

## **Auralization of an orchestra using multichannel and multisource technique (A)**

**Vigeant, Michelle C.; Wang, Lily M.; Rindel, Jens Holger**

*Published in:*  
Acoustical Society of America. Journal

*Publication date:*  
2006

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Vigeant, M. C., Wang, L. M., & Rindel, J. H. (2006). Auralization of an orchestra using multichannel and multisource technique (A). *Acoustical Society of America. Journal*, 120(5), 3009-3009.

## **DTU Library** Technical Information Center of Denmark

---

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Session 1aID****Interdisciplinary: Opening Ceremonies, Plenary Lectures**

Whitlow W. L. Au, Cochair

*Hawaii Inst. of Marine Biology, P.O. Box 1106, Kailua, HI 96734*

Sadaoki Furui, Cochair

*Tokyo Inst. of Technology, Dept. of Computer Science, 2-12-2 O-okayama Meguro-ku, Tokyo 152-8552, Japan***Opening Ceremony—8:00****Chair's Introduction—8:10***Invited Paper***8:15****1aID1. How can auditory presence be generated and controlled?** Masayuki Morimoto (Environ. Acoust. Lab., Faculty of Eng., Kobe Univ., Nada, Kobe 657-8501 Japan)

This paper reviews key results of many listening tests on auditory presence, especially auditory localization (AL) and auditory spatial impression (ASI) by the author. The author gave the first demonstration that AL in any direction can be simulated through two loudspeakers using head-related transfer functions (HRTFs). However, individual differences in HRTFs affect the accuracy of AL. It is basically possible to localize sound images in any direction using median-plane HRTFs combined with interaural differences. Furthermore, HRTFs can be simplified by the combination of only some spectral peaks and notches. Meanwhile, the author demonstrated that ASI comprises at least two perceptual components. One is auditory source width (ASW), defined as the width of a sound image fused temporally and spatially with the direct sound image. The other is listener envelopment (LEV), which is the degree of fullness of sound images around the listener, excluding the sound image composing ASW. A listener can perceive these two components separately. The perception of ASW and LEV has a close connection with the law of the first wavefront. Acoustic components under the upper limit of the law contribute to ASW, and acoustic components beyond the upper limit of the law contribute to LEV.

**Chair's Introduction—9:05***Invited Paper***9:10****1aID2. Therapeutic ultrasound.** Lawrence Crum (Ctr. for Industrial and Med. Ultrasound, Appl. Phys. Lab., 1013 NE 40th St., Seattle, WA 98105, lac@apl.washington.edu)

The use of ultrasound in medicine is now quite commonplace, especially with the recent introduction of small, portable, and relatively inexpensive, hand-held diagnostic imaging devices. Moreover, ultrasound has expanded beyond the imaging realm, with methods and applications extending to novel therapeutic and surgical uses. Among these applications are tissue ablation, acoustocautery, lipoplasty, site-specific and ultrasound mediated drug activity, extracorporeal lithotripsy, and the enhancement of natural physiological functions such as wound healing and tissue regeneration. A particularly attractive aspect of this technology is that diagnostic and therapeutic systems can be combined to produce totally noninvasive, image-guided, bloodless surgery. This general lecture will review a number of these exciting new applications of ultrasound and address some of the basic scientific questions and future challenges in developing these methods and technologies for general use in our society. We shall particularly emphasize the use of high-intensity focused ultrasound (HIFU) in the treatment of benign and malignant tumors as well as the induction of acoustic hemostasis, especially in organs that are difficult to treat using conventional medical and surgical techniques. [Work supported in part by the NIH, NSBRI, ONR, and DARPA.]

**Session 1aAA****Architectural Acoustics: Computer Modeling of Room Acoustics I**

Lily M. Wang, Cochair

*Univ. of Nebraska Lincoln, 200B Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681*

Shinichi Sakamoto, Cochair

*Univ. of Tokyo, Inst. of Industrial Science, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan***Chair's Introduction—10:55*****Invited Papers*****11:00****1aAA1. A transparency model and its applications for simulation of reflector arrays and sound transmission.** Claus Lyngge Christensen and Jens Holger Rindel (Odeon A/S, Oersted-DTU, Bldg. 352, 2800 Lyngby, Denmark)

The paper describes a new method for simulating the frequency-dependent reflection and transmission of reflector arrays, and the frequency-dependent airborne sound insulation between rooms by means of a room acoustic computer model. The method makes use of a transparency method in the ray-tracing process. In the first step of the calculation the rays hitting the relevant surfaces may either be reflected or transmitted, using a probability of 50%. In the next step the impulse responses in the receiver positions are calculated using a frequency-dependent correction to account for the reflected or transmitted energy. The method applied for the reflector array is based on a theoretical model that takes into account the dimensions of the reflecting surface, path lengths, and angle of incidence. The transmission calculation is based on the users' data for the frequency-dependent transmission loss of the partition, and this is useful for the auralization of sound transmission through different building constructions. The acoustic properties like volume, reverberation time, and the area of the transmitting surfaces are included in the simulation.

**11:20****1aAA2. Numerical determination of scattering coefficients of wall surfaces for geometrical room acoustic simulation.** Tetsuya Sakuma, Yoshiyuki Kosaka, and Yuki Tachioka (Grad. School of Frontier Sci., Univ. of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8563, Japan, sakuma@k.u-tokyo.ac.jp)

There exist a dozen geometrical room acoustic simulation programs which include scattering coefficients of wall surfaces to generate more realistic sound energy reflection. In the light of this utilization, a measurement method of the random-incidence scattering coefficient has been recently standardized by ISO 17497-1, while a numerical technique with BEM is developed to determine the scattering coefficient. First, one case study is done to investigate the behavior of scattering performance of periodical surfaces with sinusoid, triangles and rectangles, with changing the height of surface roughness. As a result, it is seen that the height-to-period ratio of about 30% maximizes the scattering coefficient in the range of middle and high frequency. Second, another case study of geometrical room acoustic simulation is done for a rectangular room composed of uneven reflective walls and absorptive ceiling and floor. Geometrical simulation is performed on a variety of conditions with changing the scattering coefficient, and the reverberation times simulated are compared with those given by wave-based simulation at every octave bands. From the correspondence in reverberation time, the scattering coefficients of the walls are estimated, and its frequency characteristics are compared with those given by the former numerical determination.

**11:40****1aAA3. Edge diffraction in computer modeling of room acoustics.** U. Peter Svensson (Acoust. Group, Dept. of Electron. and Telecommunications, Norwegian Univ. of Sci. and Technol., NO-7491 Trondheim, Norway, svensson@iet.ntnu.no) and Paul T. Calamia (Rensselaer Polytechnic Inst., Troy, NY)

Computer modeling in room acoustics is typically based on geometrical acoustics techniques. Limitations with such methods include, among other things, a lack of diffraction modeling, which primarily leads to inaccuracies at low frequencies. The inclusion of diffraction modeling is quite straightforward for first-order diffraction, which can be combined with specular reflections of any order. One impractical aspect, however, is that the number of diffraction components can be extremely high, and grows faster (with the reflection order) than the number of specular reflections does. At the same time, the importance, or magnitude, of the diffraction components will differ over an immense range. This variation can be exploited by estimating the importance of each diffraction contribution by the magnitude of its onset, and skipping the remainder of the calculations for those that are deemed too weak. This will be demonstrated for some typical geometrical cases including a set of overhead reflectors, an orchestra pit, and a convex hall shape for which diffraction is less critical. Results indicate that "diffraction culling" can provide a significant reduction in computation time with only small effects on the overall responses for the tested geometries. [This research has been supported by the Research Council of Norway].

## Session 1aAB

## Animal Bioacoustics and Underwater Acoustics: Remote Monitoring of Biological Sounds I

Marc O. Lammers, Chair

Hawaii Inst. of Marine Biology, P.O. Box 1106, Kailua, HI 96734

## Contributed Papers

10:30

**1aAB1. Assessing marine mammal abundance, density and habitat use with acoustics.** Robert Gisiner (Office of Naval Res., 875 North Randolph St., Ste. 1425, Arlington, VA 22203)

Counting marine mammals is challenging: when seen they usually cannot be heard, and when heard they usually cannot be seen. Both methods miss some fraction of the population, but how does one reconcile two sampling methods that are uncalibrated themselves and will be difficult to calibrate against each other? Three approaches to calibration are reviewed. The one-way calibration approach treats one method as correct and derives a correction function for estimating population data from the other method. The parallel independent approach derives sources of error for each method independently and then compares population estimates: close agreement would strengthen confidence in both methods. The concurrent approach attempts to reconcile each visual detection to an acoustic detection and vice versa, with estimated total numbers derived from the intersection of the two sets. Examples of each method will be provided to illustrate the challenges and potential of each approach.

10:45

**1aAB2. Reducing source localization errors: A visualization method to help guide the design, calibration, and use of widely separated acoustic sensor arrays.** Catherine L. Berchok, Gerald L. D'Spain, and John A. Hildebrand (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 93940-0701, cberchok@ucsd.edu)

The use of acoustic arrays to obtain source positions of vocalizing animals has become quite common in field studies of wild animal populations, but the results presented rarely include error bounds, leading to a false sense of confidence in the animal locations. As research from the acoustics community has shown, these source localization errors vary with respect to bearing and distance from the array and are generated by measurement uncertainties associated with the propagation environment, field equipment, and data processing/analysis methods. However, these errors can be reduced by making informed decisions about the design, calibration, and deployment of the arrays and by including measurements of amplitude ratios and vector sensor bearings in addition to time-of-arrival differences. A numerical method is presented here that allows for visualization and quantification of the impact of measurement uncertainties on source localization error bounds for widely spaced acoustic sensor array. This method will be used to show how quick comparisons of source localization error bound maps can help determine appropriate array designs for a particular field application. Among the parameters compared are array geometry, beamforming techniques, and measurement of various combinations of properties of the received acoustic field. [Work supported by a Hunt Postdoctoral Fellowship.]

11:00

**1aAB3. Detection distances of the sounds of large whales recorded on hydrophones in the offshore Gulf of Alaska.** Kathleen Stafford, Sue Moore (Appl. Phys. Lab, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98115, stafford@apl.washington.edu), and David Mellinger (Oregon State Univ., Newport, OR 97365)

The current status of most species of endangered baleen whales, including blue, fin, and humpback whales, in the Gulf of Alaska is unknown due to a lack of basic information on distribution and seasonal abundance. Remote passive acoustic monitoring can provide this information for vocal whales. However, to begin to estimate an index of abundance of calling animals, the distance at which they can be detected needs to be determined. In order to estimate transmission loss, a parabolic equation acoustic propagation model was used to provide mean loss estimates along four transects at 0.1a intervals to 5a to the N, E, S, and W from each of four receivers moored in the sound channel for species-specific frequencies and at depths at which the animals are thought to produce sound. For all species and locations, the detection range was largely determined by the choice of ambient noise levels. This suggests that masking due to anthropogenic noise could limit the range over which these animals can be detected by the moored instrument and, more importantly, the range over which they might communicate.

11:15

**1aAB4. An archive of odontocete sounds for automatic call recognition.** David K. Mellinger, Sharon L. Nieuwkerk, and Sara L. Heimlich (Cooperative Inst. for Marine Resources Studies, Oregon State Univ., 2030 SE Marine Sci. Dr., Newport, OR 97365, david.mellinger@oregonstate.edu)

Acoustic remote monitoring using automatic detection of cetacean sounds is becoming widely used to study species' seasonal distributions and to monitor and mitigate human impacts on cetaceans. To date, most work on automatic detection has focused on baleen whales and a handful of odontocetes, principally sperm whales and harbor porpoises. To further the development of detection methods for other odontocetes (particularly beaked whales, which have become the focus of much interest), it would be helpful to have standardized training and testing datasets. In preparing such datasets, several questions must be addressed: What recordings are available? How certain are the species IDs for these, and how should differing levels of certainty be handled? What is the recording quality? How should recordings be presented and annotated to make them most useful for automatic detection research? Are there any recordings known to be made on-axis with respect to the animal? How are multiple overlapping animals handled? How are mixed-species groups handled? Are data needs different for clicks and whistles? What metadata will be most useful for detection? For localization? How should recordings and metadata be stored and made them available on the net? Partial or full solutions to these questions will be presented.

**1aAB5. Automatic vocal individual recognition of acorn woodpecker (*Melanerpes formicivorus*) based on hidden Markov models.** Yuan Yao (Dept. of Ecology and Evolutionary Biol., Univ. of California, Los Angeles, Los Angeles, CA 90095, yaoyuan@ucla.edu), Ying Lin (Univ. of Arizona, Tucson, AZ 85721), Andreas Ali, and Charles Taylor (Univ. of California, Los Angeles, Los Angeles, CA 90095)

The acorn woodpecker (*Melanerpes formicivorus*) is a highly social cooperatively breeding species. A variety of conspecific interactions inside and between family groups is mediated by vocalizations. In the current study, acoustic sensor networks are used to monitor vocal behaviors of acorn woodpecker. In order to identify the callers in our behavioral study, we present a method for the automatic vocal individual recognition using hidden Markov models (HMMs). Field recordings from eight woodpeckers in two family groups were made in northern California. Individual identification accuracy is 92.65% for isolated waka calls, 84.31% for isolated syllable wa, 73.04% for isolated syllable ka, using 7-state HMMs with MFCC\_E\_D parametrization. For the continuous recordings, the recognition rate is 66.67%–100.00%. We studied the influence of two major factors in the performance of the HMMs: (i) different structures of HMMs and parametrizations of data, (ii) different qualities of signals, including degraded calls after propagation and calls from two simultaneous sources split by using beamforming of sensor networks. These results suggest that HMMs, combined with sensor networks, are a promising tool for data collection in vocal behavioral studies of animals.

**1aAB6. Cetacean vocalization tracking using an adaptive Poisson rate estimator.** Owen P. Kenny and Craig R. McPherson (Dept. of Elec. and Comput. Eng., James Cook Univ., Douglas 4811, Queensland, Australia)

This paper presents algorithms for tracking the formant structure of cetacean vocalizations using auditory modeling and adaptive filtering to estimate the Poisson intensity of nerve firing rates. These rates are in turn related to the instantaneous frequency of the formant vocalisation. The auditory model considered in this paper produces a process described as a spike train representing the auditory nerve firing, the Poisson intensity of which is determined by an underlying modulation process. Previous work [Kenny *et al.*, *Adaptive Filter for Speech Enhancement Using Poisson Rates from an Auditory Model*, paper presented at the WESPAC VIII, Melbourne, Australia (2003); C. McPherson, *Development of a Toolkit for Analysing and Localising Underwater Acoustic Signals* (James Cook University, Townsville, 2005)] has shown that this modulation process can be evaluated from the conditional expectation of the Poisson rate given interspike arrival time observations. The paper also addresses the application of adaptive filters to update the estimator parameters for different acoustic environments, which is unknown. These algorithms have been applied to real field recordings of interacting false killer whales (*Pseudorca crassidens*) and results from experiment of shown individualisation and extraction of individual narrow-band vocalizations.

TUESDAY MORNING, 28 NOVEMBER 2006

MAUI ROOM, 11:00 A.M. TO 12:05 P.M.

### Session 1aAOa

## Acoustical Oceanography and Underwater Acoustics: Acoustic Scattering by Aquatic Organisms I

Kenneth G. Foote, Cochair

*Woods Hole Oceanographic Inst., Woods Hole, MA 02543*

Masahiko Furusawa, Cochair

*Tokyo Univ. of Marine Science and Technology, 4-5-7 Konan, Minato, Tokyo, 108-8477, Japan*

Chair's Introduction—11:00

### Invited Papers

11:05

**1aAOa1. A broadband acoustic system for resonance classification of swimbladder-bearing fish.** Timothy K. Stanton, Dezhang Chu (Appl. Ocean Phys. & Eng. Dept., Woods Hole Oceanogr. Inst., M.S. #11, Woods Hole, MA 02543, tstanton@whoi.edu), J. Michael Jech (Northeast Fisheries Sci. Ctr., Woods Hole, MA 02543), and James D. Irish (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

A commercial system originally designed for surveying the seafloor has been adapted for measuring the swimbladder resonance of swimbladder-bearing fish. The new towed system has three broadband active acoustic echosounders with nearly continuous coverage of the frequency band 2–110 kHz (with some gaps). The lower frequency bands are used for measuring the resonance, while the highest band can detect zooplankton. A pulse-compression approach (cross correlating the transmit waveform with the echoes) is used to exploit the bandwidth through significantly improving the range resolution to 20, 10, and 2 cm (low-, med-, and high-frequency bands, respectively). The system has been successfully deployed as deeply as 200 m and observed resonances at approximately 3 kHz for herring. Because of the 10–20-cm range resolution at the low frequencies, fish near the towbody could be resolved from one another and fish swimming within 1 m of the seafloor could be distinguished from the seafloor. This system is compared with previous approaches of measuring swimbladder resonance. The potential of the resonance classification method for significantly reducing ambiguities associated with interpreting acoustic echo data compared with traditional high-frequency narrow-band systems (10's to 100's of kHz) is discussed.

11:25

**1aAOa2. Effects of scattering layer composition, animal size, and numerical density on the frequency dependence of volume backscatter.** Kelly Benoit-Bird (College of Oceanic and Atmospheric Sci., Oregon State Univ., Corvallis, OR 97331)

The mesopelagic boundary community around the Hawaiian Islands is a land-associated, sound scattering layer that undergoes diel migrations with both a vertical and horizontal component. A video camera system was developed to quantitatively examine the numerical density, size, and taxonomic composition of micronekton. The camera system was combined with a four-frequency vessel-mounted echosounder system (38, 70, 120, and 200 kHz) to document the full migration range of micronekton and describe the changes in composition and density throughout their diel migration. Migrating animals split into multiple, distinct layers at night with differences in micronekton density, composition, and size. These differences were correlated with differences in the frequency response of volume backscatter. The relationship between these variables and backscatter intensity relationship between frequencies is complex. The results suggest strong partitioning of habitat by these animals in space and time, which, along with the high densities of animals, indicate that competition is important in driving the behavior and structuring the community.

11:45

**1aAOa3. Probing the deep: Acoustic characterization of the Mid-Atlantic Ridge ecosystem.** John K. Horne, Cairistiona I. H. Anderson (School of Aquatic and Fishery Sci., Univ. of Washington, Box 355020, Seattle, Wa 98195), Olav Rune Godoe, and Ruben Patel (Inst. of Marine Res., Nordnes 5817, Bergen, Norway)

Implementation of the ecosystem approach to resource management potentially increases the use of acoustics during assessment surveys. A Census of Marine Life survey on the mid-Atlantic ridge during summer 2004 was an opportunity to develop a practical approach. The Norwegian vessel, *G. O. Sars*, is an acoustically-quieted platform equipped with a five-frequency (18, 38, 70, 120, and 200 kHz) echosounder, acoustic Doppler current profilers (ADCPs), multibeam sonars, and a deep-towbody. Paucity of information on species composition limited the ability to conduct a traditional acoustic survey. Acoustic structure independent of biological sampling was quantified and then integrated acoustic density and target strength observations with biological community composition and length frequency data. Species were not assigned backscatter thresholds or water column regions at the onset of analysis. Near-real-time products monitored biomass distributions: daily echograms, target strengths, and frequency-differenced echograms. Persistent biological layers at different depths occurred in the water column. Layers or components migrated toward the surface during dark hours. Modes of target strength frequency distributions and trawl catch compositions differed among layers. The use of acoustic and net technologies can be integrated with near-real-time analytic results to quantitatively characterize pelagic ecosystems. [Work supported by ONR, IMR, and NOAA Fisheries.]

TUESDAY MORNING, 28 NOVEMBER 2006

KAUAI ROOM, 11:00 A.M. TO 12:05 P.M.

### Session 1aAOB

## Acoustical Oceanography and Signal Processing in Acoustics: Acoustic Tomography for Coastal and Deep Water Applications I

James A. Mercer, Cochair

*Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6698*

Arata Kaneko, Cochair

*Hiroshima Univ., Graduate School of Engineering, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8527, Japan*

Chair's Introduction—11:00

### Invited Papers

11:05

**1aAOB1. Current structure measurements by the coastal acoustic tomography.** Arata Kaneko, Keisuke Yamaguchi, Ju Lin, Noriaki Gohda (Grad. School of Eng., Hiroshima Univ., 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan), Hong Zheng (SEA Corp., Ichikawa, Chiba 272-0127, Japan), and Yoshio Takasugi (Chugoku Ctr. of Natl. Inst. of Adv. Industrial Sci. and Technol., Kure, Hiroshima 737-0197, Japan)

The coastal acoustic tomography (CAT), coastal-sea application of deep-sea acoustic tomography (Munk and Wunsch, 1978), is proposed as an advanced technology to map current structures in the coastal sea. A sequence of successful experiments was recently carried out in the coastal seas with various oceanographic conditions around Japan (Yamaguchi *et al.*, 2005; Kaneko *et al.*, 2005; Lin *et al.*, 2005). In the Tokyo Bay experiment of November 2002, the 2-D tidal currents inducing a clockwise residual circulation were mapped by eight CAT systems, located at both the eastern and western coasts of the bay. The vortex-embedded tidal currents with a maximum velocity of  $5 \text{ ms}^{-1}$  were targets in the Kanmon Strait experiment of March 2003. In the Hiroshima Bay experiment of

September 2003, the stratified sea was shifted to the homogenized sea after the passage of a typhoon. The transition of tidal current structures before and after the typhoon was well measured by seven CAT systems, located at the northern coast and on the southern oyster rafts. The procedure of 2-D current mapping for the coastal seas with depths shallower than about 30 m is well established through a sequence of CAT experiments mentioned above.

11:25

**1aAOb2. Shallow-water tomography: Is it a good idea?** James F. Lynch and Timothy F. Duda (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, Jlynch@whoi.edu)

There seems to be a feeling among ocean acousticians that anything that can be measured acoustically should be measured acoustically. However, there are two drawbacks to that point of view, comforting as it might be to us acousticians. First, many things that one would think are measurable with acoustics are actually not measurable, or are only barely so. And second, acoustics may turn out to be a poor or cost-ineffective measurement compared to other techniques. In this paper, we will look at whether acoustic tomography is viable, can produce high-quality measurements, and is cost effective in shallow water. We will use feature resolution, cost, reliability, intrusiveness, averaging properties, and other measures to obtain a weighted description of how effective tomography is for typical shallow-water (coastal) applications and then compare this to the same measure for other (conventional) oceanographic measurement systems. [Work supported by ONR.]

11:45

**1aAOb3. Ocean acoustic tomography using high-frequency systems.** Michael B. Porter, Katherine H. Kim, Martin Siderius (Heat, Light, and Sound Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130, mikeporter@hlsresearch.com), Mohsen Badiey (Univ. of Delaware, Newark, DE 19716), Thomas Folegot (NATO Undersea Res. Ctr., 19138 La Spezia, Italy), and the KauaiEx Group (NUWC)

New system concepts have emerged in which large numbers of acoustic and environmental sensors are networked together in littoral areas. The spatial scales for these networks are much finer—features such as linear and nonlinear internal waves/tides can affect both network and sensor performance. A vision readily comes together in which ocean acoustic tomography is used to monitor the finescale, using the network to relay travel-time information in real time and also to probe the environment. Further, travel-time information may be assimilated into an ocean circulation model together with direct current and temperature measurements from fixed and mobile platforms to provide nowcasts and forecasts of the local ocean weather. The acoustic tomography in this vision uses a much higher frequency band (e.g., 8–16 kHz) than has traditionally been used and there has been some skepticism about the feasibility of doing tomography in this band. Using data from two large experiments in Hawaii (KauaiEx and MakaiEx), we show that (1) HF signals can be used to clearly measure the channel impulse response, (2) changes in the impulse response are directly linked to changes in the oceanography, and, therefore, (3) HF signals can be used to invert for the ocean structure.

TUESDAY MORNING, 28 NOVEMBER 2006

KAHUKU ROOM, 10:30 A.M. TO 12:30 P.M.

### Session 1aBB

## Biomedical Ultrasound/Bioresponse to Vibration: Ultrasound Enhancement of Drug Activity

Constantin C. Coussios, Cochair

*Univ. of Oxford, Dept. of Engineering Science, Inst. of Biomedical Engineering, 43 Banbury Rd., Oxford, OX2 6PE, U.K.*

Shin-ichiro Umemura, Cochair

*Kyoto Univ., School of Health Science, Kyoto, 607-8507 Japan*

### Invited Papers

10:30

**1aBB1. Ultrasound-enhanced thrombolysis.** Christy K. Holland, Saurabh Datta, Sampada S. Vaidya (Dept. of Biomed. Eng., MSB 6167, 231 Albert Sabin Way, Cincinnati, OH 45208), Jason M. Meunier, George J. Shaw (Dept. of Emergency Medicine), and Constantin-C. Coussios (Univ. of Oxford, Oxford OX1 3PJ, UK)

Substantial enhancement of recombinant tissue plasminogen activator (rt-PA) thrombolysis can be achieved with ultrasound, suggesting its use as an adjunctive treatment in thrombolytic therapy for stroke. This enhancement may be due to acoustic cavitation, thermal effects, forced convection of rt-PA, or biochemical enhancement. The effect of US duty cycle on the lytic efficacy of pulsed 120-kHz ultrasound in a human *in vitro* clot model was investigated, using a time-lapse microscopic imaging technique. The initial lytic rate increased linearly with duty cycle. In addition, a passive cavitation detection scheme was used to explore stable and inertial cavitation as potential mechanisms for ultrasound-enhanced thrombolysis. Porcine clots were exposed to pulsed ultrasound and rt-PA *in vitro* for 30 min using four separate ultrasound treatment regimes: (i) no cavitation (0.15 MPa); (ii) stable cavitation alone (0.24

MPa), (iii) stable and inertial cavitation combined (0.36 MPa); or (iv) no ultrasound. Percent clot mass loss after each treatment was used to determine thrombolysis efficacy. Clots exposed to stable cavitation and rt-PA exhibited the highest mass loss. Significant enhancement of thrombolysis correlates with presence of cavitation and stable cavitation appears to play a more important role in the enhancement of thrombolysis.

10:50

**1aBB2. *In situ* microbubble generation from liquid precursor for diagnostic and therapeutic application.** Ken-ichi Kawabata, Akiko Yoshizawa, Takashi Azuma, Hideki Yoshikawa, Hiroko Watanabe, Kazuaki Sasaki, Koji Hirata, and Shin-ichiro Umemura (Central Res. Lab., Hitachi, Ltd., Japan)

Stabilized microbubbles show high echogenicity and characteristic nonlinear acoustic responses and also work as “sensitizers” for HIFU therapy. Accordingly, if microbubbles were selectively placed in targets inside the body, targeted diagnosis and therapy would be possible. However, unfortunately, microbubbles are too large to apply tissues other than in vessels. For tumor detection and therapy, we propose the use of a nano-sized liquid precursor of microbubbles, which is small enough to accumulate in tumor tissues and generate microbubbles with ultrasound pulses. We have developed emulsion-based precursors containing perfluoropentane and found that microbubbles can be produced with a 10  $\mu$  pulse of 2-MHz ultrasound at intensities of more than about 150 W/cm<sup>2</sup> in water. The intensity threshold was controllable by the added nonvolatile fluorocarbons. The microbubble generation was also observed in mice tumors at the same ultrasound intensity range. It was also found that when microbubbles were generated in tumor tissues, CW ultrasound exposure resulted in tumor damage. Our results indicate the potential usefulness of our precursor for targeted ultrasound diagnosis and therapy. [Part of this work was supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology through a grant-in-aid for the creation of innovations through business-academic public sector cooperation.]

11:10

**1aBB3. Developing novel therapeutic strategies through biological effects of ultrasound.** Sumit Paliwal and Samir Mitragotri (Dept. of Chemical Eng., Univ. of California, Santa Barbara, CA 93106)

Ultrasound-mediated cavitation has been shown to facilitate delivery of drugs across biological barriers in cells and tissues. However, these nonthermal mechanisms can also trigger various biochemical pathways in the biological milieu, opening up doors to interesting ultrasonic therapeutic applications. For example, a brief application of ultrasound followed by treatment with quercetin (a poorly potent chemotherapeutic drug) becomes selectively cytotoxic towards prostate and skin cancer cells. As opposed to the increased delivery of drugs to cancer cells, this treatment methodology is hypothesized to be mediated through inhibition of heat shock protein-hsp72, a vital stress-protein essential for the survival of cancer cells. Additional examples demonstrating the use of ultrasound’s biological effects for therapeutic applications will be discussed. For example, ultrasound is demonstrated to work as a vaccination adjuvant by activating the Langerhans cells (immune cells present in skin) and leading to a robust immune response against tetanus toxoid. Overall, this presentation will demonstrate the increasing importance of ultrasound’s role as a biological sensitizer enabling novel therapeutic strategies, a role which is beyond its conventional use as a drug delivery tool.

### Contributed Papers

11:30

**1aBB4. Ultrasound-mediated nanoparticle drug delivery in an *in vivo* tumor model.** Balasundar I. Raju, Christopher S. Hall (345 Scarborough Rd., Briarcliff Manor, NY 10510; balasundar.raju@philips.com), Michael S. Hughes, Samuel A. Wickline, and Gregory M. Lanza (Washington Univ. School of Medicine, MO 63130)

One of the exciting possible future clinical uses of ultrasound is site-targeted drug delivery using focused acoustic energy. In an earlier study, it was shown that ultrasound exposure promoted lipid delivery from nanoparticles into melanoma cells in culture [Crowder *et al.*, UMB 2005]. In order to extend the study to *in vivo* situations, a novel, combined ultrasound therapeutic and imaging system was developed. The system consisted of a therapeutic transducer (−6-dB BW 1.1 to 2.3 MHz; natural focus 8 cm; *f*-number 1) with a centrally located and coregistered imaging transducer connected to a commercial ultrasound imaging system. The system provided capability for arbitrary acoustic excitation, electronic axial steering, and mechanical translation of the therapy transducer as well as real-time acquisition of images. The device was tested on bilateral murine tumor models (MDA435). Tumor images provided real-time feedback for treatment planning. The animals were injected with perfluorooctyl bromide nanoparticles targeted to  $\alpha_v\beta_3$  integrin and with a radioactive complex. Initial results showed that application of therapeutic ultrasound affected the delivery of the radioactive complex as measured by radioactivity 24 h postinjection. Experimental results and experiences will be

presented detailing the importance of imaging for *in vivo* treatment guidance and planning.

11:45

**1aBB5. Ultrasound-assisted penetration of rt-PA into clots to enhance thrombolysis.** Saurabh Datta, Sampada Vaidya (Dept. of Biomed. Eng., Univ. of Cincinnati, 231 Albert Sabin Way, Cincinnati, OH 45267-0586), Shauna Beiler (VAMC 410, Cincinnati, OH 45220), Gabrielle De Courten-Myer, and Christy K. Holland (Univ. of Cincinnati, Cincinnati, OH)

Enhanced rt-PA thrombolysis has been shown in the presence of pulsed ultrasound. Porcine clots were exposed to ultrasound (120 kHz, 80% duty cycle, 1667-Hz PRF, 30 min) and rt-PA at exposure levels exhibiting no cavitation (0.15 MPa), stable cavitation (SC) alone (0.24 MPa), and inertial and stable (IC+SC) combined (0.37 MPa). A separate group of clots was also exposed to rt-PA alone (sham). Percent clot mass loss after each treatment was used to determine thrombolysis efficacy. Cavitation activity was monitored throughout the exposure. After the treatment protocol, penetration of rt-PA and plasminogen was evaluated by immunofluorescence imaging. The highest percent mass loss was observed for SC exposure (26.8%) followed by the IC+SC exposure (19.8%) and the lowest for no cavitation exposure (14.7%). The penetration of rt-PA was distinctly higher for clots exposed to SC compared to unexposed clots. Enhancement of thrombolysis is related to presence of cavitation and SC exposure appears to enhance thrombolysis by facilitating more rt-PA penetration into the clot.



**1aBB6. Enhancement of recombinant tissue-plasminogen activator (rt-PA) activity with 2-MHz transcranial Doppler ultrasound.** Tyrone M. Porter, Christy K. Holland (Dept. of Biomed. Eng., Univ. of Cincinnati, 231 Albert Sabin Way, Cincinnati, OH 45267-0586), Jason M. Meunier, and George J. Shaw (Univ. of Cincinnati, Cincinnati, OH 45267-0769)

A video microscopy technique was developed to monitor the lytic activity of recombinant tissue plasminogen activator (rt-PA) in combination with 2-MHz transcranial Doppler ultrasound (TCD). Human whole blood clots formed on silk sutures were treated with TCD pulses alone [pulse length=10 cycles, pulse repetition frequency (PRF)=10.5 kHz,  $P_{\text{neg}}=0.18-0.47$  MPa], rt-PA alone (3  $\mu\text{g/ml}$ ), and rt-PA with TCD pulses in human fresh frozen plasma for 30 min. An inverted optical microscope was used to acquire images of the clot during treatment. The extent of clot lysis was determined by assessing clot width as a function of time. The average percent change in clot width at 30 min was comparable ( $\sim 25\%$ ) for clots treated with rt-PA and TCD pulses for all output pressures. However, the initial lytic rate (within the first 5 min) was directly related to the TCD pressure output. The initial lytic activity of rt-PA and 2-MHz TCD pulses ( $P_{\text{neg}}=0.18$  MPa, PRF=10.5 kHz, duty cycle =13%) was similar to the initial lytic activity of rt-PA and 120-kHz ultrasound pulses ( $P_{\text{neg}}=0.18$  MPa, PRF=1667 Hz, duty cycle=20%).

**1aBB7. Ultrasound and biodegradable nanoparticles for anticancer drug and gene delivery.** Rinat Esenaliev, Olga Chumakova, Inga Ciceaite (Ctr. for Biomed. Eng., UTMB, Galveston, TX), Shilla Chakrabarty, Mark Evers (UTMB, Galveston, TX), and Valery Andreev (Moscow State Univ., Moscow, Russia)

Poor penetration of macromolecular drugs and genes in tumor cells substantially limits efficacy and safety of cancer chemo- and biotherapy. Cavitation induced by interaction of ultrasound with nanoparticles accumulated in tumors due to the enhanced permeability and retention effect may provide efficient drug and gene delivery in tumor cells. Biodegradable polymer (PLGA) air-filled nanoparticles manufactured in our laboratory were used for enhancement of drug and gene delivery *in vitro* and *in vivo* in human breast, prostate, or ovarian tumors of nude mice (two tumors in each animal). Our specially designed confocal ultrasound system allowed for irradiation and cavitation detection *in vitro* and *in vivo*. It was found that the nanoparticles substantially decrease cavitation threshold and increase cavitation activity *in vitro* and *in vivo*. Moreover, we obtained three-fold increase of tumor cell transfection *in vivo* compared to non-irradiated control tumors of same mice. Substantial tumor necrosis was achieved when macromolecular anticancer drug anti-IGFBP-2 was used in combination with the nanoparticles and ultrasound. Our results indicate that biodegradable polymer nanoparticles and ultrasound can be used for enhancement of drug and gene delivery in tumors. [These studies were supported by the DoD Prostate Cancer Research Program (Grant No. W81XWH-04-1-0247) and Breast Cancer Research Program (Grant No. DAMD17-01-1-0416).]

TUESDAY MORNING, 28 NOVEMBER 2006

IAO NEEDLE/AKAKA FALLS ROOM, 10:30 A.M. TO 12:00 NOON

## Session 1aMU

### Musical Acoustics: Music Performance

Lydia Ayers, Cochair

*Hong Kong Univ. of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

Akira Nishimura, Cochair

*Tokyo Univ. of Information Sciences, Dept. of Media and Cultural Studies, Faculty of Informatics, 1200-1 Yatoh-cho, Wakaba-ku, Chiba 265-8501, Japan*

### Contributed Papers

10:30

**1aMU1. Anticipatory motion in piano performance.** Werner Goebel and Caroline Palmer (Dept. of Psych., McGill Univ., 1205 Dr. Penfield Ave., Montreal, QC H3A 1B1, Canada, werner.goebel@mail.mcgill.ca)

Recent motor control studies of piano performance address the anticipatory movements of pianists fingers. The reported study investigated anticipatory motion under manipulations of performance rate or tempo. It was tested whether faster tempi require larger anticipatory preparation, as suggested in piano pedagogy literature. Sixteen skilled pianists repeatedly performed short isochronous melodies from memory at four different tempi (from 500 to 143 ms inter-onset intervals). A passive 3-D motion capture system tracked the movements of 40 markers on the hand, fingers, and the piano keys. The melodies manipulated both the repositioning of certain fingers (thumb and pinkie finger) on the keyboard and the distance between repeating finger movements (three or six tones). Functional data analysis techniques were applied to the analysis of motion trajectories. Data analyses are presented that indicate that the repositioning movement of a finger toward its next goal (keypress) begins sooner at faster tempi than at slower tempi in the  $x$  plane (sideways motion on the keyboard).

Applications to piano practice will be discussed. [Work supported by the Austrian Science Fund, NSERC, and the Canada Research Chairs program.]

10:45

**1aMU2. Effects of altered auditory feedback on piano performance.** Noriyuki Takahashi (Kyoto City Univ. of Arts, 13-6 Kutsukake-cho, Oe, Nishikyo-ku, Kyoto, 610-1197, Japan, m03904@kcua.ac.jp) and Minoru Tsuzaki (Kyoto City Univ. of Arts, Kyoto 610-1197, Japan)

Some studies in the field of developmental psychology have suggested that a complex task executed by an expert includes the automated and sequential execution of some subordinate skills. In this study, we regard the development of musical performance as the length of motor control of a musical instrument carried out with a feedforward process. We then verify this assumption through a performance experiment using skilled and unskilled pianists. Both pianists play one musical figure repeatedly in the experiment. In the middle of the performance, a mapping function between the key velocity and its sound output is altered. It is expected that

pianists would try to correct their performance when receiving this altered auditory feedback. By checking the point at which they start to correct, we can infer the length of the unit used with the feedforward process. It is expected that the correction would occur at a specific point in the musical figure if skilled pianists have a certain length unit of automated motor control. On the other hand, the point of correction would have less consistency in the case of unskilled pianists. [Work supported by the Grant-in-Aid for Scientific Research A, No. 16200016, JSPS.]

