ for $k = N_y - N_x/2 + 1$ to $N_y - 1$. It should be fully understood that the formulation of $\hat{Y}(k)$ is based on the spectrum distribution of input sequence. Therefore, the estimated output spectrum $\hat{Y}(k)$ always deviates from the true spectrum $Y(k)$.

In summary, the operations specified in (1) and (2) are ignoring or inserting some frequency components of small values. The other frequency components may also have some small differences from those of the expected ones. All of these variations contribute to the conversion errors that may be considered as perturbations. Now the question is, given the input signal sequence $x(n)$ and the estimated $\hat{Y}(k)$, how the conversion errors can be minimized. This article presents a positive answer to this question by introducing the SR concept into the design of frequency-domain SRC.

Calibrated SRC in frequency domain

Let us denote the DFT in (1) or (2) as the perturbed DFT and assume that the error calibrated DFT, $Y'(k)$, is obtained by subtracting perturbation $\tilde{Y}(k)$ from $\hat{Y}(k)$, i.e.,

$$Y'(k) = \hat{Y}(k) - \tilde{Y}(k), \quad (3)$$

where $\hat{Y}(k)$ is obtained from (1) or (2). In particular, the DFT perturbation $\tilde{Y}(k)$ can be obtained by our proposed calibration method to be presented and subsequently the calibrated DFT $Y'(k)$ can be obtained. The corresponding calibrated time-domain signal, $y'(n)$, can be obtained from $Y'(k)$ by an IDFT. In a vector form, it is expressed as

$$\mathbf{y}' = \mathcal{F}^{-1}\mathbf{Y}' = \mathcal{F}^{-1}(\hat{\mathbf{Y}} - \tilde{\mathbf{Y}}), \quad (4)$$

where $\mathcal{F}^{-1} \in \mathbb{C}^{N_y \times N_y}$ represents the IDFT matrix. Since the perturbed DFT, $\hat{\mathbf{Y}}$, is already obtained from (1) or (2), our task becomes properly estimating the perturbation $\tilde{\mathbf{Y}}$ to obtain improvement in accuracy.

Loss function

An intuitive strategy to reduce the conversion errors in the time domain is to minimize the loss function that is defined to measure the difference between the ideal and estimated outputs. Because the ideal output, i.e., $y(n)$ in our case, is unavailable, we use the accessible input sequence, $x(n)$, to define the loss function. In particular, the Euclidean distance is employed to measure the differences between the samples selected from the original input $x(n)$ and the error-calibrated output samples $y'(m)$. Let us denote the indices selected from original input as n_x and the indices selected from error calibrated output as n_y, where the condition $n_x/F_x = n_y/F_y$ should be satisfied. In other words, these selected samples of $y'(n_y)$ should be as close to the input samples $x(n_x)$ as possible. Using the equality $n_x I = n_y D$, the following loss functions for decimation and interpolation can be obtained.

- For decimation $(I < D)$, the loss function can be expressed as

$$\sum_{n_y = 1, n_y/I \in \mathbb{N}}^{N_y - 1} \left[x\left(\frac{n_y D}{I}\right) - y'(n_y) \right]^2. \quad (5)$$

- For interpolation $(I > D)$, the loss function can be expressed as

$$\sum_{n_x = 1, n_x/D \in \mathbb{N}}^{N_x - 1} \left[x(n_x) - y'\left(\frac{n_x I}{D}\right) \right]^2. \quad (6)$$

Substituting (4) into (5) and (6), the loss functions for decimation and interpolation can be respectively obtained. For brevity, a vector form of the loss functions can be expressed as

$$\| \mathbf{x} - \tilde{\mathcal{F}}^{-1}(\hat{\mathbf{Y}} - \tilde{\mathbf{Y}}) \|_2^2, \quad (7)$$

where $\mathbf{x}$ represents the vector form of the selected input sequence x, and $\tilde{\mathcal{F}}^{-1}$ represents the row sampled IDFT matrix. In particular, the row-sampled IDFT matrix $\tilde{\mathcal{F}}^{-1}$ is different for decimation and interpolation, where the rules for performing row sampling can be easily obtained in (5) and (6). For brevity, the full expressions are omitted here. Based on (5) and (6), the number of summation terms in (7) should be equal to

$$\min\left(\left\lfloor\frac{N_x}{D}\right\rfloor,\left\lfloor\frac{N_y}{I}\right\rfloor\right)=\min\left(\left\lfloor\frac{N_x}{D}\right\rfloor,\left\lfloor\frac{1}{I}\frac{IN_x}{D}\right\rfloor\right)=\min\left\lfloor\frac{N_x}{D}\right\rfloor, \quad (8)$$

where $\lfloor x\rfloor$ represents the nearest integer less than or equal to x. However, in (7), the number of parameters $N_y = IN_x/D$ is larger than the number of summation terms N_x/D given that the up-sampling factor I is larger than one, which is often satisfied in arbitrary fractional-factor SRC. In other words, this problem is ill-posed because the number of parameters is larger than the number of observations. The SR technique can desirably solve the ill-posed problem by proper exploitation of sparsity. In our problem, however, the SR technique cannot be directly applied since sparsity assumption of $\tilde{\mathbf{Y}}$ is not satisfied. Therefore, careful design and proper modification should be carried out to utilize the SR technique.

Parameter regularization

The key ingredient in the SR technique is the proper utilization of sparsity, where an appropriate sparsity domain should be identified. In the frequency-domain SRC, the following observations and formulations can be made.

1) As discussed in [2], manipulating $X(k)$ in the frequency-domain SRC induces errors. These errors are mainly located at both ends of the time-domain sequence $\hat{y}(n)$. More specifically, most of the errors locate only in both ends of the converted signal. An important observation is that the number of error locations is often much smaller than the length of the sequence, which can be considered to be sparse or compressible. To obtain a reasonable solution of $\tilde{\mathbf{Y}}$, the sparsity regularization is used and the following constraint can be naturally obtained:

$$\|\tilde{\mathcal{F}}^{-1}\tilde{\mathbf{Y}}\|_1 \le \varepsilon_1. \quad (9)$$

As a matter of fact, the sparsity level depends on the input length and the fraction I/D. It would be difficult to express it with a closed-form expression. Moreover, the choice of ϵ_1 not only depends on the sparsity level but also on the average amplitude of the errors. Similar to all other SR problems, the value of ϵ_1 should be properly chosen. However, in our experiment, it is found that the performance of the optimization is not very sensitive to the value of ϵ_1 for various input lengths and values of I/D. As we commented previously, this sparsity pattern is unique for frequency-domain SRC technique, which is rather different from the sparsity pattern exploited in SR-based calibration problems [5].

2) In many applications, such as audio and speech signal processing systems, the signal is often real valued and its DFT is conjugate symmetric. Similarly, we can conveniently assume that the perturbation parameter $\tilde{\mathbf{Y}}$ is also conjugate symmetric. This symmetrical relationship can be explicitly exploited as

$$\tilde{Y}(k) = [\tilde{Y}(N_y - k)]^*, \quad k = 1, \ldots, N_y/2 - 1. \quad (10)$$

The use of this property effectively decreases the number of parameters to be optimized. For complex input sequences, this conjugate symmetric constraint is not valid and should be removed from the optimization.

Of course, more constraints can be imposed on the perturbation parameter $\tilde{\mathbf{Y}}$ to limit the solution space for different application purposes. However, these two above-described constraints are effective enough to obtain desirable performance improvement compared with the uncalibrated frequency-domain SRC.

Optimization formulation

By combining the loss function and constraint terms, the optimization problem can be formulated as

$$\underset{\tilde{\mathbf{Y}}}{\text{minimize}} \text{ loss function in (7)} \quad \text{subject to constraints in (9) and in (10).} \quad (11)$$

The objective of the optimization problem (11) is to search for the perturbation parameter $\tilde{\mathbf{Y}}$ in a feasible region constructed by the constraints, which can minimize the loss function. In our method, the sparsity is properly exploited to obtain a robust estimate of the parameter. One additional advantage is that the optimization scheme is able to calibrate the phase perturbations of the manipulated DFT as well, which will be shown in detail in the section "Experimental Results." It is noted that the phase information in the DFT is extremely important in some signal processing applications, such as multimedia watermarking [6]. Therefore, the frequency-domain SRC-based methods can be applied to reduce performance degradation due to phase perturbation.

In summary, our proposed frequency-domain-based SRC method can be carried out by the following steps:

- Step 1: Compute the N_x-point DFT of input sequence.
- Step 2: Manipulate the DFT according to (1) for decimation or (2) for interpolation.
- Step 3: Estimate perturbation parameters by solving (11) and calculate the calibrated DFT by (3).
- Step 4: Compute the N_y-point IDFT of the calibrated DFT obtained in the last step.

Compared with the uncalibrated approach in [2], the proposed scheme can substantially reduce the conversion errors but requires additional computational costs due to the optimization in (11). For the optimization method, it is inevitable that the computational complexity becomes high if the input sequence is long. For real-time implementation, it is possible to segment the long input sequence into many short segments, where the optimization can be performed for each segment consecutively. Moreover, the optimization can be more conveniently and efficiently carried out by resorting to other computationally efficient methods, such as fast iterative soft-thresholding [7] and

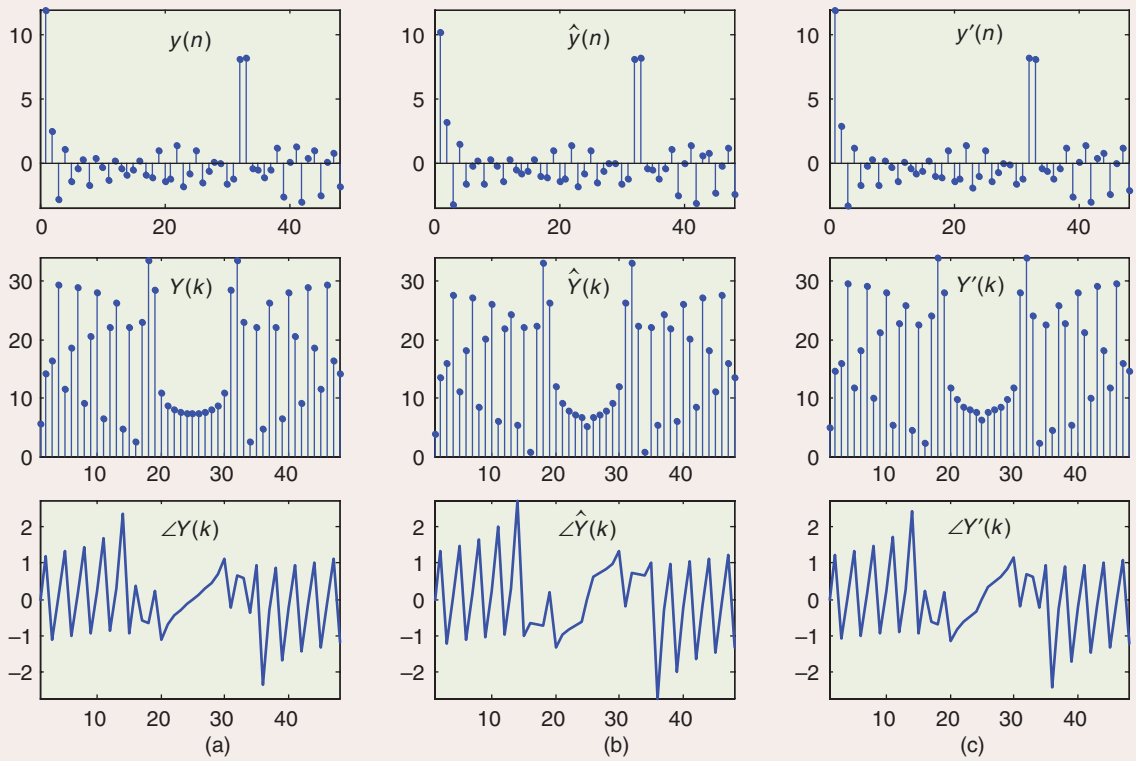


FIGURE 2. An example of the uncalibrated and calibrated SRC performance ($N_x = 64$, I/D = 3/4). (a) The time-domain ideal output sequence $y(n)$, the amplitude, and the phase of its corresponding DFT; (b) the uncalibrated time-domain output sequence $\hat{y}(n)$, the amplitude, and the phase of its corresponding DFT; and (c) the calibrated time-domain output sequence $y'(n)$, the amplitude, and the phase of its corresponding DFT.

alternating direction method of multipliers [8], which will not be discussed further for brevity.

Experimental results

In this section, we evaluate the accuracy performance of the proposed optimization scheme in terms of the MSE. The amplitude MSE is defined as

$$\mathrm{MSE_a} = 10\log\frac{\|\mathbf{y} - \mathcal{F}^{-1}\mathbf{Y}'\|_2^2}{N_y} \quad (12)$$

and the phase MSE of the spectrum is defined as

$$\mathrm{MSE}_\phi = 10\log\frac{\|\,\mathrm{angle}(\mathbf{Y}) - \mathrm{angle}(\mathbf{Y}')\|_2^2}{N_y}. \quad (13)$$

In the following experiments, CVX package [9] is used for the optimization. In our experiments, the ϵ_1 can all be set to the value within 3–4 for various cases. Figure 2 presents an example to demonstrate the outputs of the SRC before and after the calibration process in the time and frequency domains. As seen from this figure, the proposed calibration process can obtain a more accurate estimation for both the time-domain amplitude and frequency-domain phase. In particular, the calibration method can reconstruct the DFT phase of the calibrated output. In the last row of Figure 2, the phase of the calibrated DFT, $\angle Y'(k)$, is much similar to the ideal one, $\angle Y(k)$, compared to that of the uncalibrated DFT, $\angle \hat{Y}(k)$.

