

Session 5aAB**Animal Bioacoustics: Session Honoring William Watkins**

Peter L. Tyack, Chair

*Department of Biology, Woods Hole Oceanographic Institution, 45 Water Street, Woods Hole, Massachusetts 02543***Chair's Introduction—8:30*****Invited Papers*****8:35****5aAB1. The development and use of an implantable radio whale tag from 1970 to the present.** Douglas Wartzok (Florida Intl. Univ., University Park PC236, Miami, FL 33199) and Romaine R. Maiefski (Ontogen Corp., Carlsbad, CA 92009)

William A. Watkins, Woods Hole Oceanographic Institution (WHOI), led the development and refinement of the WHOI Radio Whale Tag. The initial tags were developed in the 1960s at WHOI, improved in the 1970s and 1980s by Ocean Applied Research, Inc., and adapted to satellite telemetry in the 1990s by Microwave Telemetry, Inc. Through this entire period, the tag delivery and attachment methods were continually refined. The same delivery and attachment methods were used with acoustic tags, radio tags, combined radio and acoustic tags, and satellite tags. Scientific studies using these tags have been conducted on finback, humpback, Brydes, bowhead, and sperm whales. These field studies have ranged from Venezuela to the Canadian arctic and Alaska, and from the Pacific to the Mediterranean. Every tagging experiment led to new information about the animals, and in most cases the information from radio and acoustic tagged whales showed that the best available data based on surface observations of untagged animals were deficient. Highlights from some of these findings will be presented.

8:55**5aAB2. Bill Watkins and pinniped bioacoustics.** Ray G. Carleton (Dept. of Environ. Sci., Univ. of Virginia, 291 McCormick Rd., Box 400123, Charlottesville, VA 22904)

Bill's work on pinnipeds began with a 1963 paper by Schevill, Watkins, and myself in *Science* on the underwater sounds of pinnipeds, which resulted from his recordings in the highly unnatural waters of the New York Aquarium. It was written partly in reaction to assumptions being made that seals, like cetaceans, possessed sonar. Following that mode, there followed a paper by the same team in the same location on the remarkable bell tones and knockings of captive walruses. Recording from natural situations, with Schevill in McMurdo Sound, Southern Ocean, resulted in the first analysis of Weddell seal sounds. In 1969, he provided the first technical description of a marine mammal song (in the behavioral meaning of that term), that of the bearded seal, and in 1975 a detailed description of walrus sounds in association with their reproductive behavior. Sounds of other polar pinnipeds, ribbon seals of the Arctic and Ross seals of the Southern Ocean, bring the history up to date. Pinnipeds possess a suite of underwater sounds that rivals any other group. They make up for apparent lack of echo-location ability by musical quality of voice.

Contributed Paper**9:15****5aAB3. High-frequency harmonics and source level of humpback whale songs.** Whitlow Au (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734), Darling James (West Coast Whale Res. Foundation, Vancouver, BC V6R 2L3, Canada), and Kimberly Andrews (Hawaii Inst. of Marine Biol., Kailua, HI 96734)

The songs of 8 male humpback whales were recorded at ranges varying from 20 to 40 m with a calibrated system that had a flat frequency response to 22 kHz. The songs consisted of bursts of sounds called units. Units were organized into phrases and phrases into themes. Most of the units had mean duration between 1 and 2 s and mean silent periods be-

tween units between 1 and 2 s. Many of the recorded songs contained units that had high-frequency harmonics that extended beyond 15 kHz. The amplitudes of the higher-frequency harmonics of some units were within 18–24 dB of the fundamental or the highest level harmonic at frequencies out to 13.5 kHz. These harmonic results suggest that humpback whale songs have a broadband quality not previously reported and may provide some insights on the high-frequency limit of hearing in these whales. The source levels of the different songs were also estimated by considering the root mean square sound pressure level referenced to 1 m for the unit with the largest level for each different phrase within a song. Source levels varied between 171 and 189 dB *re*: 1 μ Pa.

Invited Paper**9:30****5aAB4. Bill Watkins listens to all the whales in the North Pacific Ocean.** Robert Gisiner (Office of Naval Res., 800 N. Quincy St., Arlington, VA 22217-5660)

In the early 1990s portions of the Navy's classified underwater listening systems (IUSS) were made available for research. Bill Watkins had long been familiar with this system and eagerly went to work listening to whales in the North Pacific Ocean. A decade later the work that Bill Watkins has amassed, and its impact on marine mammal research, is truly remarkable. Each month Bill's team

produces a summary of hundreds, if not thousands, of whale calls detected and localized using the Navy's listening arrays. These data have pinpointed seasonal centers of vocal activity in unexpected places, leading to changes in our strategies for monitoring several populations of endangered large whales. Bill has tracked the migratory paths of blue, fin, and humpback whales and has charted the shifts in migratory timing caused by climatic phenomena such as ENSO. Bill has taken a complex technology created for other purposes and crafted a process to get it to deliver simple, clear, and unclassified information for an entire ocean: we have not yet plumbed the limits of what we can do with this information.