11:00

**1aMU3. An analysis of note deviation manner in piano music.** Yukiko Toda, Yosuke Nakayama, Takuya Okura, Naoto Yasuda, and Shigeki Okawa (Chiba Inst. of Technol., 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan)

Human music performance usually contains “out-of-score” features such as players’ physical characteristics and emotional expressions besides the on-score information (notes, rests, and various marks). Although those features are not written on scores, they have a very important role to add an artistic sense to the music. As one of the “out-of-score” features, the deviation manner of each note between the score and the human performance is focused on. In this study, among various genres and forms of music, F. Chopin’s mazurkas (a form of Polish folk dance), which have triple time and the accent on the second or third beat, are used. By limiting the music genres, it becomes simplified to compare the tendency of players. For the experimental dataset (three pianists, 51 mazurkas), after extracting the fluctuations (time differences) from the regular position of each note, the Fourier transform is applied to the sequence. As a result, a significant difference among the players and the playing modes is obtained, which show the temporal and the rhythm patterns. As a future work, by using the patterns, it is possible to create a virtual performer who has/imitates a certain player or a certain playing mode.

11:15

**1aMU4. A technical look at swing rhythm in music.** Kenneth A. Lindsay and Peter R. Nordquist (Southern Oregon Univ., 1250 Siskiyou Blvd., Ashland, OR 97520, nordquip@sou.edu)

We investigate swing rhythm in music using computer analysis techniques. Swing is not a genre of music, rather a style of performance. The same musical piece (data) can be played in swing or straight time. Musical notes and structure are identical in both performances, but the notes temporal patterns have slight, significant differences between straight and swing performances. We demonstrate a technical approach analyzing these differences, and show examples of several styles of swing: American Swing, Brazilian Samba, Jamaican Reggae. Prior research into swing rhythm discovered swing ratio representing the temporal variation of the music as performed compared to the music as written. Our work shows that in some cases swing ratio is inadequate to properly model the real patterns of temporal variation. We have extracted details showing swing exists hierarchically, i.e., the patterns of temporal variation played at one time scale may differ from the types of patterns of temporal variation played at a slower or faster time scale. We also show examples of different instruments playing swing differently from each other, locking together at certain canonical time locations that represent standard counting and subdivision of the musical meter. This latter phenomenon has been previously investigated as ensemble swing.

11:30

**1aMU5. A jazz arrangement system yielding a guitar chord-form sequence.** Koji Sawayama, Norio Emura, Masuzo Yanagida (Doshisha Univ., 1-3, Tatara-Miyakodani, Kyo-tanabe, Kyoto, 610-0321, Japan, dtf0751@mail4.doshisha.ac.jp), and Masanobu Miura (Ryukoku Univ., Otsu, 520-2194, Japan)

Many guitar scores of rock or pop music arranged in jazz style are available all over the world. The amount of software for arranging given sequence of chord names and books of arrangement in jazz style has been growing for these past decades. Using this software, even amateur guitarists can play their favorite tunes in jazz style without knowledge of jazz theory. Most of these software systems, however, do not take musical appropriateness into account. So, sounds of chord-form sequence output by these software systems often mismatch the input melody. Proposed here is a system that generates a chord-form sequence in jazz style for a given set of melody and chord-name sequence. The system analyzes an input set of given melody and chord-name sequence based on jazz theory, and yields the chord-form sequence considering the melody line. It is confirmed that the chord-form sequence yielded by this system meets user requirements and is more preferable than that by software systems on the market. [Work supported by Academic Frontier Project, Doshisha Univ., Knowledge Cluster Project, MEXT, Grant-in Aid for Scientific Research (No.16700154), MEXT, HRC, Ryukoku Univ.]

11:45

**1aMU6. A modular system yielding jazz-style arrangement for a given melody and sequence of chord names.** Norio Emura, Masuzo Yanagida (1-3, Tatara-Miyakodani, Kyo-Tanabe, Kyoto, 610-0321 Japan, etf1702@mail4.doshisha.ac.jp), and Masanobu Miura (Otsu 520-2194, Japan)

There are many music systems available on the market, such as ones for automatic arrangement of music pieces given as note sequences for solo piano into a piano score in a specific style. Those systems, however, are usually designed to generate music by concatenating existing arrangement patterns, so we cannot expect those systems to meet user requirements. This paper proposes a system that arranges a given set of melody expressed as a note sequence and a chord name sequence into a piano score in modern jazz style based on so-called “jazz theory.” The system is implemented in the classical theory of harmony and arrangement techniques described in jazz theory in a modular system. Proposed is a system that gives arrangement by integrating several modules according to structure of input music and user requirements. Performance of the proposed system is evaluated by comparing results given by the proposed system with those given by popular arrangement systems available on the market. Experimental results show that arrangement by the proposed system is significantly superior to arrangement by other systems available on the market. [Work supported by Academic Frontier Project, Doshisha Univ., Knowledge Cluster Project, MEXT, Grant-in Aid for Scientific Research (No. 16700154), MEXT, HRC, Ryukoku Univ.]

**Session 1aSC****Speech Communication: Communicative Speech Synthesis and Spoken Dialog**

Alan Black, Cochair

*Carnegie Mellon Univ., Language Technologies Inst., 5000 Forbes Ave., Pittsburgh, PA 15213*

Yoshinori Sagisaka, Cochair

*Waseda Univ., GITI, 297 Building, 1-3-10 Nishi Waseda, Shinjuku-ku, Tokyo, 169-0051, Japan***Chair's Introduction—10:30*****Invited Papers*****10:35****1aSC1. Hidden Markov model-based speech synthesis as a tool for constructing communicative spoken dialog systems.** Keiichi Tokuda (Dept. Computer Sci. and Eng., Nagoya Inst. of Technol., Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan)

For constructing spoken dialog systems which can realize natural human-computer interaction, speech synthesis systems are required to have an ability to generate speech with arbitrary speaker's voice and various speaking styles and/or emotional expressions. Although state-of-the-art speech synthesis systems based on unit selection and concatenation can generate natural-sounding speech, it is still difficult to synthesize various voices flexibly because they need a large-scale speech corpus for each voice. In recent years, a kind of corpus-based speech synthesis system based on hidden Markov models (HMMs) has been developed, which has the following features: (1) original speaker's characteristics can easily be reproduced because all speech features, not only spectral parameters but also fundamental frequencies and durations, are modeled in a unified framework of HMM, and then generated from the trained HMMs themselves; (2) using a very small amount of adaptation speech data, voice characteristics can easily be modified by transforming HMM parameters by a speaker adaptation technique used in speech recognition systems. From these features, the HMM-based speech synthesis approach is expected to be used as a tool for constructing communicative spoken dialog systems: keeping such a viewpoint in mind, basic algorithms and techniques for HMM-based speech synthesis are reviewed.

**10:55****1aSC2. Modeling prosody in speech processing.** Yang Liu (Computer Sci. Dept., Univ. of Texas at Dallas, MS EC-31, Box 83068, Richardson, TX 75083-0688)

Prosody is an important aspect of speech. In spoken language processing, effective prosody modeling helps to identify additional information beyond words (i.e., "how it is said" instead of "what is said") and thus better understand speech. In this talk, we will discuss how prosodic information is utilized in various speech processing tasks. Prosodic features are extracted to represent duration, pitch, and energy, with different normalization, and modeled using machine learning techniques. Research has shown that prosody provides valuable information for tasks such as automatic identification of important events in spoken language (e.g., sentence boundaries or punctuation, disfluencies, discourse markers, topics, and emotions in dialog). These phenomena are important for enriching speech recognition output and helping downstream language processing modules. Modeling prosodic variation across speakers is also useful for these tasks, as well as for developing speaker recognition systems. Additionally, some issues in machine learning techniques in prosody modeling will be discussed. Understanding how prosody is used to signal interesting events in speech will help to build better synthesis models for generating more natural and expressive speech.

**11:15****1aSC3. Expressive speech synthesis using multilevel unit selection.** Esther Klabbbers and Jan van Santen (Ctr. for Spoken Lang. Understanding, OGI School of Sci. & Eng., OHSU, 20000 NW Walker Rd., Beaverton, OR 97006, klabbbers@cslu.ogi.edu)

Generating natural sounding and meaningful prosody is a central challenge in text-to-speech synthesis, especially when generating expressive speech. Recently, we proposed a multilevel unit sequence synthesis approach, based on the general superpositional model of intonation, which describes a pitch contour as the sum of component curves that are each associated with different phonological levels, specifically the phoneme, foot, and phrase. During synthesis, segmental perturbation curves, accent curves, and phrase curves are extracted from the acoustic signal and are combined into target pitch curves; these target curves are then imposed on the acoustic unit sequences using standard pitch modification methods. This approach represents an attempt to combine the strengths of the two dominant approaches to speech synthesis: unit selection synthesis, which preserves all details of natural speech but struggles with coverage of the very large combinatorial space of phoneme sequences and prosodic contexts, and diphone synthesis, which addresses coverage by generating rule-based synthetic target prosody and imposing it on acoustic units using signal modification methods. This approach minimizes prosodic modification artifacts, optimizes the naturalness of the target pitch contour by using quasi-natural contours, yet avoids the combinatorial explosion of standard unit selection synthesis.

**1aSC4. Realization of rhythmic dialogue on spoken dialogue system using paralinguistic information.** Shinya Fujie and Tetsunori Kobayashi (Dept. Comput. Sci., Waseda Univ., 3-4-1 Okubo, Rm. 55N-509, Shinjuku-ku, Tokyo 169-8555, Japan)

Paralinguistic information recognition and its application to spoken dialogue system are discussed. Paralinguistic information brings various useful cues that cannot be obtained from linguistic information. The system may know, for example, who should speak next, how the user really feels, and so on. In this paper, we implement several paralinguistic information recognition systems, such as prosody-based attitude recognition system, back-channel feedback generation/recognition system, head gesture recognition system, and so on. Finally, we combine them into the spoken dialogue system on the humanoid robot, and realize a rhythmic dialogue.

TUESDAY MORNING, 28 NOVEMBER 2006

WAIALUA ROOM, 10:30 A.M. TO 12:00 NOON

### Session 1aUW

## Underwater Acoustics: Time Reversal Methods

Mark Stevenson, Cochair

*NATO Saclantcen Research Centre, CMR 426 Box 714, APO, AE 09613-0714*

Shunji Ozaki, Cochair

*Ok Electric Industry Co., Ltd., 7-12 Toranomon, 1-Chome, Minato-Ku, Tokyo 105-8460, Japan*

### Contributed Papers

10:30

**1aUW1. The focused acoustic fields time reversal experiment series.**

Mark Stevenson, Piero Guerrini (NATO Undersea Res. Ctr., La Spezia, Italy), Tuncay Akal, Heechun Song, Karim Sabra, Philippe Roux, William S. Hodgkiss, and William A. Kuperman (Scripps Inst. of Oceanogr., La Jolla, CA)

We review a progression of at-sea experiments that brought the concept of underwater, acoustic time-reversal (TR) focusing in the sea from an unverified hypothesis to reality. When the experiments began, it was completely unknown whether the time-varying nature of the oceanic environment would support sufficient temporal stability for the method to succeed, even at long ( $\lambda > 3$  m) acoustic wavelengths. A 10-year progression of experiments showed that not only was this possible, but that various applications could be implemented using higher frequency sound (up to 18 kHz), including phase coherent digital communications. Our most recent experiment, FAF-06, investigated multiuser applications for passive TR communications using small transmit and receive arrays. We demonstrated throughput of 10 kbps at several km range in a shallow-water propagation channel where the transfer function had an ISI span of several hundred symbols. [Work supported by NATO and ONR.]

10:45

**1aUW2. Beam sweeping in sediment layer using phase-conjugated pseudo sound sources.**

Yoshiaki Tsurugaya, Toshiaki Kikuchi<sup>a)</sup>, Ryoichi Iwase<sup>b)</sup>, and Koich Mizutani<sup>c)</sup> (NEC Corp., 1-10 Nishinchou Fuchu-city, Tokyo, 183-8501, y-tsurugaya@bp.jp.nec.com)

When an object that is buried in a sediment layer is detected from the diagonal, detection is hampered by reflection from the bottom surface. Stated in different terms, the sound-wave energy that reaches the object is small because of the reflection and refraction that occurs at the bottom surface when a sound wave is applied to the object in the sediment for diagonal detection. To alleviate that problem, a pseudo sound source can be set up in sediment layer. The sound wave radiated from the pseudo sound source is then received by a time-reversal array. The sound wave can be injected into the sediment layer by reversing the received signal with time and radiating it again from the time-reversal array. The beam can hit the target through movement of this pseudo sound source along the bottom surface. The sound wave that hits the target is reflected and returns in water. The sound wave is subsequently reflected many times at the

surface and bottom and is diffused over time. The passive-phase conjugate processing is then given to the received diffusing signal to reduce this diffusion. The pulse is compressed by this processing, thereby obtaining the target position. <sup>a)</sup>Currently at NDA; <sup>b)</sup>Currently at JAMSTEC; <sup>c)</sup>Currently at Univ. of Tsukuba.

11:00

**1aUW3. Synchronized time-reversal focusing with application to virtual array formation, beam forming, and remote imaging.**

Shane C. Walker, Philippe Roux, and W. A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093)

By sampling and time reversing the field from a distant probe source, a time-reversal mirror (TRM) can produce an intense acoustic focus at the probe source location. Extending this idea to a collection of sources, this work demonstrates that a TRM can generate a virtual array at a downrange location, which can act as a remote platform for time delay beam forming and other coherent wavefront techniques useful for imaging in the vicinity of the virtual array. The technique requires no *a priori* information about the environment, only the knowledge of the source fields and a means for synchronizing them. In one promising application the TRM is used to coherently image an extended region around a distant passive array. The geometry of the source array is arbitrary. Recent advances in self-adaptive acoustic Green's function determination may provide opportunities to apply this method without need for a source array. Simulation and experimental results are discussed.

11:15

**1aUW4. Robust wideband time-reversal using time-domain principal component analysis.**

W. J. Higley, Philippe Roux, W. S. Hodgkiss, W. A. Kuperman, Heechun Song, Tuncay Akal (Marine Physical Lab., SIO, UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0238), and J. Mark Stevenson (NATO Undersea Res. Ctr., La Spezia, Italy)

Time reversal (TR), a two-step process, involves time reversing the measured transfer functions from a source and then retransmitting this signal to produce a focus back at the source position. When attempting TR in a fluctuating ocean environment, it is desirable to alter the transmitted signal such that it produces a more robust focus. This has typically been done with work based on frequency domain analysis. In contrast, this

work, using oceanic data measured during FAF05, is broadband and the development is totally in the time domain. Different definitions of robustness give different optimal answers to the question of how to design the most robust transformation of the time-reversed data of the first step in the TR process. Two definitions of robustness are studied, one related to outage probability and another related to average power. For each definition, a different approach to a practical approximation is taken. In the case of outage probability, we iteratively solve a minimax optimization problem. In the case of average power, principal component analysis is performed on time-domain data.

11:30

**1aUW5. Identification and classification of resonant targets with iterative, single-channel time reversal in a noisy environment.**

Zachary J. Waters, R. Glynn Holt, Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215), and Benjamin R. Dzikowicz (NSWC-Panama City, Panama City, FL 32407)

The identification and classification of mines resting on and buried beneath the sea floor is an important and challenging problem. Previous numerical results indicate that iterative time reversal using a single-element transducer increases the signal-to-noise ratio of the backscattered return while simultaneously converging to a narrow-band signal characteristic of the scattering target [D. M. Pierson, "Buried object detection using time-reversed acoustics," Ph.D. dissertation, North Carolina State University 2003]. This technique offers an inexpensive and simple means for maximizing signal level in a noisy or cluttered environment. We investigate this phenomena through experiments performed in a water tank using a 6.35-mm-diam hollow steel sphere and broadband  $Q \sim 1.5$ ) piston transducers operating in the 100-kHz to 2-MHz frequency range. Time-reversal convergence of the interrogation pulse was readily achieved using a variety of broadband (0.1–2 MHz) pulses to initiate the process. Results

indicate that the spectrum of the echo rapidly converges to the dominant resonant mode of the target, with a significant enhancement in signal level even in the presence of broadband noise and coherent clutter returns. [Work supported by Office of Naval Research Award No. N000140610044.]

11:45

**1aUW6. Experiment of time-reversal communication in deep ocean.**

Takuya Shimura, Yoshitaka Watanabe, and Hiroshi Ochi (JAMSTEC/MARITEC, 2-15, Natsushima-cho, Yokosuka-City, 237-0061, Japan, shimurat@jamstec.go.jp)

The basic study on the long horizontal acoustic communication using time-reversal waves have been carried out by the authors, and the method of combining time reversal and adaptive filter was proposed to enhance the communication accuracy and to achieve the communication with a moving source-receiver. This time-reversal technique will be applied to the communication in the deep ocean with an autonomous underwater vehicle, "Urashima," which is being developed in JAMSTEC. The experiment was executed to demonstrate the effectiveness of the proposed method in the area at the depth of 1000 m in Suruga-bay. In the experiment, the 500-Hz projectors were used, and the range between the focus and the time-reversal array (TRA) was 10 km. The measurement system of the TRA side was moored with the underwater winch to compose the virtual array by changing its depth. Unfortunately, the projector of the TRA side had a breakdown, so it was abandoned to execute active time-reversal communication. However, experiments of passive time-reversal communication could be achieved. It was verified that the proposed method makes it possible to assure the communication channel with the TRA that consists of 20 elements. [This work was supported by MEXT KAKENHI 16686050.]

TUESDAY AFTERNOON, 28 NOVEMBER 2006

LANAI ROOM, 1:00 TO 6:15 P.M.

**Session 1pAA**

**Architectural Acoustics: Computer Modeling of Room Acoustics II**

Lily M. Wang, Cochair

*Univ. of Nebraska–Lincoln, 200B Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681*

Shinichi Sakamoto, Cochair

*Univ. of Tokyo, Inst. of Industrial Science, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan*

**Chair's Introduction—1:00**

***Invited Papers***

1:05

**1pAA1. Numerical analysis of sound propagation in rooms using the finite difference time domain method.** Shinichi Sakamoto, Ayumi Ushiyama, and Hiroshi Nagatomo (Inst. of Industrial Sci., The Univ. of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan, sakamo@iis.u-tokyo.ac.jp)

Impulse response is an important quantity for room acoustics. To predict the impulse response, several types of calculation methods based on wave theory, such as the boundary element method (BEM), finite element method (FEM) and finite difference method (FDM), are being investigated. Among them, the authors are exploring application of the finite difference time domain (FDTD) method to calculate room sound fields. Using the method, by which the acoustical quantities at discrete grid points of a closed field are successively solved step-by-step according to vector operations, the impulse responses can be obtained directly in a time domain using little computer memory storage. Furthermore, changes of spatial distributions of room sound fields in time can be obtained easily by storing the sound pressure at grid points and displaying them by animation. In this presentation, several examples of the FDTD analyses of room impulse responses are reported and computer animation of their room sound fields is introduced. In addition, examination of the applicability of prediction of the impulse responses by comparison with *in situ* measurement of a small hall is discussed.

1:25

**1pAA2. Applications of large-scale finite-element sound-field analysis utilizing Architectural Institute of Japan-benchmark platform.** Toru Otsuru, Noriko Okamoto, and Takeshi Okuzono (Faculty of Eng., Oita Univ., 700 Dannoharu, Oita 870-1192, Japan, otsuru@cc.oita-u.ac.jp)

A computational system, namely large-scale finite-element sound-field analysis(LsFE-SFA), has been developed by the authors. It is based on the finite-element method (FEM) with an iterative solver of linear equation that can reduce memory requirement of the system drastically. To show the effectiveness of LsFE-SFA, sound fields in a reverberation room and in a music hall were analyzed. The problem settings are provided by the web site of "AIJ-Benchmark Platform for Computational Acoustics," which is run by a subcommittee of Architectural Institute of Japan (AIJ). First, sound-pressure level distribution maps at about 300 points in the reverberation room computed by LsFE-SFA were compared with those by measurement to show the accuracy of LsFE-SFA to be within 1-dB difference. The frequencies investigated were in 1/3-octave bands with the center frequency ranging from 63 to 250 Hz. In the computations, boundary absorption is modeled by use of normal impedance that was measured *in situ* by use of the authors' newly proposed method, or EA-Noise method. Next, the diffuser shapes on the walls in the music hall were examined by LsFE-SFA to give better diffuseness. The investigations showed the effectiveness of LsFE-SFA.

1:45

**1pAA3. The application of boundary integral equation methods to the numerical calculation of transient waves in rooms.** Yasuhito Kawai (Dept. of Architecture, Kansai Univ., 3-35, Yamatecho 3, Suita, Osaka, 564-8680 Japan)

The boundary integral equations derived from Kirchhoff's formula and its normal derivative form are available for the numerical calculation of transient waves in rooms. The two equations are compared in view of the stability and accuracy in the numerical calculation of responses in a rigid cubic room in which rigorous responses can be obtained by the image method. Some source waves and some problems appear in the numerical implementation such as the singular integrals, and the treatment of absorbent surfaces is also investigated. Finally, it is shown that the boundary integral equation method is a powerful tool to obtain transient waves in rooms.

2:05

**1pAA4. Fogg Art Museum lecture room: A calibrated recreation of the birthplace of room acoustics.** Brian F. G. Katz (Percept. Située, LIMSI-CNRS, BP 133, F91403 Orsay, France, brian.katz@limsi.fr) and Ewart A. Wetherill (28 Cove Rd., Alameda, CA 94502))

At the end of the 19th century, Wallace Clement Sabine undertook the task of correcting the acoustics of the Fogg Art Museum lecture hall. His work on this acoustically difficult (semi-circular, domed, arched) auditorium was the basis for his monumental work on reverberation time and absorption, beginning the science of room acoustics as it is now known. The room underwent various renovations over 70 years, concluding with its demolition in 1973. Just prior to this event, one author was fortunate to have measured several room impulse responses. Through the use of architectural plans and the few existing photos, a computational room acoustical model was created. This model was calibrated using the historically important measurements as a reference. The geometrical model incorporates the room's various architectural phases. Using computer auralizations, the room is now accessible in acoustical terms, with the ability to "perceive" an educated rendering of the acoustics of the room as Sabine found it in 1895. The results of this study, particularly the geometrical model, are intended for public use and are made generally available for room acoustics students and researchers who chose to follow in Sabine's footsteps ([www.catt.se/FoggArtMusum.htm](http://www.catt.se/FoggArtMusum.htm)). Portions of this presentation were presented at Forum Acousticum 2005.

2:25

**1pAA5. Auralization of an orchestra using multichannel and multisource technique.** Michelle C. Vigeant, Lily M. Wang (Architectural Eng. Prog., Univ. of Nebraska—Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, mvigeant@mail.unomaha.edu), and Jens Holger Rindel (Tech. Univ. of Denmark, DK-2800 Kgs. Lyngby, Denmark)

Previous work has shown the importance of including source directivity in computer modeling for auralizations. A newer method to capture source directivity in auralizations is the multichannel technique, which uses multichannel anechoic recordings. In this study, five-channel anechoic recordings were obtained for every orchestral part of two symphonies at the Technical University of Denmark. Five-channel auralizations were then created for each instrument, located at its typical position on-stage in a concert hall, by convolving five impulse responses from sources that each represent approximately one-fifth of a sphere with the corresponding anechoic recording channel and mixing these together. The multichannel auralizations from each instrument were subsequently combined to produce a full multichannel auralization of the entire orchestra. The results from listening tests will be presented, illustrating the perceived changes in realism, source width and depth, and source separation between the output of this full orchestral mix of multichannel auralizations and other auralization methods, including the following: (i) a mix of single-channel auralizations from each instrument, and auralizations made with a single channel orchestral anechoic recording using (ii) a surface source and (iii) a single omni-directional source. [Work supported by the National Science Foundation.]

2:45

**1pAA6. Prediction of scattering coefficients for use in room-acoustic simulation.** Wolfgang Ahnert, Stefan Feistel, and Steffen Bock (Ahnert Feistel Media Group, Berlin, Germany)

While absorption coefficients have become easily available and published worldwide, computer modeling of room acoustics suffers until today under the lack of data for scattering surfaces. As a result, in most cases rules of thumb have to be used. In this work a new approach of predicting scattering coefficients is introduced. Based on the geometrical parameters of the surface, the complex reflection properties are calculated utilizing an elementary-wave method. In this respect, computed data are compared afterwards with measurement data and its applicability for use in acoustic simulation is investigated.

3:00

**1pAA7. A comparison of computer modeling auralization and binaural room recordings for a multipurpose high school auditorium utilizing an orchestral shell, overhead panels, and clouds..** Ron Freiheit (Wenger Corp., 555 Park Dr., Owatonna, MN 55060)

The development of auralization techniques has provided a new method to demonstrate potential acoustical benefits of changes in design or the application of acoustic treatments in specific spaces. It's useful to compare the results from the auralization process with actual binaural recordings in the physical space to discover similarities and differences. A typical high school proscenium-style, multipurpose auditorium was modeled, including an orchestral shell, overhead panels, and clouds. Auralizations were made of the auditorium at various locations in the audience area and on stage, with and without the shell and overhead panels, in place. These were compared with binaural recordings made at the same locations within the auditorium. Auralization metrics will be compared with the binaural recordings, highlighting the changes made with the shell and overhead panels in place, along with the similarities and differences between the two methods.

3:15–3:30 Break

3:30

**1pAA8. Optimal dispersion/dissipation-error finite-difference method for room impulse response simulation.** Yusuke Naka (Dept. of Aerosp. and Mech. Eng., Boston Univ., 677 Beacon St., Boston, MA 02215, ynaka@bu.edu), Assad A. Oberai (Rensselaer Polytechnic Inst., Troy, NY 12180), and Barbara G. Shinn-Cunningham (Boston Univ., Boston, MA 02215)

The finite difference method is suitable for the calculation of acoustic impulse responses because of its computational simplicity and efficiency. However, for room acoustics applications, where reverberation makes it necessary to calculate the room impulse response and acoustic energy over a long duration, most standard finite-difference formulations are not practical because of the accumulation of dispersion and dissipation errors. In order to calculate very long impulse responses, a new, efficient, high-order finite-difference method is developed in this study. The new method is optimized to minimize the total dispersion and dissipation errors and hence the error accumulated in each time step resulting from discretization in both space and time. The error of the new method is significantly reduced so that the impulse response can be calculated over the whole reverberation time (about 1 to 3 s), using a relatively coarse spatial (6 grid points per wavelength) and a moderate temporal (CFL=1.0) resolutions. [Work supported by AFOSR.]

3:45

**1pAA9. Temporal decay of broadband sound fields in enclosures using a boundary element method.** Donald Bliss, Ben Manning (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708, donald.bliss@duke.edu), and Jerry Rouse (Sandia Natl. Labs, Albuquerque, NM 87185-0847)

A method to predict time-dependent behavior of sound fields in enclosures with diffuse reflection boundaries is described. A boundary element formulation utilizes broadband energy-intensity boundary sources, including propagation time delays. Temporal behavior is expressed in terms of a higher-dimensional eigenvalue/eigenmode problem, with source strength distributions expressed as eigenmodes. A temporal interpolation scheme is used to reexpress the actual delays as a set of integer-multiple delays. The interpolation scheme produces a discrete set of primary eigenvalues and eigenvectors, the total number of which is an integer multiple  $M$  of the number of panels  $N$ . The value of  $M$  is set by the number of interpolation subdivisions of the enclosure wave transit time. For decay from steady-state conditions, the initial value problem involves specifying the initial panel strengths, and  $M-1$  time derivatives of each panel strength set to zero. In temporal decay from steady state, solutions exhibit rapid short-time spatial redistribution of energy, followed by long-time decay of a predominant spatial eigenmode. Long-time decay depends on the behavior of the most slowly decaying eigenmode and the relative source-panel strengths do not depend on initial conditions. Short-time adjustment and decay depends on initial source characteristics and the distribution of wall absorption.

4:00

**1pAA10. Boundary element prediction of steady-state broadband interior sound fields using absorption-based scaling.** Donald Bliss (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708, donald.bliss@duke.edu) and Jerry Rouse (Sandia Natl. Labs, Albuquerque, NM 87185-0847)

Energy-intensity boundary elements are used in combination with the method of absorption-based scaling to predict steady-state broadband sound fields in enclosures with diffuse reflection boundaries. The wall absorption is expressed in terms of an overall absorption parameter—the spatially averaged value—and spatial variations around this mean. Boundary element strengths are expressed in a power series of the overall absorption, treated as a small parameter, thereby giving a separate problem at each order. The first problem has a uniform mean-square pressure proportional to the reciprocal of the absorption parameter. The second problem gives mean-square pressure and intensity distributions that are independent of the absorption parameter and are primarily responsible for the spatial variation of the reverberant field. This problem depends on the location of sources and the spatial distribution of absorption, but not overall absorption. Higher order problems proceed in powers of the absorption parameter, as corrections to the primary spatial variation. Boundary element solutions are obtained for each order, and the simplest model for spatial variation gives fairly accurate predictions. The method gives physical understanding of the causes of the spatial variation of enclosure sound fields and provides specific insights into the behavior and design of acoustic spaces.

4:15

**1pAA11. Comparison of sound fields in rooms computed by finite-element method and by cone beam method.** Noriko Okamoto, Toru Otsuru, Reiji Tomiku (Oita Univ., 700 Dannoharu, Oita-city, Oita 870-1192, Japan, nokamoto@cc.oita-u.ac.jp), Takeshi Okuzono, Kouta Isobe (Oita Univ., Oita, Japan), and Hiroshi Furuya (Kyushu Kyoritsu Univ., Fukuoka 807-8585, Japan)

Sound fields in realistic rooms were simulated by different kinds of computational methods, and comparisons were conducted. One is based on the wave acoustics and the other is based on the geometrical acoustics, i.e., the finite-element method (FEM) and the conical beam method

(CBM). As for the software applied here, the authors have developed their own FEM code, namely LsFE-SFA, to conduct FEM computations, and they employed commercial software, i.e. RAYNOISE, to conduct CBM computations. First, theoretical basis of LsFE-SFA was given briefly. Next, sound-pressure distributions in a reverberation room with the volume of 165 cubic meters were computed by both methods to compare the agreement with measured sound-pressure levels. The boundary conditions due to the sound absorption were changed to examine the relation between the absorbent condition and the agreements. The agreement between FEM and measurement were found to be better throughout the examination. Next, both FEM and CBM were applied onto the analysis of large-scale sound fields in a multipurpose hall with the volume of 37 000 cubic meters to discuss the agreement and difference in detail.

4:30

**1pAA12. Absorption and aperture size effects on subjective response to complex computer modeled coupled volume systems.** David T. Bradley (BRC Acoust. & Technol. Consulting, 1741 1st Ave. S, Ste. 401, Seattle, WA 98134, dbradley@brcacoustics.com) and Lily M. Wang (Univ. of Nebraska-Lincoln, Omaha, NE 68182-0681)

A typical coupled volume concert hall consists of a large, absorbent main volume linked to a smaller, reverberant auxiliary space through a series of acoustically transparent openings. This type of space can exhibit a nonexponential sound decay, a phenomenon that is referred to as the double slope effect (DSE). An imagined, realistic coupled volume concert hall has been computer modeled, and room impulse responses of the space have been obtained. Several configurations of the hall were conceived in which the absorption and aperture size were modified. Paired-comparison tests were conducted to determine listener preference for the sound fields developed from auralizations of the configurations. The results from this subjective testing were statistically analyzed using multidimensional scaling methods. The analysis showed that listener preference increased with decreasing DSE, with highest preference levels for low and medium DSE. These results will be presented and contrasted with objective parameters calculated from the room impulse responses of the computer-modeled configurations.

4:45

**1pAA13. Numerical analysis for attenuation due to sound absorption in small and large closed spaces using finite-element method.** Takao Yamaguchi (Dept. of Mech. Systems Eng., Gunma Univ., 1-5-1 Tenjin-cho, Kiryu, Gunma 376-8515, Japan), Junichi Kanazawa (Musashino Damping Technol., Tachikawa 190-0002, Japan), and Katsuji Akamatsu (Machinery Acoust., Nada-ku, Kobe 657-0015, Japan)

A finite element for media having sound absorption is derived and a code is developed. For this element, an unknown parameter is pressure, and both effective density and bulk modulus of the media have complex quantity to represent damped sound fields. This model for the media having sound absorption was adopted by H. Utsumo's research group for their boundary element formulation [Utsumo *et al.*, AIAA J. **28**(11), 1870–1875 (1990)]. Firstly, a three-dimensional sound field in a small closed space ( $306 \times 255 \times 204 \text{ mm}^3$ ), containing a sound-absorbing material, is calculated using the finite element method. Validity of the calculated response is verified by comparison with the corresponding experimental result carried by H. Utsumo. Secondly, a two-dimensional sound field in relatively large closed space (about  $7 \times 3 \text{ m}^2$ ) is calculated varying damping of its medium. The shape of the large space is rectangle, and the rectangular space has a roof. Distribution of sound pressure is calculated when the velocity of the roof is given as white noise. Under the low damping condition for the medium, distribution of sound pressure is almost uniform like the diffused sound field. Under the high damping condition for the medium, sound pressure become larger when the observation point is closer to the roof.

5:00

**1pAA14. Real-time audio rendering system for virtual reality.** Tobias Lentz, Dirk Schroeder, Gottfried K. Behler, and Michael Vorlaender (Inst. fuer Technische Akustik, RWTH Aachen Univ., D-52056 Aachen, Germany)

Systems of virtual acoustics are more and more coming into the focus of research and application. Creating a virtual sound scene with spatially distributed sources requires a technique for adding spatial cues to audio signals and an appropriate reproduction. A real-time audio rendering system is introduced that combines room acoustics, dynamic crosstalk cancellation, and multitrack binaural synthesis for virtual acoustical imaging. The room acoustics component takes into account the full up-to-date algorithmic approach of specular reflections and scattering, but with real-time processing. The real-time performance of the software was reached by introduction of a flexible framework for the interactive auralisation of virtual environments. The concept of scene graphs for the efficient and flexible linkage of autonomous-operating subscenes by means of so-called portals has been incorporated into the existing framework, combined with an underlying BSP-tree structure for processing geometry issues very fast. Using this framework enables a significant reduction of computation time for both applied algorithms (deterministic image sources and a stochastic ray tracer). This enables the simulation of indoor or outdoor spatial distributed sources and, in addition to that, near-to-head sources for a freely moving listener in room-mounted virtual environments without the use of headphones.

5:15

**1pAA15. Effects of peg's positions of low-registered strings on acoustic gain by a stage riser: Theoretical model analysis.** Shinsuke Nakanishi (Dept. of Integrated Architecture, Hiroshima Intl. Univ., Hirokoshingai, Kure, Hiroshima 737-0112 Japan, s-nakani@it.hirokoku-u.ac.jp)

This paper theoretically analyzes acoustic gain by a stage riser, which has been observed when playing the low strings, e.g., a violoncello and a double bass, on the riser in concert auditoria [Beranek *et al.*, J. Acoust. Soc. Am. **36**, 1247–1262 (1964)]. The previous paper reported that “the tone quality increased in richness with the risers, particularly when the pegs of the instruments were placed about 4 in. from the edge” as musician's comments. The author discusses effects of the peg's positions of the low strings on the acoustic gain, through numerical examples of theoretical model analysis. The solution of a sound field is obtained when playing the low string on the riser placed on the rigid floor, which receives a spherical incident wave and mechanical excitation force from the low string. Numerical examples show that peg's positions of the low string affect peaks and dips of the acoustic gain, which are caused by eigenvibrations of a top plate of the riser. And, the acoustic gain becomes a positive value at more frequencies as the standing position of the low string shifts towards the riser's edge.

5:30

**1pAA16. The program battle at Ganz Hall.** Candice Tillery and Jay Pleckham (Columbia College Chicago, 33 E. Congress, Chicago, IL 60605, listen\_42@sbcglobal.net)

Current modeling programs offer a wide range of analysis and prediction tools, as well as detailed modeling capabilities and user friendly formats. However, these tools are only useful if they can do their job correctly. The accuracy of the program, its cost, and its relevance to the environment in question are all vitally important when selecting software for a project. This paper addresses these issues through a comparison of two of today's most popular modeling programs: CARA and EASE. An acoustical analysis was performed on Roosevelt University's Ganz Hall, which seats 195 people and is used primarily for student recitals. Data were acquired concerning the hall's reverberation and frequency response. Numerous calculations were performed to determine the room's musical clarity, speech clarity, and early decay time. The hall was then modeled in the two programs in question, and the results were compared to determine each program's accuracy.



**1pAA17. Parallel computing on the sound fields by bread slicing method.** Yasuhiro Tokita (Waseda Environ. Res. Inst., Waseda Univ., 1011 Okuboyama, Nishitomida, Honjou, Saitama, 367-0035, Japan), Yasuhiro Oikawa, and Yoshio Yamasaki (Waseda Univ., Honjou, Saitama 367-0035, Japan)

Numerical analyses by the explicit scheme of the finite difference method greatly reduce the number of known variables that are needed to solve an unknown variable in comparison with the implicit methods. This fact means that the amount of communications decreases among each node of the computer cluster on parallel computing. Generally, on the parallel computing of the distributed memory system, the communicating time occupies most of the total calculating time. Therefore, the reduction to the amount of the communication directly causes the reduction of the calculating time. The parallel computing by the explicit scheme of the finite difference method can enable the building up of high-performance algorithms. In order to reduce the amount of communication among each node, one has proposed a method to divide the field of numerical analyses. It is called the "bread slicing method." With this method, the fields must

be divided into pieces parallel to the section that has the least area. The results showed that the method in this study enabled the total amount of communication in the computer cluster to be reduced and the geometries of communication to be made simpler.

6:00

**1pAA18. Theoretical analysis of statistical properties of the sound-pressure field and sound energy density field in rectangular enclosures.** Buye Xu and Scott D. Sommerfeldt (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

The sound-pressure field and sound energy density (ED) field for rectangular enclosures are obtained theoretically by solving the Helmholtz equation and Euler's equation with low-energy absorption boundaries. Based on those results, both analytic and numerical methods are applied to compute some statistical properties. The results of the statistical analysis show that the spatial variation of the ED field is generally much less than that of the pressure field in reverberant enclosures. In addition, a modified ED quantity is developed and will be discussed.

TUESDAY AFTERNOON, 28 NOVEMBER 2006

KOHALA/KONA ROOM, 1:30 TO 5:30 P.M.