The quantitative performance evaluation is presented in Figures 3 and 4 for decimation and interpolation, respectively. For both decimation and interpolation, the MSE of amplitude decreases as input signal length increases. This is because the number of available samples used for calibration in (7) increases with the sequence length to enhance the calibration effects. The MSE of the phase, however, remains about the same level for different input signal lengths. Since the number of the manipulated frequency-domain parameters increases with the increase of signal length by a similar amount, the MSE of the phase is not decreased with the increase of sequence length. In general, the proposed calibration method can consistently obtain significant reduction of conversion errors in both the time-domain amplitude and frequency-domain phase, which validates its robustness of conversion accuracy performance.

The accuracy is degraded when the SRC factor decreases from 3/4 to 3/8. This can be explained by considering the number of samples that are ignored or inserted according to (1) or (2) has increased. For example, the number of points being truncated by (1) or being inserted with a value of C_2 in (2) is doubled so that the errors produced by $\hat{y}(k)$ are accordingly increased. Because the calibration process has to cope with more errors produced by the ignoring or inserting

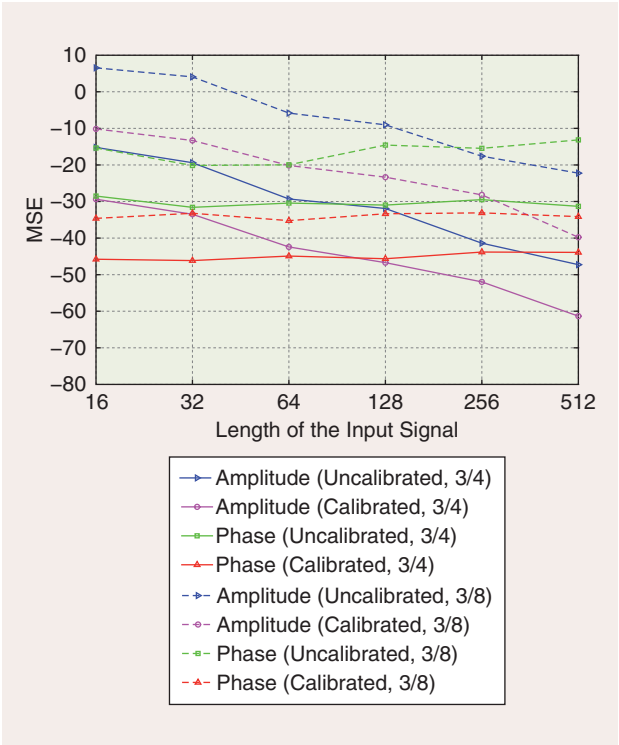


FIGURE 3. The MSE performances of decimation in terms of input signal length.

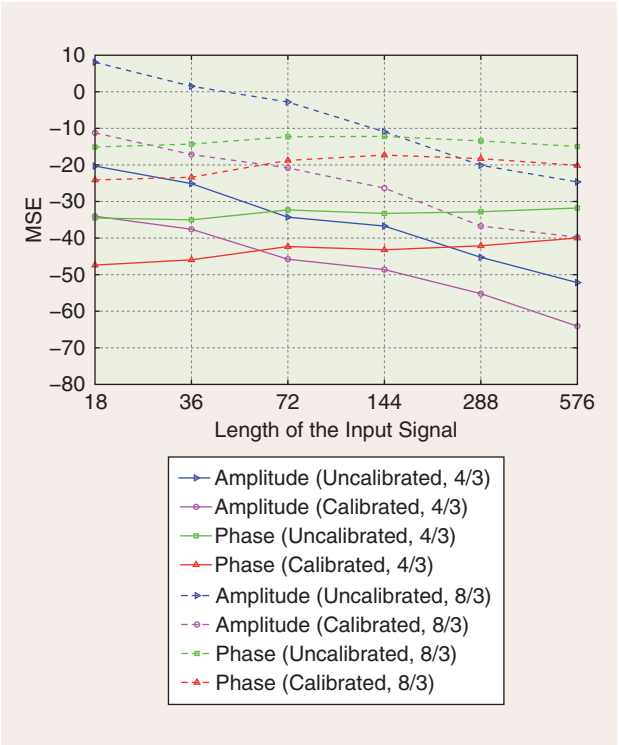


FIGURE 4. The MSE performances of interpolation in terms of input signal length.

operations, the overall effectiveness of the calibration process is decreased.

Similarly, in the interpolation case, the MSE of amplitude decreases with the increase of input signal length. In contrast, with the increase of signal length, the phase errors produced by the uncalibrated approach remain around a constant level and the one with calibration still achieves significant improvement on the estimation accuracy. The accuracy obtained by both the frequency-domain-based SRC and the calibrated one is decreased when the SRC ratio changes from 3/4 to 3/8 in Figure 3, or from 4/3 to 8/3 in Figure 4.

Let us compare the performance of the proposed SRC with the time-domain SRC methods. In our experiments, the filter used in the time-domain SRC is a poly-phase finite impulse response filter, whose coefficients are calculated by the Parks–McClellan algorithm. In the experiments, it is found that the performance of the time-domain SRC cannot be substantially improved when the order of the filter is higher than 500. Therefore, we compare the time-domain methods with the filter orders of 100, 300, and 500, respectively. Since the performances of interpolation and decimation are similar, the results of decimation ($I/D = 3/4$) are presented in Table 1 for brevity. The time-domain methods can decrease the MSE with an increase in the length of the filter. When the length of 500 is used, it can achieve comparable accuracy with the uncalibrated frequency-domain SRC. Among all of the compared methods, the calibrated frequency-domain SRC can significantly improve the conversion accuracies of amplitude and phase.

The computational complexities needed by the time-domain method is linearly proportional to the data length for a given order of a low-pass filter. For a length of N data, the uncalibrated and calibrated frequency-domain methods require the computational complexities in the orders of $N \log N$ and N^3, respectively. Figure 5 shows the required computation time by the time-domain method and the uncalibrated and calibrated frequency-domain methods. The measurements are made from the simulations by using the SRC function and fast Fourier transform function given by the MATLAB software running on a PC with 3.4-GHz central processing unit. In addition, the calibrated frequency-domain method also uses the CVX MATLAB tool box [9] for calibration. The calibrated frequency-domain

Table 1. The performance comparison of different algorithms for decimation.

Length (N_x)		Time-Domain SRC			Frequency-Domain SRC	
		100	300	500	Uncalibrated	Calibrated
64	MSEσ (dB)	−6.33	−19.24	−22.69	−23.61	**−43.84**
	MSE$_\phi$ (dB)	3.36	−26.90	−31.9	−30.44	**−45.97**
256	MSEσ (dB)	−7.01	−35.53	−36.75	−39.20	**−56.54**
	MSE$_\phi$ (dB)	−8.13	−21.02	−22.05	−29.37	**−44.51**
1,024	MSEσ (dB)	−6.74	−36.73	−43.31	−53.53	**−71.68**
	MSE$_\phi$ (dB)	−1.92	−18.17	−21.73	−28.81	**−45.42**

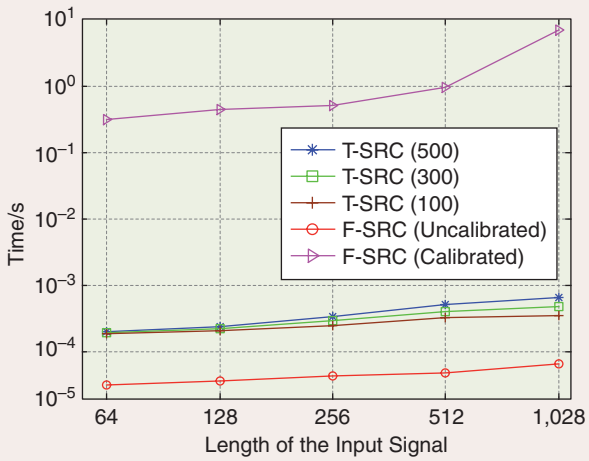


FIGURE 5. The computation time for time-domain SRC (T-SRC), uncalibrated, and calibrated frequency-domain SRC (F-SRC).

method requires significantly more computation time than the other two methods. It should be noted that utilizing other optimization techniques can speed up the optimization, which is out of the scope of this article. Remarkably, the proposed method is particularly useful for applications where higher conversion accuracies are most desirable.

Conclusions

In this article, we present a method for calibrating frequency-domain SRC to significantly minimize conversion errors. This method can be formulated as a sparsity regularized problem whose solution has been widely available in the literature. In addition, to effectively decrease the amplitude MSE of the time-domain SCR output, the proposed method also desirably reduces the phase error of the SRC output in the DFT domain.

Acknowledgment

This research work is partially supported by the National Natural Science Foundation of China (number 61571174) and the Zhejiang Provincial Natural Science Foundation of China (number LY15F010010). Xiumei Li is the corresponding author of this article.

Authors

Lifan Zhao (zhao0145@e.ntu.edu.sg) received his B.Sc. degree in electronic engineering from Xidian University, China, in 2010 and the Ph.D. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore, in 2016. He is currently a research fellow with NTU. His research interests include statistical signal processing, machine learning, compressed sensing, and Bayesian methods and their applications.

Xiumei Li (lixiumei@pmail.ntu.edu.sg) received her B.Sc. degree in electrical engineering from Lanzhou University, China, in 1999, her M.Sc. degree in information and signal processing from the Institute of Acoustics, Chinese Academy of Sciences, in 2002, and the Ph.D. degree in information engineering from Nanyang Technological University, Singapore, in 2010. She is currently an associate professor with the School of Information Science and Engineering, Hangzhou Normal University, China. Her research interests include time-frequency analysis and its applications as well as compressive sensing and its applications.

Lu Wang (wanglu@nwpu.edu.cn) received her B.Sc. degree in electronic engineering from Xidian University in 2007 and M.Eng. degree in signal processing from the National Key Lab of Radar Signal Processing, Xidian University, China, in 2010. She obtained her Ph.D. degree in 2014 from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. She is currently an assistant professor with the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China. Her major research interests include sparse signal processing and array and sonar signal processing.

Guoan Bi (egbi@ntu.edu.sg) received his B.Sc. degree in radio communications, Dalian University of Technology, China, in 1982. He received his M.Sc. degree in telecommunication systems and his Ph.D. degree in electronics systems from Essex University, United Kingdom, in 1985 and 1988, respectively. Between 1988 and 1990, he was a research fellow at the University of Surrey, United Kingdom. Since 1991, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He was the chair of the IEEE Signal Processing Society, Singapore Chapter. His research interests include digital signal processing algorithms and hardware structures and signal processing for various applications including sonar, radar, and communications. He is a Senior Member of the IEEE.

References

[1] S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*, 4th ed. New York: McGraw-Hill, 2011.

[2] G. Bi and S. K. Mitra, "Sampling rate conversion in the frequency domain [dsp tips and tricks]," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 140–144, May 2011.

[3] G. Bi, S. K. Mitra, and S. Li, "Sampling rate conversion based on DFT and DCT," *Signal Process.*, vol. 93, no. 2, pp. 476–486, 2013.

[4] D. Fraser, "Interpolation by the FFT revisited-an experimental investigation," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 5, pp. 665–675, May 1989.

[5] L. Zhao, G. Bi, L. Wang, and H. Zhang, "An improved auto-calibration algorithm based on sparse bayesian learning framework," *IEEE Signal Process. Lett.*, vol. 20, no. 9, pp. 889–892, Sept. 2013.

[6] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc. IEEE*, vol. 87, no. 7, pp. 1079–1107, 1999.

[7] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[9] M. Grant and S. Boyd. (2011). CVX: MATLAB software for disciplined convex programming, version 1.21. [Online]. Available: http://cvxr.com/cvx

SP

ADVERTISING & SALES

© GRAPHIC STOCK

The Advertisers Index contained in this issue is compiled as a service to our readers and advertisers: the publisher is not liable for errors or omissions although every effort is made to ensure its accuracy. Be sure to let our advertisers know you found them through *IEEE Signal Processing Magazine*.

IEEE SIGNAL PROCESSING MAGAZINE REPRESENTATIVE

Mark David, Director, Business Development — Media & Advertising, Phone: +1 732 465 6473, Fax: +1 732 981 1855, m.david@ieee.org

COMPANY	PAGE NUMBER	WEBSITE	PHONE
IEEE ICME	CVR 2	http://www.icme2017.org	
IEEE USA	7	www.ieeeusa.org/policy/govfel	+1 202 530 8347
Mathworks	CVR 4	www.mathworks.com/wireless	+1 508 647 7040
Southern University of Science and Technology	5	sustc.edu.cn	+86 755 88018558

Digital Object Identifier 10.1109/MSP.2017.2680518

Nominations open: 2018 SPS Distinguished Lecturer and Distinguished Industry Speakers NEW!

Each year, a select group of technically diverse and geographically dispersed individuals serve as IEEE Signal Processing Society (SPS) Distinguished Lecturers and Speakers. Generally, up to five new individuals are appointed as Distinguished Lecturers each year for a two-year term. The program provides means for chapters to have access to individuals who are recognized experts in the fields of signal processing, to speak at Chapter meetings. While many IEEE Societies have similar programs, the SPS provides financial support for the SPS Chapters to take advantage of the service. SPS invites nominations for the 2018 Class of SPS Distinguished Lecturers (DLs) and the new Distinguished Industry Speakers (DIS). The deadline for nominations is **31 May 2017**. Nomination guidelines are at the following link, and please click at the respective program names: http://signalprocessingsociety.org/get-involved/awards-submit-award-nomination

NEW in 2017 The SPS Board of Governors has approved that in addition to appointing the normal five Distinguished Lecturers, this year the Society will also appoint a **special Distinguished Lecturer in the area of Signal Processing and Machine Learning or Data Science** for a two-year term. The special Distinguished Lecturer position will be available for 2018-2019, explicitly for the areas of Signal Processing and Machine Learning or Data Science. The candidates for the special Distinguished Lecturer should be IEEE Signal Processing Society members working in the Signal Processing and Machine Learning or Data Science field, who can present engaging lecture topics in that field.