9:50–10:05 Break

Contributed Papers

10:05

5aAB5. Multiyear study of blue whale calls in the north Pacific. John A. Hildebrand, Mark A. McDonald (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0205), Rex Andrew, James A. Mercer, and Bruce M. Howe (Appl. Phys. Lab, Univ. of Washington, Seattle, WA 98105-6698)

Multiyear acoustic recordings provide new insight on blue whale behavior in the Northern Pacific. We analyze six years of blue whale acoustic recordings made in the North Pacific as part of the ATOC (Acoustic Thermometry of Ocean Climate) and NPAL (North Pacific Acoustic Laboratory) projects. Blue whale calls are recorded as a component of ambient sound data, preserved as spectra with 500 Hz Nyquist frequency in 1 Hz bins, averaged over 170 s and recorded at 5 min intervals at each of 13 sites from 1994 to the present. These data verify a separation of blue whale calls into a Western Pacific version with two primary frequencies near 20 Hz and weak harmonic components, and an Eastern Pacific version with a primary near 17 Hz, a strong harmonic component near 50 Hz and a strong overtone near 90 Hz. The frequency of the Eastern Pacific blue whale call has shifted downward throughout the last 40 years, and the present data, combined with published sources, show the frequency shift to be secular and nearly linear with time. [Work supported by ONR.]

10:20

5aAB6. Dive patterns and source levels of sperm whales off New Zealand. Adam S. Frankel (Marine Acoust., Inc., 901 N. Stuart St., Ste. 708, Arlington, VA 22203)

Diving sperm whales off Kaikoura, NZ were recorded using a 4-element vertical linear hydrophone array deployed from a small boat. The boat was frequently within close proximity to the whale dove. Whales began to click almost immediately after leaving the surface. Dive patterns of whales were determined by estimating the location of each click using an advanced time-of-arrival algorithm developed at Cornell University. Analysis of dive patterns showed near vertical descents with speeds rang-

ing from 1–2 m/s. Received click levels were determined and source levels were estimated using a spherical-spreading loss term, as the ranges to the whales were less than the water depth. Preliminary estimates of peak-to-peak source levels range from 191 to 201 dB *re*: 1 Pa at 1 m (peak-to-peak). On one occasion, a sperm whale clicked rapidly at close range (~8 m) from a hydrophone element. Click rates as high as 22/s suggest that the whale was interrogating the metal weight at the end of the array. The source level of these clicks ranged from 170 to 205 dB. [Research supported by the Smithsonian Institute and the Cornell University Bioacoustics Research Program.]

10:35

5aAB7. Blue whale call intensity varies with ambient noise level. Sean M. Wiggins, Erin M. Oleson, and John A. Hildebrand (Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., #0205, La Jolla, CA 92093, swiggins@ucsd.edu)

Blue whales (*Balaenoptera musculus*) vary the intensity of their sound production level in response to varying ambient noise levels. Blue whales produce tonal, type B calls with a fundamental frequency of 16 Hz, and a strong third harmonic at 48 Hz. These calls, along with other low-frequency baleen whale vocalizations, ambient sea noise, and shipping noise have been monitored and recorded on continuous seafloor acoustic recorders near Cortez and Tanner Banks, offshore Southern California, since August 2000. By separately comparing the intensity in the 16 and 48 Hz spectral level bands with the intensity of ambient noise at and near those frequencies, we observe a correlation between the sound intensity level of blue whale type B calls and ambient noise levels. This is verified by the inspection of individual call intensity in various noise environments. Preliminary analysis indicates a less than unity proportional increase in the call intensity with increasing ambient noise. A trend of increased call intensity with an increased noise level has implications for the effect of shipping, and other natural and artificial noise sources in the marine environment on the behavior of blue whales.

10:50–11:20

Closing Remarks

Session 5aSA

Structural Acoustics and Vibration: Shells and Plates

Jeffrey S. Vipperman, Cochair

Department of Mechanical Engineering, University of Pittsburgh, 648 Benedum Hall, 3700 Ohara,
Pittsburgh, Pennsylvania 16804

Courtney B. Burroughs, Cochair

Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

Contributed Papers

8:30

5aSA1. How to label resonances from elastic shells. H. Uberall (Phys. Dept., Catholic Univ. of America) and M. F. Werby (NRL Code 7181, Stennis Space Center, MS)

A large class of resonances is observed from the excitation of elastic shells. A review of the type of resonances is presented. It is then shown how to label the resonances depending on material properties, ambient fluid, shape of object, and evacuation or fluid filling.

8:45

5aSA2. Transmission loss of a ribbed composite vessel. Jeffrey S. Vipperman, Deyu Li, and Ilya Avdeev (Dept. of Mech. Eng., Univ. of Pittsburgh, 531 Benedum Hall, Pittsburgh, PA 15228)

The transmission loss behavior of a new type of composite structure has been characterized through experimental structural and acoustic modal analyses as well as finite element analysis (FEA) modeling. The tapered cylindrical structure, which is being considered for expendable launch vehicles, is fabricated graphite-epoxy and has a series of internal tri-directional rib stiffeners. Light coupling between the structure and interior acoustic cavity exists, permitting separate analyses of the "structurally dominant" and "acoustically dominant" modes for the system. The modal analyses and FEA modeling efforts are complimentary since the former helps to determine appropriate FEA model parameters and the latter predicts the shape and ordering of the modes to expect in the experiments. These modes can be subsequently coupled through the modal-interaction approach for prediction purposes. In addition, they can be used to determine the structural acoustic coupling coefficients or radiation modes of the system in order to assess the relative contributions of various structural modes to the internal acoustic levels. A comparison between *in situ* transmission loss measurements and the prediction capabilities of the FEA models will be assessed. [Work supported by AFRL Space Vehicles Directorate, and University of Pittsburgh.]