### Session 1pAB

## Animal Bioacoustics and Underwater Acoustics: Remote Monitoring of Biological Sounds II

Robert Gisiner, Chair

*Office of Naval Research, 875 N. Randolph St., Arlington, VA 22203*

### Contributed Papers

1:30

**1pAB1. Humpback whale social sounds: Sources levels and response to playback.** Michael J. Noad, Rebecca Dunlop (School of Veterinary Sci., Univ. of Queensland, St. Lucia, Qld, 4072, Australia, mnoad@uq.edu.au), Douglas H. Cato (Defence Sci. and Tech Org., Pyrmont, NSW 2009, Australia), Dale Stokes (Scripps Inst. of Oceanogr., La Jolla, CA 92037), Patrick Miller, and Nicoletta Biassoni (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Humpback whale "social sounds" appear to be used in communication when whales interact but they have received little study in comparison to the song. During experiments as part of the Humpback whale Acoustics Research Collaboration (HARC), whales migrating past the study site on the east coast of Australia produced a wide range of social sounds. Whales were tracked visually using a theodolite and singers were tracked acoustically using an array of five widely spaced hydrophones. Source levels of social sounds were calculated from the received level of the sounds, corrected for measured propagation loss. Playbacks of social sounds were made using a J11 transducer and the consequent reactions were recorded in terms of the change in direction of the migrating whales in relation to the playback position. In one playback, a DTAG was placed on a female with calf. Playback of social sounds resulted in significant changes in the course of the migrating whales, in some cases towards the transducer while in others it was away from the transducer. From the estimates of source levels it is possible to assess the effectiveness of the playback and the range of influence of social sounds. [Work supported by ONR and DSTO.]

1:45

**1pAB2. Megapclicks: Tagged humpback whales produce click trains and buzzes during foraging.** Alison K. Stimpert, Whitlow W. L. Au (MMRP/Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734, stimpert@hawaii.edu), Mark P. Johnson (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), David N. Wiley (Stellwagen Bank Natl. Marine Sanctuary, Scituate, MA 02066, USA), and Roland Arsenault (Univ. of New Hampshire, Durham, NH 03824)

Mysticetes and odontocetes differ in many morphological and behavioral aspects related to feeding. Odontocetes echolocate using clicks to localize prey, while the mechanism through which mysticetes orient and find food is unknown. In this study, DTAGs were deployed on four humpback whales (*Megaptera novaeangliae*) observed feeding off the north-eastern United States. From two of the four tagged whales, bouts of clicks produced with a temporal pattern resembling odontocete echolocation were recorded and 102 bouts of megapclicks were identified. Spectrally, megapclicks were unlike previously reported humpback sounds or odontocete clicks, and also had much lower amplitudes (peak frequency ~1700 Hz, RL  $148 \pm 6$  dB Re: 1 uPa peak). Interclick intervals (ICI) were bimodally distributed (peaks at 25 and 80 ms), with shorter ICIs occurring at the end of megapclick bouts, in a pattern similar to an odontocete terminal buzz used in foraging. Behavioral sensor data supported a possible foraging function for megapclicks. All megapclick bouts occurred during nighttime hours, and the terminal buzz was associated with a sharp body roll of greater than 70 deg. Low source levels, among other characteristics, contraindicate an echolocation function; however, further analysis and research is required to distinguish among this and other hypotheses for sound function.

2:00

**1pAB3. Temporal, geographic, and density variations in the acoustic activity of snapping shrimp.** Marc O. Lammers, Sara Stieb, Whitlow W.L. Au, T. Aran Mooney (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734), Russell E. Brainard, and Kevin Wong (Pacific Islands Fisheries Sci. Ctr., Honolulu, HI 96814)

The sounds produced by snapping shrimp (genus *Synalpheus* and *Alpheus*) are the main contributors to the ambient sound field of shallow-water tropical and subtropical marine habitats worldwide. Despite this, relatively little is known about the temporal and spatial variability of snapping shrimp activity. To investigate this issue further, ecological acoustic recorders (EARs) were utilized to obtain time-series records of the ambient sound field from multiple locations in Hawaii and American Samoa. Strong diel trends marked by sharp increases in acoustic activity during crepuscular periods were observed at all sites. However, sites varied greatly both in absolute sound levels and relative fluxes between day and night. In addition, fluctuations in snapping shrimp sound levels were found to track lunar cycles, with higher levels occurring during the new moon. To further investigate the reason for the higher levels observed during crepuscular periods, individual shrimp were localized using a three-hydrophone line array on both a coral reef and a mud and algae habitat in Kaneohe Bay, Hawaii. The density of acoustically active shrimp was found to be higher during the sunset period than during the day or at night.

2:15

**1pAB4. An automated approach to detecting pulsive gray whale (*Eschrichtius robustus*) vocalizations within acoustic data to estimate call rate changes over time in Laguna San Ignacio.** Melania Guerra, Aaron M. Thode (Marine Physical Lab, SIO UCSD, 9500 Gilman Dr., La Jolla, CA 92037-0238), Marilyn E. Dahlheim (Natl. Marine Mammal Lab, Seattle, WA 98115), Jorge Urban (UABCS, La Paz, Mexico), and Amanda M. Koltz (Cornell Univ., Ithaca, NY 14853)

Acoustic recordings of type 1b gray whale calls were made in San Ignacio Lagoon, during the wintering months of 2005 and 2006 when these animals were present to breed and raise calves. A collection of horizontal, bottom-mounted arrays containing two or three elements monitored the mouth of the lagoon over consecutive days, targeting vocalizations below 1 kHz. Because these pulsive calls are highly variable in the timing and number of pulses, standardized detection methods involving matched-filtering or spectrogram correlation have been unsuccessful. A sequence of automated procedures programed in MATLAB and ISHMAEL was implemented to assist with the detection of calls within the dataset. Simple manual recognition based on human judgment was also performed on the shorter time series to serve as an independent check. Results for hourly call rates were computed for both seasons individually. A similar such acoustic dataset had been previously collected by Dahlheim (1987) in the years 1982–84 and analyzed manually. This dataset would be digitized and examined with the new automated analyses in order to compare changes in average call rates over the two-decade interval. A rough relationship with animals present can be drawn, as census numbers are available for both periods.

2:30

**1pAB5. Characteristics of the deep scattering layer in the Gulf of Mexico as they relate to sperm whale diving and foraging behavior.** Alyson J. Azzara (Dept. of Oceanogr., Texas A&M Univ., College Station, TX 77843-3146, alygirl@tamu.edu), Aaron Thode (Scripps Inst. of Oceanogr., 9500 Gilman Dr., San Diego, CA 92109), Tom Norris (Sci. Applications Intl. Corp., San Diego, CA), Kelly Benoit-Bird, and Douglas C. Biggs (Oregon State Univ., Corvallis, OR 97331)

In June 2005 we used a 38-kHz ADCP and two Simrad fishery echosounders to collect backscatter data and passive acoustic array recordings of sperm whale sounds, as part of the Sperm Whale Seismic Study (SWSS). Dive profiles and presumed foraging depths of the whales were derived using a 3-D tracking technique [see abstract by Thode and Norris (this conference)]. Relationships are examined between dive profiles and acoustic backscatter intensity measurements from the main deep scattering

layer (DSL) at a daytime depth of 400–600 m. Passive detections of creaks, believed to be associated with prey acquisition, are merged with dive profiles to provide insight between mid-water backscatter returns and foraging effort. Results to date indicate whales diving in three depth ranges: deeper than 800 m, between 500 and 800 m, and less than 500 m. Dive patterns indicate some whales dive through the main DSL into deeper secondary layers seen by the 38-kHz ADCP. There seems to be no difference in dive patterns between night and day. Calibration analysis of ADCP backscatter versus 70-kHz Simrad backscatter shows a positive functional relationship, with 4 units of relative acoustic backscatter intensity equal to 1 dB of Sv.

2:45

**1pAB6. Bottom-mounted hydrophones used to investigate sperm whale click and swim characteristics.** Eva-Marie Nosal and L. Neil Frazer (Dept. of Geology and Geophys., Univ. of Hawaii, 1680 East-West Rd., Honolulu, HI 96822)

A method that relies solely on the time difference between direct and reflected arrivals (DRTD) on individual hydrophones was previously used to localize a single sperm whale [E.-M. Nosal and L. N. Frazer, Appl. Acoust., Special Ed.: Detection and Localization of Marine Mammals using Passive Acoustics (in press)]. Regular clicks in recordings from five bottom-mounted hydrophones in the AUTECH range were used in the localization. In this study, the DRTD method is extended by incorporating time differences (between hydrophones) of direct arrivals to refine the track of the whale and to determine the time of generation of individual clicks. Clicks, which have a multi-pulse structure, are analyzed in detail to determine changes in interpulse intervals (IPI) with time. The resulting time, location, and IPI information are used to investigate swim patterns and directionality characteristics of regular sperm whale clicks. [Work supported by ONR. Data provided by NUWC for the 2nd International Workshop on Detection and Localization of Marine Mammals using Passive Acoustics.]

3:00

**1pAB7. 3-D tracking of sperm whale dive profiles from a towed hydrophone array in ray-refractive environments.** Aaron Thode (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238, thode@mpl.ucsd.edu), Tom Norris (Sci. Applications Intl. Corp., San Diego, CA), Alyson Azzara, Ann Jochens, and Doug Biggs (Texas A&M Univ., College Station, TX 77843)

During the past 3 years the U.S.-based Sperm Whale Seismic Study (SWSS) has supported the development of a mobile three-dimensional passive acoustic tracking system. Acoustic data were recorded on a single four-element wide-aperture hydrophone array, towed at 80–100-m depth. The array was subdivided into forward and rear two-element subarrays, separated by 200-m aperture. To fix a location of an impulsive sound, three quantities are measured: the time difference between direct-path arrivals on the forward and rear subarrays, the time difference between the direct and surface-reflected paths on the rear subarray, and the acoustic bearing measured on the rear subarray. These quantities, combined with independent measurements of hydrophone depths and cable inclination, are postprocessed using a MATLAB GUI into range-depth position fixes by implementing an efficient numerical procedure that uses precomputed look-up tables to account for ray-refraction effects caused by depth-dependent sound-speed profiles. The predicted time difference between the direct and surface-reflected paths on the forward subarray is compared with the measured data to adjust for array tilt and depth measurement error. A by-product of this processing is a semiautomated detector to detect distinctive creak sounds. [Work supported by the U.S. Minerals Management Service and the Industry Research Funders Coalition.]

**1pAB8. Direct underwater visual and acoustic observations of echolocation behavior of sperm whales around a longline.** Aaron Thode (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla CA 92093-0238), Jan Straley, Kendall Folkert, Victoria O'Connell (Univ. of Alaska Southeast, Sitka, AK 99835), Linda Behnken (Alaska Longline Fishermens Assoc., Sitka, AK 99835), and Amanda Koltz (Cornell Univ., Ithaca, NY 14853)

Sperm whales (*Physeter macrocephalus*) have learned to approach and depredate fishing vessels longlining for black cod in the Gulf of Alaska. In May 2006, off Sitka, Alaska, a video camera was placed inside an underwater housing with an external hydrophone mount, attached to a longline that had been partially hauled to the surface, and lowered between depths of 90 and 120 m, with black cod (*Anoploploma fimbria*) attached 2 to 4 m above the camera as bait. The camera was mounted facing towards the surface to exploit the ambient light and to eliminate the need for artificial lighting. An additional hydrophone was deployed at 17-m depth during the 40–60-m deployments, and a third autonomous acoustic recorder was usually mounted 1.3 m below the camera. During one encounter a whale investigated the line, producing characteristic creak sounds that were recorded on the three hydrophones, and which were subsequently time-aligned using vessel noise as a reference, in order to permit acoustic tracking. During the second encounter a whale interacted with two fish within 4 m of the camera, but only the camera hydrophone was available. Preliminary analysis of the echolocation behavior displayed by both animals is presented. [Work sponsored by the National Geographic Society and the North Pacific Research Board.]

### 3:30–3:45 Break

### 3:45

**1pAB9. Three-dimensional passive acoustic tracking of sperm whale behavior with widely separated vertical arrays in the Gulf of Alaska.** Christopher O. Tiemann (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713), Aaron M. Thode (UCSD, La Jolla CA 92093-0238), Jan Straley, Kendall Folkert (Univ. of Alaska Southeast, Sitka, AK 99835), and Victoria O'Connell (Alaska State Fish and Game, Sitka, AK 99835)

The Southeast Alaska Sperm Whale Avoidance Project (SEASWAP) was started in 2002 to study the increasing depredation encounters between sperm whales and longline fishermen near Sitka, Alaska. In 2005, the acoustic component of the study included passive acoustic data from four autonomous recorders attached to the anchor lines of two longlines spaced 3 miles apart, essentially turning them into temporary vertical arrays and allowing the monitoring of sperm whale vocal activity during fishing operations. Although the use of multipath arrival information from recorded sperm whale clicks has been a standard procedure for range-depth tracking, with the added knowledge of azimuthally dependent bathymetry relative to an array, a three-dimensional track of whale motion can be obtained from the same multipath information. The evolution of multipath arrival patterns is matched to range-, depth-, and azimuth-dependent modeled arrival patterns to generate an estimate of whale motion. The use of an acoustic propagation model can also account for all waveguide physics such as interaction with range-dependent bathymetry and ray refraction. Acoustic tracking results concurrent with whale encounters will be presented and compared to visual sightings.

**1pAB10. Variation in sound production in relation to behavior of bottlenose dolphins (*Tursiops aduncus*).** Michelle Lemon (Marine Mammal Res. Group, GSE, Macquarie Univ., Sydney, Australia, Defence Sci. and Tech Org., Pyrmont, Australia, and michelle.lemon@dsto.defence.gov.au), Douglas H. Cato (Defence Sci. and Tech Org., Pyrmont, New South Wales, Australia), Tim P. Lynch (Jervis Bay Marine Park, Huskisson, New South Wales, Australia), and Robert G. Harcourt (Macquarie Univ., Sydney, Australia)

Dolphins are social animals that inhabit an opaque environment where visual cues are often limited, thus acoustic signals play an important role in inter- and intragroup communication. Phonations may be used to signal conspecifics, in navigation, or in the detection of prey, and may vary with behavior, environmental conditions, or geographic location. Variation in the acoustic and surface behavior of two geographically separated and genetically differentiated populations of bottlenose dolphins in Southeastern Australia were investigated over a 3-year period. Two hypotheses were tested: (1) vocal repertoire varies with behavior; and (2) variation in vocal repertoire may have arisen due to geographic isolation. Overall rate of sound production by the dolphins was similar in each field site but varied significantly with behavioral activity. Call rates were highest with behavior that involved increased arousal, such as socializing and foraging. Whistle rates and the duration of echolocation click bouts were similar for each location, however, burst-pulsed sounds varied significantly with behavioral state and between the field areas. This difference may relate to differences in social behavior between populations and be driven by variability in social structure and environmental conditions. [Research supported by Australian Geographic, PADI Foundation and Macquarie University.]

### 4:15

**1pAB11. Observation studies based on echolocation click signals of Irrawaddy dolphins in Chilika lagoon.** Rajendar Bahl (Ctr. for Appl. Res. in Electron., IIT Delhi, Hauz Khas, New Delhi, India, rbahl@care.iitd.ernet.in), Tamaki Ura, Tomoki Inoue, Harumi Sugimatsu, Takeshi Sakamaki (Univ of Tokyo, Meguro, Tokyo 153-8505, Japan), Junichi Kojima (KDDI R&D Labs. Inc., Saitama 356-8502, Japan), Tomonari Akamatsu (Natl. Res. Inst. of Fisheries Eng., Ibaraki 314-0408, Japan), Hideyuki Takahashi (System Giken Co. Ltd., Kanagawa 253-0085, Japan), Ajit Pattnaik, Muntaz Khan (Chilika Dev. Authority, Orissa, India), Sandeep Behera (Freshwater and Wetlands Prog., New Delhi 110003, India), and Sudhakar Kar (Off. of Princip. CCF (Wildlife) Chief Wildlife Warden, Bhubaneswar 751012, India)

A resident population of Irrawaddy dolphins, *Orcaella brevirostris*, in Chilika lagoon in eastern India has been recently monitored based on their echolocation clicks. The free-ranging animals were monitored over several days in their natural habitat. A horizontal three-hydrophone line array was used for click data collection and off-line analysis of click trains to obtain underwater movements in shallow water of typical depth less than 3 m. The lagoon waters being placid provided a good surface echo that has been used to obtain depth information of the individual click sounds. This paper augments previous data on the broadband click characteristics [S. M. Van Parijs, G. J. Parra, and P. J. Corkeron, J. Acoust. Soc. Am. **108**, 1938–1940 (2000)] and also provides a glimpse into the individual behavior of the dolphins. An estimate of the click lag time of the animal's biosonar has also been obtained for the first time for a free-ranging animal.

### 4:30

**1pAB12. On the use of the Teager energy operator for the detection of clicks in free-ranging dolphins.** Marie A. Roch (Dept. of Comput. Sci., 5500 Campanile Dr., San Diego State Univ., San Diego, CA 92182-7720), Melissa S. Soldevilla, Jessica C. Burtenshaw, E. Elizabeth Henderson, Sean M. Wiggins, and John A. Hildebrand (Univ. of California at San Diego, La Jolla, CA 92093-0205)

The detection of individual clicks produced by odontocetes is useful for many bioacoustic tasks such as the analysis of interclick interval, classification of odontocete species, etc. Automation of this task is required for

the large data sets gathered by remote instruments that can collect terabytes of information over short time periods. A method is presented that locates individual clicks using the Teager energy operator, which is a time-domain metric designed to measure fast variations in signal energy. The Teager energy of a high-pass filtered signal is analyzed by a rule-based classifier that determines the location of individual clicks. The click start can be determined with very high accuracy. Click ends are reasonably well characterized, but the ends of clicks are difficult to determine, even by human experts, due to the influence of external factors such as acoustic reflections. The proposed method is compared with existing techniques and is demonstrated on the calls of several species of free-ranging odontocetes resident in the Southern California Bight and the Gulf of Mexico: bottlenose dolphins (*Tursiops truncatus*), short and long beaked common dolphins (*Delphinus delphis*, *Delphinus capensis*), Pacific white-sided dolphins (*Lagenorhynchus obliquidens*), and Risso's dolphins (*Grampus griseus*).

4:45

**1pAB13. The acoustic recording of bottlenose dolphin (*Tursiops truncatus*) phonations: Problems and contributions.** Carmen Bazúa-Durán and Axel Herrera-Hernández (Lab. Acústica, Facultad de Ciencias, UNAM, Ciudad Universitaria, 04510 México, D.F., México, bazua@servidor.unam.mx)

Bottlenose dolphin sounds have been studied for decades; however, more needs to be investigated. This work demonstrates that results obtained in wild dolphin studies greatly depend on the recording system used. Dolphin phonations were simultaneously recorded with two systems, a narrow-band (<24 kHz) and a broadband (<180 kHz), and were analyzed using spectral characteristics (narrow-band: 512 points, Hamming window, 50% overlap; broadband: 2048 points, Hamming window, 50% overlap) to determine the acoustic repertoire as echolocation click trains (EC), whistles (W), and burst pulses (BP). As expected, the number of BP and EC recorded was considerably larger with the broadband system than with the narrowband; whereas the number of W recorded was greater with the narrowband system than with the broadband (it was expected to obtain the same number of W with both systems). As W are less intense than BP or EC, when gain is adjusted for recording BP or EC, the recording system is not able to record the less intense W. Therefore, it is not possible to capture the same number of dolphin phonations with the same recording system. One should use at least two different recording systems to adequately describe dolphin phonations. [Work supported by CONACYT-Campeche and PAPIIT-UNAM.]

5:00

**1pAB14. Acoustic monitoring of resident, offshore, and transient killer whales off the Washington coast.** Erin M. Oleson, Sean M. Wiggins, and John A. Hildebrand (Scripps Inst. of Oceanogr., Univ. of California San Diego, 9500 Gilman Dr., #0205, La Jolla, CA 92093, eoleson@ucsd.edu)

Three acoustically distinct populations of killer whales representing each of the known ecotypes (resident, offshore, transient) were recorded in the Summer–Fall of 2004 off the southern Olympic Coast of Washington. Two high-frequency acoustic recording packages (HARPs) continuously recording at 80-kHz sample rate were deployed to assess the seasonal occurrence of vocal odontocetes in this region. From mid-July to early-October the population-specific discrete calls of killer whales were heard on 8 days and were classified to population by Volker Deecke (UBC) and John Ford (DFO-Canada) using an acoustic ID catalogue. West Coast Transient killer whales producing calls of the California dialect were heard on three occasions from August through October. Offshore killer whales were heard twice in August–September, and Northern Resident killer whales were heard once in August. Although Northern Resident killer whales have been extensively studied within Puget Sound and coastal British Columbia, they have been visually sighted only once off the northern Olympic Peninsula, making their detection at this offshore southerly location unique. Endangered Southern Resident killer whales were not heard at this site from July–October. Analysis of year-round data from a site further offshore is underway. [Funded by Chief of Naval Operations-N45.]

5:15

**1pAB15. High-frequency Acoustic Recording Package (HARP) for long-term monitoring of marine mammals.** Sean M. Wiggins, Chris Garsha, Greg Campbell, and John A. Hildebrand (Univ. California-San Diego, 9500 Gilman Dr., 0205, La Jolla, CA 92093-0205, swiggins@ucsd.edu)

Advancements in low-power and high-data-capacity computer technology during the past decade have been adapted to autonomously record sounds from marine mammals over long time periods. Acoustic monitoring has advantages over traditional visual surveys including greater detection ranges, continuous long-term monitoring in remote locations under various weather conditions, and lower cost. However, until recently, the technology required to autonomously record whale sounds over long durations has been limited to low-frequency (< 1000 Hz) baleen whales. The need for a broader-band, higher-data capacity system capable of autonomously recording toothed whales and other marine mammals for long periods has prompted the development of a High-frequency Acoustic Recording Package (HARP) capable of sample rates up to 200 kHz. The HARP design is described and data analysis strategies are discussed using examples of broadband recorded data. [Work supported by the U.S. Navy CNO-N45.]

## Session 1pAOa

## Acoustical Oceanography and Underwater Acoustics: Acoustic Scattering by Aquatic Organisms II

Kenneth G. Foote, Cochair

Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Masahiko Furusawa, Cochair

Tokyo Univ. of Marine Science and Technology, 4-5-7 Konan, Minato, Tokyo, 108-8477, Japan

Chair's Introduction—1:00

## Contributed Papers

1:05

**1pAOa1. Broadband backscatter measurements of individual Hawaiian mesopelagic boundary community animals.** Whitlow W. L. Au (Hawaii Institute of Marine Biol., Univ. of Hawaii, P.O. Box 1106, Kailua, HI 86734) and Kelly J. Benoit-Bird (Oregon State Univ., Corvallis, OR 97331-5503)

Broadband simulated dolphin echolocation signals were used to measure the *ex situ* backscatter properties of mesopelagic boundary community animals. The signal had a peak frequency of 100 kHz, 3-dB bandwidth of 35 kHz, and duration of 80  $\mu$ s. Subjects were captured by trawling with a 2-m-opening Isaacs-Kidd midwater trawl. Backscatter measurements were conducted on the deck of the ship in a 2000-L seawater plastic tank with the transducer placed on the bottom, pointed upwards. Anesthetized animals were suspended in the transducer's beam by nylon monofilament lines connected to a support device constructed out of dowels. Backscatter measurements were obtained in both the dorsal and lateral aspects for seven myctophids and only in the dorsal aspect for the other 13 myctophids, 6 shrimps, and 3 squids. The echoes from the myctophids and shrimps usually had two highlights, one from the surface of the animal nearest the transducer and a second probably from the signal propagating through the body of the subject and reflecting off the opposite surface of the animal. The squid echoes consisted mainly of a single highlight but sometimes with a low-amplitude secondary highlight. The target strength had a rippled pattern in the frequency domain showing a distinct null.

1:20

**1pAOa2. Acoustic measurements of volume backscattering strength from red tide phytoplankton, *Cochlodinium polykrikoides*.** Tae-Hoon Bok, Dong-Guk Paeng (Faculty of Marine Industry Eng., College of Ocean, Cheju Nat'l Univ., Ara-dong, jeju-si, jeju-do, Korea, 690-756), Eunhye Kim, and Jungyul Na (Hanyang Univ., Sa-dong, Ansan-si, Kyeonggi-do, Korea, 425-791)

This research is to measure the volume backscattering strength, Sv, of the *Cochlodinium polykrikoides* (the main species of red tide in Korea) in order to investigate the possibilities of early detection of red tide using high-frequency transducers (5, 10 MHz). First, the Sv from the cultured *C. polykrikoides* was measured with the population density in the laboratory. Next, ultrasonic transducers were attached to the side of a research vessel, and the Sv was measured over 9-km ship track near Yeosu in the Korean southern sea. Seawater at three stations was sampled, and the number of *C. polykrikoides* was counted to compare the theoretical Sv with the measured one. The results show that the measured Sv was in agreement with the theoretical one. The Sv when the light was on was measured to be higher than that when the light was off in the laboratory, probably due to the increase of size and density of *C. polykrikoides* during photosynthesis. In the field experiment, the Sv was measured in real time with the position

information from the Global Positioning System data. This research proposes the use of underwater ultrasonic technology for early detection of red tide in real time.

1:35

**1pAOa3. Measurements of target strength spectra of metal spheres and fish by using broadband sonar signals of dolphin.** Tomohito Imaizumi, Masahiko Furusawa (Tokyo Univ. of Marine Sci. and Technol., 4-5-7 Kounan, Minato Tokyo, 108-8477, Japan, imaizumitomohito@yahoo.co.jp), Tomonari Akamatsu (Fisheries Res. Agency of Japan, Kamisu, Ibaraki 314-0408, Japan), Yasushi Nishimori, and Shinji Ogawa (Furuno Electric Co., Ltd., Nishinomiya-city, 662-8580, Japan)

Dolphins identify prey by using their broadband sonar signals. Considering that the spectrum of fish target strength (TS) leads to identification of objects, we measured the spectrum using dolphin's sonar signals with a broadband system. The TS spectrum is obtained by dividing the echo spectrum by the incident spectrum while setting the projector and hydrophone face to face. First, we measured the broadband form function of a tungsten carbide and copper sphere in a water tank and confirmed good agreements with theoretical value. Second, we measured the spectra of TS of three species of anaesthetized fish in a water tank. Although the spectra showed nearly similar tendencies of past measurements, they changed according to the species, the individual, and the tilt angle. Third, we measured TS spectra of fish suspended by nylon monofilament line in the open sea in semi-*in-situ*. We tried to remove the directivity of transducers by a tungsten carbide sphere suspended near the fish. The change of the TS spectra should be important information for the fish identification.

1:50

**1pAOa4. Variations of echosounder transducer performance versus water temperature.** David A. Demer and Josiah Renfree (NOAA/NMFS/Southwest Fisheries Sci. Ctr., 8604 La Jolla Shores Dr., La Jolla, CA, 92037)

Electroacoustic transducers are central components of multi-frequency echosounders used in remote target identification and acoustical surveys for fish and zooplankton. While these transducers, constructed from multiple materials such as piezoelectric ceramics, polyurethanes, and metals, are designed to be insensitive to performance changes versus time, pressure, and temperature, it has been shown that appreciable changes in echosounder system gains result from shifts in transducer resonant frequency, quality factor, and electrical impedance versus water temperature. Because it is standard practice to calibrate echosounder systems in one environment and apply the resulting gains to interpret data collected over a range of sea temperatures, the survey results may be significantly biased. The bias may be different for estimates derived from each echosounder frequency. The magnitudes of these frequency-dependent biases depend upon the temperature-dependent performances of the survey transducers, and the

range of temperatures encountered during the survey. Here, the performance has been measured for ten commonly used survey transducers versus water temperature ranging from 1 to 18 °C, using four different techniques. Temperature-dependent system gain corrections are proposed that can be used to minimize bias in multifrequency target identifications and biomass estimations. [Funding was provided by the NMFS Advanced Sampling Technology Working Group.]

2:05

**1pAOa5. Calibrating split beam transducers at depth.** Egil Ona and Geir Pedersen (Inst. of Marine Res., P.O. Box 1870, 5817, Bergen, Norway)

A new echo sounder probe for target strength measurements in deep water have been used to calibrate the depth performance of several pressure-stabilized split beam transducers. The probe carries three Simrad EK60 echo sounders and can be operated on 6500 m optical cable to about 1500-m depth. The transducers are mounted on a motorized, revolving platform. Digital angle sensors monitor transducer orientation, and an electronic compass also monitors transducer rotation. Calibration includes standard sensitivity parameters for split beam transducers, but also the equivalent beam angle and angle sensitivity at different depths. Examples of target strength data on marine animals will also be shown.

2:20

**1pAOa6. Standard-target calibration of active sonars used to quantify aquatic organisms.** Kenneth G. Foote (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

The span of sonars being calibrated by the standard-target method continues to grow. In this retrospective of the past 25 years, the essence of the method is reviewed for both narrow-band and broadband systems. Applications to scientific echo sounders with operating frequencies from 18 kHz to 3.2 MHz are summarized. Applications to multibeam sonars

that provide the water-column signal, with frequencies of 90, 200, 240, and 400 kHz, are also summarized. A planned application to a parametric sonar, with primary frequencies in the range 15–21 kHz and difference-frequency band 0.5–6 kHz, is described. The method is applicable to ordinary midfrequency sonars. In each example, particular standard targets are specified. Diameters vary from 10 mm for a sphere of tungsten carbide with 6% cobalt binder for use at low megahertz frequencies, to 60 mm for a sphere of electrolytic-grade copper for use at 38 kHz, to 280 mm for a sphere of aluminum alloy for use at low-kilohertz frequencies.

2:35

**1pAOa7. New scientific multibeam systems (ME70 and MS70) for fishery research applications.** Lars Nonboe Andersen, Sverre Berg, Ole Bernt Gammelster, and Even Borten Lunde (Simrad AS, P.O. Box 111, 3191 Horten, Norway, fish\_research@simrad.com)

Fishery scientists have for many years been requesting calibrated multibeam systems specially designed for fishery research applications. Simrad AS, Norway, has, in cooperation with IFREMER, France, developed a new multibeam echo sounder (ME70) and with IMR, Norway, a new multibeam sonar (MS70) for fishery research applications. Both systems have 800 transmitting and receiving channels and are based on similar hardware and software platforms. The systems are characterized by calibrated narrow beams, low sidelobe levels, high instantaneous dynamic range, and operate in the frequency range 70–120 kHz. The echo sounder is designed for high operating flexibility, with a configurable fan of 1–45 roll/pitch stabilized split beams with opening angles  $\sim 2$  to  $\sim 7$ . In addition, two configurable reference beams are available for comparison with standard echo sounder systems. The sonar is mounted on a drop keel, looking horizontally, covering a horizontal sector of 60 and a vertical sector of 45. This coverage matrix is generated using 500 roll stabilized beams, 25 beams horizontally with a resolution of  $\sim 3$ , and 20 beams vertically with a resolution of  $\sim 4$ . Both systems are designed for fish-stock assessment and fish-behavior studies. Results from system measurements will be presented.

2:50–3:10 Break

### *Invited Papers*

3:10

**1pAOa8. First results with the new scientific multibeam echo-sounder ME70.** Verena Trenkel (Ifremer Ctr. de Nantes, BP 21105 44311, Nantes Cedex 03, France), Valerie Mazauric, and Laurent Berger (Ifremer Ctr. de Brest, Plouzane, France)

Multibeam echo-sounders and sonars are increasingly used in fisheries acoustics for studying school structure and behavior and for abundance estimation. However, data quality and a poor signal-to-noise ratio obtained with the existing multibeam echo-sounders preclude quantitative biomass estimation as carried out with the traditional single beam echo-sounders. In order to overcome these shortcomings, Ifremer and Simrad AS started in 2003 the development of the new and highly flexible calibrated multibeam echo-sounder ME70 designed for studying a wide range of scientific questions. This paper will present the full functionalities of the ME70 and preliminary results obtained with the system installed onboard the research vessel *Thalassa*. It will also provide first indications of its performance. Data have been collected on small pelagic fish species in the bay of Biscay, using simultaneously the traditional single beam echo-sounders and the ME70 with different beam configurations. Preliminary comparisons of fish school descriptors and biomass estimates based on single- and multibeam data allow us to quantify the effect of the gain in sampling coverage and the high angular resolution of the new system.

3:30

**1pAOa9. First data from sea trials with the new MS70 multibeam sonar.** Egil Ona, John Dalen, Hans Petter Knudsen, Ruben Patel (Inst. of Marine Res., P.O. Box 1870, 5817 Bergen, Norway), Lars Nonboe Andersen, and Sverre Berg (Simrad A/S, 3191, Horten, Norway)

The new multi-beam sonar, MS70, is a horizontally observing sonar yielding very high spatial resolution when operating all 500 beams, covering the frequency band 75–112 kHz. The sonar has undergone sea trials from *R/V G. O. Sars* from December 2005. The presentation includes preliminary results obtained during this period. Topics covered will be sonar performance objectives, calibration methods, and preliminary results from calibration and school data acquisition. Examples of raw and processed 3-D and 4-D data from a stationary and surveying vessel on small sprat schools and large herring schools will be shown.

**1pAOa10. Development of a quantitative scanning sonar system and sea trial for a fish school of herring.** Yasushi Nishimori (Furuno Elec. Co., Ltd., 9-52 Ashihara-cho, Nishinomiya, 662-8580, Japan), Kohji Iida, Tohru Mukai (Hokkaido Univ., Hakodate, 041-8611, Japan), Yong Tang (Dalian Fisheries Univ., Dalian, 116023, China), Masaki Ogata, Kohzoh Tokuyama, and Emi Okazaki (Furuno Elec. Co., Ltd., Nishinomiya, 662-8580 Japan)

A quantitative sonar system enabling data recording, calibration, and numerical analysis of fish schools was developed based on the commercial scanning sonar. This system is equipped with a newly developed algorithm for estimation of the abundance of a fish school using three-dimensional echo integration. Estimation was made on the condition that the target strength of all fish is identical and known and the weight of a single fish is known. The field recording was performed to evaluate the algorithm. The school of spring spawning herring were scanned three-dimensionally and recorded using two different scanning methods. Two methods were applied in different distances between the school and vessel. The volume and three-dimensional echo integration for these two recorded dates were calculated and compared to each other. It was found that the variance of three-dimensional echo integration is smaller compared with the variance of volume. A compensation of abundance estimation using average target strength was tried and its effectivity was examined.

### Contributed Papers

4:10

**1pAOa11. Measurement of scattering directivity and behavior of fish schools using multibeam sonar.** George Cutter, Jr. and David Demer (NOAA Southwest Fisheries Sci. Ctr., 8604 La Jolla Shores Dr., La Jolla, CA 92038)

Fish dispersion and abundance are commonly estimated from volume backscattering coefficients ( $sv$ ) of animals located directly beneath the survey vessel, inside a narrow transducer beamwidth. Fish density ( $\#/m^2$ ) is estimated from a ratio of integrated  $sv$  and a representative backscattering cross-sectional area. Fish biomass is estimated from the product of fish density and the area sampled. Uncertainty in biomass estimates can arise from small sampling volumes, and the scattering directivities of the fish coupled with their orientations relative to the acoustic beam. Multibeam sonars can be used to increase the observational volumes, but add the complexity of a wider range of incident angles. If fish size and packing density are constant throughout a school, then the variation in  $sv$  versus aspect angle can be attributed to fish orientation. With modeling and multibeam sonar measurements, the following null hypotheses are explored: (1) The variability of  $sv$  versus roll angle is nearly uniform; (2) For measurements made with beams at or near vertical,  $sv$  variability is primarily due to the pitch angle; (3) For measurements with nonvertical beams, such as the outer beams of multibeam echosounders,  $sv$  variability is primarily due to the yaw angle.

4:25

**1pAOa12. Advances in the evaluation of acoustic beamformed backscatter data for fisheries applications.** Chris Malzone (The RESON Group, 100 Lopez Rd. Goleta, CA 93117, cmalzone@reson.com), Matthew Wilson (SonarData Pty Ltd, Hobart TAS 7001 Australia), and Pat Iampietro (CSU Monterey Bay, Seaside, CA 93955)

Single-beam echosounders are well-established tools for fisheries research. However, single-beam echosounder geometry is limited to a very small ensonification volume located within a 3 to 15-deg cone directly beneath the vessel. Recent advances in multibeam echosounder technology provide the option to collect backscatter and water-column data during an IHO order 1 hydrographic survey. Such comprehensive data sets are valuable for scientific and management purposes since they expand the volume in which water-column data are collected, and they provide as well a potential method to link the biology to the habitat. However, water-column information collected with multibeam echosounders poses several challenges, including increased data storage requirements, a need for data decimation, and survey methodology considerations. In this paper, concurrently acquired backscatter data from a split-beam scientific echosounder and multibeam echosounder are analyzed for relative abundance, spatial distribution, and avoidance behavior of pelagic organisms, primarily the market squid *Loligo opalescens* and various finfish. Aspects of the data sets are then explored, compared, and contrasted. Proposed solutions to some of the challenges posed by multibeam sonar water-column analysis are then reviewed and placed within an overall framework for fisheries applications.

4:40

**1pAOa13. Mapping fish schools with a high-resolution multibeam sonar.** Mark Trevorrow (Defence R&D Canada Atlantic, 9 Grove St., Dartmouth, NS, B2Y 3Z7 Canada, mark.trevorrow@drdc-rddc.gc.ca)

In late 2002 an evaluation of a high-resolution multibeam sonar (RESON SeaBat 8125) for mapping mid-water fish schools was performed in Knight Inlet, BC. The ship-mounted multibeam sonar operated at 455 kHz and provided 240 one-half-degree beams in the across-track plane with one-degree along-track beam width. Simultaneous measurements were made with a conventional 200-kHz echo-sounder. Acoustic calibrations of this multibeam sonar will be discussed, showing some range- and angle-dependent variations in sensitivity. The multibeam sonar was able to provide well-resolved mid-water images of fish schools (likely juvenile herring), sampling many fish schools missed by the echo-sounder and revealing cross-track fish school shape. Comparison of near-vertical volumetric scattering strengths between the multi-beam sonar and echo-sounder show reasonable agreement. Discussions on the combined use of conventional and multibeam sonars for fish surveys will be given.