Just Launched! The IEEE Signal Processing Society is launching a new **SPS Distinguished Industry Speakers (DIS) Program**, which closely mirrors and complements the DL program. The main difference and goal of the Distinguished Industry Speakers Program are to educate and interact with Society members about topics that are of primary importance to industry and the signal processing community-at-large. The speakers will focus their presentations on industrial applications of interest to practicing engineers and stress practical issues. The Distinguished Industry Speakers will be available to all SPS Chapters, as are DLs, but lectures will be focused on current and future issues, applications and/or activities in industry.

IN THE SPOTLIGHT *(continued from page 112)*

The IEEE Signal Processing Society launched an online blog to provide an accessible briefing to a more general audience and complement the usually highly technical signal processing topics. Blog entries include the following:

- "Sensors and Sensibility" by Vladimir Pavlovic (30 January 2017)
- "How Biometric Authentication Poses New Challenges to Our Security and Privacy" by Nasir Memon (5 January 2017)
- "Audio and Acoustic Signal Processing's Major Impact on Smartphones" by Heinz Teutsch (28 October 2016)
- "E.T. Still Can't Phone Home, but Signal Processing Solves Many Technology Challenges" by Pete Wyckoff (27 July 2016)
- "Modern Communications: Signal Processing's Vital Role in Connecting Communities" by Waheed U. Bajwa and Hamid Krim (12 May 2016)

… and more!

Visit https://signalprocessingsociety.org/publications-resources/blog to read blog articles on a variety of topics.

approval decisions. One operating point of the decision rule we developed could have reduced delayed payments by almost one-third while still accepting around 70% of the total customer pool.

This project is a prime example of the "data for social good" movement: uplifting humanity by harnessing skills-based volunteering of data scientists (of which I would argue signal processing engineers and scientists are a subset) in combination with the subject matter expertise of nongovernmental organizations (NGOs), social enterprises, and other similar organizations. At IBM Research, I codirect (with Saška Mojsilović) a program that pairs student fellows with a team of researchers and an NGO partner to conduct social good projects. Through this program—the first of its kind in a corporate setting—we have evaluated the effectiveness of programs fighting childhood diarrhea, found causal factors on the innovativeness of countries, and created a recommender system on the attributes of large philanthropic projects. Additionally, we've predicted species of primates likely to be reservoirs of the Zika virus, built tools to automatically retrieve and classify news articles on humanitarian crises, and developed cognitive technologies to accelerate open scientific discovery such as cures for multiple sclerosis—all using signal processing and related techniques.

The University of Chicago pioneered the concept of the Data Science for Social Good Summer Fellowship, which continues to spread in the academic setting with similar programs now offered at Georgia Tech and the University of Washington. Corporations and research organizations, like Two Sigma and Draper Laboratory, are following suit. Bloomberg hosts an annual conference on the topic. DataKind is a nonprofit that connects NGOs with volunteer teams to conduct projects in their spare time—this is how I worked on the Simpa project and another project using satellite image analysis of Kenyan households' roofs to estimate poverty. DrivenData conducts online competitions to solve social good problems.

Fundamental signal processing theory and methods, when applied to nontraditional application areas, can change the world for the better. We have a unique opportunity to talk with NGOs, understand their most pressing needs, and use signal and information processing tools to create appropriate solutions. This is not a pipe dream but a reality.

Acknowledgment

This article originally appeared in the online blog of the IEEE Signal Processing Society on 7 February 2017.

Author

Kush R. Varshney (krvarshn@us.ibm.com) received his B.S. degree (magna cum laude) in electrical and computer engineering with honors from Cornell University in 2004. He received his S.M. degree in 2006 and his Ph.D. degree in 2010, both in electrical engineering and computer science from the Massachusetts Institute of Technology. He is a research staff member and manager in the Data Science Department at the IBM Thomas J. Watson Research Center. He also codirects the IBM Social Good Fellowship Program.

SP

Do you know? The IEEE Jack S. Kilby Signal Processing Medal was established in 1995 in honor of Jack S. Kilby, whose innovation was a monumental precursor to the development of the signal processor and digital signal processing. The award may be presented for outstanding achievements in signal processing. The achievement may be theoretical, technological, or commercial.

Learn more at http://www.ieee.org/about/awards/medals/kilby.html

Nomination deadline: **15 June**

New benefit from the IEEE Signal Processing Society

SPS Resource Center

The SPS Resource Center is the new home for the IEEE Signal Processing Society's online library of tutorials, lectures, presentations, and more. Unrestricted access to our fast-growing archive is now included with your SPS membership.

http://rc.signalprocessingsociety.org

We accept submissions, too!
Interested in submitting your educational materials?

sps-resourcecenter@ieee.org

Digital Object Identifier 10.1109/MSP.2017.2681141

DATES AHEAD

Please send calendar submissions to:
Dates Ahead, Att: Jessica Barragué, E-mail: j.barrague@ieee.org

2017

MAY

IEEE Radar Conference (RADARCONF)
8–12 May, Seattle, Washington, USA.
General Chair: Daniel J. Sego
URL: http://www.radarconf17.org

JULY

18th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)
3–6 July, Hokkaido, Japan.
General Chairs: Yasutaka Ogawa, Wei Yu, and Fumiyuki Adachi
URL: http://www.spawc2017.org/

IEEE International Conference on Multimedia and Expo (ICME)
10–14 July, Hong Kong, China.
General Chairs: Jörn Ostermann and Kenneth K.M. Lam
URL: http://www.icme2017.org/

AUGUST

25th European Signal Processing Conference (EUSIPCO)
28 August–2 September, Kos Island, Greece.
General Chairs: Petros Maragos and Sergios Theodoridis
URL: www.eusipco2017.org

14th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)
29 August–1 September, Lecce, Italy.
General Chairs: Cosimo Distante and Larry S. Davis
URL: www.avss2017.org

Digital Object Identifier 10.1109/MSP.2017.2678358
Date of publication: 26 April 2017

©GRAPHIC STOCK

The main theme of ICME 2017 is "The New Media Experience." Approximately 400 participants, mainly from Asia, Europe, and North America, will gather in Hong Kong 10–14 July to discuss the latest development in multimedia technologies and related fields.

SEPTEMBER

IEEE International Conference on Image Processing (ICIP)
17–20 September, Beijing, China.
General Chairs: Xinggang Lin, Anthony Vetro, and Min Wu
URL: http://2017.ieeeicip.org/

OCTOBER

IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)
15–18 October, New Paltz, New York, USA.
General Chairs: Patrick A. Naylor and Meinard Müller
URL: http://www.waspaa.com/

19th IEEE International Workshop on Multimedia Signal Processing (MMSP)
16–18 October, London-Luton, United Kingdom.
General Chairs: Vladan Velisavljevic, Vladimir Stankovic, and Zixiang Xiong
URL: http://mmsp2017.eee.strath.ac.uk/

NOVEMBER

Fifth IEEE Global Conference on Signal and Information Processing (GlobalSIP)
14–16 November 2017, Montreal, Canada.
General Cochairs: Warren Gross and Kostas Plataniotis
URL: http://2017.ieeeglobalsip.org

DECEMBER

Seventh IEEE Conference of the Sensor Signal Processing for Defence (SSPD)
6–7 December, Edinburgh, United Kingdom.
General Chairs: Mike Davies, Jonathon Chambers, and Paul Thomas
URL: www.sspd.eng.ed.ac.uk/

17th IEEE International Workshop on Computational Advances in Multisensor Adaptive Processing (CAMSAP)
10–13 December, Curacao, Dutch Antilles.
General Chairs: André L.F. de Almeida and Martin Haardt
URL: http://www.cs.huji.ac.il/conferences/CAMSAP17/

2018

APRIL

43rd International Conference on Acoustic, Speech, and Signal Processing (ICASSP)
22–27 April, Seoul, South Korea.
General Chair: Monson Hayes
General Cochair: Hanseok Ko
URL: http://2018.ieeeicassp.org/

OCTOBER

25th IEEE International Conference on Image Processing (ICIP)
7–10 October, Athens, Greece.

SP

IEEE GlobalSIP

IEEE Advancing Technology for Humanity

IEEE Signal Processing Society

5th IEEE Global Conference on Signal and Information Processing

November 14 - 16, 2017, Montreal, Canada

www.ieeeglobalsip.org/

Call for Papers

The 5th IEEE Global Conference on Signal and Information Processing (GlobalSIP) will be held in Montreal, Quebec, Canada on November 14-16, 2017. GlobalSIP, as a new flagship IEEE Signal Processing Society conference, focuses on signal and information processing with an emphasis on up-and-coming signal processing themes. The conference features world-class plenary speeches, distinguished Symposium talks, tutorials, exhibits, oral and poster sessions, and panels. GlobalSIP is comprised of co-located General Symposium and symposia selected based on responses to the call-for-symposia proposals. Featured symposia include:

- General symposium
- Sparse SP and deep learning
- Graph signal processing
- Distributed optimization and resource management over networks
- Control & information theoretic approaches to privacy and security
- SP for interference cancellation and full-duplex communication systems
- SP for Accelerating Deep Learning
- SP for Smart Cities & IoTs
- SP & ML in large medical datasets
- Big Data Analytics for IoT Healthcare
- Advanced Bio-SP for Rehabilitation and Assistive Systems
- Deep Learning for Intelligent Multimedia
- Knowledge-based Multimedia Computing
- Stochastic & approximate computing for signal processing and machine learning
- Signal and information processing (SIP) for smart grid infrastructure
- SIP for Healthcare Engineering
- SIP for Finance and Business

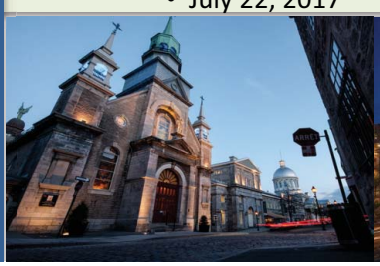
Prospective authors are invited to submit full-length papers (up to 4 pages for technical content, an optional 5th page containing only references) and extended abstracts (up to 2 pages, for paper-less industry presentations and Ongoing Work presentations). Manuscripts should be original and written in accordance with the standard IEEE 2-column paper template. Accepted full-length papers will be indexed on IEEE Xplore. Accepted abstracts will not be indexed in IEEE Xplore, however the abstracts and/or the presentations will be included in the IEEE SPS SigPort. Accepted papers and abstracts will be scheduled in lecture and poster sessions.

Conference Highlights

- **Plenary Talks within the general symposium and Distinguished Symposium Presentations/Talks within the 16 thematic symposia surveying emerging topics in SIP**
- **Panel discussions on funding opportunities, trends and targeted topics**
- **Enhanced industry program: paperless industry presentations, panels, demos and exhibitions**
- **Exciting student program: Ongoing Work tracks, Student-Industry Luncheon, 3MT competition**
- **Great venue with vibrant cultural, educational, and scientific identity, combining the modern buzz of a North American city and a specific European charm**

Important Dates:

- **May 15, 2017** : Paper submission due
- June 30, 2017 : Final acceptance notifications sent to all authors
- July 22, 2017 : Camera-ready papers due



Digital Object Identifier 10.1109/MSP.2017.2681140

Kush R. Varshney

Signal Processing for Social Good

Communication, speech processing, seismology, and radar are well-known applications of signal processing that contribute to the betterment of humanity. But is there a more direct way that signal and information processing can reduce poverty, hunger, inequality, injustice, ill health, and other causes of human suffering? The member states of the United Nations ratified 17 sustainable development goals in 2015, which, if achieved by the targeted year 2030, will end or greatly curtail these problems. Achieving the global goals, however, will require cooperation from all, including the signal processing community. Let me tell you how.

In December 2016, I visited Barauli, a small village in District Aligarh, India (Figure 1), where I had tea at the well-lit home of a Simpa Networks customer. A social enterprise, Simpa provides pay-as-you-go solar panel systems to households with inadequate access to power from the grid. The company installs a solar panel with lights and a fan—an alternative to unhealthy and unsafe kerosene lamps—for a very small down payment. Through a simple financing plan, customers can fully repay the cost of the system in monthly installments over two or three years. As a result, clean, reliable energy (and the well-being and economic opportunity that comes with it) is now within reach of people at the bottom of the pyramid—like the customer I visited, a driver and smallholder farmer.

Digital Object Identifier 10.1109/MSP.2017.2673878
Date of publication: 26 April 2017

FIGURE 1. Kush R. Varshney visits Barauli, a small village in District Aligarh, India.