9:00

5aSA3. Piezoelectric sensors on internally fluid loaded elastic shells. Peter R. Stepanishen (Dept. of Oceanic Eng., Univ. of Rhode Island, Narragansett, RI 02882) and Donald L. Cox (Naval Underwater Warfare Ctr. Div. Newport, Newport, RI)

A general analytical approach is presented to evaluate the harmonic response of electro-mechanical-acoustical systems consisting of external piezoelectric sensors on internally fluid loaded elastic shells with electrical, mechanical and/or acoustical excitations. The approach utilizes the *in-vacuo* eigenvectors of the shell as the basis functions for the solution of the fluid loaded shell and sensor interaction problem. Fluid loading on the shell is included in the approach via internal mode and frequency-dependent radiation impedances and multipoint models are used to represent the sensors. As a specific case of interest, the response of conformal piezoelectric sensors on an internally fluid loaded spherical shell is ad-

ressed using both the analytical approach and the FEM method. Numerical results are presented to illustrate the importance of acoustic and elastic interaction effects among sensors on the shell with and without internal fluid. Agreement between the FEM and analytical results is excellent over a broad frequency range. General characteristics of the electro-mechanical-acoustical responses are discussed. [Work supported by ONR.]

9:15

5aSA4. Backward projection of harmonic pressure measurements using the generalized internal source density method and SVD. Peter R. Stepanishen and Irsan S. Brodjoneboro (Dept. of Oceanic Eng., Univ. of Rhode Island, Narragansett, RI 02882)

The Generalized Internal Source Density (GISD) method [P. R. Stepanishen, J. Acoust. Soc. Am. **102**, 1955–1963 (1999)] is used to backward project acoustic harmonic pressure fields from an external measurement surface of revolution enclosing the finite three-dimensional vibrating body that generated the field. Both the normal velocity and pressure fields on the vibrating surface can be readily reconstructed using the method. The GISD method is based on first decomposing the acoustic field on the measurement surface into a modal summation of circumferential orders, where each modal pressure field is associated with an internal linear distribution of ring sources on the axis of revolution of the measurement surface. Each modal linear distribution of ring sources is evaluated using Singular Value Decomposition (SVD) methods. A brief review of the method is presented and discussed and then numerical results are presented to illustrate the accuracy of the method for a spherical shell and finite cylindrical shells with hemispherical end caps. The results show excellent agreement with the prescribed normal velocities and associated pressure fields over a wide frequency range.

9:30

5aSA5. Vibration of *in vacuo* hemiprolate spheroidal shells. Sabih I. Hayek (Dept. of Eng. Sci. and Mech., Pennsylvania State Univ., University Park, PA 16802) and Jeffrey E. Boisvert (Naval Undersea Warfare Ctr. Div. Newport, Newport, RI 02841)

The equations of motion for nonaxisymmetric vibration of hemiprolate spheroidal shells of constant thickness were derived using Hamilton's principle. The thin shell theory used in this derivation includes shear deformations and rotatory inertias. The shell is clamped at the equator and is excited by mechanical surface force fields. The displacements and rotations were expanded in infinite series of comparison functions. These include associated Legendre functions in terms of the prolate spheroidal angular coordinate and circular functions in the azimuthal angle coordinate. Five-branched frequency spectra were computed for several shell thickness-to-length ratios ranging from 0.005 to 0.1, and for various diameter-to-length ratios, including the limiting case of a spherical shell. Numerical results were obtained for the frequency spectra of the driving

and transfer mobilities of these shells due to surface force excitations. Some comparisons of the dynamic response of hemispherical and full spherical shells are presented. [Work supported by ONR and the Navy/ASEE Summer Faculty Program.]

9:45

5aSA6. Low-frequency resonances and scattering enhancements of water-filled cylindrical shells open at each end. Curtis F. Osterhoudt and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

At low frequencies, the contents and termination of cylindrical steel shells can significantly affect the scattering properties. We have observed a giant resonance associated with the limiting case of a water-filled cylindrical shell open at both ends. The resonance, which is analogous to the fundamental mode of an open organ pipe, occurs because of the strong low-frequency reflection of internal acoustic waves from the open ends of the shell [H. Levine and J. Schwinger, *Phys. Rev.* **73**, 383–406 (1948)]. Since the wavelength is approximately twice the length L of the cylinder, the wave number radius product ka may be much less than unity when L/a is large. The mode was excited by an external acoustic transducer and was detected either as a backscattering enhancement or (in the case of a sufficiently large cylinder) with a hydrophone within the cylinder. By tuning the frequency of an incident tone burst to the mode frequency, the echo is like the response of a high- Q harmonic oscillator, showing a gradual buildup and decay. [Work supported by ONR.]