4:55

**1pAOa14. Multibeam assessments of Australian seabed habitats.** Iain Parnum, Alexander Gavrilov, Paulus Siwabessy, and John Penrose (Ctr. for Marine Sci. and Technol., Curtin Univ. of Technol., G.P.O. Box UU1987, Perth, Western Australia 6845)

In the period 2003–2006 a major program of shallow seabed habitat assessment has been undertaken in Australian coastal seas. As part of the Coastal Water Habitat Mapping (CWHM) project of the national Cooperative Research Centre for Coastal Zone, Estuary and Waterway Management, research into optimal use of acoustic techniques for habitat assessment has been undertaken. Echo sounder, side scan, and multibeam acoustic systems have been used and evaluated in this program. Multibeam systems have been primarily employed with particular attention being given to combining topography data with associated measures of backscatter intensity. The emphasis on shallow waters, generally 50-m depth or less, and high-frequency multibeam systems with up to 240 beams has provided data sets with relatively high levels of spatial resolution. Measures of backscatter variation with incident angle have been incorporated into the interpretation process and a new angular dependence correction algorithm developed for production of backscatter imagery. The processing techniques developed in the CWHM project have been evaluated using towed video data for a variety of tropical and temperate coastal environments around Australia.

**1pAOa15. Acoustic identification and density estimate of *Diaphus theta* (myctophidae) off Hokkaido, Japan.** Hiroki Yasuma, Kazushi Miyashita (Field Sci. Ctr. for the Northern Biosphere, Hokkaido Univ., 3-1-1, Minato, Hakodate, Hokkaido, 041-8611, Japan), and Orio Yamamura (Hokkaido Natl. Fisheries Res. Inst., Kushiro, 085-0802, Japan)

*Diaphus theta* is the most abundant micronektonic fish in the north-western Pacific. Field acoustic data and biological samples were obtained both day and night off of eastern Hokkaido, Japan in September 2006 to estimate the fish density and diel changes of its vertical distribution. Differences of acoustical target strength (TS) values between at 38 and 120 kHz were applied to identify *D. theta* echos using field echo data. Theoretical estimation using swimbladder acoustic scattering models showed that the TS difference (TS 120-38 kHz) of larger fishes (over 60 mm) was between -4 and 1 dB; that for smaller fishes (less 60 mm) was less than .4 dB. These values differed from those of other major component species of the deep-scattering layer, such as krill and pollock, suggesting that the echo from *D. theta* is acoustically identifiable. Diel vertical distributions of *D. theta* were estimated after identification. Relatively dense schools were observed at around 400-m depth during daytime, although the schools were scattered widely above 100-m depth during nighttime. These results matched those obtained by biological sampling. In the surface layer (above 100 m), the estimated fish densities were about tenfold higher at nighttime than in daytime.

**1pAOa16. Measurement of fish school abundances in shallow sea using omnidirectional multibeam sonar.** Yong Tang (Dalian Fisheries Univ., 52 Heishijiao-jie Shahekou, Dalian, 116023, China; tang@dlfu.edu.cn), Kohji Iida, Tohru Mukai (Hokkaido Univ., Hakodate, 041-8611, Japan), and Yasushi Nishimori (Furuno Elec. Co., Ltd., Nishinomiya, 662-8580 Japan)

Previous studies had been constructed in the use of commercial scanning sonar to estimate the abundance of pelagic fish school. However, a relatively narrow dynamic range and lack of linear signal processing limited the application of sonar. In this research, quantitative omnidirectional multi-beam sonar (FSV30 research version, 24 kHz, Furuno) was used to investigate a fish school in a shallow area. The raw volume backscattering strength signal (RSV, output signal with  $20 \log r$  TVG and coefficient of multiple echo) of the fish school echoes were recorded and then its biomass was confirmed using purse seine capture. In postprocessing, the fish school echoes were extracted from the sonar echogram using the image processing method (polygon and a low-pass filter). The relationships between a school's biomass, its echo area, the maximum, and the average RSV of each ping were considered for various species. Results showed a positive correlation between the sonar-measured school area and school biomass, but no correlation between the RSV and school biomass. Also, the echo integration method for sonar survey using RSV in broad area and methods for averaging the target strength (TS) in three dimensions with respect to echo integration method were also discussed.

TUESDAY AFTERNOON, 28 NOVEMBER 2006

KAUAI ROOM, 1:00 TO 5:20 P.M.

### Session 1pAOB

## Acoustical Oceanography and Signal Processing in Acoustics: Acoustic Tomography for Coastal and Deep Water Applications II

James A. Mercer, Cochair

*Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6698*

Arata Kaneko, Cochair

*Hiroshima Univ., Graduate School of Engineering, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8527, Japan*

Chair's Introduction—1:00

### Invited Papers

1:05

**1pAOB1. Precise measurement of travel time difference for reciprocal transmission of Central Equatorial Pacific tomography.** Hiroyuki Hachiya (Dept. of Information & Image Sci., Chiba Univ., 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan), Toshiaki Nakamura (Natl. Defense Acad., Kanagawa 239-8686, Japan), Hidetoshi Fujimori, and Iwao Nakano (Japan Agency for Marine-Earth Sci. and Technol., Yokosuka 237-0061, Japan)

We present the estimation results of the travel time difference in reciprocal transmission data collected during the tomography experiment performed by Japan Agency for Marine-Earth Science and Technology (JAMSTEC) in the Central Equatorial Pacific. A precise estimation of travel time is essential for measuring ocean current using ocean acoustic tomography. We estimate the variation of the correlation of reciprocal propagated sound transmissions. Generally, reciprocal propagation sound waveforms that are transmitted at the same time are similar mutually, and it is considered that their correlation is high. However, it is found that reciprocal sound waveforms sometimes become different in the experiment. The vertical arriving angle of ray is measured using the phase differences between the receivers of a 30-m-long receiving array. The theoretical analysis of phase deviation using the Rice distribution is presented. From the ray identification results, the travel time difference between corresponding rays of reciprocal transmissions can be estimated. Since the phase of a ray signal is very stable, the precision of the travel time difference using phase difference is 6 times higher than the conventional estimation. The combination of the amplitude and phase information is effective for observing the ocean structure change.



**1pAOb2. Ocean acoustic thermometry and the seasonal cycle of temperature in the North Pacific Ocean.** Brian D. Dushaw, R. Andrew, B. Howe, J. Mercer (Appl. Phys. Lab., Univ. of Washington, 1013 N.E. 40th St., Seattle, WA 98105-6698, dushaw@apl.washington.edu), B. Cornuelle, M. Dzieciuch, W. Munk, P. Worcester (Scripps Inst. of Oceanogr., La Jolla, CA), T. Birdsall, K. Metzger (Univ. of Michigan, Ann Arbor, MI), D. Menemenlis, and R. Spindel (California Institute of Technol., Pasadena, CA)

With several years of long-range (several Mm) acoustic propagation data obtained during the Acoustic Thermometry of Ocean Climate (ATOC) and North Pacific Acoustic Laboratory (NPAL) projects, the seasonal cycle of ocean temperature in the North Pacific can be examined. Acoustic transmissions have been made from a source located off the northern Californian coast and from a source located north of Kauai, HI to several receivers of opportunity located in the North Pacific Basin. The acoustic data are a high signal-to-noise measure of large-scale temperature with excellent temporal resolution. Although only a few realizations of the seasonal cycle are available, it is clear that inter- and intraannual variabilities are large contributions to the time series, with signal amplitudes comparable to the seasonal cycle. Such variabilities are likely advective in origin. The time scales for some of the changes in temperature are short, sometimes of order weeks. Not all available acoustic paths are suitable for assessing the seasonal cycle, however. Near Hawaii, the acoustic sampling does not extend to the near-surface waters, so seasonal variations there are not measured. The acoustic results will be compared to measures of the seasonal cycle by satellite altimetry, profiling floats, and the ECCO numerical ocean model.

**1pAOb3. The Long-range Ocean Acoustic Propagation EXperiment (LOAPEX): An overview.** James A. Mercer, Bruce M. Howe, Rex K. Andrew, Michael A. Wolfson (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, mercer@apl.washington.edu), Peter F. Worcester, Matthew A. Dzieciuch (Univ. of California, La Jolla, CA 92093), and John A. Colosi (Naval Postgrad. School, Monterey, CA 93943)

The Long-range Ocean Acoustic Propagation EXperiment (LOAPEX), conducted in the NE Pacific Ocean, provided acoustic transmissions from a ship-suspended source at eight widely separated stations, and from a cabled acoustic source near the Island of Kauai, HI. The transmissions were received on several bottom-mounted horizontal hydrophone arrays distributed about the NE Pacific Ocean Basin and two, nearly collocated, vertical hydrophone line arrays spanning roughly 3500 m of the water column. Ranges varied from 50 km to several Mm. The goals of the experiment are (i) to study the evolution, with distance (range), of the acoustic arrival pattern and in particular the dependence of the spatial and temporal coherence; (ii) to investigate the nature of the deep caustics and the associated arrivals well below their turning depths; (iii) to analyze the effects of the ocean bottom near the bottom-mounted acoustic source cabled to Kauai; and (iv) to produce a thermal snapshot of the NE Pacific Ocean. The experiment goals, design, and methods are described as well as preliminary data results. [Work supported by ONR.]

**1pAOb4. The SPICEX (Spice Experiment) component of the 2004 North Pacific Acoustic Laboratory (NPAL) experiment: An overview.** Peter F. Worcester, Matthew A. Dzieciuch, Lora J. Van Uffelen, Daniel L. Rudnick, Bruce D. Cornuelle, and Walter H. Munk (Scripps Inst. of Oceanogr., Univ. of California at San Diego, La Jolla, CA 92093)

The objectives of the SPICEX component of the 2004 North Pacific Acoustic Laboratory (NPAL) experiment were (i) to elucidate the relative roles of internal waves, ocean spice, and internal tides in causing acoustic fluctuations; (ii) to understand the acoustic scattering far into the geometric shadow zone beneath caustics previously seen on bottom-mounted SOSUS receivers (shadow-zone arrivals); and (iii) to explore the range dependence of the fluctuation statistics. Two vertical line array receivers deployed in close proximity in the central North Pacific Ocean, together spanning much of the water column, recorded transmissions from broadband 250-Hz sources moored at the sound channel axis (750 m) and at surface conjugate depth (3000 m) at ranges of 500 and 1000 km from June through November 2004. Extensive environmental measurements were made along the acoustic path. The acoustic data from SPICEX reveal the vertical structure of the deep shadow zone arrivals. Internal waves are able to account for much of the observed vertical scattering in the reception finale, but other mechanisms seem necessary to account for the early shadow zone arrivals. Receptions from the abyssal sources show scattering occurring predominantly along the acoustic time fronts, as previously suggested by simulations, rather than across them.

**1pAOb5. An overview of BASSEX (Basin Acoustic Seamount Scattering Experiment).** Arthur B. Baggeroer (MIT, Cambridge, MA 02139), Kevin D. Heaney (OASIS, Inc., Lexington, MA), Peter F. Worcester, Matthew Dzieciuch (UCSD, La Jolla, CA 92037), James Mercer, Rex Andrew, and Bruce Howe (Univ. of Washington, Seattle, WA 98105)

BASSEX exploited the signals transmitted from the SPICEX, LOAPEX, and NPAL Kauai sources to examine the forward scattering from the Kermit-Roosevelt Seamounts near the Mendocino Fault in the northeast Pacific. The receiver was the FORA (four octave research array) from Pennsylvania State University. This array was towed at 300 m and had approximate resolutions of 2 and 6 deg for the 250- and 75-Hz signal transmissions. These were beamformed and pulse compressed for analyzing the multipath receptions from the sources. Source-receiver ranges covered from 250 to 1600 km provide a very rich data set. Data for forward- and backscattering from the slope near the NPAL source north of Kauai were also obtained. The analysis so far has demonstrated forward-scattering shadow convergence zones as well as diffraction from the seamounts. The upslope/downslope path structure near the Kauai source has also been analyzed. Both ray path and full wave acoustic modeling have been used for path identification as well as a theoretical model for modal scattering from a cone within an acoustic wave guide. [Work supported by ONR Code 321, Ocean Acoustics.]

**1pAOB6. Ray and finite frequency kernel stability in the presence of realistic eddies.** Bruce D. Cornuelle and Matthew A. Dzieciuch (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr, La Jolla CA 92093, bcornuelle@ucsd.edu)

Many recent studies have focused on the instability of ray paths (and therefore ray sampling kernels) in an ocean sound-speed field perturbed by internal waves. In many parts of the world's oceans there is synoptic scale eddy and planetary wave activity that produces sound-speed anomalies of larger scale than the internal waves, and of equal or larger magnitude. These are the features to be imaged by ocean tomography experiments, but the linearized forward problem that relates the travel time anomaly of the observed rays to the sound-speed anomalies may be invalid for realistic levels of eddy variability. This problem has been studied in the past for ray kernels by a number of authors, some of whom have concluded that nonlinearity requires iteration of the inversion process. We will compare and contrast the linearity of ray travel time kernel, finite frequency travel time kernels, and finite frequency kernels for other observables, such as pulse shape, in the presence of realistic variability for the N. Philippine sea, as estimated by ocean models and historical data. The travel time signals and nonlinearity due to the synoptic variability will be compared to that for internal waves. [Work supported by ONR.]

3:05–3:20 Break

### Contributed Papers

3:20

**1pAOB7. Seafloor hydrophone and vertical geophone observations during the North Pacific Acoustic Laboratory/Long-range Ocean Acoustic Propagation Experiment (NPAL/LOAPEX).** Ralph A. Stephen (Woods Hole Oceanog. Inst., Woods Hole, MA 02543-1542, rsteph@whoi.edu), James A. Mercer, Rex K. Andrew (Univ. of Washington, Seattle, WA 98105-6698), and John A. Colosi (Naval Postgrad. School, Monterey, CA 93943)

Four OBSs (ocean bottom seismometers), consisting of one hydrophone and one vertical component geophone channel each, were deployed below the deep vertical line array (DVLA) on the North Pacific Acoustic Laboratory/Long-range Ocean Acoustic Propagation Experiment (NPAL/LOAPEX). The OBSs were deployed at 2 km offset north, south, east, and west from the nominal DVLA location. Data, sampled at 500 sps, were recorded continuously from the deployment in September 2004 for over 100 days. The OBSs were on the seafloor in about 5000 m of water and would be in the shadow zone for long-range, ducted propagation in the sound channel. The goals of the deployment were (1) to quantify the amount of energy that leaks out of the sound channel into the shadow zone; (2) to measure the relative sensitivity (signal-to-noise) of seafloor hydrophones and vertical component geophones to long-range signals; and (3) to study the physics of earthquake generated T-phases. Leakage was observed out to 3200 km on the vertical component geophone and out to 1000 km on the hydrophone. The ratio of the pressure to vertical velocity varies between arrivals on the same trace. [The OBS deployments were funded by contributions from ONR, NSF, and WHOI. Data analysis funded by ONR.]

3:35

**1pAOB8. Mean acoustic field in long-range ocean acoustic propagation experiment (LOAPEX).** Rex K. Andrew, Frank S. Henyey, James A. Mercer, Bruce M. Howe (Appl. Phys. Lab., Univ of Washington, 1013 NE 40th St., Seattle, WA 92105), Peter F. Worcester, and Matthew A. Dzieciuch (Univ. of California at San Diego, La Jolla, CA 92093)

We extract the low-frequency (70-Hz) mean acoustic pressure field, and the mean acoustic field up to time shifts, from the LOAPEX receptions on the vertical line arrays, for transmissions at various ranges. Means are taken over time intervals of order a day. Source and receiver motion is removed from the data, and possibly tides and other slow phenomena. The experimental results are compared to predictions from theoretical calculations assuming scattering by internal waves. The theoretical calculations make the Markov approximation that the sound-speed fluctuation correlations can be replaced by an operator at a single range. The calculations use modes rather than rays, because of the very low frequency, thus differing from the version of the Markov approximation that assumes delta-correlated sound-speed fluctuations.

3:50

**1pAOB9. Mixed layer effects on acoustic arrival structure.** Lora J. Van Uffelen, Peter F. Worcester, Matthew A. Dzieciuch, and Daniel L. Rudnick (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238, lvanuffe@ucsd.edu)

Oceanic mixed layers develop as the summer sun warms the upper layer of the water column and turbulence from early autumn storms stirs this heat downward, forming a well-mixed layer with constant temperature and salinity. The mixed layer is therefore thinnest in summer, when the surface is warm and the winds are relatively weak, and deepest in late winter. Receptions from moored 250-Hz acoustic sources at ranges of 500 and 1000 km in the North Pacific were recorded on vertical line array receivers from June through November 2004, spanning the early development of the winter mixed layer. The time-evolving character of the acoustic arrival structure throughout this transmission period will be investigated to explore the effect of a changing mixed layer environment on long-range acoustic propagation. Particular consideration will be given to variability in the early branches of the acoustic timefronts, which correspond to steep rays interacting with the upper ocean.

4:05

**1pAOB10. Analysis of low-mode arrivals at 500 and 1000-km ranges in the North Pacific.** Kathleen E. Wage (George Mason Univ., 4400 Univ. Dr., MSN 1G5, Fairfax, VA 22030, kwage@gmu.edu), Peter F. Worcester, and Matthew A. Dzieciuch (UCSD, La Jolla, CA 92093)

Experimental measurements of low-mode signals in the deep ocean are needed to verify statistical models and to design improved processing algorithms for these signals. Several long-range propagation experiments provided estimates of low mode statistics at ranges between 3 and 5 megameters [Wage *et al.*, *J. Acoust. Soc. Am.* **117**, 1565–1581 (2005)]. The 2004–2005 SPICE experiment included measurements at shorter ranges and a higher frequency than previous experiments. A 40-element vertical line array (VLA) received signals from two broadband sources at ranges of 500 and 1000 km. Both sources were moored at depths near the sound-channel axis; thus, they directly excited the lowest modes. The VLA was designed to spatially resolve 20 modes at the center frequency of 250 Hz. In addition to hydrophones the SPICE array included temperature and salinity sensors, which can be used to infer sound speed, and hence mode shapes, at the VLA. This talk describes an analysis of the low-mode arrivals in the SPICE data set. The presentation will include a discussion of mode filter design, focusing on how often the filter coefficients need to be updated to avoid problems with environmental mismatch. [Work supported by ONR.]

**1pAOB11. Seamount geo-acoustic inversion from an acoustic scattering experiment (NPAL-04).** Kevin D. Heaney (OASIS Inc., 11006 Clara Barton Dr., Fairfax Station, VA 22039), Arthur B. Baggeroer (MIT, Cambridge, MA 02139), Peter F. Worcester, and Matthew A. Dzieciuch (Univ. of California, La Jolla, CA 92093)

It is thought that the deep ocean has thick, soft sediment due to millennia of build up from detritus. An open question is the sediment thickness on vast seamounts that rise many km from the ocean floor. Acoustic scattering from the Kermit-Roosevelt seamount (39.20 N 146.3 O W) in the central Pacific during the Basin Acoustic Seamount Scattering Experiment (BASSEX-04) in August of 2004 is used to address this question. The Kermit-Roosevelt seamount rises from a floor of 5400 m to a maximum height of 977 m below the surface. Broadband transmissions (100 Hz bandwidth, 250 Hz center frequency, 12.276 s period length) from two sources suspended (800 m depth, ranges of 521 and 625 km) were received on towed array (300 m depth, 200 m aperture) in the vicinity of the seamount. Combining array gain and time-bandwidth gain from the matched filter it is possible to clearly identify energy that has scattered from the seamount in 15 transmissions. These receptions are used to infer the sediment thickness and type as a function of the water depth. An efficient PE computational approach is applied as well as a reflection-coefficient ray-based approach.

**1pAOB12. Mode processing of transient signals using a deficient vertical receiving array, with application to LOAPEX measurements.**

Ilya A. Udovychenkov, Michael G. Brown (RSMAS, AMP, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149), Peter F. Worcester, Matthew A. Dzieciuch, Lora J. Van Uffelen (Univ. of California at San Diego, La Jolla, CA, 92093), James A. Mercer, Bruce M. Howe, and Rex K. Andrew (Univ. of Washington, Seattle, WA 98195)

Mode processing of transient signals in a deep ocean environment using a deficient vertical receiving array is considered. Here, "deficient" implies that the vertical array is too sparse to resolve modes and/or has gaps. It is well known that mode processing of cw signals using a deficient array leads to significant modal cross-talk, i.e., nonzero projections of mode  $n$  energy onto mode  $m$ ,  $m \neq n$ . It is shown, using simulated wave fields, that when phase-coherent processing is employed, transient signals are less sensitive to modal cross-talk than their cw counterparts. An explanation of this behavior is provided. Mode processing of a subset of the measurements made during the LOAPEX experiment at approximately 1 Mm range with  $f_0 = 75$  Hz,  $\Delta f = 35$  Hz is shown to give robust estimates of modal group time spreads for mode numbers  $0 \leq m \leq 70$ . [Work supported by ONR.]

**1pAOB13. Horizontal coherence of tracked arrivals in the North Pacific Acoustic Laboratory98 (NPAL98).** Matthew A. Dzieciuch (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., La Jolla, CA 92093, mad@ucsd.edu), Michael D. Vera (The Univ. of Southern Mississippi, Hattiesburg, MS), and The NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester)

The NPAL98 experiment offered a unique opportunity to study horizontal coherence because of the deployment of a billboard array. The billboard consisted of five vertical line arrays with a 3600-m horizontal aperture listening to a 75-Hz acoustic signal 3500 km away. Original estimates of horizontal coherence were complicated by the bottom-interacting sound field, which made it difficult to track individual arrivals. Thus, previous studies of the horizontal coherence in this experiment were limited to the arrival pattern in aggregate. By using a turning-point filter and a rake correlator, it is now possible to track individual arrivals during the NPAL98 experiment. We can thus measure the coherence of individual early non-bottom-interacting arrivals. Results are compared to recent calculated predictions using both Monte Carlo and path-integral methods for the same environment. [Work supported by ONR.]

**1pAOB14. The range evolution of the mean wavefront intensity for the long-range ocean acoustic propagation experiment (LOAPEX) off-axis source transmissions.**

Jinshan Xu (MIT/WHOI Joint Program in Oceanogr. and Oceanog. Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), John A. Colosi (Naval Postgrad. School, Monterey, CA 93943), James A. Mercer, Bruce M. Howe, Rex K. Andrew (Univ. of Washington, Seattle, WA 98105), Peter F. Worcester, and Matthew Dzieciuch (UCSD, La Jolla, CA 92093)

One of the main objectives of the NPAL 2004 experiment, LOAPEX (Long-range Ocean Acoustic Propagation EXperiment), which was conducted between 10 September and 10 October 2004, was to better understand the roles of scattering and diffraction in general. The LOAPEX measurement provided acoustic transmission data for ranges of 50, 250, 500, 1000, 1600, 2300, and 3200 km. By placing the source off-axis in order to avoid exciting low-order modes, we are able to study phenomena of the significant in-filling of acoustic energy into the finale region. Our focus will be on the transmissions for the off-axis source location (nominally 350-m depth), and the acoustic receptions as recorded on the 1400-m-long axial receiving array. The observation of the mean intensity of the wavefront arrival pattern at each range will be compared to deterministic ray and parabolic equation calculations. The following questions will be addressed here: (1) How does high angle acoustic energy from an off-axis source transfer energy to low angles in the axial region of the waveguide? (2) What are the relative contributions from diffraction and scattering? (3) How does this energy transfer scale with range? [Work supported by ONR.]

**Session 1pBB****Biomedical Ultrasound/Bioresponse to Vibration and Signal Processing in Acoustics: Acoustic Tomography in Tissue**

Jeffrey A. Ketterling, Cochair

*Riverside Research Inst., 156 William St., New York, NY 10038-2609*

Iwaki Akiyama, Cochair

*Shonan Inst. of Technology, 1-1-25 Tsujido-nishikaigan, Fujisawa 251-8511, Japan***Chair's Introduction—1:35*****Invited Papers*****1:40****1pBB1. Acoustic computed tomography: Waves of approaches.** James F. Greenleaf (Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55901)

Computed tomography (CT) using acoustic waves was invented at about the same time as MRI. The method of computing attenuation from projections was a direct analog of the x-ray CT approach. However, the method of computing speed from the time of flight of the acoustic ray was a larger logical jump. The field started with straight ray, earliest arrival time methods and has matured to what can be termed the full wave equation inversion method. The extension from two dimensions to three, although essential in the body, did not find a practical solution until very recently. The inverse problem represented by this field has found the attention of many investigators over the years. Practical medical implementations continue being investigated by several investigators more than 30 years after the first invention of the method. Acoustic tomography has been applied in nonmedical areas including oceanography, seismic, and solar sciences. Time awaits its success in medicine.

**2:00****1pBB2. From laboratory to clinical trials: An odyssey of ultrasound inverse scattering imaging for breast cancer diagnosis.** Steven A. Johnson, David T. Borup, James W. Wiskin, Michael Berggren, Barry Hanover, Frank Setinsek, Scott Olsen, and Karleen Callahan (Techniscan, Inc., 1011 East Murray Holladay Rd., Salt Lake City, UT 84117)

Techniscan has completed a local phase of clinical trials of a unique inverse scattering breast cancer scanner. The scanner is based on transmission of plane waves from 0.3 to 2 MHz through the breast to a receiver on the opposite side of the breast. The receiver consists of a 2-D array of six rows by 160 columns. These elements are multiplexed by a factor of 12 into 80 analog to digital converters. The dual transducer-head design allows the inversion of the transmission data to create two independent quantitative images, one of speed of sound and a second of attenuation. The two heads are mechanically attached to a hollow rotating shaft, to provide 360 deg of mechanical motion. Typically, 180 to 360 steps per 360-deg rotation and 2-mm vertical steps are used to create a data set that is inverted into a 3-D image of 240 by 240 by 70 voxels. A unique set of custom circuit boards was constructed to implement scanner electronics. The next scanner version with anticipated inverse scattering resolution of 1 mm or less and reflection tomography resolution of 0.5 mm or less should improve performance. Clinical images to date and performance will be discussed.

**2:20****1pBB3. Image reconstruction using eigenfunctions of a scattering operator.** Robert Waag (Dept. of Elec. and Comput. Eng. and Dept. of Imaging Sci., Univ. of Rochester, Rochester, NY 14627)

A multiple-frequency iterative inverse scattering method is reviewed. In the method, scattered acoustic fields are described using a scattering operator with eigenfunctions that correspond to far-field patterns of an effective source distribution. Incident-wave patterns determined using the eigenfunctions focus on the distribution. The focusing properties of the eigenfunctions are employed to reconstruct an object by using products of numerically calculated fields defined by weighted products of the eigenfunctions. The full range of frequencies present in an incident pulse waveform is employed. Iteration using a linearized version of the method and calculated scattering from an estimate of the object permits reconstruction of large-scale high-contrast objects. The method is illustrated using calculated and measured data. In the calculations, an exact solution for scattering from nonconcentric lossy cylinders with parallel axes was employed to obtain the scattered field. In the measurements, a novel ring-transducer system was used to obtain the incident and total fields. The results of simulations and experiments show that the method converges and is accurate for high-contrast large-scale tissue-mimicking objects. Reconstructions using measured scattering from random media have considerably less speckle than conventional b-scan images, implying that the inverse scattering method can clearly demonstrate small high-contrast inclusions.

**1pBB4. Optical coherence-based imaging and sensing in biomedicine and biotechnology.** Joseph A. Izatt (Duke Univ., Campus Box 90281, Durham NC 27708)

Optical coherence-based imaging techniques including optical coherence tomography (OCT), optical coherence microscopy (OCM), and spectral domain phase microscopy (SDPM) use low-coherence spectral interferometry to obtain nanometer to micron-scale measurements of structure, motion, and molecular composition in living cells, tissues, and organisms. OCT has become a standard diagnostic tool in clinical ophthalmology and is under investigation for other human diagnostic applications including cancer detection and evaluation of cardiovascular disease. Within the past few years, dramatic technology advances have increased the performance of OCT and OCM systems manyfold, and they are now capable of micron-scale two- and three-dimensional functional and molecular imaging noninvasively in living systems. Applications of these new technologies for high-throughput noninvasive phenotyping and rapid 3D imaging in small animals and developmental biology models is particularly compelling. Related technology advances have enabled the design of highly phase-stable interference microscopes capable of resolving nanometer-scale structures and motions in living cells with ms temporal resolution. These new capabilities are being used to probe cellular internal and external surfaces and their responses to chemical and mechanical stimuli. I will review our ongoing work at Duke in these areas. [Work supported by NIH Grants RR019769, EY013516, EB006338, and EY017393.]

**1pBB5. Whole breast imaging with ultrasound tomography: First clinical results at the Karmanos cancer institute.** Neb Duric, Peter Littrup, Earle Holsapple, and Olsi Rama (Karmanos Cancer Inst., 4100 John R, Detroit, MI 48201)

Although mammography is the gold standard for breast imaging, it leads to a high rate of biopsies of benign lesions and a significant false negative rate for women with dense breasts. A clinical breast-imaging device, based on the principles of ultrasound tomography, is being developed in response to this imaging performance gap. The computed ultrasound risk evaluation (CURE) system has been designed with the clinical goal of providing fast, whole-breast, operator-independent imaging, capable of differentiating breast masses. In an initial assessment of its *in vivo* performance, 50 patients were examined with CURE. The following results were demonstrated: (i) tomographic imaging of breast architecture, (ii) in-plane spatial resolution of 0.5 mm in reflection and 4 mm in transmission, (iii) imaging of breast masses in reflection and transmission mode (sound speed and attenuation), and (iv) differentiation of breast masses was shown to be feasible. These initial results indicate that operator-independent, whole-breast imaging and the detection and characterization of breast masses are feasible. Future studies will focus on improved detection and differentiation of masses with the long-term goal of improving the specificity of breast exams. [The authors acknowledge financial support from the Karmanos Cancer Institute and the NIH (grant R43 CA108072-01).]

**1pBB6. Advanced tomographic algorithms and their potential applications to ultrasound imaging.** Xiaochuan Pan, Samuel LeRoque, Emil Sidky (Dept. of Rad., MC2026, The Univ. of Chicago, 5841 S. Maryland Ave., Chicago, IL 60637, xpan@uchicago.edu), and Mark Anastasio (Illinois Inst. of Technol.)

The success of computed tomography has led to the development of a wide variety of tomographic imaging techniques. There has always been a great interest in ultrasound-based tomographic imaging because it can provide detailed, quantitative information about the imaged subject, which can be used for diagnosis, evaluation, and monitoring of cancer and other diseases. Tomographic imaging relies not only upon the imaging system for data acquisition but also upon the algorithms for image recovery from the data. In the last decade or so, there has been a tremendous development of innovative algorithms for tomographic imaging. One important breakthrough is that an artifact-free ROI image within the subject can be obtained from, e.g., truncated data. This breakthrough has led to the development of the so-called targeted-ROI imaging strategy in which one can substantially reduce imaging effort and artifacts. There have also been significant advances in reconstructing images from highly incomplete data acquired, e.g., at a small number of views and/or over a limited angular range. In this paper, we will discuss these new developments and their potential applications to ultrasound-based imaging. We will use examples to illustrate how they can be exploited for reconstructing images in different types of ultrasound-based tomographic imaging.

### Contributed Papers

**1pBB7. On the use of inverse scattering tomographic methods for quantitative ultrasound techniques.** Roberto J. Lavarello and Michael L. Oelze (Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801)

The application of the distorted Born iterative method (DBIM) is proposed to construct tomographic images of small scale structures. Specifically, DBIM is proposed as a tool to construct better ultrasonic models of cells in order to improve the accuracy of quantitative ultrasound (QUS) analysis. An alternative implementation of DBIM valid for measurement configurations where the receiver positions are fixed is discussed. In addition, a method for the selection of the regularization parameter is deter-

mined, based on the Rayleigh quotient iteration. QUS imaging has been used to differentiate malignant from benign tissue. Recently, cells have been hypothesized as an important source of scattering in tissue. Generating tomographic images of sound speed and density of cells would allow better models to be constructed. Because the contrast mechanism in cells is unknown, the effect of variable density is explored. Stage II–III Xenopus oocytes are proposed as an initial cell model for scattering, with typical sizes ranging between 300 and 600  $\mu\text{m}$ . Based on measurements of the propagation delay through oocytes, it was determined that their diameters correspond to at most 24 resolution cells when using single-frequency DBIM. Hence, frequency hopping may be required to discriminate smaller features inside the oocytes.

**1pBB8. Linear sampling method for acoustic inverse scattering in breast microcalcification detection.** Matthew A. Lewis, Peiyong Liu, Edmond Richer, Peter P. Antich (Dept. of Radiol., UT Southwestern Medical Ctr. at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75390-9058, matthew.lewis@utsouthwestern.edu), and Stuart G. Johnson (UT Austin, Austin, TX)

In women under age 40, the breast may be too dense for mammographic screening and microcalcification detection. A nonionizing method would be preferred for frequent screening in patients with significant risk factors for developing breast cancer. Ultrasonic detection of microcalcifications has been problematic since frequencies above 10 MHz suffer from appreciable attenuation in soft tissues. Transmission diffraction tomography is by construction a low-pass filter and insensitive to scattering caused by small, hard inhomogeneities. A more general form of acoustic inverse scattering is therefore needed for microcalcification detection and localization by ultrasound. An advanced scalar inverse scattering theory developed by Colton, Kirsch, and others [D. Colton, J. Coyle, and P. Monk. *SIAM Review* **42**(3), 369–414 (2000)] in the electromagnetic inverse scattering community for determining the location and shape of scatterers with size on the order of the wavelength (the so-called resonance region) has been simulated in the context of breast microcalcification detection. This method was shown to be related to both multiple signal classification and time-reversal algorithms [M. Cheney. *Inverse Problems* **17**, 581–585 (2001)] [S.K. Lehman and A.J. Devaney. *J. Acoust. Soc. Am.* **113**(5), 2742–2753 (2003)]. Translation of this method to a heterogeneous soft-tissue background is a significant challenge.

**1pBB9. Full wave nonlinear inverse scattering for transmission breast imaging.** James Wiskin (Techni Scan Medical Systems, 1011 Murray Holladay Rd., Salt Lake City, UT, 84117, jwiskin@techniscanmedical.com), David Borup, and Steven Johnson

Ultrasound has been proposed as an adjunct diagnostic modality for 25+ years. However, initial studies were hampered by relatively low resolution. Over the past 20 years there has been a tremendous amount of research carried out in inverse scattering, the mathematical foundation behind quantitative determination of the coefficients of a partial differential equation. In particular, the research at TechniScan, Inc., and the Dept. of Bioengineering, Univ. of Utah, has culminated in algorithms that give approximately millimeter resolution, and quantitative accuracy, in 2-D and 3-D problems. We invert the full nonlinear acoustic inverse scattering problem (no Born/Rytov approximation) in 2-D and 3-D. We reconstruct a phantom with known inclusions. The image space is  $241 \times 241 \times 135 = \sim 7.8$  Megapixels or larger. Each pixel is  $0.8 \times 0.8 \times 1.0 \text{mm}^3$  in size. The data is collected on a 6 by 160 element array, positioned 180 views surrounding the breast. The transmitted wave is a plane wave chirp. Reconstruction time is  $\sim 2$  h on a six-node cluster of Pentium Ms. This paper will discuss theoretical algorithm development, practical issues that result in images, and a device presently being used in a clinical setting. We show several reconstructed images from  $\sim 400$  volunteers.

**1pBB10. Radial reflection diffraction tomography.** Sean K. Lehman (Lawrence Livermore Natl. Lab., L-154, 7000 East Ave., Livermore, CA 94550) and Stephen J. Norton (Geophex, Ltd., Raleigh, NC 27603)

A wave-based tomographic imaging algorithm based upon a single, rotating, radially outward oriented transducer has been developed. At successive angular locations at a fixed radius, the transducer launches a primary field and collects the backscattered field in a “pitch/catch” operation. The hardware configuration, operating mode, and data collection method are identical to that of most medical intravascular ultrasound (IVUS) systems. The goal is to develop a wave-based imaging algorithm using diffraction tomography techniques. Given the hardware configuration and the imaging method, this system is referred to as “radial reflection diffraction tomography.” Two hardware configurations are considered: a multimono-static mode using a single transducer as described

above, and a multistatic mode consisting of a single transmitter and an aperture formed by multiple receivers. The probe is mounted at the end of a catheter or snaking tube inserted into a part or medium with the goal of forming images of the plane perpendicular to the axis of rotation. An analytic expression for the multimono-static inverse is derived but ultimately a Hilbert space inverse wave algorithm is used to construct images using both operating modes. Applications include improved IVUS imaging, bore hole tomography, and nondestructive evaluation (NDE) of parts with existing access holes.

**1pBB11. Ultrasonic reflection imaging with a heterogeneous background for breast cancer detection and diagnosis.** Lianjie Huang (Los Alamos Natl. Lab., M.S. D443, Los Alamos, NM 87545), Youli Quan (Stanford Univ., Stanford, CA 94305), Nebojsa Duric, and Cuiping Li (Karmanos Cancer Inst., Detroit, MI 48201)

Sound speeds and densities of breast tissue vary in space, and those of tumors are different from the surrounding tissues. These differences in mechanical properties cause scattering of ultrasonic waves, particularly in dense breasts. An ultrasonic reflection imaging method using a heterogeneous background is presented to properly account for scattering off heterogeneities within the breast. The method is based on downward continuation of ultrasonic wave fields in the frequency-space and frequency-wave-number domains. Synthetic ultrasonic data for scatterers within a heterogeneous breast model were used to investigate the imaging resolution of the method. Imaging examples using ultrasonic data acquired using a clinical prototype, consisting of an ultrasound scanner with a ring transducer array, are presented. The heterogeneous background sound-speed models were obtained using transmission tomography. The imaging results are compared with those obtained using a homogeneous background. The study demonstrates that imaging with heterogeneous backgrounds significantly improves image quality and resolution. [This work was supported through the DOE Laboratory-Directed Research and Development program at Los Alamos National Laboratory.]