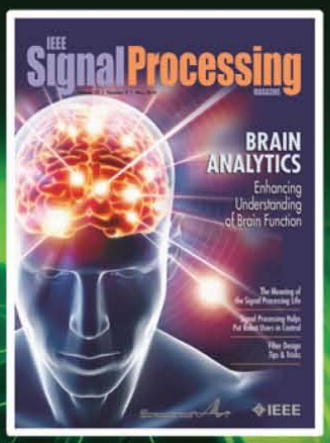
Signal processing is the science behind Simpa's sustainability. Two years ago, I led a team on a pro bono project to develop a predictive model of customer repayment behaviors based on signals captured in an application form. Specifically, we wanted to assess the risk of an applicant's system being eventually repossessed due to payment delinquency, thus allowing Simpa to make better

(continued on page 108)

January 2016
Feature Article Collection

March 2016
Assistive Living Technologies

May 2016
Brain Signal Analytics

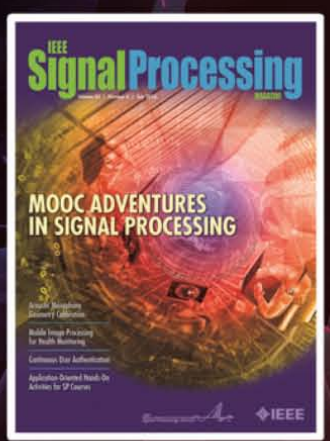


Publish with IEEE Signal Processing Magazine

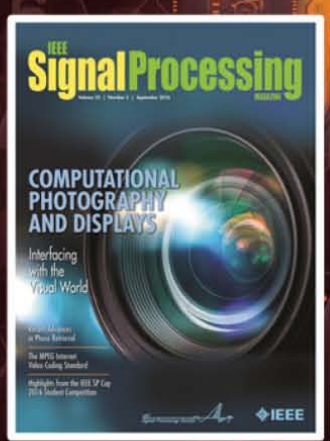
HIGH IMPACT among all electrical engineering publications

REACH a broad signal processing audience worldwide

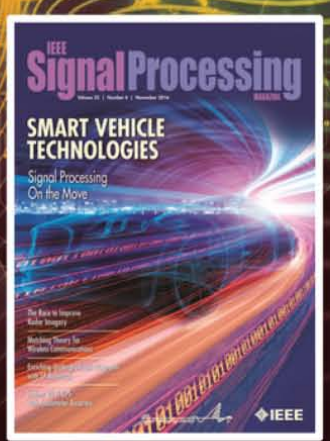
WELCOME proposals for Special Issues and Feature Articles, and contributions to Columns



July 2016
Feature Article Collection



September 2016
Computational Photography & Display



December 2016
SP for Smart Vehicle

Digital Object Identifier 10.1109/MSP.2017.2681138

MATLAB SPEAKS WIRELESS DESIGN

You can simulate, prototype, and verify wireless systems right in MATLAB. Learn how today's MATLAB supports RF, LTE, WLAN and 5G development and SDR hardware.

mathworks.com/wireless

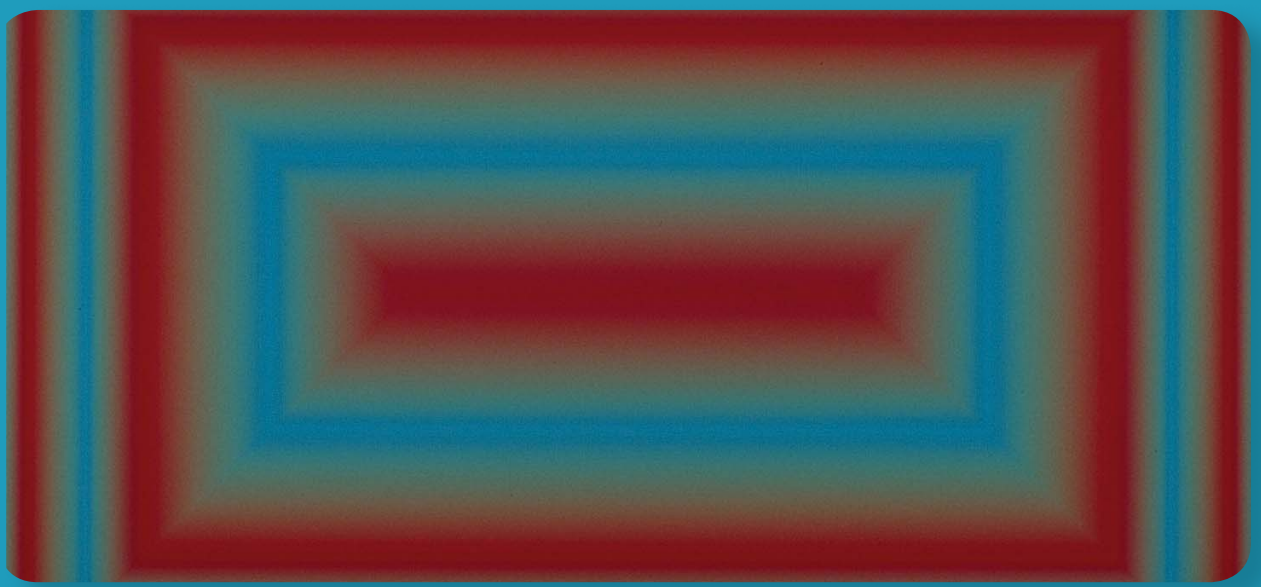
MathWorks®

IEEE SIGNAL PROCESSING SOCIETY

ContentGazette

MAY 2017

ISSN 2167-5023



T-SP May 1 Vol. 65 #9
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7870794

T-SP May 15 Vol. 65 #10
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7885148

T-ASLP April Vol. 25 #4
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7886449

T-IP April Vol. 26 #4
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7882754

T-IFS April Vol.12 #4
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7879996

T-IFS May Vol. 12 #5
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7885623

T-MM April Vol. 19 #4
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7879498

T-SPL April Vol. 24 #4
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7885179

T-CI March Vol. 3 #1
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7847494

T-SIPN March Vol. 3 #1
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7847498

http://www.ieee.org/publications_standards/publications/authors/author_ethics.html

IEEE Signal Processing Society

IEEE

IEEE ASRU 2017

Okinawa, Japan, December 16-20, 2017

General Chairs:

John R. Hershey, MERL
Tomohiro Nakatani, NTT

Important Dates:

Paper Submission:
June 29, 2017

Paper Notification:
August 31, 2017

Early Registration Period:
August 31 - Oct 5, 2017

Camera Ready Deadline:
Sept 21, 2017

More Information:

http://asru2017.org
info@asru2017.org

IEEE Automatic Speech Recognition and Understanding Workshop

The biennial IEEE ASRU workshop has a tradition of bringing together researchers from academia and industry in an intimate and collegial setting to discuss problems of common interest in automatic speech recognition, understanding, and related fields of research. The workshop includes keynotes, invited talks, poster sessions and will also feature challenge tasks, panel discussions, and demo sessions.

We invite papers in all areas of spoken language processing, with emphasis placed on the following topics:

- Automatic speech recognition
- ASR in adverse environments
- New applications of ASR
- Speech-to-speech translation
- Spoken document retrieval
- Multilingual language processing
- Spoken language understanding
- Spoken dialog systems
- Text-to-speech systems

IEEE

IEEE Signal Processing Society

IEEE TRANSACTIONS ON
SIGNAL PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed MEDLINE U.S. National Library of Medicine

MAY 15, 2017 VOLUME 65 NUMBER 10 ITPRED (ISSN 1053-587X)

REGULAR PAPERS

IEEE

MMSP 2017

2017 IEEE Workshop on Multimedia Signal Processing

October 16-18, 2017, London-Luton, UK

http://mmsp2017.eee.strath.ac.uk

CALL FOR PAPERS

MMSP 2017 is the IEEE 19th International Workshop on Multimedia Signal Processing. The workshop is organized by the Multimedia Signal Processing Technical Committee of the IEEE Signal Processing Society. This year's event has a theme of '**Multimedia Processing for Healthcare and Assisted Living**'.

Recent advances in multimedia processing and communications have potential to significantly advance current healthcare and assisted living services by enabling remote health monitoring, remote diagnostics, increased patient privacy, robotic-assisted surgery, and home-based treatment. A huge diversity of multimedia processing techniques, ranging from image/audio sensing, compression and networking, denoising, feature extraction, security, distributed processing, depth image processing, cloud and social computing, visualization and multimedia big data analytics, can all find their applications in future healthcare and elderly-care. However, to fully realize this potential and embed multimedia solutions into day-to-day healthcare practice, many challenges need to be overcome that call for significant engineering innovation, which can only happen through close interdisciplinary effort. The MMSP-2017 Workshop will bring together experts from different fields, including signal processing, computer science, communications, medicine, rehabilitation, psychology, to exchange ideas on how multimedia research can support advancements of future healthcare and how best to facilitate interactions between multimedia researchers and healthcare and assisted living providers.

Papers are solicited in (but not limited to) the following areas, covering not only this year's workshop theme, but also the general scope of multimedia signal processing:

- Multimedia big data analytics
- Distributed multimedia for body networks
- Deep learning for health-specific event detection and classification
- Streaming, security and privacy for healthcare
- Sparsity-based and low-rank based sensing of human vital signs
- Multimedia for smart homes and elderly care
- Multimedia processing for tele-rehabilitation
- Computational imaging for healthcare applications
- Healthcare monitoring applications using wearable technologies
- Image/video/speech/audio coding and processing
- Multimedia networking
- Multimedia traffic, communications and heterogeneous interactions
- Multimedia quality assessment
- Internet of Things (IoT)-based multimedia systems and applications
- Multimedia hardware design
- Augmented, mixed and virtual reality

Important dates:

Proposals for Special Sessions and Tutorials: **April 1, 2017**
Notification of Acceptance for Special Session and Tutorial Proposals: **April 15, 2017**
Submission of Regular and Special Session Papers: **June 1, 2017**
Notification of Acceptance for Regular and Special Session Papers: **July 15, 2017**
Submission of Sketch and Demo Papers: **August 1, 2017**
Camera Ready Deadline: **August 1, 2017**

General Co-chairs:
Vladan VELISAVLJEVIC, *University of Bedfordshire, UK*
Vladimir STANKOVIC, *University of Strathclyde, UK*
Zixiang XIONG, *Texas A&M University, TX, USA*

TPC Co-chairs:
Homer H. CHEN, *National Taiwan University, Taiwan*
Frederic DUFAUX, *CNRS, France*
Raouf HAMZAOUI, *De Montfort University, UK*

Finance Co-chairs:
Ching-Te CHIU, *National Tsing Hua University, Taiwan*
Lina STANKOVIC, *University of Strathclyde, UK*

Special Session & Panel Co-chairs:
Gene CHEUNG, *National Institute for Informatics, Japan*
Tasos DAGIUKLAS, London South Bank University, UK

Keynote Co-chairs:
Nikolaos BOULGOURIS, *Brunel University London, UK*
Yiannis ANDREOPOULOS, *University College London, UK*

Industry & Demo Chair:
Marta MRAK, *BBC R&D, UK*

Publication Chair:
Maria MARTINI, *Kingston University London, UK*
Safak DOGAN, *Loughborough University London, UK*

Publicity Chair:
Ivana TOSIC, *Ricoh, CA, USA*

Asia & Australia Liaison:
Tao MEI, *Microsoft Research Beijing, China*

North & South America Liaison:
Jacob CHAKARESKI, *University of Alabama, AL, USA*

Local Arrangements Chair:
Nicholas GARDNER, *University of Bedfordshire, UK*

University of Bedfordshire | IEEE | IEEE Signal Processing Society | University of Strathclyde Glasgow

IEEE/ACM TRANSACTIONS ON

AUDIO, SPEECH, AND LANGUAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society | acm Association for Computing Machinery | www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed | MEDLINE U.S. National Library of Medicine

APRIL 2017 | VOLUME 25 | NUMBER 4 | ITASFA | (ISSN 2329-9290)

REGULAR PAPERS

IEEE

IEEE TRANSACTIONS ON

IMAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed MEDLINE U.S. National Library of Medicine

APRIL 2017 VOLUME 26 NUMBER 4 IIPRE4 (ISSN 1057-7149)

PAPERS

IEEE

COMMENTS AND CORRECTIONS

MLSP2017

IEEE International Workshop on

MACHINE LEARNING FOR SIGNAL PROCESSING

SEPTEMBER 25-28, 2017 ROPPONGI, TOKYO, JAPAN

MLSP2017.CONWIZ.DK

Organizing committee

General Chairs	**Tomoko Matsui** The Institute of Statistical Mathematics, Japan **Jen-Tzung Chien** National Chiao Tung University, Taiwan
Program Chairs	**Naonori Ueda** NTT Communication Science Laboratories, Japan **Shinji Watanabe** Mitsubishi Electric Research Laboratories, Japan
Finance Chair	**Motoi Okamoto** The Institute of Statistical Mathematics, Japan
Publicity Chair	**Marc van Hulle** *KU Leuven, Belgium*
Publication Chair	**Jan Larsen** *Technical University of Denmark, Denmark*
Program Advisory Committee	**Shigeru Katagiri** Doshisha University, Japan **Tülay Adali** University of Maryland, Baltimore County, USA **Vince Calhoun** University of New Mexico, USA **Raviv Raich** Oregon State University, USA

Important Dates and Deadlines

Special session proposals	April 3, 2017
Paper submissions	May 28, 2017
Decision notifications	July 5, 2017
Camera-ready papers and advance author registration	August 7, 2017

Sponsored by

IEEE Signal Processing Society

Supported by

Research Organization of Information and Systems
The Institute of Statistical Mathematics

Copyright by Moyan Brenn, CC-BY-2.0

CALL FOR PAPERS

The 27th MLSP workshop in the series of workshops organized by the IEEE Signal Processing Society MLSP Technical Committee will present the most recent and exciting advances in machine learning for signal processing through keynote talks, tutorials, as well as special and regular single-track sessions. Prospective authors are invited to submit papers on relevant algorithms and applications including, but not limited to:

- Bayesian learning and signal processing
- Cognitive information processing
- Deep learning techniques
- Dictionary learning
- Graphical and kernel methods
- Independent component analysis
- Information-theoretic learning
- Learning theory and algorithms
- Pattern recognition and classification
- Bounds on performance
- Subspace and manifold learning
- Sequential learning and decision methods
- Source separation
- Tensor and structured matrix methods
- Machine learning from big data
- Scalable learning algorithms
- Applications including: speech, audio & music, image & video, biomedical signals & images, communications, bioinformatics, biometrics, systems biology, computational intelligence, genomic signals & sequences, social networks, games, smart grid, security & privacy

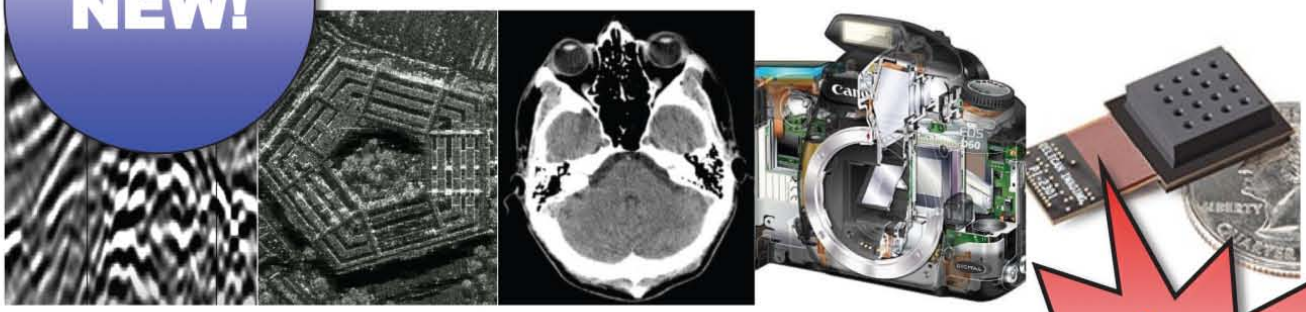
MLSP 2017 seeks proposals for **Special Sessions** that will address research in emerging or interdisciplinary areas of particular interest, not covered already by traditional MLSP sessions.