10:00–10:15 Break

10:15

5aSA7. Impedance tuning of an electromechanical acoustic liner. Stephen B. Horowitz, Toshikazu Nishida (Dept. of Elec. and Computational Eng., Univ. of Florida, 460 NEB Bldg. 33, Gainesville, FL 32611-6130, steveh23@ufl.edu), Louis N. Cattafesta, III, and Mark Sheplak (Dept. of Aerosp. Eng., Mech., and Eng. Sci., Univ. of Florida, 231 Aero Bldg., Gainesville, FL 32611-6250)

The results of a theoretical and experimental study of the impedance tuning capabilities of an electromechanical acoustic liner are presented. An electromechanical acoustic liner utilizes an array of Helmholtz resonators, each containing an electroactive, compliant-walled cavity, to couple the electrical, mechanical and acoustical energy domains. Due to this coupling, the acoustic impedance of this liner is a function of elements from all of these energy domains. Therefore, the acoustic impedance of the liner can be modified via an electrical network. In this study, a prototype of a single cell of an electromechanical acoustic liner is developed and tested. The prototype consists of a cylindrical Helmholtz resonator with a circular, piezoelectric composite backplate, consisting of a thin brass disk bonded to a disk of PZT. An electrical network of resistors, inductors, and capacitors was connected across the terminals of the PZT to implement tuning. Normal-incidence acoustic impedance measurements were taken in an impedance tube from 1 to 6.4 kHz for various values and combinations of electrical components. The measured impedance spectra are then compared to theoretical impedance spectra based on lumped element modeling.

10:30

5aSA8. Dynamic stability of a beam excited by moving masses. Seroj Mackertich (The Pennsylvania State Univ., Harrisburg, Middletown, PA 17057, oct@psu.edu)

The dynamic stability of an elastically supported Timoshenko beam excited by constant velocity equally spaced traveling masses has been investigated. The regions of dynamic stability are determined for different values of the elastic foundation stiffness. Floquet theory is utilized to study the parametric regions of stability and instability, which are displayed in graphical form. Since the occurrence of this dynamic instability reduces the axial buckling load of the beam, the result is important for the study of buckling of a continuous beam.

10:45

5aSA9. Variable structure control of vibration of the simply-supported plates. Jiaqiang Pan (Dept. of Instrument Sci. and Eng., Zhejiang Univ., Hangzhou 310027, PRC) and Shuwen Pan (Dept. of Civil and Environ. Eng., Univ. of California, Irvine, CA 92697)

Robust active vibration control technology is still on its way to develop, though there has been a great number of studies on active vibration control (AVC) topic published since 1980s. The key is to find a good control strategy fitting AVC use and with strong control robustness, in addition to other qualities. As a newer member of robust control strategies, variable structure control (VSC) has attracted control researchers to study its application potentiality at different engineering areas for recent years. Few of them, however, have already paid attention to use in AVC area. In this paper, VSC design methodology used for AVC of a simply-support rectangular plate is investigated. The control design process at the modal state-space is deduced and the matching conditions introduced. Numerically simulated results by using VSC under scheme of the independent modal space control are obtained and compared to those obtained by using least quadratic control and the poles allocation method. It shows that VSC has advantages over the traditional control strategies at control robustness and can control more vibratory modes. [Work supported by the Zhejiang Science and Technology plan fund, PRC.]

11:00

5aSA10. Hydrofoil singing on fast ferries. Konstantin I. Matveev (Caltech, MC 301-46, Pasadena, CA 91125, matveev@caltech.edu)

Hydrofoil singing phenomenon is sometimes observed on new and repaired fast ferries with hydrofoil systems. Hydrofoil singing is a tonal sound caused by interaction of the foil trailing edge with incoming flow. The intensity of the excited tone often dominates the total noise. Hydrofoil singing has very negative effects: first, the noise level is unacceptable for passenger transportation; second, vibration of the foil structure participating in singing leads to fatigue of the system and crack formation on foils and struts. The problem of hydrofoil singing first emerged on ferries with steel foils, but especially critically it was manifested on hydrofoil systems made from titanium. Experimental results obtained on a recent sea-going hydrofoil vessel are reported. Based on the structural peculiarities of the foil system, vortex shedding process and vibration of the trailing edge are discussed, and a lift-oscillator model is used for mathematical modeling. A new method of the special trailing edge profiling is considered; it allows elimination of the tone without degrading the hydrodynamic characteristics of the high performance vehicles.

11:15

5aSA11. Application of feedback control design to improve the vibrator distortion. Bai-Tang Lee and Yun-Hui Liu (Bldg. 16, 321 Kuang Fu Rd., Sec. 2, Hsinchu, Taiwan 300, ROC)

In the beginning, the power amplifier with vibrator combination of the low frequency primary vibration calibration system has total harmonic distortion (THD) of acceleration higher than 2% at most frequencies. To get better calibrated values and to meet the requirements of the International Organization for Standardization (ISO) 16063-11; a method has been developed to reduce the THD. The feedback control method is designed to reduce the signal distortion from the vibrator. The vibrator signal is measured by an accelerometer located on a moving platform that is driven by the vibrator. This signal is fed into a controller at which the difference between the signal and a pure sine wave is calculated and a modified waveform is obtained. After signal processing on the controller, this modified signal is fed back as the input signal of the vibrator to reduce the THD. As expected, the THD measured is below 2%.