**1pBB12. Photoacoustic imaging with a planar measurement array: Improving the resolution using image sources.** Benjamin T. Cox (Dept. of Med. Phys. and Bioengineering, Univ. College London, Gower St., London, WC1E 6BT, UK, bencox@mpb.ucl.ac.uk), Simon. R. Arridge, and Paul C. Beard (Univ. College London, Gower St., London, WC1E 6BT, UK)

Biomedical photoacoustic (PA) imaging is a soft-tissue imaging modality that combines the high spatial resolution of ultrasound (US) with the contrast and spectroscopic opportunities afforded by imaging optical absorption. Exact reconstruction algorithms for PA imaging have been described for spherical, cylindrical, and planar sensor arrays. Many applications to date have used a cylindrical measurement geometry because a single US transducer may be scanned straightforwardly in a circle around a sample. However, planar US arrays composed of piezo or optical elements with small element sizes and fast acquisition times are readily available, making them an attractive option for imaging applications. To capture sufficient data for an exact PA reconstruction with a planar geometry requires an infinitely wide array. In practice, of course, it will be finite, resulting in a loss of resolution and introducing artefacts into the image. To overcome this limitation, it is proposed that acoustic image sources are used to generate a periodically repeating sound field that can be reconstructed exactly from measurements made over a finite aperture. [This work was funded by the EPSRC, UK]

**1pBB13. An acoustic technique to assess particle distribution following transcutaneous needle-free injection.** Jamie Condliffe, Jamie R. T. Collin, George Costigan, and Constantin-C. Coussios (Inst. of Biomed. Eng., Dept. of Eng. Sci., Univ. of Oxford, 17 Parks Rd., Oxford OX3 0DH, UK, constantin.coussios@eng.ox.ac.uk)

Needle-free injection is a novel technique enabling pain-free transdermal drug and vaccine delivery. The present work aims at developing a noninvasive, ultrasonic technique to assess distribution of drug particles delivered in this manner beneath an optically opaque layer, such as the skin. Drug particles are modeled as a monodisperse distribution of 50- $\mu\text{m}$  polystyrene spheres that are injected into an agar-based tissue-mimicking material, which they penetrate to a mean depth of 500  $\mu\text{m}$ . A 15-MHz broadband focused immersion transducer, driven in pulse-echo mode, was used to scan the surface of the agar phantom. By assuming that the mean-squared voltage of the received waveform is proportional to the number of particles encountered by the incident pulse, a 3-D map of particle distribution was generated. The resulting acoustic microscopy data were validated by examining phantom slices of thickness similar to the acoustic pulse length under a conventional optical microscope. The acoustic technique was found to provide a good prediction of the mean particle penetration depth and an accurate map of transverse particle distribution in planes normal to the direction of injection. It thus shows great promise as a noninvasive means of assessing the efficacy of needle-free injection, both *in vitro* and *in vivo*.

**1pBB14. Speckle reduction of echographic image by transmitting ultrasonic waves of double peak spectrum.** Iwaki Akiyama (Shonan Inst. of Technol., 1-1-25 Tsujido-nishikaigan, Fujisawa 251-8511 Japan), Shigemi Saito (Tokai Univ., Shimizu-ku, Shizuoka, 424-8610 Japan), and Akihisa Ohya (Univ. of Tsukuba, Tsukuba, Ibaraki, 305-8573 Japan)

Speckle noise of ultrasonic echography is a major issue. To reduce the speckle noise a method of summation after logarithmic compression of higher order harmonic images is proposed. The feasibility is confirmed by the phantom experiments using a mechanical scanner system equipped with a bi-layered structure transducer of PZT and PVDF. The dynamic range of harmonic components higher than the third, however, is not enough for effective speckle reduction. In this study a method of using difference and sum frequency components as well as harmonic components by transmitting pulsed waves of double peak spectrum is proposed. To compare the dynamic range of those components the numerical simulation of transmitting the ultrasonic waves of 2 and 6 MHz and those of 2 and 8 MHz in absorbing media is carried out. An attenuation coefficient of 1.0 dB/cm/MHz is assumed. As a result, complex beam patterns are formed due to degeneration between second harmonic and difference frequency components of 4 MHz in the former case. In the latter case three focal beam patterns of 4-, 6-, and 8-MHz components within 40 dB in dynamic range are formed. Consequently this method has the potential to apply for the biological tissues. [Work supported by a Grant-in-Aid for Scientific Research.]

TUESDAY AFTERNOON, 28 NOVEMBER 2006

OAHU ROOM, 1:00 TO 4:00 P.M.

### Session 1pEA

## Engineering Acoustics, Acoustical Oceanography and Underwater Acoustics: Acoustic Sub-Bottom Profiling Technology

Kenneth M. Walsh, Chair

*K. M. Engineering Ltd., 51 Bayberry Ln., Middletown, RI 02842*

Chair's Introduction—1:00

### Invited Papers

1:05

**1pEA1. Subbottom profiling using synthetic aperture processing.** Steven Schock (Dept. of Ocean Eng., Florida Atlantic Univ., 777 Glades Rd., Boca Raton, FL 33431, schock@oe.fau.edu)

Using synthetic aperture processing, a chirp sonar generates reflection profiles of the seabed. Synthetic aperture processing is implemented by calculating the vertical change in the positions of acoustic transducers between transmission and using the offsets to time shift the acoustic data so that horizontal reflectors are summed coherently. The calculated changes in the positions of the acoustic transducers between transmission events are based on Doppler velocity log and inertial measurement unit measurements of vehicle motion. Imagery generated from reflection profiles of seabeds with various sediment types are used to quantify the improvement in the spatial resolution and the SNR of reflection profiles as a function of the length of the synthetic aperture. [Work supported by Office of Naval Research.]

1:25

**1pEA2. Chirp seismic surveying of shallowly buried fluvial systems on the outer New Jersey continental shelf: Mapping channel morphology and internal stratigraphy in 3D.** John Goff, Sylvia Nordfjord, James Austin, Jr. (Inst. for Geophys., Jackson School of Geosciences, Univ. of Texas, Austin, TX 78759), and Steven Schock (Florida Atlantic Univ.)

The outer New Jersey continental shelf has been the focus of Office of Naval Research-sponsored acoustic studies for more than a decade. As part of this effort, ONR has funded a number of field programs to characterize the geologic history and sedimentary properties of the shallow seabed. We report here on a chirp seismic campaign conducted in 2001, with particular focus on characterizing the morphology and internal stratigraphy of shallowly buried fluvial channel systems that are widespread in this region. These channels were formed during the last glacial lowstand, and then filled and buried by sediments during sea level rise. Mapping the channel morphology in 3D presented a significant challenge. Although coverage (200-m line spacing) was dense, application of

standard interpolation techniques resulted in poor representation of channel continuity along sinuous pathways. We developed a technique for transforming the seismic horizon data into a channel-oriented reference frame, wherein the channel is effectively straight and simple interpolation techniques can be used to maintain channel continuity. Seismic units within the channel fill could be discerned based on bounding reflectors and acoustic character and mapped throughout the 3D channel systems. These units provide a record of sea level change and drowned-river estuary development.

1:45

**1pEA3. Removing multiples and depth imaging and inverting primaries without knowing or determining the velocity model: A response to pressing seismic challenges.** Arthur B. Weglein (Phys. Dept., Univ. of Houston, 4800 Calhoun, Houston, TX 77004, aweglein@uh.edu)

The isolated task specific subseries of the inverse scattering series represents a direct response to many pressing seismic exploration challenges. Those challenges often involve removing multiply reflected events and locating and identifying targets when the subsurface properties and boundaries are rapidly varying and beyond all current methods to adequately predict and estimate. The goal of these fundamentally new concepts and algorithms is to make currently inaccessible targets accessible and accessible targets better defined. We will exemplify these new algorithms with synthetic and field data. Open issues and plans will be discussed.

2:05

**1pEA4. Seismic imaging of subduction zone deformational structures using 3D seismic profiling.** Nathan L. Bangs (Inst. for Geophys., Univ. of Texas, 4412 Spicewood Springs Rd., Austin, TX 78759, nathan@ig.utexas.edu)

In the past 15–20 years, considerable efforts have been made to image complex deformational structures beneath the seafloor at subduction zones using seismic imaging techniques. The goals have been to: 1) accurately image structures to determine the geometry of thrust faults controlling the deformation and tectonics of plate collision at subduction zones, and 2) to determine the rock physical properties from seismic reflections to infer fault zone mechanical properties and examine conditions leading to fault rupture and earthquake generation. Subduction zones generate the world's largest and most destructive earthquakes and tsunamis and accurate seismic images of faults to depths of 6–7 km below seafloor offer an opportunity to examine conditions at these inaccessible depths. Three-dimensional seismic techniques have proven to be critical for sampling the reflected wavefield at the sea surface sufficiently to be able to migrate reflections to their proper subsurface positions and construct accurate images. I will present examples of these 3-D seismic reflection images, analyses of these data, and future directions. [Work supported by NSF.]

2:25

**1pEA5. Application of CHIRP data to neotectonic, sediment dispersal and sea level studies.** Elizabeth Johnstone and Neal Driscoll (Scripps Inst. of Oceanogr., Univ. of California San Diego, La Jolla, CA 92093-0244)

CHIRP seismic data will be presented in order to demonstrate the importance of subbottom imaging in addressing fundamental problems in Earth sciences. Examples from actively deforming regions in California place important constraints on the rupture history along major fault systems as well as tectonic control on sediment preservation and exposure of hardgrounds on the seafloor. Tectonic deformation creates a rich tapestry of geologic substrate exposed on the seafloor, ranging from mobile sands to lithified hardgrounds. It is this heterogeneity of the seafloor that creates different biological ecosystems and consequent increased biodiversity. CHIRP data have also provided new insights into sediment dispersal and clinoform development in the Gulf of Papua. In the northern gulf, oblique clinoforms suggest the rate of sediment supply has outstripped the formation of accommodation, whereas in the south the sigmoidal package indicates that accommodation increased faster than sediment supply. The morphology and internal geometry of the clinoform itself are inherently three-dimensional. The underlying physiography also appears to play an important role in governing sediment dispersal and distribution.

2:45–3:00 Break

### Contributed Papers

3:00

**1pEA6. A multibeam subbottom profiler.** Jerald W. Caruthers (Univ. of Southern Mississippi, 1020 Balch Blvd., Stennis Space Center, MS 39529) and Sean Griffin (Omni Technologies, Inc., Stennis Space Center, MS 39529)

A multibeam subbottom profiling system (MSBPS) is being built with funding from the Defense University Research Instrumentation Program (DURIP) in coordination with the Office of Naval Research. The MSBPS is being constructed from three T70 Neptune Sonar transducer modules in a line. Each T70 has five 11.5-kHz elements in each of two parallel lines. The resulting system is composed of 15 elements in each line with a total length of 1.3 m separated by about 8.5 cm. Each line has full beamwidths to the 3-dB points of about  $5 \times 70$  deg. The system will be used in an interferometric mode between the two lines. [This work is supported by ONR 3210A.]

3:15

**1pEA7. Using ambient noise and vector sensor arrays to form a subbottom profiler.** Paul Hursky and Martin Siderius (Heat, Light, and Sound Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130)

Surface-generated ambient noise is distributed in time, but its correlation produces a strong impulse that is excellent for imaging the ocean floor. This technique has been experimentally demonstrated using vertical line arrays of pressure phones. The up-looking endfire beam (for ambient noise traveling down) is cross-correlated with the down-looking endfire beam (for the ambient noise reflected from the bottom, traveling up). Each layer in the bottom produces a distinct reflection. Previous work has required somewhat large vertical arrays. A towed horizontal array would be much better and allow us to make measurements with a mobile platform. The problem is that a horizontal line array cannot distinguish up from down—a broadside beam is symmetric with respect to the axis of the array. Vector sensors, sensing both pressure and particle velocity, have additional degrees of freedom for overcoming such ambiguities. We will show how vector sensors make it possible to implement such an ambient



noise subbottom profiler, either in a greatly reduced vertical array or in a horizontal (towed) array configuration. We will also show results of processing data collected on a four-element vector sensor array and a much larger pressures-only array during the Makai experiment.

3:30

**1pEA8. Development of practical broadband magnetohydrodynamic transducers for undersea digital communication.** Eric Berkenpas and Corey Jaskolski (Hydro Technologies, 700 Automation Dr., CO 80550)

Demand for broadband underwater digital communication is being driven by development of tetherless autonomous underwater vehicles (AUVs). Current communications methods employ typically piezoelectric transducers. While efficient, and broadband as receivers, these transducers exhibit significant frequency dependence when transmitting, thus limiting their utility in digital communications. Magnetohydrodynamic (MHD) transducers generate acoustic waves in seawater by applying time-varying Lorentz force to moving ionic charges in the presence of a magnetic field. MHD transducers are ultracompliant, exhibit wide-frequency bandwidth, and are therefore attractive for high-speed undersea digital communications. This paper reports on exploratory development MHD transducers with the goal of improved transduction efficiency and bandwidth. Novel MHD transducer structures were fabricated with permanent magnet, parallel electrode structures, and planar interdigitated electrode structures. To evaluate transducer performance, an automated transducer test setup was assembled incorporating a 7500-L saltwater test environment, a time-gated high-current signal source, and a wideband hydrophone. Use of acoustic

baffles and waveguides was also investigated. A theoretical discussion is also included. This work represents the practical employment of MHD transducers to undersea digital communication. [Work supported by General Atomics.]

3:45

**1pEA9. Influence of nonlinear propagation on calibration of hydrophone sensitivity using a two-transducer reciprocity method.** Masahiro Yoshioka and Sojun Sato (Natl. Metrology Inst. of Japan, Natl. Inst. of Adv. Industrial Sci. and Technol., Tsukuba Central 3, 1-1-1 Umezono, Tsukuba, Ibaraki, 305-8563 Japan)

This paper describes the phenomenon by which the measured value depends on the ultrasound pressure in the measurement field when the hydrophone sensitivity has been calibrated using the two-transducer reciprocity method. This dependence is attributable to nonlinear ultrasonic propagation. However, the international standard, IEC 60866:1987, "Characteristics and calibration of hydrophones for operation in the frequency range 0.5 MHz to 15 MHz," provides a correction factor for diffraction and attenuation derived assuming linear propagation of sound. We attempted to derive the correction factor from the ultrasound pressure distribution calculated using the Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation, which is well known as a nonlinear ultrasonic model. Results demonstrated that the dependence of the calibrated sensitivity value on measured sound pressure is quantitatively explainable. The hydrophone sensitivity can be calibrated properly by compensating for the influence of ultrasonic nonlinear propagation using the KZK equation.

TUESDAY AFTERNOON, 28 NOVEMBER 2006

IAO NEEDLE/AKAKA FALLS ROOM, 1:30 TO 4:15 P.M.

### Session 1pMU

#### Musical Acoustics: Singing Voices

Peter L. Hoekje, Cochair

*Baldwin-Wallace College, Dept. of Physics and Astronomy, 275 Eastland Rd., Berea, OH 44017*

Naotoshi Osaka, Cochair

*Tokyo Denki Univ., Dept. of Information Systems and Multimedia Design, School of Engineering, 2-2 Kandanishiki-cho, Chiyoda-ku, Tokyo, 101-8457, Japan*

#### Contributed Papers

1:30

**1pMU1. Voice quality of artistic expression in Noh: An analysis-synthesis study on source-related parameters.** Hideki Kawahara (Faculty of Systems Eng., Wakayama Univ., 930 Sakaedani, Wakayama, 640-8510 Japan), Osamu Fujimura (Intl. Inst. for Adv. Studies, Sorakugun, Kyoto 619-0225 Japan), and Yasuyuki Konpaku (Konparu School of Noh, Nara, Japan)

High-quality recordings of Noh vocalizations, produced by one of the authors, were made in an anechoic chamber, comparing vocal styles and emotional expression, using the same sample text from classical Noh. The recorded signals were subject to analysis, synthesis, and morphing by using STRAIGHT. The morphing procedure generated mean parameter settings for comparing two subtly differentiated voicing styles, traditionally called "wagin" soft vocalization, and "gougin," strong vocalization, to investigate the relation between signal parameters and communicative functions. STRAIGHT extracts three parameters from recorded speech signals: fundamental frequency, smoothed spectrogram, and a time-frequency aperiodicity map. The fundamental frequency and the time-frequency aperiodicity, in other words, source-related parameters, were systematically manipulated to investigate their role in artistic expression.

1:45

**1pMU2. Spectro-temporal analysis of Palauan chants.** Birgit Abels (Ruhr Univ., Imam Husnu Sok. 1/1, IStambul Bochum, Germany) and Jonas Braasch (Rensselaer Polytechnic Inst., NY)

A field study was carried out to determine how songs from the repertoire of chants in Palau have changed over the last century. To this end, several chanters were recorded during recent fieldwork; these sound documents were then compared to items from several historic collections of sound recordings from the Phonogram Archives Berlin, Germany, and the Pacific Collection, Library of the University of Hawai'i at Manoa. The available recordings cover the time span from 1909–2006. Analytical results were derived in the form of spectrograms both for recent recordings from the years 2005/2006 and historic ones from 1963. They confirmed the auditory impression that chanting in Palau today has moved away from a recitative style, relying predominantly on the speaking voice, towards a performance practice the sound ideal of which is the singing voice. Also, the range of pitch employed in these chants has expanded. Spectral analysis also revealed how, in particular genres, the musical form of group chants is structured by the changing contours of a frequency band that is

resolved at phrasal ends, when the singers narrow the range of this tonal cluster. This unique technique is one of the core musical elements of the Palauan singing tradition.

2:00

**1pMU3. Spectral modification technique in conversion of musical notes or tempos for singing voice synthesis system.** Hideki Banno, Kumi Ohta, Masato Kawazoe, and Fumitada Itakura (Faculty of Sci. and Technol., Meijo Univ., 1-501, Shioyamaguchi, Tenpaku-ku, Nagoya-shi 468-8502, Japan, banno@ccmfs.meijo-u.ac.jp)

The STRAIGHT time-frequency representations (spectrograms) of singing voice signals in various musical notes and various tempos are observed to develop a high-quality synthesis system of singing voice. The spectrogram of STRAIGHT, which is a very high-quality analysis-synthesis system, can represent the vocal tract information accurately. A conversion system of a musical note or a tempo of an input singing voice signal has been implemented based on the observation. As a result of the observation, the frequency warping of the STRAIGHT spectrogram based on a DP matching algorithm has been introduced into the system. It was found that the method using a differential of a smoothed spectrum as a spectral distance measure in the frequency warping produces subjectively better quality than that using a smoothed spectrum directly. It was also found that the method without spectral modification, i.e., only with pitch/tempo modification in the conversion, produces better quality than that using a differential of a smoothed spectrum. This can be caused by the destruction of naturalness in the method using a differential of a smoothed spectrum.

2:15

**1pMU4. Physical modeling vocal synthesis.** David Howard, M. Jack Mullen, and Damian Murphy (Audio Lab., Intelligent Systems Res. Group, Dept. of Electron., Univ. of York, Heslington, York YO10 5DD, UK. dh@ohm.york.ac.uk)

Physical modeling music synthesis produces results that are often described as being “organic” or “warm.” A two-dimensional waveguide mesh has been developed for vocal synthesis that models the adult male oral tract. The input is either the LF glottal source model or a user-provided waveform file. The mesh shape, which is based on MRI data of a human vocal tract, can be changed dynamically using an impedance well approach to allow sounds such as diphthongs to be synthesized without clicking. The impedance well approach enables the mesh shape to be varied without removing or adding elements, actions that cause audible clicking in the acoustic output. The system has been implemented as a real-time MIDI-controlled synthesizer, taking its inspiration from the von Kempelin speaking machine, and this will be demonstrated live as part of the presentation. The system is set up to allow a continuous glottal source to be applied that includes vibrato, and thus the real-time output it currently produces is close to being vocalized. It should be noted, though, that appropriate variation of the lip opening does produce a voiced bilabial plosive, demonstrating the potential for moving towards a full speech synthesis system in the future.

2:30

**1pMU5. Analysis of acoustic features affecting singing-voice perception and its application to singing-voice synthesis from speaking-voice using STRAIGHT.** Takeshi Saitou (ATR Cognit. Information Sci. Lab, 2-2-2 Hikaridai Seika-cho Souraku-gun, Kyoto 619-0288, Japan), Masashi Unoki, and Masato Akagi (School of Information Sci. Japan Adv. Inst. of Sci. and Technol., Nomi Ishikawa 923-1292 Japan)

A singing-voice synthesis method that can be transformed from a speaking voice into a singing voice using STRAIGHT is proposed. This method comprises three sections: the F0 control model, spectral sequence control model, and duration control model. These models were constructed by analyzing characteristics of each acoustical feature that affects singing-voice perception through psychoacoustic experiments. The F0 control

model generates a singing-voice F0 contour through consideration of four F0 fluctuations: overshoot, vibrato, preparation, and fine (unsteady) fluctuation that affect the naturalness of a singing voice. The spectral sequence control model modifies the speaking-voice spectral shape into a singing-voice spectral shape by controlling a singer’s formant, which is a remarkable peak of a spectral envelope at around 3 kHz, and amplitude modulation of formants synchronized with vibrato. The duration control model stretches the speaking-voice phoneme duration into a singing-voice phoneme duration based on note duration. Results show that the proposed method can synthesize a natural singing voice, whose sound quality resembles that of an actual singing voice.

2:45–3:00 Break

3:00

**1pMU6. Automatic detection of head voice in sung musical signals via machine learning classification of time-varying partial intensities.** Ryan J. Cassidy and Gautham J. Mysore (Dept. of Music, Ctr. for Comput. Res. in Music and Acoust., 541 Lasuen Mall, Stanford, CA 94305)

The automatic detection of portions of a musical signal produced according to time-varying performance parameters is an important problem in musical signal processing. The present work attempts such a task: the algorithms presented seek to determine from a sung input signal which portions of the signal are sung using the head voice, also known as falsetto in the case of a male singer. In the authors’ prior work [Mysore *et al.*, Asilomar Conf. Signal. Sys. Comp. (2006) (submitted)], a machine learning technique known as a support vector classifier [Boyd and Vandenberghe, 2004] was used to identify falsetto portions of a sung signal using the mel-frequency cepstral coefficients (MFCCs) of that signal (computed at a frame rate of 50 Hz). In the present work, the time-varying amplitudes of the first four harmonics, relative to the intensity of the fundamental, and as estimated by the quadratically interpolated fast Fourier transform (QIFFT) [Abe and Smith, ICASSP 2005], are used as a basis for classification. Preliminary experiments show a successful classification rate of over 95% for the QIFFT-based technique, compared to approximately 90% success with the prior MFCC-based approach. [Ryan J. Cassidy supported by the Natural Sciences and Engineering Research Council of Canada.]

3:15

**1pMU7. A proposal of discrimination method between voice and music using Gaussian mixture models and similarity in a music selection.** Takuma Yoshida, Masaaki Ishigame, Yoshiaki Ito, and Kazunori Kojima (Iwate Prefectural Univ., Sugo 152-52, Takizawa, Iwate, 020-0193, Japan, g231e038@edu.soft.iwate-pu.ac.jp)

We have proposed a method for automatic music boundary detection using similarity in a music selection. The method is able to capture a whole position of a consecutive music selection, although the boundary position of a music selection is likely to be roughly estimated. In this paper, we propose a new method that combines a GMM (Gaussian mixture model) discrimination method between music and voice with the previously proposed method. The GMM enables us to determine strict boundary positions. On the other hand, the GMM often misdetects boundaries that are not actual boundaries, such as changing points of music instruments, the point of modulation in a music selection, and so on. We, therefore, exclude the GMM misdiscrimination in a music selection by the previously proposed method, and also realize the precise detection of music boundaries by the GMM. We conducted various experiments using open music selections that are provided by RWC, and the results showed the proposed method could improve the performance of music boundary detection, compared with the previous method and the method using only the GMM.

3:30

**1pMU8. Acoustical analysis of nonlinguistic sounds.** Yell Inverso, Corine Bickley, and Charles Limb (Gallaudet Univ., 800 Florida Ave. NE, Washington, DC 20011, danielle.inverso@gallaudet.edu)

The perception of speech is an undisputed important goal for cochlear implantation; however, the reception of nonlinguistic sounds (NLS) is also important. Nonlinguistic sounds are important for one's safety and they are considered to have importance to a person's sense of connection to and welfare in one's environment. NLS are different from speech, and the perception and acoustic characteristics of NLS have not been studied adequately to allow clinicians to fit CIs for optimal recognition of both speech and NLS. The specific categories of nonlinguistic sounds to be evaluated in this study are (1) human vocal/nonverbal, such as a baby crying or a person coughing; (2) mechanical/alerting, such as a telephone or an alarm clock buzzing; (3) nature/inanimate, such as weather sounds; (4) animal/insect, such as a dog barking; and (5) musical instruments such as the strum of a guitar. In this ongoing study, 50 listeners (half adults with normal hearing and half adult post-ling deafened CI users) are being evaluated with respect to their ability to identify the sources of a variety of common, everyday NLS. In this poster, an acoustic analysis of a representative set of the stimuli will be presented.

3:45

**1pMU9. A psycho-acoustical analysis of vocal effort by opera singers.** Konrad Skirlis, Densil Cabrera, and Anna Connolly (P.O. Box A2106, Sydney South 1235 NSW, Australia, audiodarnok@yahoo.com)

A singer's vocal output will be affected by a number of factors such as the size of the auditorium used for the performance. In this paper, the "large hall" and "small hall" voices of singers were explored in terms of the specific loudness pattern, which represents the distribution of loudness

in the auditory system. Recordings of seven professional opera singers were made in an anechoic room and in a small recital hall. Singers were asked to imagine they were in a small hall for one set of recordings and in a large hall for another set. Each singer sang the same song under these two "conditions." The recordings were then analyzed to determine spectral and temporal changes. Spectral analysis was used to calculate the strength of the singers formant under these conditions, and this was compared with the specific loudness patterns. Results help us to understand why the singer's formant is so important for vocal projection.

4:00

**1pMU10. The effect of seating arrangement on choral sound in a nonselected mixed collegiate choral ensemble.** Wei-Chun Wang (St. Johns Univ., 499, Sec. 4, Tam King Rd., Tamsui, Taipei, 25135 Taiwan, R. O. C.)

With both macroarrangement concerns and microarrangement concerns, choral directors use singer's placement as an essential technique to find compatible voices by listening to individual voices in various combinations. This study investigated the effect of seating arrangements, musical textures, and language familiarity on the choral balance and choral blend in an inexperienced collegiate choral ensemble. After ten rehearsals directed by the investigator, 48 nonselected choristers of a curricular SATB choral ensemble sang four musical excerpts, with the combination of two musical textures (homophonic and polyphonic compositions) and two types of languages (native and foreign language), a cappella in three formations (a sectional block, a mixed quartet, and a hybrid formation) in a chapel. Five experienced judges evaluated the screened live performance by rating the choral balance and choral blend on seven-point scales. The process was repeated. For blend only, sectional and hybrid formations were favored for polyphonic excerpts. The excerpts sung in native language were rated higher than the foreign excerpts, in particular for blend.

TUESDAY AFTERNOON, 28 NOVEMBER 2006

NIIHAU ROOM, 1:00 TO 4:40 P.M.

## Session 1pPAa

### Physical Acoustics and Engineering Acoustics: Infrasound

Milton A. Garces, Chair

*Univ. of Hawaii, Infrasound Lab., 73-4460 Queen Kaahumanu Hwy., Kailua-Kona, HI 96740-2638*

Chair's Introduction—1:00

#### Invited Papers

1:05

**1pPAa1. Contribution of continuous infrasound monitoring to scientific research.** Alexis Le Pichon, Karl Antier (CEA/DASE BP12, Bruyeres-le-Chatel, France), Lars Ceranna (Federal Inst. of Geosciences and Natural Resources, 30655 Hannover, Germany), Doug Drob (NRL, Washington, DC 20375), and Michel Lardy (IRD Ctr., 98848 Noumea Cedex, New-Caledonia)

Continuous automatic processing of the data is being performed at the French NDC from a number of IMS-type infrasound stations in order to detect and characterize coherent infrasonic waves recorded on the arrays. A large number of detections are associated with natural phenomena. Infrasonic waves generated by strong earthquakes occurring within region of high mountains are of great interest. Such events allow the evaluation of the relative contributions of the different source mechanisms involved in large earthquakes, and offer an opportunity to improve our understanding of the amplification of ground displacement caused by the topography. Microbaroms as well as mountain waves are quasipermanent sources of infrasonic waves detected on a global scale. Their monitoring reveals clear periodic trends in the detected bearings and signal amplitude, providing further confirmation that long-range propagation strongly depends on the atmospheric conditions. Signals from volcanic activities also offer a unique opportunity for atmospheric studies. At large propagation ranges from the volcanoes, infrasound measurements can be used as input of inversion procedures to evaluate more precisely the vertical structure of the wind in a range of altitude inaccessible to ground-based or satellite measurements.

**1pPAa2. Infrasound associated with stratospheric ash injection, window-shattering explosions, and incandescent avalanches at Tungurahua Volcano, Ecuador.** Milton Garces, David Fee, Sara McNamara (Infrasound Lab., Univ. of Hawaii, Manoa, 73-4460 Queen Kaahumanu Hwy., #119, Kailua-Kona, HI 96740-2638, Milton@isla.hawaii.edu), David McCormack (Geological Survey of Canada, Ottawa, ON K1A 0Y3, Canada), and Henry E. Bass (NCPA, University, MS 39688)

On July 14, 2006, Tungurahua volcano in Ecuador initiated a dramatic eruptive sequence that caused the evacuation of villagers, the destruction of crops, and the reports of ash clouds above 15 km. Acoustic signals have been recorded by the prototype infrasound acoustic-surveillance system (ASHE) currently operating in Ecuador. The infrasonic signals recorded by the ASHE included explosions, pyroclastic flows, and a sustained low frequency vibration of the ground. We present the features of the different types of eruption signals, and discuss the physics of eruptive processes and volcanic ash injection into the atmosphere. We also compare the July 2006 ash-rich eruption sequence with the May 2006 eruptive episode, which injected very little ash into the atmosphere. Results are of significance for the future operational use of acoustic monitoring of volcanic activity for aviation safety.

1:45

**1pPAa3. Infrasound from tornadoes and other severe storms.** John M. Noble (U.S. Army Res. Lab., 2800 Powder Mill Rd., Attn: AMSRD-ARL-CI-ES, Adelphi, MD 20783, jnoble@arl.army.mil) and Alfred J. Bedard, Jr. (NOAA/Earth System Res. Lab./PSD/CIRES, Boulder, CO 80305)

Infrasound arrays have been used for a long time to detect and monitor man-made and natural acoustic sources below 20 Hz. Much of the earlier work on infrasound from severe weather was below 0.1 Hz and was related to severe weather penetrating the tropopause. It was noticed when violent thunderstorms passed near an array that a coherent signal would be detected coming from the same direction as the storm. By correlating the measurements with visual, satellite, and Doppler radar observations, the characteristics of the storm could be studied to determine what type of storm would cause the production of infrasound. Most of the storms that would produce an infrasonic signal were the strong supercell-type storms that produced intense gust fronts and tornadoes. More recent measurements at frequencies above 0.1 Hz have been associated with tornadoes and tornadic storms. Following on these observations, NOAA established an infrasonic network to evaluate the use of infrasound for tornado detection and warning. This network has operated since 2003 and is still being optimized and evaluated. This talk will cover the studies that have been conducted and numerical models developed to try and explain this phenomenon.

2:05

**1pPAa4. High-altitude infrasound propagation experiment.** Bill Andre (U.S. Army Space and Missile Defense Command, P.O. Box 1500, Huntsville, AL 35807, andreb@smdc.army.mil) and Henry E. Bass (Nat'l. Ctr. for Physical Acoust., University, MS 38677)

Experimental verification of propagation predictions and sensor performance for infrasound frequencies that propagate long distance (greater than 100 km) are difficult to perform due to the rarity of controlled sources. If the source is elevated, the lower atmospheric pressure allows an explosion of a given charge weight to expand further during the supersonic phase, resulting in a lower fundamental frequency. A 50-lb charge exploded at 40 km yields a fundamental frequency similar to a 500-ton explosion at ground level. During the fall of 2005 and the winter and summer of 2006, experiments were conducted at White Sands Missile Range, NM, where an Orion rocket lifted a 50-lb explosive charge to altitudes between 30 and 40 km. Launches occurred near 0200 and 0600 MST. Portable infrasound arrays to record events were deployed in New Mexico, Texas, Arizona, and California to complement permanent arrays. Infrasound recordings were supplemented with extensive meteorological measurements. Data analysis is in a preliminary stage. The High Altitude Infrasound Propagation Experiment resulted from collaboration between several U.S. Army and Navy commands, University of Mississippi, Southern Methodist University, University of Hawaii, University of California at San Diego, University of Alaska at Fairbanks, BBN Technologies, SAIC, and Los Alamos National Laboratory.

### Contributed Papers

2:25

**1pPAa5. Modeling the propagation of volcanic infrasound with a 3-D parabolic equation.** Joseph F. Lingeitch, Geoffrey F. Edelmann, Douglas P. Drob, and Michael D. Collins (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375)

Active volcanoes are significant sources of infrasound in the 1–4-Hz frequency band and are detectable at ranges of hundreds of kilometers. The propagation is affected by the atmospheric temperature, wind, and attenuation profiles, which are range and time dependent. In this paper, a 3-D atmospheric parabolic equation method that accurately incorporates the effects of wind is applied to modeling volcanic infrasound propagation. A parabolic equation is derived using a narrow angle operator approximation in azimuth and a wide angle approximation in depth to account for wind-induced refraction that can trap or disperse sound energy, leading to large variations of propagation with wind conditions. The atmospheric specification for this model is derived from the attenuation coefficients of Sutherland and Bass [J. Acoust. Soc. Am. **115**, 1012–1032 (2004)] and the temperature/wind profiles from the Naval Research Laboratory Ground to Space model [J. Geophys. Res. **108**, 4680–4691 (2003)], which provides global atmospheric profiles from the ground to 200 km

altitude updated four times daily. Signal amplitude and direction of arrival estimates from the model are compared with measurements of volcanic infrasound at listening stations several hundred kilometers from the source. [Work supported by ONR.]

2:40

**1pPAa6. Infrasonic scattering studies using time-domain parabolic equation (TDPE) and gravity wave models.** David E. Norris (BBN Technologies, 1300 N. 17th St, Ste. 400, Arlington, VA 22209)

The time-domain parabolic equation (TDPE) model is useful in predicting infrasonic waveforms, as it can account for diffraction and scattering mechanisms. In this study, a split-step Fourier (SSF) implementation of the TDPE is used to study infrasonic propagation. Scattering is modeled using a spectral gravity wave model, and multiple gravity wave realizations are used to quantify the uncertainty in waveform properties introduced by scattering from the gravity wave inhomogeneities. TDPE predictions are compared to specific infrasonic events and conclusions regarding the dominant propagation mechanisms are presented.

**1pPAa7. The deep sound of one wave plunging.** Milton Garces, David Fee, Sara McNamara (Infrasound Lab., Univ. of Hawaii, Manoa, 73-4460 Queen Kaahumanu Hwy., #119, Kailua-Kona, HI 96740-2638), Jerome Aucan, and Mark Merrifield (Univ. of Hawaii, Manoa, Honolulu, HI 96822)

The ability to provide infrasound estimates of breaking ocean wave height and period in shallow reefs, steep rocky coastlines, and sand beaches has been demonstrated in previous work. Yet defining the source process and isolating the source pressure function remained elusive because of ambiguity introduced by complex coastlines and multiple breaker zones. Due to the steep bathymetry and its proximity to land, the Temae reef in the northeast coast of Moorea island, French Polynesia, provided a well constrained experimental environment where individual breaking waves could be identified and recorded. Synchronous wave height, infrasonic, seismic, and visual recordings of individual waves breaking against the shallow reef ledge were made and correlated. We characterize a possible source mechanism for surf infrasound and demonstrate the capability to acoustically track alongshore traveling (peeling) plunging waves. We also introduce preliminary results on remote infrasonic monitoring of the surf zone in the North Shore of Hawaii during the 2006-07 Winter high surf season.

### 3:10–3:25 Break

#### 3:25

**1pPAa8. Low-frequency acoustic propagation in a shallow water internal wave field.** Geoffrey F. Edelmann, Joseph F. Lingeitch, and David C. Calvo (U.S. Naval Res. Lab., 4555 Overlook Ave. SW, Code 7145, Washington, DC 20375)

A fully 3-D parabolic equation is used to simulate low-frequency acoustic propagation through a nonlinear internal wave field. Previous simulations have shown that acoustic propagation is dependent on the relative orientation between the direction of the internal wave train and the acoustic signal. When the acoustic and internal waves are parallel, a simple 2-D vertical refraction is the dominant interaction. As the two waves become perpendicular, horizontal refraction of the acoustic signal due to azimuthal coupling has been suggested [Frank *et al.*, *J. Acoust. Soc. Am.* **118**, 723–734 (2005)]. Comparisons of 3-D acoustic propagation through an internal wave field with and without azimuthal coupling will be made. The parabolic equation used makes a narrow angle operator approximation in azimuth and a wide angle approximation in depth. We will extend the previous work such that coupling in both the horizontal and vertical direction will be calculated simultaneously. [This work is supported by the Office of Naval Research.]

#### 3:40

**1pPAa9. Infrasonic wave observations at Fairbanks, AK, of Augustine volcano eruptions during January 2006.** Charles Wilson, Steve McNutt, John Olson, and Guy Tytgat (Geophysical Inst., Univ. of Alaska, Fairbanks, AK)

The recent Augustine eruptions, from the 11th to the 28th of January 2006, have produced a series of ten infrasonic signals observed at the I53US array. The eruption times for the signals were provided by the Alaska Volcanic Observatory at UAF using a Chaparral microphone present on Augustine Island a few kilometers from the crater. The bearing and distance of Augustine from I53US are respectively 207.8 deg and 675 km. The analysis of the signals is done with a least-squares detector/estimator that calculates, among other things, the horizontal trace-velocity and the azimuth of arrival of the signal. The average values of the trace-velocity and azimuth for all ten Augustine signals are  $0.336 \pm 0.0136$  km/s and  $208.7 \pm 1.5$  deg. respectively. The celerity for each signal was calculated using the range 675 km and the individual travel times to I53US. The average celerity for all ten eruption signals was  $0.287 \pm 0.018$  km/s. Ray tracing studies, using wind speed and temperature profiles supplied by Dr. Doug Drob at NRL, have shown that both stratospheric and thermospheric ray paths are present in the data set.