The **MLSP Best Student Paper Award** will be granted to the best paper for which a student is the principal author and presenter.

Prospective authors are invited to submit a double column paper of up to six pages using the electronic submission procedure at http://mlsp2017.conwiz.dk. Accepted papers will be published on a password-protected website that will be available during the workshop. The presented papers will be published in and indexed by IEEE Xplore.

NEW!

IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING



30% FASTER TURN AROUND THAN TIP!

Editor-in-Chief
W. Clem Karl
Boston University

Technical Committee
Charles Bouman
Eric Miller
Peter Corcoran
Jong Chul Ye
Dave Brady
William Freeman

The IEEE Transactions on Computational Imaging publishes research results where computation plays an integral role in the image formation process. All areas of computational imaging are appropriate, ranging from the principles and theory of computational imaging, to modeling paradigms for computational imaging, to image formation methods, to the latest innovative computational imaging system designs. Topics of interest include, but are not limited to the following:

Computational Imaging Methods and Models

- Coded image sensing
- Compressed sensing
- Sparse and low-rank models
- Learning-based models, dictionary methods
- Graphical image models
- Perceptual models

Computational Image Formation

- Sparsity-based reconstruction
- Statistically-based inversion methods
- Multi-image and sensor fusion
- Optimization-based methods; proximal iterative methods, ADMM

Computational Photography

- Non-classical image capture
- Generalized illumination
- Time-of-flight imaging
- High dynamic range imaging
- Plenoptic imaging

Computational Consumer Imaging

- Mobile imaging, cell phone imaging
- Camera-array systems
- Depth cameras, multi-focus imaging
- Pervasive imaging, camera networks

Computational Acoustic Imaging

- Multi-static ultrasound imaging
- Photo-acoustic imaging
- Acoustic tomography

Computational Microscopy

- Holographic microscopy
- Quantitative phase imaging
- Multi-illumination microscopy
- Lensless microscopy
- Light field microscopy

Imaging Hardware and Software

- Embedded computing systems
- Big data computational imaging
- Integrated hardware/digital design

Tomographic Imaging

- X-ray CT
- PET
- SPECT

Magnetic Resonance Imaging

- Diffusion tensor imaging
- Fast acquisition

Radar Imaging

- Synthetic aperture imaging
- Inverse synthetic aperture imaging

Geophysical Imaging

- Multi-spectral imaging
- Ground penetrating radar
- Seismic tomography

Multi-spectral Imaging

- Multi-spectral imaging
- Hyper-spectral imaging
- Spectroscopic imaging

For more information on the **IEEE Transactions on Computational Imaging** see *http://www.signalprocessingsociety.org/publications/periodicals/tci/*

NOW WITH NO PAGE CHARGES!

IEEE Signal Processing Society | EMB | GRSS

General Chairs

André L. F. de Almeida
andre@gtel.ufc.br
Federal University of Ceará, Brazil

Martin Haardt
martin.haardt@tu-ilmenau.de
Ilmenau University of Technology, Germany

Technical Program Chairs

Xiaoli Ma
xiaoli@gatech.edu
Georgia Institute of Technology, USA

Shahram ShahbazPanahi
shahram.shahbazpanahi@uoit.ca
University of Ontario Institute of Technology, Canada

Finance Chair

Cihan Tepedelenlioglu
cihan@asu.edu
Arizona State University, USA

Publicity and Publications Chair

Ami Wiesel
ami.wiesel@huji.ac.il
Hebrew University of Jerusalem, Israel

Local Arrangements Chair

Geert Leus
g.j.t.leus@tudelft.nl
Delft University of Technology, The Netherlands

Robert W. Heath Jr.
rheath@utexas.edu
University of Texas at Austin, USA

Join Twitter Conversation
#CAMSAP2017

IEEE

CAMSAP 2017

Call for Papers

The Seventh IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing

Curaçao, Dutch Antilles
December 10-13, 2017
http://www.cs.huji.ac.il/conferences/CAMSAP17



Following the success of the first six editions of the IEEE workshop on Computational Advances in Multi-Sensor Adaptive Processing, we are pleased to announce the seventh workshop in this series. IEEE CAMSAP 2017 will be held in Curaçao, Dutch Antilles, and will feature a number of plenary talks from the world's leading researchers in the area, special focus sessions, and contributed papers. All papers will undergo peer review in order to provide feedback to the authors and ensure a high-quality program.

Topics and applications of interest for the workshop include, but are not limited to, the following.

TOPICS OF INTEREST

- Array processing, waveform diversity, space-time adaptive processing
- Convex optimization and relaxation
- Computational linear & multi-linear algebra
- Computer-intensive methods in signal processing (bootstrap, MCMC, EM, particle filtering, etc.)
- Signal and information processing over networks
- Sparse signal processing

APPLICATIONS

- Big data
- Biomedical signal processing
- Communication systems
- Computational imaging
- Radar
- Sensor networks
- Smart grids
- Sonar

Submission of Papers: Prospective authors are invited to submit original full-length papers, with up to four pages for technical content including figures and references, using the formatting guidelines on the website for reviewing purposes. All accepted papers must be presented at the workshop to appear in the proceedings. Best student paper awards, selected by a CAMSAP committee, will also be presented at the workshop.

Special Session Proposals: In addition to contributed sessions, the workshop will also have a number of special sessions. Prospective organizers of special sessions are invited to submit a proposal form, available on the workshop website, by e-mail to the Special Sessions Chair.

IMPORTANT DEADLINES

Submission of proposals for special sessions	March, 2017
Notification of special session acceptance	March 15, 2015
Submission of papers	July, 2017
Notification of paper acceptance	September, 2017
Final paper submission	

IEEE Signal Processing Society

IEEE TRANSACTIONS ON
INFORMATION FORENSICS AND SECURITY

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

APRIL 2017 VOLUME 12 NUMBER 4 ITIFA6 (ISSN 1556-6013)

REGULAR PAPERS

IEEE

IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

MAY 2017 | VOLUME 12 | NUMBER 5 | ITIFA6 | (ISSN 1556-6013)

REGULAR PAPERS

IEEE

IEEE TRANSACTIONS ON MULTIMEDIA

A PUBLICATION OF
THE IEEE CIRCUITS AND SYSTEMS SOCIETY
THE IEEE SIGNAL PROCESSING SOCIETY
THE IEEE COMMUNICATIONS SOCIETY
THE IEEE COMPUTER SOCIETY

http://www.signalprocessingsociety.org/tmm/

APRIL 2017 VOLUME 19 NUMBER 4 ITMUF8 (ISSN 1520-9210)

PAPERS

IEEE JOURNAL OF
SELECTED TOPICS IN SIGNAL PROCESSING

IEEE Signal Processing Society

www.ieee.org/sp/index.html

MARCH 2017 VOLUME 11 NUMBER 2 IJSTGY (ISSN 1932-4553)

ISSUE ON TIME/FREQUENCY MODULATED ARRAY SIGNAL PROCESSING

EDITORIAL

PAPERS

IEEE

IEEE

SIGNAL PROCESSING LETTERS

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.ieee.org/sp/index.html

APRIL 2017 VOLUME 24 NUMBER 4 ISPLEM (ISSN 1070-9908)

LETTERS

IEEE TRANSACTIONS ON

COMPUTATIONAL IMAGING

A PUBLICATION OF
IEEE SIGNAL PROCESSING SOCIETY
IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY
IEEE CONSUMER ELECTRONICS SOCIETY

TECHNICALLY CO-SPONSORED BY
IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY

MARCH 2017 VOLUME 3 NUMBER 1 ITCIAJ (ISSN 2333-9403)

PAPERS

IEEE TRANSACTIONS ON

SIGNAL AND INFORMATION PROCESSING OVER NETWORKS

A PUBLICATION OF
THE IEEE SIGNAL PROCESSING SOCIETY
THE IEEE COMMUNICATIONS SOCIETY
THE IEEE COMPUTER SOCIETY

IEEE Signal Processing Society | IEEE COMMUNICATIONS SOCIETY | IEEE computer society

MARCH 2017 | VOLUME 3 | NUMBER 1 | ITSIBW | (ISSN 2373-776X)

PAPERS

IEEE

Call for Papers

The 5th IEEE Global Conference on Signal and Information Processing (GlobalSIP) will be held in Montreal, Quebec, Canada on November 14-16, 2017. GlobalSIP, as a new flagship IEEE Signal Processing Society conference, focuses on signal and information processing with an emphasis on up-and-coming signal processing themes. The conference features world-class plenary speeches, distinguished Symposium talks, tutorials, exhibits, oral and poster sessions, and panels. GlobalSIP is comprised of co-located General Symposium and symposia selected based on responses to the call-for-symposia proposals. Featured symposia include:

- General symposium
- Sparse SP and deep learning
- Graph signal processing
- Distributed optimization and resource management over networks
- Control & information theoretic approaches to privacy and security
- SP for interference cancellation and full-duplex communication systems
- SP for Accelerating Deep Learning
- SP for Smart Cities & IoTs
- SP & ML in large medical datasets
- Big Data Analytics for IoT Healthcare
- Advanced Bio-SP for Rehabilitation and Assistive Systems
- Deep Learning for Intelligent Multimedia
- Knowledge-based Multimedia Computing
- Stochastic & approximate computing for signal processing and machine learning
- Signal and information processing (SIP) for smart grid infrastructure
- SIP for Healthcare Engineering
- SIP for Finance and Business

Prospective authors are invited to submit full-length papers (up to 4 pages for technical content, an optional 5th page containing only references) and extended abstracts (up to 2 pages, for paper-less industry presentations and Ongoing Work presentations). Manuscripts should be original and written in accordance with the standard IEEE 2-column paper template. Accepted full-length papers will be indexed on IEEE Xplore. Accepted abstracts will not be indexed in IEEE Xplore, however the abstracts and/or the presentations will be included in the IEEE SPS SigPort. Accepted papers and abstracts will be scheduled in lecture and poster sessions.

Conference Highlights

- Plenary Talks within the general symposium and Distinguished Symposium Presentations/Talks within the 16 thematic symposia surveying emerging topics in SIP
- Panel discussions on funding opportunities, trends and targeted topics
- Enhanced industry program: paperless industry presentations, panels, demos and exhibitions
- Exciting student program: Ongoing Work tracks, Student-Industry Luncheon, 3MT competition
- Great venue with vibrant cultural, educational, and scientific identity, combining the modern buzz of a North American city and a specific European charm

Important Dates:

- **May 15, 2017** : Paper submission due
- June 30, 2017 : Final acceptance decisions notifications sent to all authors
- July 22, 2017 : Camera-ready papers due



CALL FOR PAPERS

IEEE Journal of Selected Topics in Signal Processing: Special Issue on Machine Learning for Cognition in Radio Communications and Radar (ML-CR²)

While machine learning is achieving ground breaking success in speech recognition, computer vision, natural language processing and business analytics, its impact on radio communications, and on the associated problem area in signal processing, has been less pronounced mainly due to the lack of 'big data' and big applications. However, in the era of the Fifth Generation (5G) cellular systems and Internet-of-Things (IoT), some significant changes are under way. For example, as 5G cellular systems demand huge capacity, massive connectivity, high reliability and low latency, acquiring adequate resources to operate such systems is difficult and novel models and algorithms are needed to help improve spectrum utilization by leveraging large-scale databases, full of context and information. These databases can be sourced from handheld devices, network infrastructure, and the environment, such as typical user trajectories provided by vehicular traffic management systems. In addition, government agencies are now willing to share their spectrum with commercial users. The 3550-3650 MHz band is identified for spectrum sharing between military radars and communication systems. This requires cognition both in communication systems and radars. There is also a general trend toward cognitive radars as the next generation of environment-adaptive radars with unprecedented spectral and behavioral agility. A natural approach to handling all this is the development of efficient machine learning algorithms, which, combined with traditional signal processing methods, will allow for the automation of cognitive functionality both in radars and radio communication networks. There are nontrivial challenges and open questions in the application of machine learning to RF environments starting with the fact that, as opposed to speech recognition and computer vision where the output of machine cognition can be readily compared and verified against human auditory and visual perception, no such option is available for radio signals. The main goal of this Special Issue is to raise awareness of this emerging interdisciplinary research area, and to showcase the existing state-of-the-art and its current and future challenges. Topics of interest include (but are not limited to):

- Machine learning for blind channel and signal characterization
- Machine learning for source separation
- Deep learning for RF signal classification
- Machine learning for channel decoding
- Machine learning for RF-based geolocation and signal association
- Distributed multi-agent learning in collaborative radio networks
- Machine learning-based antenna selection
- Reinforcement Learning in wireless networks
- Machine learning of the topology and structural properties of radio networks
- Joint optimization and learning of spectrum usage dynamics and spectrum access control
- Privacy-preserving machine learning for cognitive radio
- Machine learning for cognitive technologies in 5G cellular networks
- Non-parametric Bayesian machine learning for temporal clustering of spectral activity
- Machine learning for activity recognition of partially observable wireless network nodes
- Machine learning in cognitive radars for spectrum sharing with communication devices
- Machine learning for passive radars
- Machine learning for Bayesian target characterization
- Machine learning for cognitive radar characterization and for radar waveform design

Prospective authors should follow the instructions given on the IEEE JSTSP webpages: https://signalprocessingsociety.org/publications-resources/ieee-journal-selected-topics-signal-processing and submit their manuscript with the web submission system at: https://mc.manuscriptcentral.com/jstsp-ieee.