Session 5aSC

Speech Communication: Speech Recognition and Synthesis (Poster Session)

Astrid Schmidt-Nielsen, Chair

Naval Research Laboratory, Code 5513, 4555 Overlook Avenue, SW, Washington, D.C. 20375-5337

Contributed Papers

All posters will be on display from 8:30 a.m. to 11:00 a.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 9:45 a.m. and contributors of even-numbered papers will be at their posters from 9:45 a.m. to 11:00 a.m.

5aSC1. Speaker identification with dynamic synapses. Alireza Afshordi Dibazar, Hassan Heidari Namarvar, Jim-Shih Liaw, and Theodore Berger (Dept. of Biomed. Eng., Univ. of Southern California, 3650 McClinton St., Los Angeles, CA 90089-1451, dibazar@usc.edu)

In this paper, we have proposed a new method of presenting a speech signal to the dynamic synapses neural network (DSNN) and optimal selection of input neurons for the speaker identification task. The DSNN developed by Liaw and Berger (1996) was employed to capture the speaker information from the action potentials, which was generated by wavelet filter banks. In order to sufficiently capture the speaker information, the time-delayed temporal patterns of action potentials are used as the inputs of the network. The optimal number of input neurons obtained by minimizing the total output error of the network while identifying a target speaker and rejecting the other unknown speakers. The genetic algorithm was employed to design a supervised learning rule for training of the network. The TI-46 dataset was used in both the training and testing phase. Comparative results with previous methods showed a better performance of the correct classification for the model-based dynamic synapses speaker identification system. [Work supported by DARPA and ONR.]

5aSC2. Discrete type dynamic synapses neural network: Speech recognition application. Hassan Heidari Namarvar (Dept. of Biomed. Eng., Univ. of Southern California, OHE-500, Los Angeles, CA 90089-1451, heidarin@usc.edu), Jim-Shih Liaw, and Theodore W. Berger (Univ. of Southern California, Los Angeles, CA 90089-1451)

A discrete type of the Dynamic Synapses Neural Network (DSNN) has been developed and applied to speech recognition. In order to speed up the training time of the network, a new discrete time implementation of the original DSNN [J.-S. Liaw and T. W. Berger, 1996] has been introduced based on the impulse invariant transformation. The new architecture of the network was trained with the Genetic Algorithms [H. H. Namarvar *et al.*, 2001] and tested against the continuous-type DSNN. The overall speed of the new algorithm with discrete-time difference equation set is about 13 times faster than the same algorithm with the continuous differential equation set. This significant reduction of processing time not only decreases the training time but also makes the system better suited for real-time speech recognition tasks. [Work supported by DARPA.]

5aSC3. VUI aide to character recognition. Qiguang Lin, Mario Munich, and Kun Xia (VocalPoint Technologies, 847 Howard St., San Francisco, CA 94103)

Recognizing characters over the telephone is a challenging task, not only for computers but even for human beings. The reason is because there exist several highly confusable pairs/groups, e.g., the E-set, M/N, F/S, and A/J/K. Human beings resolve errors by saying phrases like: "F as in

Frank." In this report, a new character recognizer is first described, and then a novel voice-user-interface (VUI) is suggested to achieve better user experience with the task. The recognizer provides a hidden Markov model (HMM) for each character, augmented with two context-independent silence models. Furthermore, state-tying is made so that only discriminative states in individual HMMs are unique, while other states are shared. The next step is to design an effective VUI to facilitate character recognition. The VUI asks the user to say one character at a time. It then plays back the best result of recognition for confirmation. If the result is correct, the user proceeds; otherwise, the user says a keyword such as "correction," and the recognizer retrieves an alternative candidate. Candidate selection can be made, for instance, according to an N-best list or a prior confusion matrix. It is found that the proposed VUI effectively disambiguates pairwise confusions such as F/S.

5aSC4. Tactical communications bring new challenges to automatic speech recognition. Thomas H. Crystal (Consultant, 2312 S. Joyce St., Arlington, VA 22202-2103, thcrystal@ieee.org), Astrid Schmidt-Nielsen, and Elaine Marsh (Naval Res. Lab., Washington, DC 20375-5337)

Automatic speaker-independent recognition of conversational speech has been making significant progress in recent years. Military communications challenge the robustness of current recognition systems. Compared to telephone conversations or news broadcasts, tactical communication has a reduced vocabulary. It consists of short utterances, limited to the task at hand, with occasional chat words. What makes recognition difficult is high levels of background noise in tactical environments (helicopters, tanks) and the degradation of the signal by military microphones, including noise canceling microphones. Of perhaps more importance, people alter their speech to overcome these degradations and the loss of intelligibility from communicating over vocoders. The DARPA-sponsored SPINE program is exploring these issues using speech from pairs of participants performing a collaborative task while communicating between separate sound booths. Each person sits in an accurately reproduced military background noise environment and uses the type of microphone and ear protection that would normally be found in that environment. The communication is conducted through vocoders. This presentation will review results from the SPINE evaluations of state-of-the-art speech recognition systems and from analyses of the transcripts in an attempt to understand how environmental factors and speaking changes affect recognizer performance. [Work supported by DARPA.]