**1pPAa10. Infrasound signals detected by the observation system in Isumi, Japan.** Takahiko Murayama, Motohiro Honma, Mami Nogami, Takayuki Otsu, Takafumi Amano, Yasuharu Saitoh, Nobuo Arai, and Toyomi Sakamoto (Res. Dept., Japan Weather Assoc., 3-1-1 Higashi-Ikebukuro, Toshima-ku, Tokyo 170-6055 Japan)

An infrasound observation system is installed in Isumi, Japan (35.3N, 140.3E) as the component of the International Monitoring System for CTBT. The observation system was deployed on November 2004. It comprises a six-element array. Five elements of the array are deployed as a pentagon with sides approximately 1.2 km long. The sixth element is located within the pentagon-shaped array. Until now, some interesting infrasound signals were observed. One of them was the infrasound wave produced by the large earthquake that occurred off the Pacific coast of North-East Japan on 16 August 2005. The infrasound wave-train, which had a spindle-shaped envelope, was detected following the arrival of the seismic wave. The arrival direction of the infrasound wave, which had maximum amplitude, indicated the direction of the land area, which had maximum intensity. The time difference between the on-set time of that infrasound wave and the origin-time of the earthquake coincided with the travel time for propagation in the atmosphere between the area that had maximum intensity and the observation site. Therefore, it was estimated that the observed infrasound waves were produced by shaking ground. In this paper, some remarkable infrasound signals including the above are discussed.

#### 4:10

**1pPAa11. Infrasound wind noise contributions.** Richard Raspet, Jiao Yu, and Jeremy Webster (Univ. of Mississippi, University, MS 38677)

In a recent study [*J. Acoust. Soc. Am.* **119**(2), 834–843 (2006)] predictions of various noise contributions were compared to measured wind noise from various devices and windscreens. The fluctuating stagnation pressure, the turbulence-turbulence interaction pressure, and the self-noise contributions in the inertial range were calculated from the measured wind velocity spectrum and compared to measurements. In that study the bare microphone agreed well with the predicted stagnation pressure spectrum, and a heavily windscreened microphone agreed well with the predicted turbulence-turbulence interaction spectrum. In this paper we extend the analysis and the measurements into the source region of the spectra—well into the infrasound range. In addition we predict the spectral contribution due to the turbulence-mean shear interaction. The prediction displays a steep slope in the inertial range and may dominate at low frequencies. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

#### 4:25

**1pPAa12. Comparative study of wind noise reduction devices.** Jeremy Webster, Richard Raspet, and Jiao Yu (Univ. of Mississippi, University, MS 38677)

Spherical porous foam windscreens are commonly used to reduce wind noise in acoustic measurements. We have argued [*J. Acoust. Soc. Am.* **119**(2), 834–843 (2006)] that the wind noise measured outdoors will lie between the predicted fluctuating stagnation pressure generated by the microphone interaction with the incoming air flow for bare microphones and the turbulence-turbulence interaction pressure for a very large wind screen. The pressure level measured in a moderate size screen will lie between these two extremes. Different transducer geometries and arrangements have been used in aerodynamic and meteorological studies to measure the “fluctuating static pressure” without interference from transducer generated wind noise. In this paper we compare measurements of wind noise using bare microphones, microphones in foam windscreens, streamlined probes, the Quad Disk, and microphones mounted in flat plates. The purpose of this study is to clarify the principles involved in wind noise generation by transducer-turbulence interaction. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

## Session 1pPAb

## Physical Acoustics: Sound Speeds, Phonons, and Thermodynamics of Condensed Matter

Keiji Sakai, Cochair

Univ. of Tokyo, Inst. of Industrial Science, 4-6-1 Komaba, Meguro-Ku, Tokyo, 153-8505, Japan

Cristian Pantea, Cochair

Los Alamos National Lab., MST-NHMFL, MS E536, Los Alamos, NM 87545

Chair's Introduction—1:00

## Invited Papers

1:05

**1pPAb1. Determination of elastic properties and interatomic force constants using inelastic neutron scattering.** Robert McQueeney (Dept. of Phys. and Astron., Ames Lab., Iowa State Univ., Ames, IA 50011)

Neutron scattering is the most powerful technique for measuring vibrational excitation spectra. With single-crystal samples, the complete phonon dispersion relation can be measured. At small phonon wavelengths, the slope of the dispersion curve is related directly to the elastic constants of the material. Information from shorter wavelength phonons can be used to determine parameters for interatomic force models or compared to first-principles calculations. Model parameters can then be used to calculate phonon densities-of-states and consequently connect to the thermodynamics. Polycrystalline samples yield less information, but the data is directly related to the phonon density-of-states for Bravais lattices. Measurements on polycrystalline samples have the advantage of being fast and are more useful for parametric studies of the lattice dynamics (with temperature, for example). These ideas will be used to discuss several interesting systems, most notably Pu metal and Fe-Ga magnetostrictive alloys. [This work is supported by the U.S. Department of Energy.]

1:25

**1pPAb2. Elastic constants, Blackman diagrams, and new lanthanide-actinide insights.** Hassel Ledbetter and Albert Migliori (Los Alamos Natl. Lab., Los Alamos, NM 87545, hledbet@lanl.gov)

Using monocrystal elastic constants and Blackman diagrams, one can infer material interconnections, physical-property trends, and knowledge about interatomic bonding. After reviewing briefly the fcc-metal case, focus shifts to Ce, Pu, Th, with known monocrystal fcc elastic constants, the  $C_{ij}$ . For fcc U and a few other cases, one can deduce the  $C_{ij}$ . From these results about the actinides, there follow several conclusions: (1) elastic anisotropy ranges widely; (2) Poisson ratios are low; (3) interatomic bonding varies widely; (4) all show unusual *negative* Cauchy pressures; (5) a strong covalent-bonding component occurs, perhaps related to localized *f*-electrons; (6)  $\delta$ -Pu behaves most oddly; (7)  $\delta$ -Pu shares some bonding features with fcc alkali metals. [Work supported by DOE, NSF, State of Florida.]

1:45

**1pPAb3. Elastic response of exotic ferromagnets?** Veerle Keppens, Yanbing Luan, Sriparna Bhattacharya, Ivan Sergienko (The Univ. of Tennessee, Knoxville TN 37934), David Mandrus, and Brian Sales (Oak Ridge Natl. Lab., Oak Ridge, TN 37831)

Resonant ultrasound spectroscopy (RUS) measurements have been carried out as a function of temperature (2–300 K) and magnetic field (0–2 T) for a series of exotic ferromagnetic materials, including  $\text{Eu}_8\text{Ga}_{16}\text{Ge}_{30}$  ( $T_c \approx 32$  K),  $\text{EuFe}_4\text{Sb}_{12}$  ( $T_c \approx 82$  K),  $\text{Au}_4\text{V}$  ( $T_c \approx 46$  K), and  $\text{Yb}_{14}\text{MnSb}_{11}$  ( $T_c \approx 53$  K). The different materials show remarkable differences in their elastic response near the Curie temperature, revealing a sudden softening at  $T_c$  in  $\text{EuFe}_4\text{Sb}_{12}$ , and a stiffening of the elastic moduli at  $T_c$  in  $\text{Yb}_{14}\text{MnSb}_{11}$ . Ongoing attempts to theoretically model the data using Landau theory will be discussed. [Work was supported by the National Science Foundation under Grant No. DMR0506292. Oak Ridge National Laboratory is managed by UT-Batelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.]

2:05

**1pPAb4. Measuring elastic constants of novel materials and thin films using small sample resonant ultrasound spectroscopy.** J. D. Maynard (The Penn State Univ., University Park, PA 16802)

One of the most important undertakings for materials is the measurement of the elastic constants. As derivatives of the free energy with respect to atomic displacements, the elastic constants are closely connected to the thermodynamic properties of the material, including the specific heat, the Debye temperature, the Grüneisen parameter, etc. The elastic constants are a sensitive probe of the lattice environment in which all solid state phenomena occur, particularly in the vicinity of a phase transition. Novel materials to which RUS might be applied are often available only in small samples, and for these cases we have developed a small sample RUS

apparatus that can measure bulk samples only a few hundred microns in size and thin films a few hundred nanometers in thickness. Thin films are particularly interesting because substrate induced strains in the lattice may produce dramatic changes in the elastic and thermodynamic properties. In this talk, small sample RUS measurements on the new superconductor magnesium diboride and colossal magnetoresistance materials, in bulk and thin film forms, will be discussed.

2:25

**1pPAb5. Ultrasonic studies of hydrogen absorbing materials: Quasicrystals, nanocrystals, and intermetallic compounds.**

Robert G. Leisure (Dept. of Phys., Colorado State Univ., Fort Collins, CO 80523-1875)

Resonant ultrasound spectroscopy was used to study several materials—nanocrystalline palladium, quasicrystalline Ti-Zr-Ni, and a number of intermetallic compounds—that have the ability to absorb considerable amount of hydrogen, and thus are of potential use for hydrogen storage. Elastic energies are relevant to the thermodynamics of such systems, which are often described by a lattice gas model. Elastic interactions between dissolved H atoms lead to a critical temperature and precipitation of a hydride phase below the critical temperature. The energetics of hydride precipitation depend on the elastic constants. Thus, it is important that the elastic properties of such systems be measured. In addition, these materials are of interest for several other reasons. The elastic constants and internal friction for the hydrogen-free materials were measured over a temperature range of 3 to 300 K, and in a few cases to higher temperatures. These results will be discussed and interpreted. [This work was supported by the U.S. National Science Foundation under Grant DMR-0070808.]

2:45

**1pPAb6. Osmium's full elastic tensor between 5 K and 300 K.** Cristian Pantea (Los Alamos Natl. Lab., Los Alamos, NM 87545)

Reported values of the zero-pressure bulk modulus of Os are both above and below that of diamond, creating a controversy as to which is the least compressible material. This work is another such study, but using two measurement techniques on two types of Os specimens. The full elastic tensor for osmium was measured using resonant ultrasound spectroscopy between 5 K and 300 K on polycrystalline and monocrystalline specimens and confirmed at ambient temperature using a pulse-echo-overlap measurement. It was found that osmium is more compressible than diamond (bulk modulus: 405 GPa for Os, versus 442 GPa for diamond). The independent elastic moduli of osmium are in relatively good agreement with the theory of Fast [Phys. Rev. B **51**, 17431 (1995)]. However, the Debye temperature determined from low-temperature monocrystal elastic-moduli measurements extrapolated to 0 K disagrees with most previous reports and is in good agreement with one specific-heat measurement.

3:05–3:20 Break

*Contributed Papers*

3:20

**1pPAb7. Excitation of elastic and capillary waves by electric field tweezers and its application to material characterization.**

Maiko Hosoda, Hideo Ogawa, and Hiroyasu Nomura (Tokyo Denki Univ., Ishizaka, Hatoyama, Saitama 350-0394, Japan)

We recently developed the electric tweezers system to excite the surface vibration for the measurement of local mechanical properties of soft condensed materials. A periodical electric field applied onto the material surface by a conductive thin needle induces the Maxwell stress and picks up the sample surface. We can determine the shear elasticity of the sample from the absolute displacement of the deformation, and the viscosity from its dynamic response, respectively. The numerical simulation predicts that the electric voltage of 100 V applied to the gap of 100  $\mu\text{m}$  between the needle edge and surface induces the deformation of several nm, which was experimentally confirmed by the optical lever technique. The remarkable feature of the system is that the characteristic scale of the surface deformation, such as the depth and width of the medium motion, is dominated by the gap and, therefore, we can uniquely determine the complex mechanical impedance and the surface energy by observing the dynamic surface response in a wide range of the gap and oscillation frequency. The system was applied to the gelatin gel at various concentrations and we found the results obtained are well explained by the previously known complex elasticity of the material.

3:35

**1pPAb8. Direct observation of phonon relaxation by optical beating light scattering spectroscopy.**

Keiji Sakai (Inst. of Industrial Sci., Univ. of Tokyo, 4-6-1 Komaba, Meguro-Ku, Tokyo 153-8505, Japan)

A novel dynamic light scattering system to observe elastic relaxation phenomena with hyper frequency resolution was developed. The principle of the measurement is based on the theory, which describes the dynamic structure factor of fluid under the condition of the frequency-dependent

compressibility. The dynamic structure factor, which is usually composed of the Brillouin and Rayleigh triplet, is modulated and shows an additional central component that directly reflects the whole aspect of the relaxation. In the experiment, the output from a frequency-doubled cw-YAG laser was incident into the liquid sample and the power spectrum of the light scattered into the backward direction was analyzed by the optical beating spectroscopy technique. The sample is liquid acetic acid that is known to show a strong ultrasonic relaxation around 1 MHz due to the molecular association process. We could find, in the observed spectrum, the central component introduced by the phenomenon, whose relaxation frequency and strength can be obtained from the width and the intensity of the observed central peak, respectively. The results show very good agreement with those previously obtained by the conventional ultrasonic spectroscopy technique. The recent experiment of the thermal phonon resonance will also be shown.

3:50

**1pPAb9. High resolution Brillouin scattering in solid and gases.**

Takeshi Yogi and Keiji Sakai (Dept. of Appl. Phys., Univ. of Tokyo, 4-6-1 Komaba Meguro-ku, Tokyo, 153-8505, Japan, yogitake@iis.u-tokyo.ac.jp)

The experimental system of our optical beating Brillouin spectroscopy was improved, which can observe the phonon propagation with the resolution of 1 kHz. The increased sensitivity and frequency resolution were demonstrated in the observation of phonon spectrum in solids of fused silica and also in gases, nitrogen, and freon  $\text{CHClF}_2$ ; for their weak light-scattering ability, these substances are very tough specimens for Brillouin scattering studies, and the classical spectrometer of Fabry-Pérot etalon has been so far used for the frequency analysis. The phonon peaks were observed with the present light beating system in nitrogen and freon near 300 K over the wave-number range from  $2\pi = 6.20 \times 10^3$  to  $3.09 \times 10^5 \text{ m}^{-1}$ , the pressure range from 14.8 to 177 kPa. The observed spectra were well fitted with the Lorentzian curve representing the phonon lifetime, and the phase velocity. The phase velocity of phonon in fused silica agrees well

with that obtained by ultrasonic measurement, while the phonon dispersions obtained for gases show the relaxational behavior due to the coupling to the rotational and vibrational motion of molecules.

4:05

**1pPAb10. Hypersonic phononic crystals.** Taras Gorishnyy, Chaitanya K. Ullal, Martin Maldovan, Edwin L. Thomas (Dept. of Mater. Sci. and Eng., MIT, Cambridge, MA 02139), and Ji-Hyun Jang (MIT, Cambridge, MA 02139)

Phononic crystals are structures with spatially periodic variations in density and sound velocities. The most fascinating feature of phononic crystals is the existence of stop bands or band gaps in their dispersion relation, where no phonons are allowed to propagate. As a result, phononic crystals have great promise for achieving precise control over propagation of sound and other mechanical waves. In this talk our recent investigation of 2-D and 3-D hypersonic phononic crystals with band gaps in GHz frequency range will be presented and their potential applications as acoustic mirrors, lenses, waveguides, etc. will be discussed. High-quality single-crystalline structures were fabricated using interference lithography. Their phonon dispersion relation was directly measured with Brillouin light scattering. Finite-element analysis was employed to compute theoretical dispersion relation and provide interpretation for the experimentally observed propagation modes [T. Gorishnyy *et al.*, Phys. Rev. Lett. **94**, 115501 (2005); T. Gorishnyy *et al.*, Physics World **18** (12), 24–29 (2005)]. [Work supported by NSF Grants DMR-0308133 and DMR-0414974.]

4:20

**1pPAb11. Acoustical properties of two-dimensional sonic crystals.** Francisco Cervera, Elisabeth Schwarz, Daniel Torrent, Suitberto Cabrera, and José Sánchez-Dehesa (Wave Phenomena Group, Nanophotonics Technol. Ctr., Polytechnic Univ. of Valencia, C/ Camino de vera s/n, E-46022 Valencia, Spain)

The transmittance and reflectance spectra of two-dimensional sonic crystals made of hexagonal arrays of cylindrical bars in water is reported. The balance of sound fluxes is analyzed by studying angle-resolved spectra of transmitted and reflected sound for all directions in space. A characterization of the acoustic band structure along the two high symmetry directions in the lattice (i.e.,  $\Gamma X$  and  $\Gamma J$  directions) in a large region of frequencies will be presented. It is shown that, for example, the shape of the stop band features in different types of acoustical spectra is interlinked and must be studied jointly to understand acoustical phenomena in these materials. Moreover, the dependence of acoustical properties as a function of the number of layers is also studied. The experiments are compared with theoretical simulations based on multiple scattering theory. [Work supported by MEC of Spain.]

4:35

**1pPAb12. Design of resonant ultrasound systems for use with very small samples.** Ben Migliori (UCSD, 9500 Gilman Dr., San Diego, CA 92122), Albert Migliori, and Jon Betts (NHMFL-LANL, Los Alamos, NM 87545)

The design of resonant ultrasound (RUS) systems for small samples, whether used in small-bore cryostats at low temperature and high field or at room temperature, presents several unique challenges. These include alignment and containment of the sample, thermal contraction of construction materials, ease of use, and robustness in the presence of large ( $\approx 10$  T) magnetic fields. Such problems may be resolved through the use of novel

(and often easily obtained) materials as well as unique designs. At the National High magnetic Field Laboratory (NHMFL) we have addressed these problems and built a sub-4 K cryostat RUS measurement system capable of reliably measuring samples  $\approx 1$  mm in diameter. Examples of such a cryostat systems, designed for conditions ranging from room temperature and zero field to  $\approx 4$  K and  $\approx 10$  T, as well as explanations of the necessary design parameters and recommended materials are presented. Examples of RUS measurements with this system are presented.

4:50

**1pPAb13. Nondiffractive propagation in sonic crystals.** Victor J. Sanchez-Morcillo, Isabel Perez-Arjona, Javier Redondo, Victor Espinosa (Dept. de Física Aplicada, Universidad Politécnica de Valencia, Ctra. Nazaret-Oliva s/n 46730 Grao de Gandia, Spain), and Kestutis Staliunas (Universitat Politècnica de Catalunya, E-08222 Terrassa, Spain)

Sonic crystals are media with a periodic modulation of the acoustic parameters, such as the density and the bulk modulus. They have recently attracted a great interest, because of their potential applications in the control of sound propagation, used as reflectors, focusers, or waveguides. All these properties are related with the dispersion introduced by the crystal anisotropy. We report on the nondiffractive propagation of sound in two-dimensional sonic crystals. It is shown that, for given frequencies and directions of incidence, a narrow sonic beam can propagate without diffractive broadening. Such nondiffractive sonic beams exist in crystals with perfect symmetry and do not require the presence of defects, differently from other waveguiding phenomena reported previously. The cancellation of diffraction has been predicted using the plane-wave expansion method to evaluate the dispersion surfaces of the crystal and the spatial dispersion (isofrequency) curves. It occurs for frequencies and wave vectors for which dispersion curves have zero curvature, denoting a transition between focusing and defocusing regimes. By means of perturbative techniques, a simple analytical expression for the nondiffractive conditions has been obtained. The phenomenon is also demonstrated by numerical integration of the acoustic equations using FDTD techniques. [Support from Spanish MEC, Project FIS2005-07931-C03-01/03, is acknowledged.]

5:05

**1pPAb14. Strobe-photoelastic visualization of Lamb-type waves with negative group velocity in a solid/fluid/solid trilayer.** Ken Yamamoto (Kobayasi Inst. of Physical Res., 3-20-41 Higashi-Motomachi, Kokubunji, Tokyo, 185-0022, ken-yamamoto@kobayasi-riken.or.jp)

An elastic wave with negative group velocity transports energy in the direction opposite to the phase velocity. This phenomenon exists in certain modes of Lamb waves propagating in a solid plate, but only over a very narrow range of frequency-thickness products and Poisson's ratios of a plate material. A minimum value of the negative group velocity of a mode is dominated by the Poisson's ratio of a material. A similar phenomenon also appears in Lamb-type waves that are guided in a symmetric solid/fluid/solid trilayer. Derivation of the dispersion equation of Lamb-type waves emphasizes the existence of the negative group velocity in certain modes. The existence of negative group velocity depends not only on the modes, frequency-thickness product, and Poisson's ratio, but also on the fluid layer thickness. The magnitude of negative group velocity is controlled by changing the fluid layer thickness. Direct evidence is presented of the negative group velocity that is obtained by means of stroboscopic photoelastic visualization of pulsed Lamb-type waves propagating in a glass/water/glass trilayer. A time-sequence photoelastic image shows the wave-phase behavior and the pulse envelope propagating in the opposite direction.



## Session 1pSC

## Speech Communication: Synthesis and Recognition of Speech (Poster Session)

Om D. Deshmukh, Cochair

*Univ. of Maryland, Electrical and Computer Engineering Dept., A.V. Williams Bldg., College Park, MD 20742*

Koichi Shinoda, Cochair

*Tokyo Inst. of Technology, Graduate School of Information Science and Engineering, Dept. of Computer Science, 2-12-1 W8-81, Ookayama, Meguro-ku, Tokyo 152-8552, Japan*

## Contributed Papers

All posters will be on display from 2:00 p.m. to 6:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 2:00 p.m. to 4:00 p.m. and contributors of even-numbered papers will be at their posters from 4:00 p.m. to 6:00 p.m.

**1pSC1. Improving quality of small body transmitted ordinary speech with statistical voice conversion.** Hidehiko Sekimoto, Tomoki Toda, Hiroshi Saruwatari, and Kiyohiro Shikano (Nara Inst. of Sci. and Technol., 8916-5, Takayama-cho, Ikoma-shi, Nara, 630-0192 Japan)

The explosive spread of cellular phones enables us to communicate with each other at any time or place. Although cellular phones are convenient, there are still some problems. For example, it is difficult to send intelligible speech under noisy conditions, which is a fatal problem especially when talking privately using small speech in crowds. To improve the quality of small speech under such situations, a new speech communication style is proposed using a nonaudible murmur (NAM) microphone [Nakajima *et al.*, Eurospeech (2003), pp. IV-2601–2604]. The NAM microphone is robust to eternal noise, although body transmission causes quality degradation. In this paper, to improve the sound quality of small body transmitted ordinary speech (SBTOS), which is small speech recorded with a NAM microphone, two conversion methods that reflect a statistical voice conversion method based on Gaussian mixture models (GMMs) [Toda *et al.* Interspeech (2005), pp. 1957–1960] are conducted. One conversion method is from SBTOS to ordinary speech (SBTOS-to-SP), and the other is from SBTOS to small speech (SBTOS-to-SSP). SBTOS-to-SSP has more consistent correspondence of voiced/unvoiced segments between input and output speech than SBTOS-to-SP. The results of objective and subjective evaluations show that SBTOS-to-SSP outperforms SBTOS-to-SP.

**1pSC2. Evaluation of eigenvoice conversion based on Gaussian mixture model.** Yamato Ohtani, Tomoki Toda, Hiroshi Saruwatari, and Kiyohiro Shikano (Grad. School of Information Sci., Nara Inst. of Sci. 8916-5, Takayama-cho, Ikoma-city, Nara, Japan 630-0192)

Eigenvoice conversion (EVC) has been proposed as a new framework of voice conversion (VC) based on the Gaussian mixture model (GMM) [Toda *et al.*, “Eigenvoice Conversion Based on Gaussian Mixture Model,” ICSLP, Pittsburgh, Sept. 2006]. This paper evaluates the performance of EVC in conversion from one source speaker’s voice to an arbitrary target speakers’ voices. This framework trains canonical GMM (EV-GMM) in advance using multiple parallel data sets consisting of utterance pairs of the source and many prestored target speakers. This model is adapted to a specific target speaker by estimating a small number of free parameters using a few utterances of the target speaker. This paper compares spectral distortion between converted and target voices in EVC with conventional VC based on GMM when varying the amount of training data and the number of mixtures. Results show EVC outperforms conventional VC when using small amounts of training data. EVC can effectively train a

complex conversion model using the information of many prestored speakers. By contrast, conventional VC needs a large-sized parallel data set for training. It also shows the results of subjective evaluations of speech quality and conversion accuracy for speaker individuality.

**1pSC3. Evaluating naturalness of speeches morphed by independently using the interpolation ratios of the time-frequency axes and amplitude.** Toru Takahashi, Masanori Morise, and Toshio Irino (Faculty of Systems Eng., Wakayama Univ., 930, Sakaedani, Wakayama, Japan, 640-8510, tall@sys.wakayama-u.ac.jp)

In speech morphing procedure [Kawahara *et al.*, Proc. IEEE-ICASSP 2003, Vol. I, pp. 256–259], two parameters exist for controlling the degree of morphing between the source and target: the interpolation ratios of the time-frequency axes and amplitude. Conventional morphing methods use only a single path in which both ratios are identical. It has, however, been reported that perception of naturalness degrades around a morphing ratio of 0.5. It was assumed that there would be better combination of the ratios with maintaining the quality. It was surveyed that the naturalness contour is within a two-dimensional morphing space. The morphed speech sounds are synthesized in combinations of 11 axis ratios (between 0 and 1 by 0.1 step) and 11 amplitude ratios. Synthetic sounds (121 in all) were presented to subjects to judge the naturalness rate. The optimum path will be described at a presentation to be given at this conference. [This research was partly supported by the “e-Society Leading Project” of the Ministry Education, Culture, Sports, Science, and Technology.]

**1pSC4. Building an English speech synthetic voice using a voice transformation model from a Japanese male voice.** Akemi Iida (School of Media Sci., Tokyo Univ. of Technol., 1404-1, Katakura-cho, Hachioji, Tokyo, 192-0982, Japan, ake@media.teu.ac.jp), Shimpei Kajima, Kiichi Yasu, Takayuki Arai (Sophia Univ., Tokyo, 102-8554, Japan), and Tsutomu Sugawara (Sophia Univ., Tokyo, 102-8554, Japan)

This work reports development of an English speech synthetic voice using a voice transformation model for a Japanese amyotrophic lateral sclerosis patient as part of a project of developing a bilingual communication aid for this patient. The patient, who had a tracheotomy 3 years ago and had difficulty in speaking, wishes to speak in his own voice in his native language and in English. A Japanese speech synthesis system was developed using ATR CHATR 6 years ago and the authors have worked in developing a diphone-based synthesis using FESTIVAL speech synthesis system and FESTVOX by having the patient read the diphone list. However,

it was not an easy task for the patient to phonate and, moreover, to pronounce words in a foreign language. We therefore used a voice transformation model in FESTIVAL to develop the patient's English speech synthetic voice which enables text-to-speech synthesis. We trained using 30 sentences read by the patient and those synthesized with an existing FESTIVAL diphone voice created from a recording of a native English speaker. An evaluation including a listening experiment was conducted and the result of this voice conversion showed that the synthesized voice successfully reflected the voice quality of the patient.

**1pSC5. An MRI-based time-domain speech synthesis system.** Tatsuya Kitamura, Hironori Takemoto, Parham Mokhtari (NICT/ATR Cognit. Information Sci. Labs., 2-2-2 Hikaridai, Seikacho, Sorakugun, Kyoto, 619-0288, Japan), and Toshio Hirai (Arcadia, Inc., Minoushi, Osaka, 562-0003, Japan)

A speech synthesis system was developed based on Maeda's method [S. Maeda, *Speech Commun.* **1**, 199–229 (1982)], which simulates acoustic wave propagation in the vocal tract in the time domain. This system has a GUI interface that allows fine control of synthesis parameters and timing. In addition, the piriform fossae were included to the vocal tract model, resulting in antiresonances in speech spectra at the frequency region from 4 to 5 kHz. The system can produce all the Japanese phonemes using vocal tract area-functions (VTAFs) extracted from 3-D cine-MRI obtained during production of VCV or CVCV sequences for a male speaker. The system can be used to synthesize Japanese sentences with high naturalness and intelligibility by concatenating segmental units and controlling the glottal source using the GUI interface. Since a time-varying VTAF is obtained by interpolating between VTAFs, the dataset size of the system is significantly smaller than that of corpus-based speech synthesizers. The speaker-specific VTAFs and inclusion of the piriform fossae permit us to reproduce speaker-specific spectral shapes, not only the lower formants but also higher frequency regions that contribute to the perception of speaker individualities. [Work supported by NICT, SCOPE-R, and Grant-in-Aid for Scientific Research of Japan.]

**1pSC6. Database size and naturalness in concatenative speech synthesis.** H. Timothy Bunnell, James T. Mantell, and James B. Polikoff (Speech Res. Lab., A. I. duPont Hospital for Children, 1600 Rockland Rd., Wilmington, DE 19803)

Unit concatenation TTS systems seek to maximize perceived naturalness by minimizing the amount of signal processing applied to the recorded speech on which they are based. To generate distinct suprasegmentals for a given segmental sequence (e.g., to convey variation in focus or emotion), it is necessary to record and store multiple instances of the same segments that vary in fundamental frequency and voice quality. At the expense of naturalness, concatenative systems can store a minimal segmental inventory and synthesize suprasegmental factors by manipulating  $f_0$  and voice quality via signal processing. Classic diphone synthesis (where only a single instance of each diphone sequence is stored) represents the limiting case of this strategy. The present study explores aspects of the trade-off between perceived naturalness and segmental inventory size using the ModelTalker TTS system. Twenty-five speakers each recorded about 1650 utterances. From these, databases were constructed that limited the maximum number of alternate diphones usable for synthesis in five conditions to 1, 5, 10, 20, and 40. Sentences were synthesized from these databases using either full or no  $f_0$  control. Results of listening tests wherein subjects rate the naturalness of each sentence will be presented. [Work supported by NIDCD.]

**1pSC7. Optimization of target cost weights in concatenative speech synthesis with very short segments of 5-ms duration.** Toshio Hirai (Arcadia, Inc., 3-1-15 Nishishoji, Mino, Osaka 5620003, Japan, thirai@arcadia.co.jp)

If a concatenative speech synthesis system uses more short speech segments, it increases the potential to generate natural speech because the concatenation variation becomes greater. Recently, a synthesis approach was proposed in which very short (5 ms) segments are used [T. Hirai and S. Tenpaku, 5th ISCA Speech Synthesis Workshop, pp. 37–42 (2004), <http://www.ssw5.org/>]. In that approach, the target cost (how close a database segment is to desired segment) was defined as the simple average of the root mean squares (rms's) of the difference between the features of the database segment and the desired segment for simplicity. Therefore, it has been expected to optimize the weight of each rms. A Japanese speech database was used to optimize the weights and to evaluate its effects. For corpus construction, 150 utterances were selected from the database. Ten other utterances were selected randomly for generation of feature time series as natural targets for synthesis. Half of them were used for optimization using many weight combinations in synthesis to determine the optimal weight set that shows the minimum concatenation distortion. Distortion for the other half of the utterance synthesis with optimized weights was reduced 34.4% compared to the former approach.

**1pSC8. Improving Japanese syllable unit selection using search space with small acoustical features variance.** Takaaki Moriyama and Seiichi Tenpaku (Arcadia, Inc., 3-1-15, Nishishoji, Mino, Osaka, 5620003 Japan, taka@arcadia.co.jp)

Usually in concatenative text-to-speech systems, suitable units are selected according to acoustical features. However, a method to reduce calculation costs has been proposed, which is using syllable notations and positions along with the mora length [Murakami *et al.*, *Trans. IEICE J85-D-II(7)*, 1157–1165 (2002)]. An enlargement of the corpus to synthesize longer phrases lowers the quality of synthesized speech because of increased feature variance within syllable units with identical notation. Methods to produce detailed groups in which the feature's variance is smaller than syllable units, and to synthesize speech using these groups, are proposed. To produce the groups, a mel-cepstrum distance is used as a measure for evaluation. In this method, the quality of synthesized speech that is given an ideal group sequence might improve, as the distance between units is close within the same group. As a first step to evaluate the proposed method, five groups were generated from a Japanese corpus; then, speech that was synthesized to replace one unit in natural speech to others in the same groups was examined. Results of this listening experiment indicated that an average of the naturalness for these speech types was greater than 80%.

**1pSC9. Toward hidden Markov model-based spontaneous speech synthesis.** Tatsuya Akagawa, Koji Iwano, and Sadaoki Furui (Dept. of Comput. Sci., Tokyo Inst. of Technology, 2-12-1-W8-77 Ookayama, Meguro-ku, Tokyo, 152-8552 Japan, akagawa@furui.cs.titech.ac.jp)

This paper investigates several issues of spontaneous speech synthesis. Although state-of-the-art synthesis systems can achieve highly intelligible speech, their naturalness is still low. Therefore, much work must still be done to achieve the goal of synthesizing natural, spontaneous speech. To model spontaneous speech using a limited amount of data, we used an HMM-based speech synthesizer based on three features: cepstral features modeled by HMMs, and duration and fundamental frequency features modeled using Quantification Theory Type I. The models were trained with approximately 17 min of spontaneous lecture speech, from a single speaker, which was extracted from the Corpus of Spontaneous Japanese (CSJ). For comparison, utterances by the same speaker, reading a transcription of the same lecture, were used to train analogous models for read speech. Spontaneity of the synthesized speech was evaluated by subjective pair comparison tests. Results obtained from 18 subjects showed that the preference score for the synthesized spontaneous speech was significantly

higher than that for the synthesized read speech. This implies that HMM-based speech synthesis using actual spontaneous utterances for model training is effective at producing natural speech. Additional subjective evaluation tests were also conducted to analyze the effects of individual features on the impression of spontaneity.

**1pSC10. Clustering strategies for hidden Markov model-based polyglot synthesis.** Javier Latorre, Koji Iwano, and Sadaoki Furui (Tokyo Inst. of Technol., Ookayama 2-12-1, Meguro-ku, Tokyo, 152-8552, Japan, latorre@furui.cs.titech.ac.jp)

A speaker-adaptable polyglot synthesizer is a system that can synthesize multiple languages with different voices. In our approach, we mix data from multiple speakers in different languages to create an HMM-based multilingual average voice. This average voice is then adapted to a target speaker with a small amount of speech data from that speaker. Using this adapted voice our system is able to synthesize any of the languages for which it was trained, with a voice that mimics that of the target speaker. The main technical difficulty of this approach is how to cluster the acoustic models to create a voice that sounds equal for all the languages. In this paper we analyze different possible strategies for clustering the HMMs and compared them on a perceptual test. The results showed that questions that refer only to the phonetic features and ignore the original language of the sounds produced voices with more foreign accent but that sound clearly more consistent across languages. Surprisingly, this foreign accent affected neither the intelligibility nor the naturalness of the synthetic speech. Therefore for polyglot speech synthesis the overall better performance was obtained by clustering strategies based only on the phonetic features of the sounds.

**1pSC11. Concept-to-speech conversion for reply speech generation in a spoken dialogue system for road guidance and its prosodic control.** Yuji Yagi, Seiya Takada, Keikichi Hirose, and Nobuaki Minematsu (School of Eng., Univ. of Tokyo, Bldg. No. 2, 7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan, yagi@gavo.t.u-tokyo.ac.jp)

Because of advancements of speech and language processing, a number of spoken dialogue systems have been constructed. However, because most of them adopt existing text-to-speech synthesizers to generate output speech, it is rather difficult to reflect all the linguistic information that is obtained during the reply sentence generation. To resolve this situation, a framework must correctly reflect higher-level linguistic information, such as syntactic structure and discourse information, on the prosody of output speech: concept-to-speech conversion, where reply sentences are generated from information (to be transmitted) and converted into speech in a unified process. A spoken dialogue system for road guidance was constructed, and concept-to-speech conversion was realized in the system. The linguistic information of the generated sentence is handled in tag LISP form to retain the syntactic structures throughout the process. Moreover, a new method of sentence generation from concept was realized with this system: it handles a concept in phrase units and aggregates them to form a sentence. It is tested whether the linguistic information could be reflected properly on the prosody of output speech. Results of listening experiments verified the effectiveness of our proposed method.

**1pSC12. Vector-based language generation for associatively evoking environmental images.** Kiyooki Aikawa and Akemi Iida (School of Media Sci., Tokyo Univ. of Technol., 1404-1 Katakuracho, Hachioji, Tokyo 192-0982, Japan)

A novel language generation method was proposed for spoken dialogue systems. The method selected verbal expressions for sensory inputs using a vector-based method. A weather-reporter system was implemented with seven input categories of weather parameters including temperature, humidity, wind velocity, and cloud cover. The system selected the most appropriate short phrase for representing the measured weather parameters

to let the listener associatively imagine the reporter's weather condition descriptions. The system included 149 short phrases for weather expressions, each represented by a 62-element vector called the template vector. Given the 62-element input vector for a set of weather parameters, one or two phrases were selected based on the similarity measured between the input and template vectors. The system was able to produce an expression by combining a couple of phrases. A dissimilarity measuring algorithm was proposed to assign the conjunction "and" or adversative "but" automatically to conjoin two phrases. A bilingual capability was also added for enabling the system to output English and Japanese phrases. For English outputs, sentence-initial phrases were added to the selected phrases depending on the similarity value, while sentence-finals were added for Japanese outputs. Evaluation experiments showed that automatic conjunction assignment can function with high reliability.

**1pSC13. Development project for screen reader interface dynamically hastens speech while giving emphasized information to the tactile sense.** Tohru Ifukube (Res. Ctr. for Adv. Sci. and Technol., The Univ. of Tokyo, Tokyo, 153-8904, Japan) and Shinta Kimura (Animo Ltd., Naka-ku, Yokohama-shi, 231-0015, Japan)

In present screen readers for blind people, users are still required to continue listening to synthesized voices and also they might easily miss important points of a document. A development project of a new screen reader interface has been promoted in cooperation with Japan's Ministry (METI). The interface named TAJODA can dynamically control the speech rate while giving rich texts to the tactile sense. The speech rate can be changed according to word units by clicking a button of the interface using a thumb. From evaluation tests using sightless people, the maximum speech rate was determined at around 2000 morae/minute. Seven vibrotactile patterns were selected for presenting rich texts onto an index finger using a tactile matrix display (2×8). Both speech and tactile information are synchronized automatically through a USB interface. It was also ascertained that users can read some documents two to three times faster using the TAJODA interface [Asakwa *et al.*, IEICE Trans. **E87-D**(6) (2004)].