Dates:

Manuscript submission: June 30, 2017

First review completed: August 15, 2017

Revised Manuscript Due: September 15, 2017

Second Review Completed: October 31, 2017

Final Manuscript Due: December 15, 2017

Publication: February 2018

Guest Editors (ge.ml.crcr@gmail.com):

Maria Sabrina Greco, University of Pisa

Silvija Kokalj-Filipovic, U.S. Naval Research Laboratory

H. Vincent Poor, Princeton University

George Stantchev, U. S. Naval Research Laboratory

Liang Xiao, Xiamen University

The **IEEE International Workshop on Information Forensics and Security (WIFS)** is the primary annual event organized by the IEEE Information Forensics and Security (IFS) Technical Committee with the technical sponsorship of the IEEE Biometrics Council. Its major objective is to bring together researchers from relevant disciplines to exchange new ideas and the latest results and to discuss emerging challenges in different areas of information security. The 9th edition of WIFS will be held in Rennes, France, from December 4 to December 7, 2017. WIFS 2017 will feature keynote lectures, tutorials, technical & special sessions, and also demo and on-going work sessions.

Topics of interest include, but are not limited to:

Forensics: Multimedia forensics | Counter Forensics | Acquisition Device Identification | Evidence Validation | Benchmarking

Biometrics: Single or Multi-Modalities Systems | Security and Privacy | Spoofing | Performance Evaluation

Security and Communication: Covert Channels | Physical Layer Security | Steganography | Secret Key Extraction | Digital Watermarking

Multimedia Security: Cryptography for multimedia | Near duplicate detection | Data Hiding | Authentication | Forensics

Information theoretic security: Differential Privacy | Adversarial Machine Learning | Game theory | Communication with Side Information

Cybersecurity: Model and validation | Cloud Computing | Distributed Systems with Byzantines | Social Networks | Rumors and Alternative Facts

Hardware security: New primitives | Physical Unclonable Functions | Anti-Counterfeiting | Side Channels Attacks | Forensics

Surveillance: Tracking | Object / Person Detection | Behavior Analysis | Anti-Surveillance and De-identification | Privacy

Network Security: Intrusion Detection | Protocols | Traffic Analysis | Anonymity | Mobile Ad-hoc Networks | Internet of Things

Applied cryptography: Processing in the encrypted domain | Multiparty computation | traitor tracing | property preserving encryption

Submission of papers: Prospective authors are invited to submit six-page papers, including figures and references. All submitted papers will go through double-blinded peer review process. The WIFS Technical Program Committee will select papers for the formal proceedings based on technical quality, relevance to the workshop, and ability to inspire new research. Accepted papers will be presented in either lecture tracks or poster sessions. Authors of the accepted papers are required to present their papers at the conference. For all questions contact WIFS'17 Technical Program Chairs at tpc@wifs2017.org.

Warning: Papers are reviewed on the basis that they do not contain plagiarized material and have not been submitted to any other conference at the same time (double submission). These matters are taken very seriously. IEEE takes action against any author who has engaged in either practice. http://www.ieee.org/web/publications/rights/Plagiarism_Guidelines_Intro.html http://www.ieee.org/web/publications/rights/Multi_Sub_Guidelines_Intro.html

IEEE IEEE Signal Processing Society IEEE Biometrics Council

(continued on next page)

Tutorial Proposals: Up to four tutorials are scheduled on Monday December 4, 2017. Prospective tutorial contributors are encouraged to submit a tutorial proposal with a brief CV of the presenters and the detailed structure of the tutorial to the Tutorials Chair at tutorials@wifs2017.org.

Demo and Ongoing Works Proposal: This session enables both academic researchers and industrial exhibitors to showcase innovative technologies demonstrating new ideas in the field. We encourage the submission of early research prototypes and interesting mature systems. Proposals must be accompanied by a description of the demo. Please contact the Demo Session Chair at demo@wifs2017.org.

Submission of SPL and TIFS papers: Authors of IEEE Signal Processing Letters (SPL) and IEEE Transactions on Information Forensics and Security (TIFS) papers are given the opportunity to present their work at WIFS 2017, subject to space availability and approval by the WIFS Technical Program Chairs. Proposals have to be submitted to the Technical Program Chairs at tpc@wifs2017.org.

Organizing committee:

General chairs: Teddy Furon (*Inria, France)* & Carmela Troncoso (*IMDEA Soft. Institute, Spain)*
Program chairs: Zekeriya Erkin (*TU Delft, The Netherlands)* & Patrick Bas (*CNRS, France)*
Publicity chair: Bin Li (*Shenzhen University, China)* & Wei Fan *(Dartmouth College, USA)*
Tutorial chair: Luisa Verdoliva (*University di Napoli, Italy)*
Industrial liaison and demo session chair: Gwenaël Doërr (*ContentArmor, France)*

Area Chairs:

Biometry and Authentication: Jean-Luc Dugelay *(Eurecom Institut, France)*
Computer Network Security & Forensics: Jiankun Hu *(University of New South Wales, Australia)*
Data hiding: Andrew Ker *(University of Oxford, United Kingdom)*
Forensics: Hany Farid *(Dartmouth College, USA)*
Physical Layer Security and Cryptography: Matthieu Bloch *(Georgia Institute of Technology, USA)*
Privacy-Enhancing Technologies: Andreas Peter *(University of Twente, The Netherlands)*

Important deadlines:

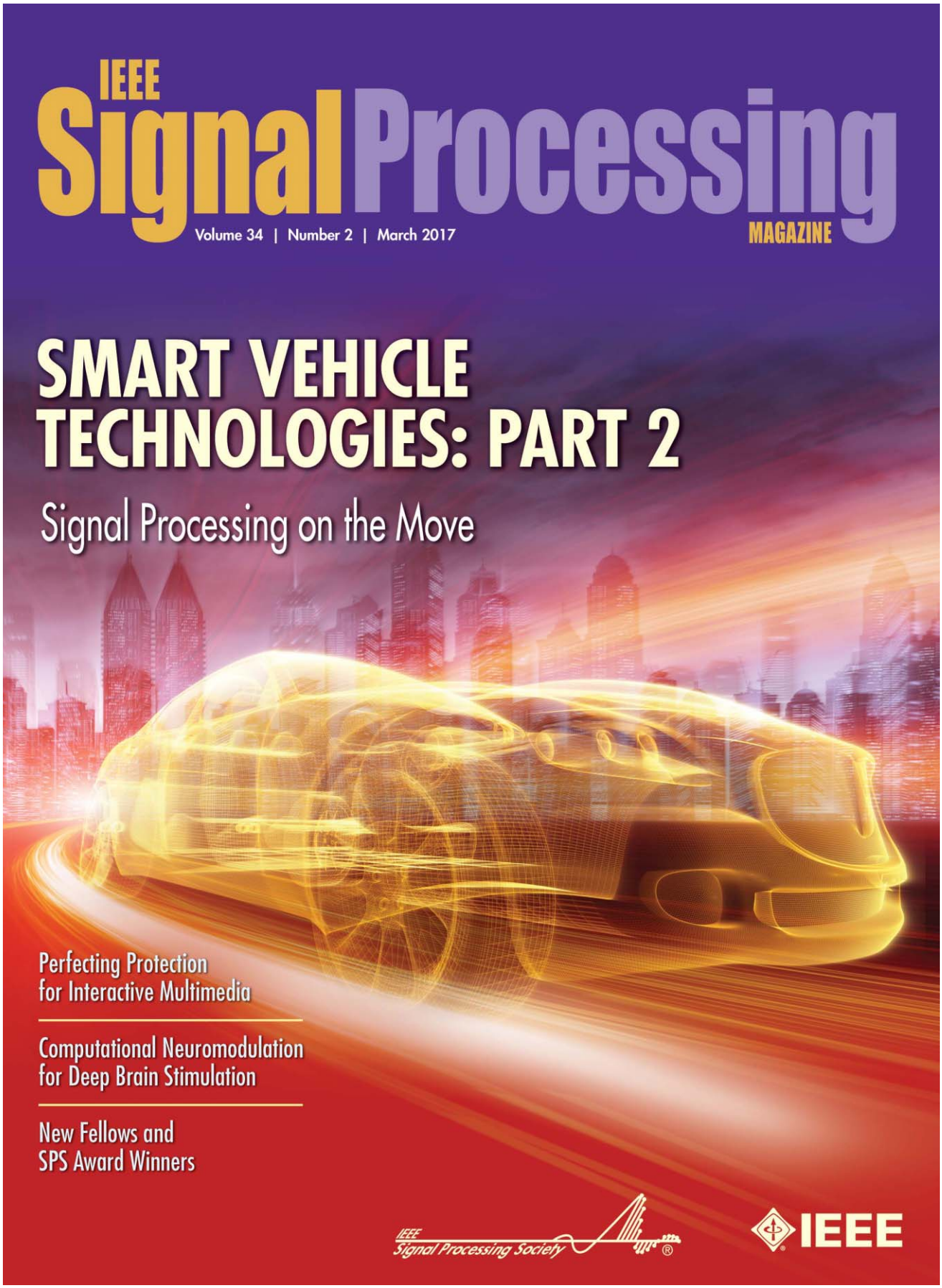
- Paper submission — June 19, 2017
- Notification of paper acceptance — September 18, 2017
- Camera-ready paper submission — October 2, 2017

- Special Session proposals — April 1, 2017
- Notification of Special Session acceptance — April 15, 2017
- Tutorial proposals — May 15, 2017
- Notification of tutorial acceptance — June 12, 2017

- Demo/on-going work proposals — September 25, 2017
- Notification of demo/on-going work — October 9, 2017

- Early registration deadline — October 22, 2017

- Workshop — December 4-7, 2017


IEEE
Signal Processing
MAGAZINE
Volume 34 | Number 2 | March 2017
SMART VEHICLE TECHNOLOGIES: PART 2
Signal Processing on the Move
Perfecting Protection for Interactive Multimedia
Computational Neuromodulation for Deep Brain Stimulation
New Fellows and SPS Award Winners
IEEE Signal Processing Society
IEEE

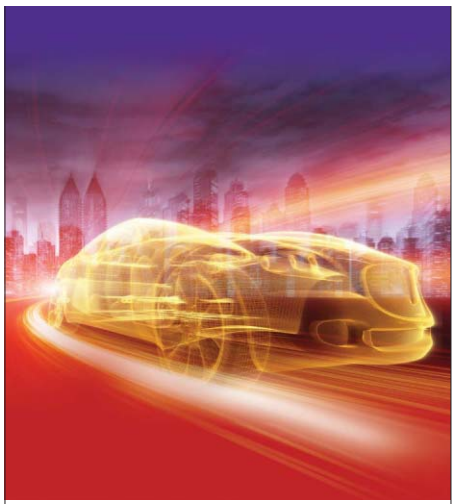
Contents

Volume 34 | Number 2 | March 2017

SPECIAL SECTION

SIGNAL PROCESSING FOR SMART VEHICLE TECHNOLOGIES: PART 2

IMAGE LICENSED BY GRAPHIC STOCK



ON THE COVER

Part 2 of the special section on signal processing for smart vehicle technologies provides an additional six articles, complementing those in Part 1, which was published in the November 2016 issue of *IEEE Signal Processing Magazine*.

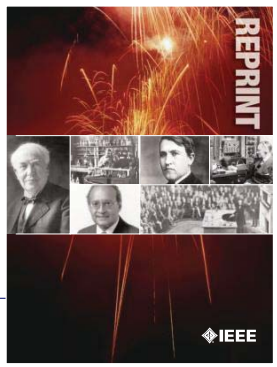
COVER IMAGE: ©ISTOCKPHOTO.COM/NADIA

FEATURE

COLUMNS

IEEE SIGNAL PROCESSING MAGAZINE (ISSN 1053-5888) (ISPREG) is published bimonthly by the Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, 17th Floor, New York, NY 10016-5997 USA (+1 212 419 7900). Responsibility for the contents rests upon the authors and not the IEEE, the Society, or its members. Annual member subscriptions included in Society fee. Nonmember subscriptions available upon request. **Individual copies:** IEEE Members US$20.00 (first copy only), nonmembers US$241.00 per copy. **Postmaster:** Send address changes to IEEE Signal Processing Magazine, IEEE, 445 Hoes Lane, Piscataway, NJ 08854 USA. Canadian GST #125634188 **Printed in the U.S.A.**

Digital Object Identifier 10.1109/MSP.2016.2642406



IEEE ORDER FORM FOR REPRINTS

Purchasing IEEE Papers in Print is easy, cost-effective and quick.

Complete this form, send via our secure fax (24 hours a day) to 732-981-8062 or mail it back to us.

PLEASE FILL OUT THE FOLLOWING

Author: _____

Publication Title: _____

Paper Title: _____

_____

RETURN THIS FORM TO:
IEEE Publishing Services
445 Hoes Lane
Piscataway, NJ 08855-1331

Email the Reprint Department at reprints@ieee.org for questions regarding this form

PLEASE SEND ME

■ 50 ■ 100 ■ 200 ■ 300 ■ 400 ■ 500 or _____(in multiples of 50) reprints.

■ YES ■ NO Self-covering/title page required. COVER PRICE: $74 per 100, $39 per 50.

■ $58.00 Air Freight must be added for all orders being shipped outside the U.S.

■ $21.50 must be added for all USA shipments to cover the cost of UPS shipping and handling.

PAYMENT

■ Check enclosed. Payable on a bank in the USA.