5aSC5. A case study in optimizing TTS for Web content. Cynthia Hagstrom and Qiguang Lin (VocalPoint Technologies, 847 Howard St., San Francisco, CA 94103, hagstrom@vocalpoint.com)

This is a case study of linguistic optimization of the Speechify text-to-speech (TTS) engine for use in the VocalPoint voice-browser and applications. To browse a website by phone the TTS engine must be able to

handle two types of text, static system messages and dynamic content. The static messages include basic browser prompts such as the Default prompt: "What would you like?" or Help messages. In addition, the system must also handle dynamic content from the websites that are being browsed. The criteria of intelligibility, naturalness, and context-sensitivity were used to evaluate all the TTS output. Every static message was tested using a Speechify dialog tool. The TTS quality for dynamic content was evaluated through pronunciation bugs reported while browsing websites such as My Yahoo. Our study found that the text normalization module needs to be expanded to cover the vast variety of text on the Web. The study also suggested that three additional functionalities would improve the caller experience: (1) make it possible to manipulate vowel length through punctuation or tagging; (2) make it possible to change intonation by manipulating pitch and phrase accents and tone breaks; and (3) add another phoneme to the sound inventory to allow flaps.

5aSC6. An introduction to the Buckeye corpus of conversational speech. William D. Raymond, Matthew J. Makashay, Robin Dauricourt, Keith Johnson (Dept. of Linguist., The Ohio State Univ., 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210, raymond@ling.ohio-state.edu), and Mark Pitt (Dept. of Psych., The Ohio State Univ., 222 Lazenby Hall, 1827 Neil Ave., Columbus, OH 43210)

Words in connected speech are subject to many types of phonological variation. Variation is both widespread and conditioned by an array of factors, including speech rate and context. However, little is known about the frequencies of variants or their acoustic realization. The variation-in-conversation (ViC) group at Ohio State has set about to quantify and measure the distributional properties of variants, with an eye toward their implications for spoken word perception. The current study reports on a corpus of approximately 40 h of conversation-style speech recorded during interviews of 43 midwestern speakers, stratified for age and sex. The speech was paired with text transcription, and is being phonetically transcribed using a combination of automatic and hand alignment. Details of the collection and transcription methodology will be presented, along with some parameters of variation that have emerged from the data. For example, in the corpus *t* is more likely to flap (16% of all underlying *t*'s) than *d* (8%). However, *d* is more likely to be deleted (36%) than *t* (18%), and almost all deletions of *d* are in the word final position (33%) unlike *t* (11%).

5aSC7. Concatenative text-to-speech synthesis based on waveform interpolation (a time frequency approach). Edmilson Morais and Grzegorz Dogil (IMS—Inst. of Natural Lang. Processing, Univ. of Stuttgart, AzenbergStrasse 12, D-70174, Stuttgart, Germany)

The time domain pitch synchronous overlap and add (TD-PSOLA) is the technique most used in commercial concatenative text-to-speech (TTS) synthesis systems. However, it is well known that TD-PSOLA presents several drawbacks. In order to overcome some drawbacks of the TD-PSOLA, this work presents a method based on time frequency interpolation (TFI) [Yair Shoham]. The method introduced here is a pitch-synchronous time-frequency approach of the waveform interpolation technique (WI) [Bastian Kleijn]. The goal of this work is to show that the TFI technique presents some important advantages to concatenative TTS synthesis. It allows pitch scale modification (PSM) independent of time scale modification (TSM) in a quite straightforward manner, and with high quality. TSM and PSM can be done in a continuous way, without any limitation of pitch period resolution. Moreover, the TFI technique allows simple, flexible, and efficient procedures to smooth diphone (or any other kind of unit) boundaries. The proposed system was evaluated using diphones and prosodies generated by the Festival system [Alan Black, Paul Taylor]. Subjective tests were performed, between the proposed TFI system and the standard TD-PSOLA system, highlighting the superior quality of the proposed system in comparison with TD-PSOLA.

5aSC8. Extraction and generation of aperiodic component in speech sounds. Hideki Kawahara (Dept. of Design Information Sci., Faculty of Systems Eng., Wakayama Univ., ATR/CREST, 930 Sakaedani, Wakayama 640-8510, Japan)

An effective method to extract and represent aperiodicity in speech sounds is proposed for making refinements on a high-quality speech analysis, modification and synthesis system STRAIGHT [Kawahara *et al.*, Speech Commun. **27**, 187–207 (1999)], based on time-frequency analyses by the use of an instantaneous frequency and group delay. The proposed signal representation consists of a frequency domain aperiodicity measure and a time domain energy concentration measure to represent source attributes. The frequency domain aperiodicity measure is defined as a ratio between the lower and upper smoothed spectral envelopes to represent the relative energy distribution of aperiodic components. The time domain measure is defined as an effective duration of the aperiodic component and event distribution pattern. These aperiodicity parameters and F_0 as time functions are used to generate the source signal for synthetic speech by controlling relative noise levels and the temporal envelope of the noise component of the mixed mode excitation signal, including fine timing and amplitude fluctuations. Relations between these parameters and temporal masking effects also will be discussed. [Work partly supported by MEXT Grant (C) 11650425 and CREST Japan.]