**1pSC14. Decision-tree-based  $F_0$  quantization for hidden Markov model-based speech coding at 100 bit/s.** Yoshihiro Itogawa, Heiga Zen, Yoshihiko Nankaku, Akinobu Li, and Keiichi Tokuda (Dept. of Comput. Sci., Nagoya Inst. of Technol., Gokiso-cho, Showa-ku, Nagoya, 466-8555 Japan)

A decision-tree-based quantization scheme for a very low bit rate speech coder based on HMMs is described. The encoder carries out HMM-based phoneme recognition and then recognized phonemes, state durations, and  $F_0$  sequence are quantized, Huffman coded, and transmitted. In the decoder, sequences of mel-cepstral coefficient vectors and  $F_0$ 's are generated from the concatenated HMM-using the HMM-based speech synthesis technique. Finally, a speech waveform is synthesized by the MLSA filter using the generated mel-cepstral coefficient and  $F_0$  sequences. In the previous system, we train an MSD-VQ codebook for each phoneme for  $F_0$  quantization. Although this scheme can quantize  $F_0$  sequences efficiently, to achieve a better speech quality, larger codebook sizes are required. It leads to an increase in the bit rate of the system. To avoid this problem, we cluster  $F_0$  sequences using phonetic decision trees and then train a codebook for each leaf node. In the encoding and decoding, codebooks to be used can be determined by tracing the decision tree. It allows us to use smaller codebook sizes since the number of codebooks can be augmented without increase in bit rate. A subjective listening test result shows that the proposed scheme improves the quality of coded speech.

**1pSC15. Speech enhancement during analysis-synthesis by harmonic plus noise model.** Parveen K. Lehana and Prem C. Pandey (EE Dept., Indian Inst. of Technol. Bombay, Powai Mumbai 400076, India, pcpandey@ee.iitb.ac.in)

Speech analysis-synthesis in codecs and voice transformation often introduces distortion in the resynthesized output, particularly if the input is corrupted by noise. The harmonic plus noise model (HNM), particularly suited for voice transformation, divides speech signal into harmonic and noise parts. The harmonic part is synthesized by taking the summation of pitch harmonics up to a frequency called maximum voiced frequency, and the spectrum above this frequency is attributed to the noise part. Removal of nonharmonic components in harmonic part may result in noise rejection, while the errors in the estimation of harmonics may lead to distortion. Investigations were conducted to assess the speech enhancement obtainable through HNM on broadband noise corrupted speech signal. Listening tests showed that both the input and the resynthesized signals were perceived as distortion-free for SNR greater than 10 dB. In the 0–10-dB range, both the signals had no loss of intelligibility, but listeners rated the resynthesized signal as having higher quality. No enhancement was obtained for SNR values lower than 0 dB. Hence, for higher noise levels, spectral subtraction or other enhancement techniques should be employed along with analysis-synthesis.

**1pSC16. Enhancement of electrolaryngeal speech by spectral subtraction with minimum statistics-based noise estimation.** Priyanko Mitra and Prem C. Pandey (EE Dept., Indian Inst. of Technol. Bombay, Powai Mumbai 400076, India, pcpandey@ee.iitb.ac.in)

Transcervical electrolarynx is a vibrator held against the neck tissue by a laryngectomy patient to provide excitation to the vocal tract, as a substitute to that provided by glottal vibrations. The electrolaryngeal speech suffers from lack of voicing and pitch control, low-frequency deficiency, and presence of background noise, caused by leakage of acoustic energy from the vibrator and vibrator-tissue interface. Pitch-synchronous application of spectral subtraction has been used for reducing the self-leakage noise, with averaging-based noise estimation on an initial segment with closed lips. But, the leakage noise spectrum varies with speech production and vibrator orientation and pressure. A dynamic estimation of noise is carried out by using minimum statistics, with two-pitch period analysis frames and one period overlap. Minimum value of each spectral sample in a set of past frames is used for dynamically estimating the magnitude spectrum of noise. This is subtracted from the magnitude spectrum of noisy speech and the signal is reconstructed using the original phase spectrum. Smoothing of the estimated noise spectrum resulted in better noise cancellation. Quality of electrolaryngeal speech is further enhanced by spectral deficit compensation of the magnitude spectrum.

**1pSC17. Evaluating the perceptual quality of speech signals enhanced using the Modified Phase Opponency model.** Om D. Deshmukh and Carol Y. Espy-Wilson (Dept. of Elec. & Comput. Eng., A. V. Williams Bldg., Univ. of Maryland, College Park, MD 20742)

Previously, we presented a speech enhancement algorithm based on the auditorily motivated MPO model. It was shown that the MPO-based speech enhancement scheme is robust to fluctuating noise. Compared to the output of several other speech enhancement techniques, the MPO-enhanced speech signals exhibit the lowest increase in Itakura-Saito, log-area-ratio, and log-likelihood-ratio-based distortion measures as the SNR reduces. In the present work, the performance of the MPO speech enhancement is compared with that of the other techniques using subjective evaluations. Three normal-hearing subjects were recruited to perform two-alternative, forced-choice preference tasks. Each sentence was presented twice: one of them was enhanced using the MPO scheme and the other was either left unprocessed or enhanced using one of three other speech enhancement techniques. Sentences are 5-digits long and corrupted by additive subway noise at  $-5$ ,  $0$ ,  $5$ , or  $15$ -dB SNR. Most of the time, the MPO-enhanced speech signal was preferred weakly. Moreover, the MPO-

processed clean speech signal was preferred over clean unprocessed speech signal just as many times as the clean unprocessed speech signal was preferred, indicating that the MPO processing does not introduce any perceptual artifacts in clean speech. [Work supported by NSF BCS0236707 DOD H9823005C0425.]

**1pSC18. Noise reduction of speech using multiband noise subtraction and adaptive filtering methods in two stages.** Md. Ekramul Hamid, Keita Ogawa, and Takeshi Fukabayashi (Grad. School of Elec. Sci. and Tech., Shizuoka Univ., 3-5-1 Jouhoku, Hamamatsu 432-856, Japan)

Real-world noise is mostly colored having nonuniform power distribution over the whole spectrum. A single-channel two-stage method with multiband processing technique is presented in this paper. Initially a noise dominant signal is estimated from the observed speech and obtained the first noise-reduced speech by subtracting the signal. The noise-dominant signal is estimated from the valleys of the spectrum whose amplitudes are tuned by the estimated noise degree. The bandwise estimated noise degree is obtained from a *a priori* function as a function of a parameter obtained from the autocorrelation of the noisy speech. The outcome of the first stage is allowed to process by the second stage in bandwise to obtain more enhanced speech. Blind source separation technique is employed that consists of two adaptive filters. The least-mean-square algorithm based on steepest-descent method is implemented to update the filter coefficients in an iterative manner. The method addresses the situations in which the input signal-to-noise ratio varies substantially and a fixed number of iteration of the least-mean-square algorithm for each signal-to-noise ratio is time consuming. Therefore, a function has been proposed that estimates the minimum iteration number required for a given value of the noise degree of noisy speech.

**1pSC19. Speech interface for operating information technological home appliances.** Masanori Enoki, Tadashige Noguchi, Tsubasa Arai, Ayako Miyake (Doshisha Univ., 1-3, Tatara-Miyakodani, Kyotanabe, Kyoto, 610-0321, Japan), Kohei Yasui, Kenko Ota, Masuzo Yanagida (Doshisha Univ., Kyoto, 610-0321, Japan), and Masaki Ida (OMRON Corp., Kyoto, 619-0283, Japan)

Information technological (IT) home appliances are becoming highly advanced by being connected to the Internet. Conventional push-button systems on remote controllers cannot suffice for IT home appliances because of their complicated requirements for versatile and complex functions. Speech input is expected to be a good substitute or supplement for push-button systems. Effects of introducing speech input to a TV system are evaluated, implementing facilities of on-line program retrieval besides ordinary functions of controlling the TV sets. One of the problems with introducing speech input is degradation of speech recognition rate due to being affected by surrounding noises and waves reflected on the surface of walls, ceiling, and so on. So, the test system introduced reduction of known noise, dereverberation scheme, delay-and-sum beamformer and blind source separation by independent component analysis as preprocessing for speech recognition. Effects of introducing these types of preprocessing are investigated comparing speech recognition rates in case of employing these types of preprocessing with those in the case of not employing these types of preprocessing. [Work supported by Knowledge Cluster Project, MEXT, and by Academic Frontier Project, Doshisha Univ.]

**1pSC20. Speech detection using Real Adaboost in car environments.** Tetsuya Takiguchi, Hiroyoshi Matsuda, and Yasuo Ariki (Dept. of Comput. and System Eng., Kobe Univ., 1-1 Rokkodai, Nada, Kobe, 657-8501, Japan, takigu@kobe-u.ac.jp)

In real noisy environments, a speech detection algorithm plays an especially important role for noise reduction, speech recognition, and so on. In this paper, a speech/nonspeech detection algorithm using Real Adaboost

is described, which can achieve extremely high detection rates. Boosting is a technique of combining a set weak classifiers to form one high-performance prediction rule, and Real Adaboost [R. E. Schapire and Y. Singer, *Mach Learn.* **37**, 3, 297–336, (1999)] is an adaptive boosting algorithm in which the rule for combining the weak classifiers adapts to the problem and is able to yield extremely efficient classifiers. The Real Adaboost algorithm is investigated for speech/nonspeech detection problem. The proposed method shows an increasing speech detection rate in car environments, comparable with that of the detector based on GMM (Gaussian mixture model), where the detection accuracy rate was 98% for the proposed method and 92% for GMM in a car at highway speed. The results of the experiments clarified the effectiveness of the proposed method.

**1pSC21. Nonrestrictive and long-term monitoring system of utterance and a laughing voice: Laughometer.** Masafumi Matsumura, Yutaka Kawabata, Ryoji Suzuki, and Masaki Yoshida (Faculty of Biomed. Eng., Osaka Electro-Commun. Univ., 1130-70 Kiyotaki, Shijonawate, Osaka 575-0063, Japan)

It is said that laughter is the best medicine. Laughter is known to be effective for relieving stress and promoting health. Therefore, this study is intended to clarify the relationship between laughter and physiological factors related to health promotion. We developed a hands-free wireless system (Laughometer) that monitors laughter (a laughing voice) in everyday conversations. This Laughometer detects vocal cord vibrations using a bone-conductive microphone attached to the skin on the neck. Based on data from repeated vocalizations of laughter, this apparatus can distinguish features of laughter in everyday conversations. Based on the mean fluctuation (standard deviation) in syllable intervals, 90% of laughs are differentiable from normal conversation. While attending a live rakugo comedy show, the number and duration of laughs were monitored by analyzing laryngeal voice (approximately 1-h duration). The mean differentiation rate for laughing voice was 94%, thereby confirming the efficacy of laughter recognition based on fluctuation in syllable intervals.

**1pSC22. Development of nonvoice dialogue interface for robot systems.** Ryuichi Nisimura and Aki Hashizume (Faculty of Systems Eng., Wakayama Univ., 930 Sakaedani, Wakayama 640-8510, Japan, nisimura@sys.wakayama-u.ac.jp)

In this presentation, a nonvoice dialogue interface scheme for robot systems to improve the flexibility of human-robot interaction is proposed. A prototype system with a microphone in a plastic doll has been developed to examine its potential for distinguishing interaction noises and human utterances. It realizes a simple, one-input one-response interaction with perception of a user's action. For example, the system can reply "Ouch!" when its body or head is beaten by a user. It is also able to make a response to a calling by the voice. This scheme would provide a new interaction style, unlike conventional spoken dialogue systems in which a user's rough actions were treated as meaningless obstacles. It identifies the source of noises based on GMM (Gaussian mixture model) noise recognition [A. Lee *et al.*, *Proc. INTERSPEECH*, 1, 173–176 (2004)]. Thirteen-class GMMs were constructed from voices and noises like slapping or stroking the head, and knocking the body, which were recorded via trial testing the prototype system. It is possible to discriminate noises and utterances by comparing acoustic likelihoods from GMMs. The experiment investigates performances of acoustic feature vectors consisting of MFCC (mel frequency cepstral coefficients) to identify interaction noises.

**1pSC23. An evaluation of hands-free voice activity detection algorithm based on weighted cross-power spectrum phase analysis and zero crossing detector.** Takamasa Tanaka, Yuki Denda, Masato Nakayama (Grad. School of Sci. and Eng., Ritsumeikan Univ., 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan, rs012019@se.ritsumei.ac.jp), and Takanobu Nishiura (Ritsumeikan Univ., Kusatsu, Shiga 525-8577, Japan)

VAD (voice activity detection) is necessary for recognizing hands-free speech. However, conventional VADs could not achieve enough detection performance in real noisy environments because they utilize only time-sequence information or only spatial-sequence information of captured signals. To cope with this problem, we have already proposed noise-robust hands-free VAD by composing the time-sequence and spatial-sequence information. The conventional ZCD (zero crossing detector) was employed to acquire the time-sequence information, and weighted CSP (cross-power spectrum phase) analysis that was a specialized DOA (direction of arrival) estimation algorithm for speech signal was also employed to acquire the spatial-sequence information. However, this algorithm can insufficiently improve the VAD performance in higher noisy environments. To overcome this problem, in this paper, we try to utilize talker direction information and attempt the environmental adaptation of our former algorithm. As a result of evaluation experiments in real acoustic environment, the proposed hands-free VAD could achieve a more effective performance than the conventional VAD and our former VAD. [Work supported by MEXT of Japan.]

**1pSC24. A study of robust omnidirectional audio-visual talker localization algorithm with microphone array and omnidirectional image.** Yuki Denda (Grad. School of Sci. and Eng., 1-1-1 Nojihigashi, Kusatsu, Japan, gr021052@se.ritsumei.ac.jp), Tanaknobu Nishiura, and Yoichi Yamashita (College of Information Sci. and Eng., 1-1-1 Nojihigashi, Kusatsu, Japan)

In video conferencing environments, it is very important to localize the talker. However, conventional audio signal-based algorithms often suffer from audio interference, and conventional visual signal-based algorithms fail in the presence of visual interference. To deal with these problems, this paper proposes a robust omnidirectional audio-visual talker localization algorithm that not only exploits audio feature parameters, but also subordinately uses visual feature parameters. To achieve omnidirectional audio-visual talker localization, paired-omnidirectional microphones are employed as an audio sensor, and an omnidirectional camera is employed as a visual sensor. For robust talker localization, audio feature parameters are extracted using weighted cross-power spectrum phase (CSP) analysis and CSP coefficient subtraction, and visual feature parameters are extracted using background subtraction and skin-color detection. The talker is finally located by the fusing of weighted audio/visual feature parameters, and the weight of this feature parameter fusion is automatically controlled based on the reliable criterion of audio feature parameters. The results of localization experiments in an actual room revealed that the proposed audio-visual talker localization algorithm is superior to that of conventional localizers using only audio or visual feature parameters, but not both. [Work supported by MEXT of Japan.]

**1pSC25. Application of harmonic plus noise model for enhancing speaker recognition.** Gidda Reddy Gangula, Prem C. Pandey, and Parveen K. Lehana (EE Dept., Indian Inst. of Technol. Bombay, Powai Mumbai 400076, India, pcPandey@ee.iitb.ac.in)

Speaker recognition systems mostly employ mel frequency cepstral coefficients (MFCC). Performance of these systems is generally affected by background noise, transmission medium, etc. Further, they do not perform well in text-independent environment with limited training data. For enhanced performance, the set of parameters used should separate the speaker-dependent information from the linguistic information. Towards this end, application of parameters of the harmonic plus noise model (HNM)-based analysis is investigated. HNM divides the speech spectrum

into harmonic and noise bands separated by a dynamically varying maximum voiced frequency. This frequency is a speaker-dependent parameter and its estimation is not affected by moderate SNR degradation. A speaker recognition system was devised using three HNM parameters, namely, maximum voiced frequency, relative noise band energy, and pitch. It gave performance comparably to that of MFCC-based recognition, for a group of ten speakers. An enhanced performance was observed by using MFCC and the three HNM parameters together, indicating the suitability of HNM for speaker recognition.

**1pSC26. Noisy speech recognition using modification of spectral envelopes and smoothing and emphasis method by image filters.** Yoichi Midorikawa and Masanori Akita (Faculty of Eng., Oita Univ., 700 Dannoharu, Oita, 870-1192, Japan)

The purpose of this study is to propose an improvement of the speech recognition under the noisy environment. Automatic speech recognition (ASR) systems are most effective in noiseless environments. We propose a method for ASR with signals contaminated with colored noise using modification of the spectral envelope shape. Our previous study has proposed a method of speech recognition for the signals polluted with color noise by modifying spectral envelope shape by rules. The proposed method is based on cepstral analysis. We have proposed the modified rules for adding valleys are defined. These methods are applied to the enhancement system of noisy signals and the performance of the system is improved. However, these rules only make use of frequency domain. Therefore, the cepstral coefficients of neighboring frames become independent of time domain. In this study, we apply the smoothing and emphasis method by image filters to the modification of spectral envelope shape. As a result, performance of speech recognition is improved.

**1pSC27. Noisy speech recognition based on codebook normalization of discrete-mixture hidden Markov models.** Tetsuo Kosaka, hidden Markov Models, Masaharu Katoh, and Masaki Kohda (Faculty of Eng., Yamagata Univ., Yonezawa-city, Yamagata, 992-8510 Japan, tkosaka@yz.yamagata-u.ac.jp)

This paper presents a new method of robust speech recognition under noisy conditions based on discrete-mixture HMMs (DMHMMs). The DMHMMs were proposed originally to reduce calculation costs. Recently, we applied DMHMMs to noisy speech recognition and found that they were effective for modeling noisy speech [Kosaka *et al.*, Proc. of ICA04 (2004), Vol. II, pp. 1691–1694]. For further improvement of noisy speech recognition, we propose a novel normalization method for DMHMM codebooks. The codebook normalization method is based on histogram equalization (HEQ). The HEQ is commonly applied for feature space normalization. The DMHMM codebooks were normalized in this study. Therefore, this method is considered as normalization not in feature space, but in model space. Some merits are inherent in choosing model space normalization. In model space normalization, a transformation function can be prepared for each acoustic model. In addition, it is not necessary to normalize input parameters frame-by-frame. The proposed method was compared with both conventional continuous-mixture HMMs (CHMMs) and DMHMMs. That comparison showed that the proposed method obtained the best performance: an average relative improvement of 29.2% over the CHMM baseline.

**1pSC28. Robust speech recognition for home environments using factorial hidden Markov models.** Agnieszka Betkowska, Koichi Shinoda, and Sadaoki Furui (Dept. of Comput. Sci., Tokyo Inst. of Technol., 2-12-1, Ookayama, Meguro-ku, Tokyo, 152-8552 Japan, agabet@furui.cs.titech.ac.jp)

In this paper we present a speech recognition system for the home environment that is robust to nonstationary sudden noise such as doors slamming, knocking, and falling objects. This noise appears rapidly and

lasts for a short period of time. Additionally the SNR is difficult to estimate. As a compensation model for this task, we investigated the use of the factorial hidden Markov model (FHMM) architecture proposed by Deoras (2004). In our method, a hidden Markov Model (HMM) for each word in a dictionary and an HMM for sudden noise are employed to build a factorial HMM to recognize noisy speech. Since the factorial HMM defined by Deoras applies only to static features of the observation vector during the recognition process, we extended it to work with dynamic features. The proposed method was evaluated with a database recorded by a personal robot, PaPeRo, in the houses of 12 Japanese families. The experiments were conducted for both synthesized speech (at different SNRs) and real noisy speech, which was recorded by the robot. In both cases, our method showed significant improvement over HMMs trained with clean speech data.

**1pSC29. A phoneme duration model considering speaking-rate and linguistic features for speech recognition.** Yuichi Ohkawa (Grad. School of Educational Informatics, Tohoku Univ., 27-1, Kawauchi, Aoba-ku, Sendai 980-8576, Japan), Akinori Ito, Motoyuki Suzuki, and Shozo Makino (Tohoku Univ., Aramaki, Aoba-ku, Sendai 980-8579, Japan)

In this paper, we proposed a method of phoneme duration modeling for speech recognition. A phoneme with extremely short or long duration often causes a decline of performance of speech recognition. In order to improve performance of recognition, an estimation of phoneme duration determined by various parameters is required. However, there was no usual method of duration modeling for speech recognition considering the influence of both speaking-rate and linguistic feature (phoneme location in sentence, part-of-speech, *et al.*), which influence phoneme duration strongly. Therefore, we modeled influence of speaking-rate by two-dimensional normal distribution of phoneme duration and local average of vowel duration. Each normal distribution is determined by tree-based clustering with various questions, which include linguistic feature. With an experiment of estimation of phoneme duration by this model, we acquired 20.8% reduction of standard deviation of estimation error. We also used the proposed duration model for rescoring of *N*-best hypothesis of speech recognition. With an experiment of rescoring of recognition results for spontaneous speech, we acquired significant reduction of 4.7% in phoneme error rate.

**1pSC30. Investigation of model-based error rate estimation techniques for speaker-independent large-vocabulary recognition tasks.** Atsuhiko Kai and Shoji Nakamori (Faculty of Eng., Shizuoka Univ., 3-5-1, Johoku, Hamamatsu, Shizuoka, 432-8561, Japan)

This paper proposes model-based error rate estimation techniques for HMM-based speech recognition systems and shows their effectiveness in several speaker-independent task conditions. The proposed techniques are based on an HMM-based error rate estimation method [C.-S. Huang *et al.*, IEEE Trans. SAP, 11, No. 6, 581–589 (2003)], which only depends on subword-based pronunciation lexicon knowledge. This study assumes the need of the dynamic change of open vocabulary. The vocabulary size is varied between 5000 and 25 000 words. Two types of formalizations are investigated. They utilize the Bhattacharyya distance as a local distance measure between two competing word classes which are modeled by syllable-unit HMMs. A misclassification measure, which is similar to the one used by minimum classification error (MCE) training methods, is used for estimating error rate. The correlation between the error rates for testing utterances by 20 speakers and the estimated error rates for varied vocabulary sizes are evaluated. Except for one speaker, the correlation coefficient above 0.8 was obtained for both proposed formulations. Moreover, when the uttered words are clustered to five coarse classes of different estimated error rates, the correct rate of relative consistency between those adjacent classes was above 90% on average.

**1pSC31. Large-scale continuous speech recognition system design using discriminative training.** Erik McDermott and Atsushi Nakamura (NTT Commun. Sci. Labs, NTT Corp., Hikari-dai 2-4, Seikacho, Soraku-gun, Kyoto-fu 619-0237, Japan, mcd@cslab.kecl.ntt.co.jp)

Discriminative training is difficult to implement but essential to attaining state-of-the-art performance in automatic speech recognition systems today. Most of the discriminative training results for large scale recognition tasks (with vocabularies well over 10 000 words) so far use the maximum mutual information (MMI) framework, but recent results for the minimum classification error (MCE) framework suggest that MCE too yields significant improvements in recognition accuracy and system compactness on large-scale tasks. MCE embodies rather well the general intuition that recognition system design should attempt to improve performance (i.e., recognition accuracy) directly, by optimizing a criterion function that is closely related to performance, rather than indirectly, by optimizing a criterion such as overall log likelihood that does not reflect performance. This presentation provides an overview of the MCE framework, and describes recent MCE speech recognition results obtained with both the MIT Galaxy system and the NTT Communication Science Labs SOLON system. The tasks examined include a 33 000-word vocabulary telephone-based spontaneous speech weather information task, a 22 000-word telephone-based name recognition task, and a 100 000-word Japanese lecture speech transcription task.

**1pSC32. Hyperparameter estimation for speech recognition based on variational Bayesian approach.** Kei Hashimoto, Heiga Zen, Yoshihiko Nankaku, Lee Akinobu (Nagoya Inst. of Technol., Dept. of Comput. Sci. and Eng., Gokiso-cho, Showa-ku, Nagoya, 466-8555 Japan, bonanza@ics.nitech.ac.jp), and Keiichi Tokuda (Nagoya Inst. of Technol., Nagoya, 466-8555 Japan)

A hyperparameter estimation technique for HMM-based speech recognition using variational Bayesian (VB) approach is proposed. In recent speech recognition systems, the ML and MDL criteria have been adopted for training HMMs and the context clustering, respectively. However, since the ML estimator produces a point estimate of parameters, the estimation accuracy is reduced with insufficient data. The MDL criterion is also affected by the amount of training data due to an asymptotic approximation. To avoid this problem, the VB method has been applied to HMM-based speech recognition. By using the VB approach, it is expected that the higher generalization ability is achieved and an appropriate model structure can be selected in the context clustering. In the Bayesian framework, the predictive distribution is obtained by integrating the likelihood function with respect to the prior density. Therefore, the performance of the Bayesian method depends on the setting of hyperparameters representing the prior distribution. However, the detailed investigation for hyperparameter estimation has not been conducted in the speech recognition field. This study proposes an estimation technique of hyperparameters based on the maximization of the marginal likelihood for HMM-based speech recognition. Experimental results show that the proposed technique can estimate appropriate hyperparameters from training data.

**1pSC33. An investigation of early reflection signal based on impulse response on automatic speech recognition.** Yoshiki Hirano, Yuki Denda, Masato Nakayama (Grad. School of Sci. and Eng., Ritsumeikan Univ., 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan, rs037021@se.ritsumei.ac.jp), and Takanobu Nishiura (Ritsumeikan Univ., Kusatsu, Shiga 525-8577, Japan)

Early reflection signals (especially, within 50 ms from direct sound arrival) are useful to speech listening for human being. On the other hand, the higher the reflection energy becomes, the more present automatic speech recognition performance decreases, subject to using clean acoustic phoneme model. However, it was not clear that early reflections are useful or useless to recognize the speech on distant-talking speech recognition, because reverberation time was employed as reverberation criterion. It is a unique time and does not depend on source position in a room. In addition,

it can only express the time in which sound attenuation becomes  $-60$  dB and cannot express the early reflection that strongly influences speech recognition performance. Therefore, it should be suitable reverberation criterion for distant-talking speech recognition. To overcome this problem, we investigate the relation between early reflections and distant-talking speech recognition performance based on impulse response between talker and microphone towards newly suitable reverberation criterion on distant-talking speech recognition. As a result of evaluation experiments in real acoustic environments, we could confirm the relation between speech recognition performance and early reflection signals. In future work, we will try to propose the newly suitable reverberation criterion. [Work supported by MEXT of Japan.]

**1pSC34. Acoustic modeling with contextual additive structure for hidden Markov model-based speech recognition.** Kazuhiro Nakamura, Heiga Zen, Yoshihiro Nankaku, and Keiichi Tokuda (Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan)

This study proposes an acoustic modeling technique based on an additive structure of context dependencies for HMM-based speech recognition. Context dependent models, e.g., triphone HMMs, are widely employed in speech recognition systems. Generally the dependency between phonetic contexts and acoustic feature distributions is direct, i.e., if a context is given, the Gaussian distribution is specified immediately. However, more complex dependencies can be considered. This study assumes an additive structure of acoustic feature components that have different context dependencies. For example, if the left and right phones are assumed as contexts of additive components, the generative process of the triphone feature is as follows: the additive components of left and right contexts are generated independently from each distribution and then added to the monophone feature. Since the output probability distribution is composed of the sum of the mean vectors and covariance matrices of additive components, a number of different distributions can be efficiently represented by a combination of fewer distributions. However, it is unknown which kinds of contexts have additive dependencies to acoustic features. Therefore, this study proposes a context clustering algorithm for the additive structure model. The proposed algorithm automatically extracts additive components by constructing multiple decision trees simultaneously.

**1pSC35. Towards an efficient archive of spontaneous speech: Design of computer-assisted speech transcription system.** Hiroaki Nanjo (Faculty of Sci. and Technol., Ryukoku Univ., Seta, Otsu 520-2194, Japan) and Tatsuya Kawahara (Kyoto Univ., Sakyo-ku, Kyoto 606-8501, Japan)

Computer-assisted speech transcription (CAST) system for making archives such as meeting minutes and lecture notes is addressed. For such a system, automatic speech recognition (ASR) of spontaneous speech is promising, but ASR results of spontaneous speech contain a lot of errors. Moreover, the ASR errors are essentially inevitable. Therefore, it is significant to design a good interface with which users can correct errors easily in order to take advantages of ASR for making speech archives. It is from these points of view that our CAST system is designed. Specifically, the system has three correction interfaces: (1) pointing device for selection from competitive candidates, (2) microphone for respeaking, and (3) keyboard. One of the most significant correction methods is selection from competitive candidates, thus more accurate competitive candidates are required. Therefore, generation methods of competitive candidates are discussed. Then, a speech recognition strategy (decoding method) based on the minimum Bayes-risk (MBR) framework is discussed. Since MBR decoding can distinguish and deal with content words and functional words, which are conventionally treated in a same manner, the approach is expected to generate more content words in competitive candidates.

**1pSC36. Acoustic modeling of spontaneous speech of Japanese preschool children.** Izumi Shindo, Tobias Cincarek, Tomoki Toda, Hiroshi Saruwatari, and Kiyohiro Shikano (Nara Inst. of Sci. and Technol., 8916-5 Takayama-cho, Ikoma-shi, Nara 630-0192, Japan)

In recent years, there is an increasing demand for speech recognition of children. However, the recognition of children's speech, especially preschool children (2 to 5 years of age), is very difficult. For example, recognition accuracy using a children's acoustic model provided by the Japanese Dictation Toolkit is only 21.4%. Many different variations of child speech with palatal sounds and pronunciation error decrease recognition performance. This paper proposes a recognition method that investigates the characteristics of preschool children's speech using experimental data and considers phonetic changes. Mapping between standard and altered pronunciations of words is determined. In experiments, a large amount of spontaneous child speech (2 to 15 years of age) was collected with the speech-oriented public guidance system, "Takemaru-kun," which is currently available. Recognition performance increases to 49.2% by acoustic model adaptation of preschool children's speech. When allowing multiple pronunciation variations per word during recognition, further improvement to 52.0% is achieved.

**1pSC37. Robust speech recognition for the control of wheelchairs by inarticulate speech of the severely disabled.** Hiroaki Kojima, Akira Sasou, Soo-Young Suk, Shi-Wook Lee, and Ken Sadohara (Nat. Inst. of Adv. Industrial Sci. and Technol. (AIST), Central 2, Umezono, Tsukuba, Ibaraki 305-8568, Japan, h.kojima@aist.go.jp)

The goal of this research is to develop voice-controlled wheelchairs that can be operated by inarticulate speech affected by severe cerebral palsy or quadriplegia, for instance. In this case, principal factors obstructing recognition performance are significant pronunciation variation caused by difficulty in stable articulation and bad influences of variety of noise in the real environment. To cope with the pronunciation variations, pronunciation lexicons that consist of multitemplates of reference patterns are utilized. The reference patterns are represented with subphonetic codes, which can describe variations of inarticulate speech more precisely than ordinary phonetic transcriptions. Pronunciation lexicons are generated by generalizing coded samples into compact representation of templates based on DP and data mining. For noise robustness, a voice activity detection method is investigated in order to circumvent friction of microphone, cough, etc. A sound source localization using a microphone array is also integrated in order to reject sounds from outside of the wheelchair. These methods are integrated into a system that can be mounted on a wheelchair. A usability test operated by a severe cerebral palsy patient in the real environment results in 95.8% accuracy within 1404 samples for the recognition of the five-command set. [Work supported by MEXT.]

**1pSC38. Selection of recognition candidates based on parallel recognition of sentences and partial words in spoken dialogue.** Makiho Kawakami (Faculty of Eng., Chiba Univ.; 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan), Hiromasa Terashi, Masafumi Nishida, Yasuo Horiuchi, and Akira Ichikawa (Chiba Univ., Inage-ku, Chiba 263-8522, Japan)

We have proposed a method that predicts a user's utterances in spoken dialogue systems by recognizing prediction sentences of each dialogue state, thereby decreasing recognition errors. However, the conventional method might repeat recognition errors because it confirms the whole sentence even if it recognizes only a part of a long word, such as a compound word, correctly. For this study, we introduce a method using decoders based on prediction sentences and partial words obtained by dividing long words. The proposed method can confirm a user's utterances by selecting a candidate to a partial matched word using recognition results of the prediction sentence and a divided partial word. We conducted experiments for 560 utterances of seven persons using a navigation system. Results showed that the word recognition accuracy was 81.1%. The rate at which recognition results by the prediction sentence and a divided

partial word were an exact or partial match was 70.5%. The rate of recognition errors when the recognition results were an exact or partial match was 0.9%. Therefore, we demonstrated that it is possible to confirm a user's utterances using the proposed method by selecting recognition candidates.

**1pSC39. Multiple utterance prediction based on a tree structure of dialogue states.** Hiromasa Terashi, Masafumi Nishida, Yasuo Horiuchi, and Akira Ichikawa (Grad. School of Sci. and Technol., Chiba Univ., 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan)

A problem of spoken dialogue systems is that recognition errors increase when the recognition-word vocabulary is large. To solve the problem, we propose a method to predict a user's utterances by recognizing prediction sentences of each dialogue state. The method can judge whether the user's utterance is within the range of prediction using decoders based on a prediction sentence and a large word vocabulary. However, the conventional method performs processing to lead in a prediction frequently when a topic changes because even an utterance within a domain is treated as an utterance of the prediction outside. In this study, we propose a new method using multiple predictions by performing recognition based on all prediction sentences in a domain. The proposed method produces a tree structure for dialogue states in a domain and regards child nodes of an arbitrary node as within the range of prediction. We conducted experiments for 451 utterances of 15 dialogues by ten persons using the university guide system. As a result, recognition accuracy of prediction sentences was 97.7% and the judgment accuracy of utterance predictions was 88.2%; high performance was obtained using the proposed method. Results demonstrate that it is possible to perform flexible dialogue control.

**1pSC40. Document-level optimization in speech recognition.** Rie Nakazato, Kugatsu Sadamitsu, and Mikio Yamamoto (Univ. of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, Ibaraki-ken, 305-8573, Japan)

Most ASR systems optimize scores at the sentence level because language models are designed to assign a meaningful probability to a sentence, but not to a document. However, in the last half-decade, NIPS people have specifically examined aspect-based global language models and developed generative text models such as the latent Dirichlet allocation (LDA), which can assign a meaningful probability to an entire document. We investigated a system to recognize all sentences in a read document, maximizing a global score including the document probability, in addition to assigning acoustic and local language probabilities. Document level optimization using global scores engenders a combinatorial problem in its implementation. Therefore, we used the rescoring framework with  $N$ -best recognition results for each sentence in a read document and the hill-climbing method to search for approximately optimal recognition results. In this study, we compare mixture of unigrams (MU), Dirichlet mixtures (DM), and LDAs as generative text models in experiments using read document speech data in Japanese Newspaper Article Sentences (JNAS) and academic presentation speech data in The Corpus of Spontaneous Japanese (CSJ).

**1pSC41. Generating search query in unsupervised language model adaptation using WWW.** Yasutomo Kajiura, Motoyuki Suzuki, Akinori Ito, and Shozo Makino (Grad. School of Eng., Tohoku Univ., Aramaki, Aoba-ku, Sendai-shi, Miyagi, 980-9579 Japan, makino@ecei.tohoku.ac.jp)

To improve the accuracy of an LVCSR (large vocabulary continuous speech recognition) system, it is effective to gather text data related to the topic of the input speech and adapt the language model using the text data. Several systems have been developed that gather text data from World Wide Web using keywords specified by a user. Those systems require the user to be involved in the transcription process. However, it is desirable to automate the entire process. To automate the text collection, we propose a method to create an adapted language model by collecting topic-related



text from World Wide Web. The proposed method composes the search query from the first recognition result, and it gathers text data from the WWW and adapts the language model. Then the input speech is decoded again using the adapted language model. As the first recognition result contains recognition errors, we developed a method to exclude the mis-recognized words using word-based confidence score and similarities between keywords. To evaluate the proposed method, we carried out adaptation experiments.

**1pSC42. Collection of multimodal data in real-world driving.** Takashi Kusakawa, Chiyomi Miyajima (Grad. School of Information Sci., Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan, kusakawa@sp.m.is.nagoya-u.ac.jp), Takanori Nishino (Nagoya Univ., 464-8603, Japan), Katsunobu Itou (Hosei Univ.), and Kazuya Takeda (Nagoya Univ., 464-8603, Japan)

Our research group has recently developed a new data collection vehicle equipped with various sensors for the synchronous recording of multimodal data including speech, video, driving behavior, and physiological signals. Driver speech is recorded with 12 microphones distributed throughout the vehicle. Face images and a view of the road ahead are captured with three CCD cameras. Driving behavior signals including gas and brake pedal pressures, steering angles, vehicle velocities, and following distances are recorded. Physiological sensors are mounted to measure the drivers' heart rate, skin conductance, and emotion-based sweating on the palm of the hand and sole of the foot. The multimodal data are collected while driving on city roads and expressways during four different tasks: reading random four-character alphanumeric strings, reading words on billboards and signs seen while driving, interacting with a spoken dialogue system to retrieve and play music, and talking on a cell phone with a human navigator using a hands-free device. Data collection is currently underway. The multimodal database will be published in the future for various research purposes such as noise-robust speech recognition in car environments, detection of driver stress while driving, and the prediction of driving behaviors for improving intelligent transportation systems.

**1pSC43. Comparison of visual features for audio-visual speech recognition using the AURORA-2J-AV database.** Takahiro Togo, Yukitaka Nimura, Takayuki Kitasaka, Kensaku Mori, Yasuhito Suenaga, Chiyomi Miyajima, and Kazuya Takeda (Dept. of Media Sci., Grad. School of Information Sci., Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan, ttogo@suenagam.is.nagoya-u.ac.jp)

Various techniques for noise-robust speech recognition have been developed for many years. However, speech recognition in nonstationary noisy environments, such as in car conditions, is still a very challenging research problem. Audio-visual speech recognition is a promising technique for improving noise robustness in such adverse acoustic conditions. Our research group has recently developed an audio-visual speech database "AURORA-2J-AV" following the AURORA2 database protocol, which has been widely used for evaluating the performance of audio-only speech recognition techniques. The AURORA-2J-AV database consists of audio speech signals and two types of image sequences: color and infrared human face images recorded when multi-digit numbers are spoken by about 100 native Japanese speakers. Visual noise as well as acoustic noise is added to the database after recording for simulating various conditions. Audio-visual speech recognition experiments are conducted using the audio speech and infrared images of the AURORA-2J-AV database. Two visual features, correlation coefficients (CCs) between two successive image frames and principal component scores (PCSs) of lip images, are compared in various signal-to-noise ratio (SNR) conditions. Experimental results show that CCs outperform PCSs at high SNRs, while PCSs function better at lower SNRs.