■ Charge my: ■ Visa ■ Mastercard ■ Amex ■ Diners Club

Account # _____ Exp. date _____

Cardholder's Name (please print): _____

_____

■ Bill me (you must attach a purchase order) Purchase Order Number _____

Send Reprints to:

_____

_____

_____

_____

Bill to address, if different:

_____

_____

_____

_____

Because information and papers are gathered from various sources, there may be a delay in receiving your reprint request. This is especially true with postconference publications. Please provide us with contact information if you would like notification of a delay of more than 12 weeks.

Telephone: _____ Fax: _____ Email Address: _____

2012 REPRINT PRICES (without covers)

Number of Text Pages

	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36	37-40	41-44	45-48
50	$129	$213	$245	$248	$288	$340	$371	$408	$440	$477	$510	$543
100	$245	$425	$479	$495	$573	$680	$742	$817	$885	$953	$1021	$1088

Larger quantities can be ordered. Email reprints@ieee.org with specific details.

Tax Applies on shipments of regular reprints to CA, DC, FL, MI, NJ, NY, OH and Canada (GST Registration no. 12534188).
Prices are based on black & white printing. Please call us for full color price quote, if applicable.

Authorized Signature: _____ Date: _____

2017 IEEE MEMBERSHIP APPLICATION

(students and graduate students must apply online)

IEEE
Advancing Technology for Humanity

Start your membership immediately: Join online www.ieee.org/join

Please complete both sides of this form, typing or **printing in capital letters**. Use only English characters and abbreviate only if more than 40 characters and spaces per line. We regret that incomplete applications cannot be processed.

1 Name & Contact Information

Please PRINT your name as you want it to appear on your membership card and IEEE correspondence. As a key identifier for the IEEE database, circle your last/surname.

☐ Male ☐ Female Date of birth (Day/Month/Year) ____/____/____

Title First/Given Name Middle Last/Family Surname

▼ Primary Address ☐ Home ☐ Business (All IEEE mail sent here)

Street Address

City State/Province

Postal Code Country

Primary Phone

Primary E-mail

▼ Secondary Address ☐ Home ☐ Business

Company Name Department/Division

Street Address City State/Province

Postal Code Country

Secondary Phone

Secondary E-mail

To better serve our members and supplement member dues, your postal mailing address is made available to carefully selected organizations to provide you with information on technical services, continuing education, and conferences. Your e-mail address is not rented by IEEE. Please check box only if you do not want to receive these postal mailings to the selected address. ☐

2 Attestation

I have graduated from a three- to five-year academic program with a university-level degree.
☐ Yes ☐ No

This program is in one of the following fields of study:
☐ Engineering
☐ Computer Sciences and Information Technologies
☐ Physical Sciences
☐ Biological and Medical Sciences
☐ Mathematics
☐ Technical Communications, Education, Management, Law and Policy
☐ Other (please specify): _____

This academic institution or program is accredited in the country where the institution is located. ☐ Yes ☐ No ☐ Do not know

I have _____ years of professional experience in teaching, creating, developing, practicing, or managing within the following field:

☐ Engineering
☐ Computer Sciences and Information Technologies
☐ Physical Sciences
☐ Biological and Medical Sciences
☐ Mathematics
☐ Technical Communications, Education, Management, Law and Policy
☐ Other (please specify): _____

3 Please Tell Us About Yourself

Select the numbered option that best describes yourself. This information is used by IEEE magazines to verify their annual circulation. Please enter numbered selections in the boxes provided.

A. Primary line of business → ☐

1. Computers
2. Computer peripheral equipment
3. Software
4. Office and business machines
5. Test, measurement and instrumentation equipment
6. Communications systems and equipment
7. Navigation and guidance systems and equipment
8. Consumer electronics/appliances
9. Industrial equipment, controls and systems
10. ICs and microprocessors
11. Semiconductors, components, sub-assemblies, materials and supplies
12. Aircraft, missiles, space and ground support equipment
13. Oceanography and support equipment
14. Medical electronic equipment
15. OEM incorporating electronics in their end product (not elsewhere classified)
16. Independent and university research, test and design laboratories and consultants (not connected with a mfg. co.)
17. Government agencies and armed forces
18. Companies using and/or incorporating any electronic products in their manufacturing, processing, research or development activities
19. Telecommunications services, telephone (including cellular)
20. Broadcast services (TV, cable, radio)
21. Transportation services (airline, railroad, etc.)
22. Computer and communications and data processing services
23. Power production, generation, transmission and distribution
24. Other commercial users of electrical, electronic equipment and services (not elsewhere classified)
25. Distributor (reseller, wholesaler, retailer)
26. University, college/other educational institutions, libraries
27. Retired
28. Other_____

B. Principal job function → ☐

1. General and corporate management
2. Engineering management
3. Project engineering management
4. Research and development management
5. Design engineering management—analog
6. Design engineering management—digital
7. Research and development engineering
8. Design/development engineering—analog
9. Design/development engineering—digital
10. Hardware engineering
11. Software design/development
12. Computer science
13. Science/physics/mathematics
14. Engineering (not elsewhere specified)
15. Marketing/sales/purchasing
16. Consulting
17. Education/teaching
18. Retired
19. Other_____

C. Principal responsibility → ☐

1. Engineering and scientific management
2. Management other than engineering
3. Engineering design
4. Engineering
5. Software: science/mngmnt/engineering
6. Education/teaching
7. Consulting
8. Retired
9. Other_____

D. Title → ☐

1. Chairman of the Board/President/CEO
2. Owner/Partner
3. General Manager
4. VP Operations
5. VP Engineering/Dir. Engineering
6. Chief Engineer/Chief Scientist
7. Engineering Management
8. Scientific Management
9. Member of Technical Staff
10. Design Engineering Manager
11. Design Engineer
12. Hardware Engineer
13. Software Engineer
14. Computer Scientist
15. Dean/Professor/Instructor
16. Consultant
17. Retired
18. Other_____

Are you now or were you ever a member of IEEE?
☐ Yes ☐ No If yes, provide, if known:

Membership Number Grade Year Expired

4 Please Sign Your Application

I hereby apply for IEEE membership and agree to be governed by the IEEE Constitution, Bylaws, and Code of Ethics. I understand that IEEE will communicate with me regarding my individual membership and all related benefits. **Application must be signed.**

Signature Date

Over Please

(continued on next page)

5 Add IEEE Society Memberships (Optional)

The 39 IEEE Societies support your technical and professional interests. Many society memberships include a personal subscription to the core journal, magazine, or newsletter of that society. **For a complete list of everything included with your IEEE Society membership, visit www.ieee.org/join.** All prices are quoted in US dollars.

Please check ✔ the appropriate box.

		BETWEEN 16 AUG 2016-28 FEB 2017 PAY	BETWEEN 1 MAR 2017-15 AUG 2017 PAY
IEEE Aerospace and Electronic Systems ⌘ 🖂	AES010	25.00 ☐	12.50 ☐
IEEE Antennas and Propagation ⌘ 🖂	AP003	15.00 ☐	7.50 ☐
IEEE Broadcast Technology ▪ ⌘ 🖂	BT002	15.00 ☐	7.50 ☐
IEEE Circuits and Systems ▪ ⌘ 🖂	CAS004	22.00 ☐	11.00 ☐
IEEE Communications ⌘ 🖂	COM019	33.00 ☐	16.50 ☐
IEEE Components, Packaging, & Manu. Tech. ▪ ⌘ 🖂	CPMT021	15.00 ☐	7.50 ☐
IEEE Computational Intelligence ⌘ 🖂	CIS011	29.00 ☐	14.50 ☐
IEEE Computer ▪ ⌘ 🖂	C016	60.00 ☐	30.00 ☐
IEEE Consumer Electronics ▪ ⌘ 🖂	CE008	20.00 ☐	10.00 ☐
IEEE Control Systems ⌘ 🖂	CS023	25.00 ☐	12.50 ☐
IEEE Dielectrics and Electrical Insulation ⌘ 🖂	DEI032	26.00 ☐	13.00 ☐
IEEE Education ⌘ 🖂	E025	20.00 ☐	10.00 ☐
IEEE Electromagnetic Compatibility ▪ ⌘ 🖂	EMC027	31.00 ☐	15.50 ☐
IEEE Electron Devices ▪ ⌘ 🖂	ED015	18.00 ☐	9.00 ☐
IEEE Engineering in Medicine and Biology ⌘ 🖂	EMB018	40.00 ☐	20.00 ☐
IEEE Geoscience and Remote Sensing ⌘ 🖂	GRS029	19.00 ☐	9.50 ☐
IEEE Industrial Electronics ⌘ 🖂	IE013	9.00 ☐	4.50 ☐
IEEE Industry Applications ⌘ 🖂	IA034	20.00 ☐	10.00 ☐
IEEE Information Theory ▪ ⌘ 🖂	IT012	30.00 ☐	15.00 ☐
IEEE Instrumentation and Measurement ⌘ 🖂	IM009	29.00 ☐	14.50 ☐
IEEE Intelligent Transportation Systems ▪ ⌘ 🖂	ITSS038	35.00 ☐	17.50 ☐
IEEE Magnetics ▪ 🖂 ♦⌘	MAG033	26.00 ☐	13.00 ☐
IEEE Microwave Theory and Techniques ⌘ 🖂	MTT017	24.00 ☐	12.00 ☐
IEEE Nuclear and Plasma Sciences ▪ ⌘ 🖂	NPS005	35.00 ☐	17.50 ☐
IEEE Oceanic Engineering ▪ ⌘ 🖂	OE022	19.00 ☐	9.50 ☐
IEEE Photonics ▪ ⌘ 🖂	PHO036	34.00 ☐	17.00 ☐
IEEE Power Electronics ⌘ 🖂	PEL035	25.00 ☐	12.50 ☐
IEEE Power & Energy ▪ ⌘ 🖂	PE031	35.00 ☐	17.50 ☐
IEEE Product Safety Engineering ▪	PSE043	35.00 ☐	17.50 ☐
IEEE Professional Communication ▪ ⌘ 🖂	PC026	31.00 ☐	15.50 ☐
IEEE Reliability ▪ ⌘ 🖂	RL007	35.00 ☐	17.50 ☐
IEEE Robotics and Automation ⌘ 🖂	RA024	9.00 ☐	4.50 ☐
IEEE Signal Processing ▪ ⌘ 🖂	SP001	22.00 ☐	11.00 ☐
IEEE Social Implications of Technology ▪ ⌘ 🖂	SIT030	33.00 ☐	16.50 ☐
IEEE Solid-State Circuits ⌘ 🖂	SSC037	22.00 ☐	11.00 ☐
IEEE Systems, Man, & Cybernetics ⌘ 🖂	SMC028	12.00 ☐	6.00 ☐
IEEE Technology & Engineering Management ⌘ 🖂	TEM014	35.00 ☐	17.50 ☐
IEEE Ultrasonics, Ferroelectrics, & Frequency Control ▪ ⌘ 🖂	UFFC020	20.00 ☐	10.00 ☐
IEEE Vehicular Technology ▪ ⌘ 🖂	VT006	18.00 ☐	9.00 ☐

Legend—Society membership includes:

⌘ One or more Society publications
▪ Society newsletter
🖂 Online access to publication
♦ CD-ROM of selected society publications

Complete both sides of this form, sign, and return to:
IEEE MEMBERSHIP APPLICATION PROCESSING
445 HOES LN, PISCATAWAY, NJ 08854-4141 USA
or fax to +1 732 981 0225
or join online at www.ieee.org/join

Please reprint your full name here

6 2017 IEEE Membership Rates (student rates available online)

IEEE member dues and regional assessments are based on where you live and when you apply. Membership is based on the calendar year from 1 January through 31 December. All prices are quoted in US dollars.

Please check ✔ the appropriate box.

RESIDENCE	BETWEEN 16 AUG 2016-28 FEB 2017 PAY	BETWEEN 1 MAR 2017-15 AUG 2017 PAY
United States	$199.00 ☐	$99.50 ☐
Canada (NB, NF, NS, and PEI HST)*	$190.35 ☐	$95.18 ☐
Canada (ON HST)	$187.37 ☐	$93.69 ☐
Canada (GST)*	$175.45 ☐	$87.73 ☐
Canada (GST and QST Quebec)	$190.31 ☐	$95.16 ☐
Africa, Europe, Middle East	$162.00 ☐	$81.00 ☐
Latin America	$153.00 ☐	$76.50 ☐
Asia, Pacific	$154.00 ☐	$77.00 ☐

*IEEE Canada Business No. 125634188

Minimum Income or Unemployed Provision

Applicants who certify that their prior year income did not exceed US$14,900 (or equivalent) or were not employed are granted 50% reduction in: full-year dues, regional assessment and fees for one IEEE Membership plus one Society Membership. If applicable, please check appropriate box and adjust payment accordingly. Student members are not eligible.

☐ I certify I earned less than US$14,900 in 2016
☐ I certify that I was unemployed in 2016

7 More Recommended Options

Proceedings of the IEEE ..... print $49.00 ☐ **or** online $43.00 ☐
Proceedings of the IEEE (print/online combination) ..... $59.00 ☐
IEEE Standards Association (IEEE-SA) ..... $54.00 ☐
IEEE Women in Engineering (WIE) ..... $25.00 ☐

8 Payment Amount

Please total the Membership dues, Society dues, and other amounts from this page:

IEEE Membership dues ⑥ ..... $_____
IEEE Society dues (optional) ⑤ ..... $_____
IEEE-SA/WIE dues (optional) ⑦ ..... $_____
Proceedings of the IEEE (optional) ⑦ ..... $_____
Canadian residents pay 5% GST or appropriate HST (BC–12%; ON–13%; NB, NF, NS, PEI–15%) on Society payments & publications only ..... TAX $_____

AMOUNT PAID ..... **TOTAL** $_____

Payment Method

All prices are quoted in US dollars. You may pay for IEEE membership by credit card (see below), check, or money order payable to IEEE, drawn on a US bank.