5aSC9. A novel method to estimate voice source and formant parameters based on an ARX speech production model. Takahiro Ohtsuka (Grad. School of Eng., Utsunomiya Univ.) and Hideki Kasuya (Faculty of Eng., Utsunomiya Univ., 7-1-2, Yoto, Utsunomiya 321-8585, Japan)

A novel method is presented to automatically estimate voice source and formant parameters from a speech utterance based on an autoregressive with exogenous input (ARX) speech production model. The method has been tested with both synthetic and natural speech materials and compared with the well-established linear prediction (LP) method in terms of the accuracy of estimated formant frequencies. The new method always revealed better performance than the LP method. In particular, the error of the proposed method was as small as 6% for the materials of an average pitch frequency of 447 Hz, while the LP method gave an error of 21%. The result clearly indicates superiority of the proposed method over the LP method for very high-pitched voices generated by females and children. Other key features of the proposed method include: (1) very low first formant frequencies of the high vowels /i/ and /u/ can be accurately estimated; (2) voicing source amplitude and open quotient of the glottal flow pulse are both reliably estimated; and (3) the implementation of an adaptive prefilter in the analysis remarkably improves the accuracy of estimated parameter values.

5aSC10. Complete optimization of excitation and model parameters in parametric speech coders. Khosrow Lashkari (DoCoMo USA Labs, San Jose, CA 95110)

In a speech coding system, synthesis error (the difference between the original speech at the encoder input and reproduced speech at the decoder output) is a more relevant measure of signal distortion than linear prediction (LP) error. By minimizing the synthesis error instead of the linear prediction error, the analysis and synthesis stages become more compatible. In a previous paper (J. Acoust. Soc. Am. **109**, 2491), an Analysis-by-Synthesis (AbS) technique for joint optimization of the excitation and filter parameters in parametric speech coders were presented. Using a gradient search in the root domain, synthesis error is minimized by reoptimizing the filter parameters for a given excitation. This paper provides a major improvement over the previous work, resulting in more than 3 dB of gain in the segmental signal-to-noise ratio. The improvement is due to a new algorithm for computing the gradient vector in the gradient search procedure. The paper describes the new algorithm and report on the im-

proved results. By adding an extra minimization step, this technique can be incorporated into the existing LPC, multipulse LPC, and CELP-type speech coding standards.

5aSC11. Estimation of vocal-tract area function from lip impulse response based on lossy vocal-tract digital filter model. Takemi Mochida and Masaaki Honda (NTT Commun. Sci. Labs., NTT Corp., 3-1 Morinosato-wakamiya, Atsugi-Shi, Kanagawa Pref. 243-0198, Japan, mochida@idea.br1.ntt.co.jp)

A method for estimating vocal-tract area function from lip impulse response based on lossy vocal-tract digital filter model is provided. Previous methods of lip impulse response [M. M. Sondhi and J. R. Resnick, **73**, 985–1002 (1983)] determines the area function from impulse response measured at lips by using the incident signal excited outside the oral cavity. This method is sensitive to the noise of measured impulse response. An alternative method [M. Honda and T. Mochida, Proc. 1st ETRW-SPM, 167–170 (1996)] is advantageous for accurate estimation under noisy acoustic condition, based on the vocal-tract digital filter model that optimizes the M sections of areas by using the impulse response longer than M samples. In this method, however, the vocal-tract model is based on the transmission line that treats the effect of loss in simplified form. The new method proposed in this study applies a mixture model of the lossless distributed circuit and the lumped circuit accounting for effects such as viscous loss, heat conduction loss, and loss due to wall vibration. The results on acoustic measurements have shown that the new method has better performance over the old method using optimization scheme based on simplified transmission line model. [Work supported by CREST, JST.]

5aSC12. The unwarranted effectiveness of smoothness as a criterion for estimating vocal-tract length from speech. Edward Neuburg (IDA-CCRP Thanet Rd., Princeton, NJ 08540)

Unambiguous quantitative characterization of speech sounds would be a useful thing. This talk will be about experiments done in an attempt to find such a characterization, by studying corpora of spoken vowel tokens all of which “sound alike.” Using Linear Predictive Coding (LPC) of order 10, 4 damped sinusoids (“formants”) were recovered from the tokens, and their bandwidths were algorithmically fixed as a function of frequency. (Other formant finders have also been tried.) Hypothesizing that a token is composed of 4 damped sinusoids is equivalent to hypothesizing that it was produced by a “vocal tract” with 8 sections; the shape (8 areas) of that tract is a function of its length, which is a free parameter of the formant-to-tract transformation. The experiments show that over a corpus of tokens of a given speech sound, simply choosing the vocal-tract

length that produces the “smoothest” tract will make for shapes that are very nearly identical over all tokens, for all talkers (in particular for men and women), and are physiologically reasonable. Such 8-parameter shapes might be candidates for the desired characterization.