**1pSC44. A method for estimating the degree of a speaker's anger using acoustic features and linguistic representation.** Yoshiko Arimoto (Grad. School of Bionics, Comput. & Media Sci., Tokyo Univ. of Technol., 1401-1 Katakura, Hachioji, Tokyo, 192-0982, Japan), Sumio Ohno, and Hitoshi Iida (Tokyo Univ. of Tech., Hachioji, Tokyo, 192-0982, Japan)

Automatic speech recognition (ASR) systems are greatly demanded for customer service systems. With advanced interactive voice response systems, humans have more opportunities to have dialogues with computers. Existing dialogue systems process linguistic information, but do not process paralinguistic information. Therefore, computers are able to obtain less information during a human-computer dialogue than a human can during a human-human dialogue. This report describes a study of the estimation method of degree of speakers' anger emotion using acoustic features and linguistic representation expressed in utterances during a natural dialogue. To record utterances expressing the users' internal anger emotion, we set pseudo-dialogues to induce irritation arising from discontentment with the ASR system performance and to induce exasperation against the operator while the user makes a complaint. A five-scale subjective evaluation was conducted to mark each utterance with a score as the actual measurement of anger emotion. As a result of this, an emotional speech corpus was produced. We examine the acoustic features and features of linguistic representation of each utterance with reference to these anger score. Then we conduct experiments to estimate automatically the degree of anger emotion using parameters selected from those features.

**1pSC45. Marking up Japanese Map Task Dialogue Corpus.** Tomoko Ohsuga, Shuichi Itahashi (Natl. Inst. of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan, osuga@nii.ac.jp), and Syun Tutiya (Chiba Univ.)

This presentation reports an outline of marking up the Japanese Map Task Dialogue Corpus. The project was conducted by an independent group of researchers from the faculties of Chiba University and other institutions. This corpus contains 128 dialogues by 64 Japanese native speakers. The basic design of the dialogues and recordings conform to the original HCRC Map Task Corpus of the University of Edinburgh. Two speakers participated in the map task dialogues; an instruction giver who has a map with a route and an instruction follower who has a map without a route. The giver verbally instructs the follower to draw the route on his/her map. The corpus is marked up according to XML-based TEI (Text Encoding Initiative) P5 format, which is developed in an effort to provide interchangeable and shareable text data. The transcriptions of TEI format are viewed as "tags" that describe the start and end times of utterances, the duration of pauses, nonverbal events, and synchronization of overlapping utterances. This corpus will be available through NII-SRC (Speech Resources Consortium).

**1pSC46. Towards the detection of potentially hazardous situations in vehicle traffic using driver speech and brake pedal operation.** Lucas Malta (Grad. School of Information Sci., Nagoya Univ., Nagoya-shi, Chikusa-ku Higashiyama-doori 5-10 Nakanishi biru 306 - Japan, malta@sp.m.is.nagoya-u.ac.jp), Chiyomi Miyajima (Nagoya Univ., Japan), Katsunobu Itou (Hosei Univ., Japan), and Kazuya Takeda (Nagoya Univ., Japan)

A method for automatic detection of potentially dangerous situations in motor vehicle traffic is introduced. Unlike precedent works, which typically relied on camera arrays or road-traffic monitoring sensors to detect collision incidents, the proposed approach specifically incorporates changes in a drivers' behavior, detected through driver speech and brake pedal operation. Experiments were performed using a large real-world multimedia driving database of 493 drivers, obtained from the Centre for Integrated Acoustic Information Research (CIAIR, Nagoya University). The drivers, who interacted verbally with a human operator, uttered expletive words to express negative feelings in 11 of the 25 situations that we selected as potentially hazardous. In 17 of them, sudden and intense com-

pression of the brake pedal was observed. The proposed lexicographical speech-feature-based method also detected 33 false alarms to detect 80% of these 11 scenes. As for the other 17 scenes, our method based on two-dimensional histograms of brake pressure and its dynamics achieved an 80% detection rate for 473 false alarms. Analyses of data recorded while drivers interacted with a machine and a Wizard of Oz system as well as a rank of the most commonly uttered words in dangerous situations are also presented.

**1pSC47. Computational complexity of a distance-based active search algorithm.** Masahide Sugiyama (The Univ. of AIZU, Tsuruga, Ikki-machi, Aizu-Wakamatsu-shi, Fukushima, 965-8580 Japan)

An efficient pattern (segment) search in a large-scale database is a very important technology in current widespread Internet society. To achieve

efficient searches, many efficient pattern search algorithms have been proposed. Active search (AS) is one of them; it is based on skipping of search calculation. Initially, AS was formulated between two histograms using a similarity measure. In fact, AS can be formulated using  $L_p$  distance, where  $p$  is greater than 1. The search efficiency can be evaluated using a number of distance calculations. The first result in this paper shows that the  $L_1$  distance in AS gives the minimum number of distance calculations among  $L_p$  distances. This property can be derived from a newly derived inequality between the  $L_p$  and  $L_1$  norm on output probability space. A new distance can be derived from positive weighting of any distance. The second result of this paper is that the average skip width for any weighed  $L_p$  distance cannot be greater than the most efficient distance. Therefore, no positive combination of  $L_p$  can be more efficient than  $L_1$ . The two results indicate that  $L_1$  distance is the most efficient for active searching.

TUESDAY AFTERNOON, 28 NOVEMBER 2006

HONOLULU ROOM, 1:00 TO 4:50 P.M.

### Session 1pSP

#### Signal Processing in Acoustics: Blind Signal Processing

Leon H. Sibul, Cochair

*Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804*

Shoji Makino, Cochair

*NTT Communication Science Lab., 2-4 Hikaridau, Seika-cho, Sorakugun, Kyoto, 619-0237, Japan*

Chair's Introduction—1:00

#### Invited Papers

1:05

**1pSP1. Frequency domain blind source separation in a noisy environment.** Ryo Mukai, Hiroshi Sawada, Shoko Araki, and Shoji Makino (NTT Commun. Sci. Labs., NTT Corp., 2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan)

A prototype system for blind source separation (BSS) in a reverberant and noisy environment is presented. Our system uses a small three-dimensional array with eight microphones and realizes BSS in various situations such as the separation of many speech signals located in a three-dimensional space, and the extraction of primary sound sources surrounded by many background interferences. The mixed signals observed by the microphone array are processed by independent component analysis (ICA) in the frequency domain. The permutation problem is solved by normalized basis vector clustering, which is a generalized version of directions of arrival (DOA) clustering and has several advantages to the DOA based method. The system estimates the DOA of the source signals as a by-product of the separation process. Moreover, the system has the ability to distinguish primary target signals and ambient noise. Our system performs an on-the-spot BSS of live recorded signals. Live demonstration will be performed at the meeting.

1:25

**1pSP2. Beyond the independent component analysis: New blind acoustic sound separation in real world via single-input multiple-output-based independent component analysis.** Hiroshi Saruwatari (Nara Inst. of Sci. and Technol., 8916-5 Takayama-cho, Ikoma, Nara, 630-0192 Japan)

This review discusses a new generation of blind acoustic-signal separation in the real world, focusing on a recently proposed ICA method, Single-Input Multiple-Output (SIMO)-signal-based ICA (SIMO-ICA). The term "SIMO" is referred to as a specific transmission system in which the input is a single source signal and the outputs are its transmitted signals observed at multiple microphones. The SIMO-ICA consists of multiple ICA parts and a fidelity controller, and each ICA runs in parallel under the fidelity control of the entire separation system. In the SIMO-ICA scenario, mixed multiple sources detected at the microphones can be separated, not into monaural source signals but into SIMO-model-based contributions from independent sources as they are at the microphones. Thus, the separated signals of the SIMO-ICA can maintain the spatial qualities of each sound source. This attractive feature of the SIMO-ICA shows the promise of applicability and connectivity with the other multichannel signal processing for achieving the higher performance beyond the conventional ICA methods. As the successful examples of SIMO-ICA's application, three kinds of systems are described: (i) combination with inverse filtering for MIMO deconvolution, (ii) combination with adaptive beamformer for higher-performance BSS, and (iii) combination with binary masking for real-time processing.

1:45

**1pSP3. How blind are we in blind signal processing?** Walter Kellermann (Univ. Erlangen-Nuremberg, Cauer-str. 7, 91058 Erlangen, Germany, wk@int.de)

Blind signal processing algorithms aim at blindly solving classical signal processing problems such as system identification, signal separation, or source localization. Here it is investigated to which extent some of the more popular blind signal processing concepts are really blind when they are used for signal acquisition in the acoustic domain. Blindness regarding the source configuration and the source signal properties, blindness with respect to channel properties, and, finally, blindness regarding the microphone configuration are investigated. It turns out that many algorithms are explicitly nonblind with respect to important properties of the source configuration, such as the number of sources and their pointlike nature. The source signals typically have to fulfill certain statistical properties or have to meet sparseness constraints. Some algorithms rely on a dominant direct acoustic path between sources and sensors, and linearity of the transmission channel model is implied with all convolutive mixture models. Finally, some popular algorithms are actually aiming at direction of arrival estimation, which always requires knowledge on the microphone array geometry. In essence, blind algorithms are wellinformed in many respects and are blind only with regard to a few, although decisive, properties.

2:05

**1pSP4. Informed acoustic source separation and localization.** Kevin H. Knuth (Dept. of Phys., Univ. at Albany, Albany, NY 12222)

Advances in Bayesian computational technology in the last decade have enabled the development of new source separation and source localization algorithms. These algorithms are greatly improved by the encoding of prior information about a specific problem in the form of the chosen relevant model parameters, the assignment of the likelihood functions, and the assignment of the prior probabilities of the model parameter values. I refer to such source separation algorithms as informed source separation for the reason that they are endowed with specific and often vital information about the problem. Furthermore, the Bayesian methodology allows source separation to be united with source localization simply by including the model parameters that are of interest to the researcher. Here, I will discuss the union of source separation and source localization under the Bayesian methodology, the incorporation of prior information, and the construction of an informed algorithm using the new computational technologies that allow us to estimate the values of the parameters that define these high-dimensional problems.

2:25

**1pSP5. Recovering the quality of speech degraded by reverberations in a room.** Masato Miyoshi (NTT Commun. Sci. Labs, 2-4 Hikaridai Seika-cho, Keihanna Sci. City, Kyoto 6190237 Japan, miyo@cslab.kecl.ntt.co.jp)

Speech signals captured by distant microphones in a room are usually reverberated due to wall reflections. Reverberation may seriously deteriorate the signal characteristics, thus damaging the quality of such applications as hands-free telephony or automatic speech recognition. To eliminate such room reverberation effects, the inverse-filtering of the room transfer functions (RTFs) between a speaker and the microphones appears to be a very promising approach. When applied to a microphone system, this processing should work blindly because no explicit information on RTFs or source signals is available. The blind design of an inverse-filter set may be roughly classified into three approaches: (1) calculating an inverse-filter set of RTF estimates obtained from observed reverberant signals based on the "subspace method," (2) calculating the filter set from observed signals and replicas of direct-sound signals using statistical properties inherent in speech signals, and (3) calculating the filter set directly from received signals, by whitening these signals without excessive degradation of the original signal characteristics. This presentation will provide an overview of these three approaches and introduce our speech dereverberation trials.

2:45

**1pSP6. Blind deconvolution for noisy dynamic channels.** Michael J. Roan (Dept. of Mech. Eng., Virginia Tech., Blacksburg, VA 24060)

It is common in acoustics to measure a signal that has been degraded by propagation through an unknown, noisy channel prior to measurement. While only the degraded measured signal is available, the data of interest are the original signal and the channel parameters. Often, it is desirable to reverse the filtering process by application of an inverse filter to recover the original signal. When neither the input signal properties nor the channel properties are known, this process is known as blind deconvolution (BDC). Typically, BDC algorithms assume noiseless, stationary propagation channels and input sources. These assumptions are usually violated in practical applications (e.g., noisy multipath propagation environments with moving source and receiver). To model these effects, predictive techniques are applied to incorporate *a priori* information about the system into the existing blind processing framework. The original contributions of this work follow. First, a novel formulation of the extended Kalman filter (EKF) is proposed. This allows incorporation of *a priori* information into gradient-based blind processing algorithms. This formulation is then applied to the natural gradient (NG) BDC algorithm. Finally, results are presented that suggest significant improvement in signal recovery performance over the NG BDC algorithm for dynamic noisy channels.

## Contributed Papers

3:20

**1pSP7. Blind dereverberation based on auto-correlation functions of framewise time sequences of frequency components.** Kenko Ota and Masuzo Yanagida (Doshisha Univ., 1-3, Tarata-Miyakodani, Kyotanabe, Kyoto, 610-0321, Japan, [etf1704@mail4.doshisha.ac.jp](mailto:etf1704@mail4.doshisha.ac.jp))

Proposed is a new blind dereverberation method based on auto-correlation function of framewise time series of each frequency component. Inverse filtering of source-microphone transfer functions is widely employed for suppressing the effects of reflected waves, but this method cannot be employed for cases where source-microphone transfer functions are not available or for time-variant cases. Several methods of spectral subtraction have been proposed to cope with these cases. However, most of them require transfer functions among sources and microphones. The proposed method, however, does not require transfer function inverse filters. Moreover, the method can handle frequency characteristics of sound paths. To realize dereverberation, it is necessary to estimate the delay time and frequency characteristics of reflection, though most conventional methods assume flat frequency characteristics. The proposed method estimates these parameters based on auto-correlation function of each time series of spectral components in frequency spectra obtained every frame-shift interval. The proposed method improves dereverberation performance, comparing with methods that assume flat frequency characteristics. The proposed method raises average segmental SNR by 3.4 dB and reduces reverberation time from 390 to 109 ms for a room employed for experiments. [Work supported by Knowledge Cluster Project, MEXT, and by Academic Frontier Project Doshisha University.]

3:35

**1pSP8. Sparseness-based 2ch blind source separation in reverberant environments based on direction-of-arrival estimation with a reliability measure.** Yosuke Izumi, Nobutaka Ono, and Shigeki Sagayama (Dept. of Information Sci. and Technol., The Univ. of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan, [izumi@hil.t.u-tokyo.ac.jp](mailto:izumi@hil.t.u-tokyo.ac.jp))

We present a novel sparseness-based 2ch blind source separation (BSS) method based on robust direction-of-arrival (DOA) estimation that enables reasonable separation of source signals even in a highly reverberant environment. DOA estimation for BSS is not easy in reverberant environments since reflected sound waves behave as coherent interferences in the observed signals and make ambiguous the sound source direction. Our key idea to solve this problem is to find and integrate clean, reliable time-frequency fragments through the observation with a reliability measure. For that, (1) based on a diffused sound field model of reverberation, a theoretical correlation matrix is introduced for whitening the coherent interferences, (2) the DOA of each time-frequency bin is estimated with reliability by utilizing information of neighboring bins, and (3) the reliability-weighted DOAs are integrated over the whole time-frequency bins. After that, separated signals are obtained by time-frequency masking or inverse filtering. Experimental results including comparison with other methods are also reported.

3:50

**1pSP9. Multilayered spatio-temporal gradient analysis for acoustic blind source separation.** Kenbu Teramoto, Md. Tawhidul Islam Khan, Seiichirou Torisu, and Akito Uekihara (Saga Univ., 1-Honjo, Saga-shi, Japan, 8408502, [tera@me.saga-u.ac.jp](mailto:tera@me.saga-u.ac.jp))

A novel blind source separation of a mixture of two or more voice signals has been proposed in the present paper. The separation system has been focused based on the spatio-temporal gradient analysis. The proposed algorithm utilizes the linearity among the signals: sound pressure of source signals, the three-dimensional ( $x$ ,  $y$ , and  $z$  directional) particle velocity vector, and its gradient of the observed signals, all of which are governed by the equation of motion. Principally, as the mechanism of blind source

separation uses no-*priori* information about the parameters of convolution, filtering as well as mixing of source signals, some simple assumptions such as the statistical independency of the linearly combined (mixed) observed signals containing zero mean as well as unit variance have been implied in the present separation algorithm. Therefore, the proposed method has successfully simplified the convoluted blind source separation problem into an instantaneous blind source separation problem over the spatio-temporal gradient spaces. An acoustic experiment with two female voices has been carried out to compare the simulated data as well. A eight-microphone array system has been adopted to evaluate the voice signals efficiently.

4:05

**1pSP10. Study for blind source separation on time-frequency domain considering phase information.** Fumio Sasaki, Masahito Yasuoka, and Osamu Tanaka (Dept. of Architecture, Tokyo Univ. of Sci., 1-3 Kagurazaka, Shinjuku-ku, Tokyo 1628601, Japan, [fsasaki@rs.kagu.tus.ac.jp](mailto:fsasaki@rs.kagu.tus.ac.jp))

The methods that can be done for the estimation of the number of source signals and separation of source signals using time-frequency information are proposed as a method of blind source separation. In these methods, some independent conditions are assumed in a time-frequency domain, and observed signals are expanded to time-frequency domain using wavelet analysis. However, these methods are not considered phase differences (difference of distance) between source signals and observed signals. A new method is proposed. The method can be done not only for the source separation but also the specification of the locations of source signals considering phase differences. In this method, it is necessary to assume a rather stronger independent condition than former methods. But, on account of this assumption, a function that becomes a real value only when phase differences are coincidence in time-frequency domain is determined. Using this function, first of all, the number of source signals is calculated, then the locations of source signals and source signals are calculated. The effectiveness of the method is shown using actual synthetic signals.

4:20

**1pSP11. Blind spatial subtraction array based on independent component analysis for speech enhancement and recognition.** Yu Takahashi, Tomoya Takatani, Hiroshi Saruwatari, and Kiyohiro Shikano (Speech and Acoust. Processing Lab., Nara Inst. of Sci. and Technol., 8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0192 Japan)

We propose a new blind spatial subtraction array (BSSA) that contains an accurate noise estimator based on independent component analysis (ICA) for the realization of noise-robust hands-free speech recognition. Many previous studies on ICA-based blind source separation often dealt with the special case of speech-speech mixing. However, such a sound mixing is not realistic under common acoustic conditions; the target speech can be approximated to a point source but real noises are often not point sources. Under the condition, our preliminary experiment suggests that the conventional ICA is proficient in the noise estimation rather than the direct speech estimation. Based on the above-mentioned findings, we propose a new noise reduction method that is implemented in subtracting the power spectrum of the estimated noise by ICA from the power spectrum of noise-contaminated observations. This architecture provides us a noise-estimation-error robust speech enhancement rather than a simple linear-filtering-based enhancement. Although nonlinear processing often generates an artificial distortion, the so-called musical noise, it is still applicable to the speech recognition system because the speech decoder is not so sensitive to such a distortion. Experimental results reveal that the proposed BSSA can improve the speech recognition rate by 20% compared with the conventional ICA.

**1pSP12. Frequency-domain independent component analysis by overlap piecewise integration of separation processing.** Tadashige Noguchi, Kenko Ota, Masuzo Yanagida (Doshisha Univ., 1-3 Tatara-Miyakodani, Kyotanabe, Kyoto, 610-0321 Japan, dtg0731@mail4.doshisha.ac.jp), and Leandro Di Persia (Universidad Nacional de Entre Ros Casa de la Universidad, Paran, Entre Ros, Argentina)

Conventional frequency-domain ICA yields the optimal separation for each frequency bin, but it suffers from the permutation problem. The authors have developed permutation-free ICA as a separation scheme by obtaining the separation matrix for a long vector consisting of temporal changes of all frequency components of the received signal. The

permutation-free ICA, however, only yields a common separation matrix for all frequency bins. So, the separation matrix obtained in the permutation-free ICA has a common directivity pattern for all frequency bins, though the method can avoid the permutation problem. Proposed in this paper is a scheme of multibin ICA that deconvolves mixed signals into original source signals by shifting piecewise integration of a set of frequency bins consisting of the frequency bin in concern and neighboring frequency bins. The proposed method can yield nearly optimal directivity in the form of the separation matrix for each frequency bin, avoiding the permutation problem. Performance of multibin ICA is compared with those of conventional frequency-domain ICA and permutation-free ICA employing segmental SNR as an evaluation index. [Work supported by Knowledge Cluster Project, MEXT, and by Academic Frontier Project Doshisha University.]

TUESDAY AFTERNOON, 28 NOVEMBER 2006

WAIALUA ROOM, 1:00 TO 5:35 P.M.

### Session 1pUW

## Underwater Acoustics: Array Processing, Sensors, and Technology

Paul Hursky, Cochair

*Heat Light and Sound Research, 12730 High Bluff Dr., San Diego, CA 92130*

Hiroyuki Hachiya, Cochair

*Chiba Univ., Research Ctr. for Frontier Medical Engineering, 1-33 Yayoi-cho, Inake-ku, Chiba, 263-8522, Japan*

Chair's Introduction—1:00

### Contributed Papers

1:05

**1pUW1. Acoustic seaglider.** Bruce M. Howe (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698, howe@apl.washington.edu)

Continually improving acoustics measurements in the ocean is a *sine qua non* for the advancement of ocean acoustics and related fields. Acoustic sensors on mobile platforms will enable basic research topics on temporal and spatial coherence and the description of ambient noise, with direct impact on the applicability of coherent processing with its associated gain to signal detection and acoustic navigation and communications within the context of distributed undersea sensor networks. We describe the integration of acoustic receiving and communication capability in gliders with results from several recent field tests. [Work supported by the ONR.]

1:20

**1pUW2. Single element synthetic aperture using an ocean glider.** Georges A. Dossot, James H. Miller, Kristy A. Moore (Dept. of Ocean Eng., Univ. of Rhode Island, 217 Sheets Bldg., Narragansett Bay Campus, Narragansett, RI 02882, georges@oce.uri.edu), Steven E. Crocker (Naval Undersea Warfare Ctr., Newport, RI 02871), Jason D. Holmes (Boston Univ., Boston, MA 02215), and Scott Glenn (Rutgers University, NJ, NJ 08901-8521)

The feasibility of using a single transducer element on an ocean glider to create a synthetic aperture is discussed. Acoustic data were collected on two Webb Slocum gliders deployed by Rutgers University during the Shallow Water Experiment (SW06) on the continental shelf off New Jersey. These gliders periodically surfaced for GPS fixes and data transfer via satellite phone. A synthetic aperture is created through coherent processing of the acoustic data as the glider travels through the water. A number of issues including changes in depth, nonsteady motion of the glider, and

clock drift can limit the performance of the processing. However, the glider provides a low-noise and low-speed platform, potentially improving the signal-to-noise ratio. The geometry of the experiment provided for near-broadside recording of the 224- and 400-Hz WHOI tomography sources. An acoustic normal mode representation of the field provides the basis for processing of the data similar to the Hankel transform approach of Frisk and Lynch [J. Acoust. Soc. Am. **76**, 205–216 (1984)] and Holmes *et al.* [J. Acoust. Soc. Am. **119**, 3346 (2006)]. Effects of spatial variations in sound speed are examined. [Work sponsored by the Office of Naval Research.]

1:35

**1pUW3. Processing data from a low-drag array towed by a glider.** Paul Hursky, Michael B. Porter, Martin Siderius (Heat, Light, and Sound Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130), Vincent K. McDonald, Mark Gillcrist, Brian Granger, Ryan Jones, Aaron Bratten, Andy Huizinga, Peter T. Sullivan, and Susan G. Briest (Space and Naval Warfare Systems Ctr., San Diego, CA 92152)

Small underwater robotic vehicles such as gliders and AUVs have limited capabilities in terms of propulsion, speed, and power consumption, so it takes great care to design suitable sensor systems such as towed arrays for these platforms. Two such arrays were deployed during the Makai experiment, off the coast of Kauai in Hawaii. A 15-element line array with an acoustic aperture of 21 m was towed by a Webb Research Slocum glider. A 40-element line array with an acoustic aperture of 60 m was towed by a small work boat. We will describe the design and construction of these arrays and present results of processing data recorded on these arrays. We will also discuss potential applications for these arrays deployed from gliders and AUVs.

**1pUW4. Synthetic stationary array processing for a towed array in maneuver.** Yung Lee (Sci. Applications Intl. Corp., 1710 SAIC Dr., McLean VA 22102), William Lee (Duke Univ.), and Andrew Lee (MIT)

When long towed arrays undergo significant maneuvering, the array shape rapidly changes within a processing interval that degrades the array beamforming performance. Towed arrays are known to behave much like pulleys through the water. Thus, hydrophones on the towed array will closely follow the motion of the towed point through the water with a time lag relative to the towing speed. From this concept, it follows that, given an appropriate time alignment, a signal snapshot from the rear section of the array can be interpreted as a signal snapshot from the front section of the array at a later time. This synthetic stationary array processing (SSAP) divides the physical array into highly overlapping sequential segments and, using these segments, forms time-lagged snapshots of a stationary array. As a result, this synthetic stationary array has a reduced size and shape that is defined by the first array segment at the beginning of a processing interval. To demonstrate the effectiveness of this process, real towed array data collected from the MAPEX 2000 experiment conducted by SCALANTCEN were processed. The SSAP results successfully demonstrate processing improvements through turns. A few selected signal tracks were simulated to validate the performance of the SSAP processing.

2:05

**1pUW5. Experimental results of a fast Fourier transform-based motion compensation of synthetic aperture sonar.** Takao Sawa, Taro Aoki, Jyunichirou Tahara (Adv. Marine Technol. Program, JAMSTEC, 2-15 Natsushima, Yokosuka, Kanagawa 2370061, Japan, sawa@jamstec.go.jp), Tomoo Kamakura, Ryouji Suzuki (Dept. of Elec. Eng., UEC), and Hitoshi Yamaguchi (Nagasaki Prefectural Inst. of Public Health and Environ. Sci.)

Synthetic aperture and its motion compensation technique became available in sonar systems owing to dramatic developments of computer performance. A bottom observation test with synthetic aperture sonar was carried out at the gulf of Katagami in November 2005. The sonar on a tow-fish had detected a bottom and floats at the test. Accelerations and angular velocities of the sonar system had been recorded with external sensors on the tow-fish, too. A relative position and attitude of sonar were calculated from recorded motion data, and a new motion compensation method based on FFT was applied to the synthetic aperture sonar algorithm. In this report, details and capacity of this new method are shown through these experimental results. Sonar images with this new compensation were better than a conventional compensation that uses full bandwidth of motion frequency. Moreover, images were improved by adjusting bandwidth. The authors would like to thank Takeshi Hamazone, a captain of the mother ship, for his maneuvering the ship even in bad weather.

2:20

**1pUW6. Higher frequency matched-field processing.** Claire Debever and William A. Kuperman (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., Mail Code 0238, La Jolla, CA 92093-0238, cdebever@ucsd.edu)

We study matched-field processing (MFP) in the 3.5-kHz shallow water regime at a range of 4 km for which MFP is typically problematic. Data are used to construct both the cross-spectral-density-matrix (CSDM) and the weight vectors for the Bartlett, minimum variance distortionless response (MV), and white noise constraint (WNC) processors. There is sufficient data to construct the weight vectors as a function of time as measured around the CSDM construction time. While the MV processor degrades significantly over a few minutes, the WNC processor remains stable for 15 min or longer. We study these experimentally derived weight vectors to further understand the origin of the increased stability of the WNC over the MV and to gain further insight into how to construct a robust MFP processor. [Work supported by ONR.]

**1pUW7. Steering vector sensor array elements with cardioids, hippopedes, and other beam shapes.** Kevin B. Smith (Dept. of Phys., Naval Postgrad. School, Monterey, CA 93943) and A. Vincent Van Leijen (Netherlands Defense Acad., Den Helder, Netherlands)

The purpose of this report is to study the impact on linear beamforming due to individual vector sensor element steering patterns. Standard, linear beamformers employ cardioid beampatterns for each vector sensor. In this work, we examine a class of vector sensor element steering patterns beyond the standard cardioid. The element weighting is nonlinear but non-adaptive, making it simple to implement in hardware processing. The new sensor steering patterns, referred to as hippopeds, are products of cardioids and various orders of hippopedes. The angular resolution of individual sensors, and the impact on angular resolution from arrays of varying aperture, will serve as the performance measure along with peak-to-sidelobe levels. An example of the differences in vector sensor steering patterns is provided using measured, DIFAR buoy data. [Work supported by the Office of Naval Research, Code 3210A, the Naval Postgraduate School, and the Netherlands Defense Academy.]

2:50

**1pUW8. Wideband acoustic communication using quadrature-phase shift keying and eight-phase shift keying.** Hiroshi Ochi, Yoshitaka Watanabe, and Takuya Shimura (Marine Technol. Ctr., JAMSTEC, 2-15 Natsushima-cho, Yokosuka, 237-0061 Japan, ochi@jamstec.go.jp)

In JAMSTEC, a short-range (up to 500 m) high-speed acoustic communication system for deep sea environment has been investigating. To widen a bandwidth, the new wideband projector was developed. It consists of two transducer elements and has a toroidal beam pattern. To reduce the acoustic radiation pattern of the rear side, the unit, which is like a lampshade, was attached to that projector. Moreover, the beam axis is tilted to the upper side by delayed input signal to the upper transducer element. When this projector is used in the vicinity of the sea bottom, a bottom-reflected wave can be reduced. Carrier frequency of this projector is 80 kHz, and its bandwidth is 40 kHz. The experiment was carried out at the depth of 1000-m area using this projector. Two omni-directional hydrophones were used for receiving. Quadrature phase shift keying (Q-PSK) and 8-PSK were used as its modulation method. The data of several transmission speeds [10 k symbol per second (sps), 12.5, 16, 20, 25, and 40 ksp] were obtained at the transmission range of 200–230 m. Under those conditions, error free communication was carried out for 60 000 bits of data.

3:05–3:20 Break

3:20

**1pUW9. The study on data transmission with short positioning pulse in deep sea.** Yoshitaka Watanabe, Hiroshi Ochi, and Takuya Shimura (Japan Agency for Marine-Earth Sci. and Technol., 2-15, Natsushima-cho, Yokosuka-shi, Kanagawa pref, Japan, yoshitakaw@jamstec.go.jp)

The ultra short base line (USBL) underwater acoustic positioning in deep sea with digital data transmission is being studied. The goal of this study is improvement of the accuracy and expansion of the functions of the USBL for autonomous underwater vehicles (AUVs) and submersibles due to utilization of the transmitted data. To achieve the goal, the reliable data transmission with short positioning pulse must be ensured. The direct sequence spread spectrum (DSSS) is applied. The frequency band width of signal is 3 kHz with a center focus on 12 kHz for the deep sea use. From the ocean experimental data, it seems that the multipaths from the neighborhood have a dominant effect on the transmission quality. The results of investigation about this matter by experiments and simulations are shown. [This study is supported by a Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology in Japan.]

**1pUW10. Development of a digital 3-D SBL underwater positioning system combined with GNSS compass.** Yoshinori Miyamoto, Masahiko Furusawa, Keiichi Uchida, Toshikazu Motoyoshi (Tokyo Univ. of Marine Sci. and Technol., miyamoto@s.kaiyodai.ac.jp), Toshihiro Yamaguchi, Katuji Miwa (Koden Electron. Co., Ltd.), and Toshiharu Kakahara (Tokyo Univ. of Marine Sci. and Technol.)

A highly digitized underwater positioning system using the SBL method was developed. This system can measure highly accurate three-dimensional positions of marine organisms equipped with an ultrasonic transmitter (pinger). This system is combined with a GNSS (GPS) compass that measures the bearing, the location (latitude and longitude), and the commotion (rolling angle and pitching angle) of a boat or a buoy. Then the pinger positions are compensated for the commotion and are translated into the underwater absolute positions. The four arbitrary frequencies of the pingers are received by four wideband hydrophones. The received signals are immediately AD converted in high speed and processed by the quadrature detection and the digital filtering with a bandwidth of 3 kHz using FPGA. Finally, the receiving time differences are measured by the cross correlation, realizing a resolution of 0.01 ms. A position measurement experiment was performed in a water tank (length 15 m, width 10 m, depth 10 m), changing the position of a pinger, and it is confirmed that almost the expected accuracy could be realized. Moreover, this system was installed in a boat and the absolute positions of a fish were pursued to confirm the practicality.

## 3:50

**1pUW11. Continued Haro Strait path data processing.** A. Tolstoy (ATolstoy Sci., 1538 Hampton Hill Circle, McLean, VA 22101) and Z.-H. Michalopoulos (NJIT, Newark, NJ)

Previous efforts have led to the estimation of candidate source-array (S-R) geometries for some Haro Strait data. In particular, nw014 (one broadband shot to the NW array) has been extensively analyzed with the result that source range, source depth, array phone localizations, water depth at the source (D1), and water depth at the array (D2) can now be restricted to sets based on the time-domain data alone, i.e., on the signal boundary reflections. Most recent efforts include FFTs of the data followed by MFP geoacoustic inversion at multiple high frequencies (480–520 Hz)—using the restricted sets—in order to estimate array depths and shapes as well as some of the bottom properties for the nw014 data. Array shape and phone localizations are extremely important but difficult to attain. This presentation will describe these new efforts.

## 4:05

**1pUW12. Near-orthogonal signals.** John Piper (NSWC-PC, 110 Vernon Ave., Panama City, FL 32407)

A least-squares approach was used to construct near-orthogonal signals that can be useful in sonar applications. For the matched-filter case the solution can be written with Lagrange multiplier terms, which reduces to a simple eigenvector problem. Decoding filters are also examined, and these results are compared to the matched-filter case.

## 4:20

**1pUW13. Wave front design by acoustic contrast control.** Jung-Woo Choi and Yang-Hann Kim (Ctr. for Noise and Vib. Control (NOVIC), 4114 ME Dept., KAIST, Sci. Town, Daejeon, Korea, jungwoochoi@gmail.com)

A way to design and generate desired wave fronts using multiple sound sources is proposed. In principle, any kind of wave front can be described as a sum of orthogonal basis functions, and the wave front of a specific shape can be generated by focusing multiple sources' energy on a single basis function. Once the sound field is decomposed into a set of orthogonal functions, we can apply a conventional focusing algorithm to generate a desired wave front. Extending the concept of energy focusing, we can also generate a wave front packet, whose energy is concentrated on

a group of orthogonal functions rather than single component. This enables us to generate a wave front within a desired bound. For this purpose, we employ acoustic contrast control, which focuses sound energy on a selected group of orthogonal functions by enhancing the energy difference between the focal region and others. Various example cases demonstrate how the proposed method can generate a group of planar and spherical wave fronts propagating to different directions. [Work supported by the NRL project of KISTEP and BK21 project initiated by Ministry of Education and Human Resources Development of Korea.]

## 4:35

**1pUW14. Predicted versus measured sensitivities of fiber-optic sensors.** Fred C. DeMetz, Sr. (Sabeus, Inc., 26610 Agoura Rd., #100, Calabasas, CA 91302)

Fiber-optic sensors are finding wide use in underwater and seismic applications due to their ease of construction, wide useful acoustic bandwidth, sensor deployment ranges from the light source of 10 km or more, robust temperature and static pressure tolerance, and immunity from requiring supporting electronics near the measurement points. The theoretical sensitivity for a number of optical sensor designs are compared with experimental data. Their signal-to-noise performance is assessed from the standpoint of minimum detectable signal, environmental background, and optical noise source.

## 4:50

**1pUW15. Fiber optic interferometric hydrophone based on fiber Bragg grating with polarization diversity receiver.** Chiaki Okawara and Kenji Saijyou (5th Res. Ctr., Tech. R&D Inst., Japan Defense Agency, 3-13-1 Nagase, Yokosuka, Kanagawa, 239-0826 Japan)

The objective of this work is to overcome polarization-induced signal fading (PSF) that occurs in fiber optic interferometric hydrophones based on fiber Bragg grating (FBG). The use of FBGs as reflectors in fiber optic interferometric hydrophones would greatly simplify the construction of the devices' wet-end portion by eliminating couplers, joints, and other connections that are required for other types of fiber optic interferometric hydrophones. Unfortunately, the visibility of the interference signal is reduced by PSF, and the demodulated signal is contaminated greatly by the phase noise. A polarization diversity receiver (PDR) is therefore applied to reduce the PSF influence. We developed a prototype of fiber optic interferometric hydrophone based on FBG with the PDR and demodulate an acoustic signal to confirm the effectiveness of the PDR application. Experimental results described here show that the phase noise is suppressed by selecting an optimum port that has high visibility. Consequently, the effectiveness of the application of the PDR is confirmed.

## 5:05

**1pUW16. A new optical-fiber-based underwater sensor.** Francois M. Guillot, D. H. Trivett, and Peter H. Rogers (George W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

A new type of underwater sensor based on optical fiber Bragg gratings is presented. This transducer can be used in a dipole or a quadrupole (shear sensor) configuration. The sensing principle of the device relies on the interference signal from two Bragg gratings written on the same fiber and illuminated by a tunable, narrow-band light source. The gratings are 5 cm apart, and they each reflect a portion of the incident light. The fiber is epoxied to two spacers separated by a 1-mm gap situated between the gratings. This assembly is then adhered to two plates connected by a hinge, which is located below the gap. One plate is held rigidly and the tip of the other (free) plate experiences transverse vibrations, when ensonified. These vibrations produce periodic gap length changes, which modulate the interference signal from the two gratings. The modulation is related to the amplitude of the sound wave and is monitored with a photodetector. Design optimization to increase the sensitivity and to reduce the noise floor is discussed, and the performance of a practical sensor is presented. [Work supported by ONR.]

**1pUW17. Development of an acoustic cloud condensation nuclei counter.** Marcellin Zahui, Abhijit Deshpande, and David Delene (Univ. of North Dakota, Grand Forks, ND)

Theoretical development of an acoustic cloud condensation nuclei counter is presented. The proposed instrument will be able to count the number of aerosols present in a sample of air flowing through a growth chamber. The air condenses around aerosols and forms water droplets. The number of aerosols is determined by measuring the sound produced underwater by these droplets when the droplets strike a water surface at the bottom of the growth chamber with an impact velocity equal to either their

terminal or maximum velocities. First, the terminal velocities of the droplets are calculated. Then, the maximum velocities that these droplets can sustain without breaking are calculated as a function of droplet diameter. Second, the sound due to droplet impact is estimated and experimentally verified. In this study the bubble sound is not considered because micro droplets falling with terminal velocities do not create bubbles. Also, when accelerated, the velocities are limited such that there is no bubble formation at impact. Finally, two approaches are considered for the instrument. One case uses one hydrophone to measure the total sound radiated by all the droplets. The other case is based on a grid of tiny hydrophones, each measuring a droplet impact sound.