☐ Check ☐ VISA ☐ MasterCard ☐ Diners Club International ☐ American Express ☐ Discover

Credit Card Number

MONTH / YEAR EXPIRATION DATE

CARDHOLDER'S 5-DIGIT ZIP CODE (BILLING STATEMENT ADDRESS) USA ONLY

Name as it appears on card

Signature

Auto Renew my Memberships and Subscriptions (available when paying by credit card).
☐ I agree to the Terms and Conditions located at www.ieee.org/autorenew

9 Were You Referred to IEEE?

☐ Yes ☐ No If yes, provide the following:

Member Recruiter Name _____

IEEE Recruiter's Member Number (Required) _____

CAMPAIGN CODE	PROMO CODE

16-MEM-031 P 6/16

Information for Authors

(Updated/Effective January 2015)

For Transactions and Journals:

Authors are encouraged to submit manuscripts of Regular papers (papers which provide a complete disclosure of a technical premise), or Comment Correspondences (brief items that provide comment on a paper previously published in these TRANSACTIONS).

Submissions/resubmissions must be previously unpublished and may not be under consideration elsewhere.

Every manuscript must:

i. provide a clear statement of the problem and what the contribution of the work is to the relevant research community;
ii. state why this contribution is significant (what impact it will have);
iii. provide citation of the published literature most closely related to the manuscript; and
iv. state what is distinctive and new about the current manuscript relative to these previously published works.

By submission of your manuscript to these TRANSACTIONS, all listed authors have agreed to the authorship list and all the contents and confirm that the work is original and that figures, tables and other reported results accurately reflect the experimental work. In addition, the authors all acknowledge that they accept the rules established for publication of manuscripts, including agreement to pay all overlength page charges, color charges, and any other charges and fees associated with publication of the manuscript. Such charges are not negotiable and cannot be suspended. The corresponding author is responsible for obtaining consent from all co-authors and, if needed, from sponsors before submission.

In order to be considered for review, a paper must be within the scope of the journal and represent a novel contribution. A paper is a candidate for an Immediate Rejection if it is of limited novelty, e.g. a straightforward combination of theories and algorithms that are well established and are repeated on a known scenario. Experimental contributions will be rejected without review if there is insufficient experimental data. These TRANSACTIONS are published in English. Papers that have a large number of typographical and/or grammatical errors will also be rejected without review.

In addition to presenting a novel contribution, acceptable manuscripts must describe and cite related work in the field to put the contribution in context. Do not give theoretical derivations or algorithm descriptions that are easily found in the literature; merely cite the reference.

New and revised manuscripts should be prepared following the "Manuscript Submission" guidelines below, and submitted to the online manuscript system, ScholarOne Manuscripts. Do not send original submissions or revisions directly to the Editor-in-Chief or Associate Editors; they will access your manuscript electronically via the ScholarOne Manuscript system.

Manuscript Submission. Please follow the next steps.

1. *Account in ScholarOne Manuscripts.* If necessary, create an account in the on-line submission system ScholarOne Manuscripts. Please check first if you already have an existing account which is based on your e-mail address and may have been created for you when you reviewed or authored a previous paper.
2. *Electronic Manuscript.* Prepare a PDF file containing your manuscript in double-column, single-spaced format using a font size of 10 points or larger, having a margin of at least 1 inch on all sides. Upload this version of the manuscript as a PDF file "double.pdf" to the ScholarOne-Manuscripts site. Since many reviewers prefer a larger font, you are strongly encouraged to also submit a single-column, double-spaced version (11 point font or larger), which is easy to create with the templates provided **IEEE Author Digital Toolbox (http://www.ieee.org/publications_standards/publications/authors/authors_journals.html)**. Page length restrictions will be determined by the double-column version. Proofread your submission, confirming that all figures and equations are visible in your document before you "SUBMIT" your manuscript. Proofreading is critical; once you submit your manuscript, the manuscript cannot be changed in any way. You may also submit your manuscript as a .PDF or MS Word file. The system has the capability of converting your files to PDF, however it is your responsibility to confirm that the conversion is correct and there are no font or graphics issues prior to completing the submission process.
3. *EDICS (Not applicable to Journal of Selected Topics in Signal Processing)*. All submissions must be classified by the author with an EDICS (Editors' Information Classification Scheme) selected from the list of EDICS published online at the at the publication's EDICS webpage (*please see the list below). Upon submission of a new manuscript, please choose the EDICS categories that best suit your manuscript. Failure to do so will likely result in a delay of the peer review process.
4. *Additional Documents for Review.* Please upload pdf versions of all items in the reference list that are not publicly available, such as unpublished (submitted) papers. Graphical abstracts and supplemental materials intended to appear with the final paper (see below) must also be uploaded for review at the time of the initial submission for consideration in the review process. Use short filenames without spaces or special characters. When the upload of each file is completed, you will be asked to provide a description of that file.
5. *Supplemental Materials.* IEEE Xplore can publish multimedia files (audio, images, video), datasets, and software (e.g. Matlab code) along with your paper. Alternatively, you can provide the links to such files in a README file that appears on Xplore along with your paper. For details, please see IEEE Author Digital Toolbox under "Multimedia." To make your work reproducible by others, these TRANSACTIONS encourages you to submit all files that can recreate the figures in your paper.
6. *Submission.* After uploading all files and proofreading them, submit your manuscript by clicking "Submit." A confirmation of the successful submission will open on screen containing the manuscript tracking number and will be followed with an e-mail confirmation to the corresponding and all contributing authors. Once you click "Submit," your manuscript cannot be changed in any way.
7. *Copyright Form and Consent Form.* By policy, IEEE owns the copyright to the technical contributions it publishes on behalf of the interests of the IEEE, its authors, and their employers; and to facilitate the appropriate reuse of this material by others. To comply with the IEEE copyright policies, authors are required to sign and submit a completed "IEEE Copyright and Consent Form" prior to publication by the IEEE. The IEEE recommends authors to use an effective electronic copyright form (eCF) tool within the ScholarOne Manuscripts system. You will be redirected to the "IEEE Electronic Copyright Form" wizard at the end of your original submission; please simply sign the eCF by typing your name at the proper location and click on the "Submit" button.

Comment Correspondence. Comment Correspondences provide brief comments on material previously published in these TRANSACTIONS. These items may not exceed 2 pages in double-column, single spaced format, using 9 point type, with margins of 1 inch minimum on all sides, and including: title, names and contact information for authors, abstract, text, references, and an appropriate number of illustrations and/or tables. Correspondence items are submitted in the same way as regular manuscripts (see "Manuscript Submission" above for instructions).

Authors may also submit manuscripts of overview articles, but note that these include an additional white paper approval process http://www.signalprocessingsociety.org/publications/overview-articles/. *[This does not apply to the Journal of Selected Topics in Signal Processing. Please contact the Editor-in-Chief.]*

Digital Object Identifier

Manuscript Length. For the initial submission of a regular paper, the manuscript may not exceed 13 double-column pages (10 point font), including title; names of authors and their complete contact information; abstract; text; all images, figures and tables, appendices and proofs; and all references. Supplemental materials and graphical abstracts are not included in the page count. For regular papers, the revised manuscript may not exceed 16 double-column pages (10 point font), including title; names of authors and their complete contact information; abstract; text; all images, figures and tables, appendices and proofs; and all references. For Overview Papers, the maximum length is double that for regular submissions at each stage (please reference http://www.signalprocessingsociety.org/publications/overview-articles/ for more information).

Note that any paper in excess of 10 pages will be subject to mandatory overlength page charges. Since changes recommended as a result of peer review may require additions to the manuscript, it is strongly recommended that you practice economy in preparing original submissions. Note: Papers submitted to the TRANSACTIONS ON MULTIMEDIA in excess of 8 pages will be subject to mandatory overlength page charges.

Exceptions to manuscript length requirements may, under extraordinary circumstances, be granted by the Editor-in-Chief. However, such exception does not obviate your requirement to pay any and all overlength or additional charges that attach to the manuscript.

Resubmission of Previously Rejected Manuscripts. Authors of manuscripts rejected from any journal are allowed to resubmit their manuscripts only once. At the time of submission, you will be asked whether your manuscript is a new submission or a resubmission of an earlier rejected manuscript. If it is a resubmission of a manuscript previously rejected by any journal, you are expected to submit supporting documents identifying the previous submission and detailing how your new version addresses all of the reviewers' comments. Papers that do not disclose connection to a previously rejected paper or that do not provide documentation as to changes made may be immediately rejected.

Author Misconduct. Author misconduct includes plagiarism, self-plagiarism, and research misconduct, including falsification or misrepresentation of results. All forms of misconduct are unacceptable and may result in sanctions and/or other corrective actions. Plagiarism includes copying someone else's work without appropriate credit, using someone else's work without clear delineation of citation, and the uncited reuse of an author's previously published work that also involves other authors. Self-plagiarism involves the verbatim copying or reuse of an authors own prior work without appropriate citation, including duplicate submission of a single journal manuscript to two different journals, and submission of two different journal manuscripts which overlap substantially in language or technical contribution. For more information on the definitions, investigation process, and corrective actions related to author misconduct, see the Signal Processing Society Policies and Procedures Manual, Section 6.1. http://www.signalprocessingsociety.org/about-sps/governance/policy-procedure/part-2. Author misconduct may also be actionable by the IEEE under the rules of Member Conduct.

Extensions of the Author's Prior Work. It is acceptable for conference papers to be used as the basis for a more fully developed journal submission. Still, authors are required to cite their related prior work; the papers cannot be identical; and the journal publication must include substantively novel aspects such as new experimental results and analysis or added theoretical work. The journal publication should clearly specify how the journal paper offers novel contributions when citing the prior work. Limited overlap with prior journal publications with a common author is allowed only if it is necessary for the readability of the paper, and the prior work must be cited as the primary source.

Submission Format. Authors are required to prepare manuscripts employing the on-line style files developed by IEEE, which include guidelines for abbreviations, mathematics, and graphics. All manuscripts accepted for publication will require the authors to make final submission employing these style files. The style files are available on the web at the **IEEE Author Digital Toolbox** under "Template for all TRANSACTIONS." (LaTeX and MS Word). Please note the following requirements about the abstract:

- The abstract must be a concise yet comprehensive reflection of what is in your article.
- The abstract must be self-contained, without abbreviations, footnotes, displayed equations, or references.
- The abstract must be between 150-250 words.
- The abstract should include a few keywords or phrases, as this will help readers to find it. Avoid over-repetition of such phrases as this can result in a page being rejected by search engines.

In addition to written abstracts, papers may include a graphical abstract; see http://www.ieee.org/publications_standards/publications/authors/authors_journals.html. for options and format requirements.

IEEE supports the publication of author names in the native language alongside the English versions of the names in the author list of an article. For more information, see "Author names in native languages" (http://www.ieee.org/publications_standards/publications/authors/auth_names_native_lang.pdf) on the IEEE Author Digital Toolbox page.

Open Access. The publication is a hybrid journal, allowing either Traditional manuscript submission or Open Access (author-pays OA) manuscript submission. Upon submission, if you choose to have your manuscript be an Open Access article, you commit to pay the discounted $1,750 OA fee if your manuscript is accepted for publication in order to enable unrestricted public access. Any other application charges (such as overlength page charge and/or charge for the use of color in the print format) will be billed separately once the manuscript formatting is complete but prior to the publication. If you would like your manuscript to be a Traditional submission, your article will be available to qualified subscribers and purchasers via IEEE Xplore. No OA payment is required for Traditional submission.

Page Charges.

Voluntary Page Charges. Upon acceptance of a manuscript for publication, the author(s) or his/her/their company or institution will be asked to pay a charge of $110 per page to cover part of the cost of publication of the first ten pages that comprise the standard length (two pages, in the case of Correspondences).

Mandatory Page Charges The author(s) or his/her/their company or institution will be billed $220 per each page in excess of the first ten published pages for regular papers and six published pages for correspondence items. (**NOTE: Papers accepted to IEEE TRANSACTIONS ON MULTIMEDIA in excess of 8 pages will be subject to mandatory overlength page charges.) These are mandatory page charges and the author(s) will be held responsible for them. They are not negotiable or voluntary. The author(s) signifies his willingness to pay these charges simply by submitting his/her/their manuscript to the TRANSACTIONS. The Publisher holds the right to withhold publication under any circumstance, as well as publication of the current or future submissions of authors who have outstanding mandatory page charge debt. No mandatory overlength page charges will be applied to overview articles in the Society's journals.

Color Charges. Color figures which appear in color only in the electronic (Xplore) version can be used free of charge. In this case, the figure will be printed in the hardcopy version in grayscale, and the author is responsible that the corresponding grayscale figure is intelligible. Color reproduction charges for print are the responsibility of the author. Details of the associated charges can be found on the IEEE Publications page.

Payment of fees on color reproduction is not negotiable or voluntary, and the author's agreement to publish the manuscript in these TRANSACTIONS is considered acceptance of this requirement.

*EDICS Webpages:

IEEE TRANSACTIONS ON SIGNAL PROCESSING:
http://www.signalprocessingsociety.org/publications/periodicals/tsp/TSP-EDICS/
IEEE TRANSACTIONS ON IMAGE PROCESSING:
http://www.signalprocessingsociety.org/publications/periodicals/image-processing/tip-edics/
IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE / ACM:
http://www.signalprocessingsociety.org/publications/periodicals/taslp/taslp-edics/
IEEE TRANSACTIONS ON INFORMATION, FORENSICS AND SECURITY:
http://www.signalprocessingsociety.org/publications/periodicals/forensics/forensics-edics/
IEEE TRANSACTIONS ON MULTIMEDIA:
http://www.signalprocessingsociety.org/tmm/tmm-edics/
IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING:
http://www.signalprocessingsociety.org/publications/periodicals/tci/tci-edics/
IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS:
http://www.signalprocessingsociety.org/publications/periodicals/tsipn/tsipn-edics/

IEEE Signal Processing Society®

IEEE