5aSC13. An audition-inspired method for producing high resolution spectrograms. Hilmi R. Dajani, Hans Kunov, and Willy Wong (Dept. of Elec. and Computer Eng., Univ. of Toronto, Toronto, ON M5S 1A1, Canada)

Inspired by the ability of the auditory system to dynamically track frequency and amplitude changes in modulated signals, an algorithm for high resolution time-varying spectral analysis is proposed. The main feature of the algorithm is an auditory-inspired filterbank followed by a peak detector. Mathematical analysis, with various types of synthetic modulated signals, demonstrates that the proposed method correctly demodulates these signals. The resulting time-frequency display, referred to as a fine structure spectrogram, shows the frequency and amplitude modulations in higher detail than is possible with conventional spectrograms. With recorded consonant–vowel syllables, the fine structure spectrogram detects small frequency and amplitude modulations in the initial glides and in the steady vowel formants. These modulations are not clearly seen in conventional time-frequency representations.

5aSC14. A nonstationary speech production model based on ARMAX analysis. Jun Huang and Stephen Levinson (Beckman Inst., Univ. of Illinois, Urbana, IL 61801)

Speech production is a very complicated process which includes both the linear propagative mode and the nonlinear flow mode. The plane wave propagation mode has been modeled by an autoregressive (AR) process and resulted in the widely used linear predictive (LP) model in speech production. In this work, we introduce an autoregressive moving average exogenous (ARMAX) model for speech production. We model the additional turbulence-induced acoustic sources and sinks by a nonstationary exogenous noise process. First, we classify the English phonemes into five broad phonetic categories, namely, vowels, semi-vowels, stops, fricatives, and stops. Then we identify the model orders of each phonetic category from the acoustic observation. The model parameters are estimated by the Gauss–Newton algorithm. Finally, the prediction of the speech signal is performed through a recursive way after the model parameters of each phoneme are estimated. We tested this speech production model by synthesizing English phonemes and continuous sentences. Experimental results show that we can synthesize almost humanlike, natural quality speech signal based on this nonstationary speech production model.

Session 5aUW

Underwater Acoustics: Object Scattering and Imaging

Preston S. Wilson, Chair

Department of Aerospace and Mechanical Engineering, Boston University, Room 101, 110 Cummington Street,
Boston, Massachusetts 02215

Chair's Introduction—8:25

Contributed Papers

8:30

5aUW1. Sub-critical insonification of buried spherical shells. Irena Veljkovic and Henrik Schmidt (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

The detection and classification of buried elastic targets presents a significant challenge to the shallow water mine countermeasures (MCM) community. Recent experimental and modeling studies have shown that in the case of subcritical insonification, significant evanescent field components can be converted into vertically propagating field components making the plane-wave, ray-tracing approach to the wave propagation inadequate. There has also been empirical evidence that the modal response of the target may be stronger than the direct backscattered return due to the evanescent nature of the sub-bottom field. The GOATS98 experiment had as one of its main objectives to explore the possibility of classifying features of the 3-D acoustics scattering by buried and proud objects using bistatic configurations such as autonomous underwater vehicle (AUV) and horizontal line array (HLA). An unexpected physical phenomenon of a strong delayed flexural return of a flush buried spherical shell was observed in the AUV and the HLA data at frequencies much higher than predicted. A scenario similar to the GOATS98 experiment has also been modeled using OASES-3D target modeling framework to further investigate scattering mechanisms of the flush buried spherical shell under evanescent insonification, as well as to validate the modeling capability of the package. [Work supported by ONR.]

8:45

5aUW2. Bistatic synthetic aperture imaging of buried objects from an AUV. Kevin D. LePage (SACLANT Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy, lepage@saclantc.nato.int) and Henrik Schmidt (Dept. of Ocean Eng., MIT, 77 Massachusetts Ave., Cambridge, MA 01238, henrik@keel.mit.edu)

During SACLANTCEN's GOATS98 experiment an assortment of proud and buried objects were insonified above and below critical angle by a parametric array and their scattered fields were observed bistatically on the nose array of a MIT Odyssey class AUV. Here the challenges of synthetic aperture sonar (SAS) processing for bistatic data sets are addressed and the detection of buried objects at subcritical insonification angles is demonstrated. In addition, the differences between the measured resonant characteristics of spherical target scattering at below and above critical grazing angles is discussed.

9:00

5aUW3. Buried target classification with bistatic synthetic aperture sonar. Joseph R. Edwards and Henrik Schmidt (MIT, 77 Massachusetts Ave., 5-204, Cambridge, MA 02139, jre@mit.edu)

Buried target classification is of paramount importance in mine countermeasures (MCM) applications. Traditional methods use optimal detection techniques, e.g., monostatic synthetic aperture sonar (SAS) imaging, to classify objects by target shape. The shape of a target, however, does

not necessarily indicate whether it is manmade (mine) or natural (rock), thus creating the possibility of an unacceptably high false alarm rate. Furthermore, the limited resolution available at the low frequencies required for seabed penetration may prevent the shape-based classification altogether, as the imaging wavelength is typically on the order of the target size. For these reasons, additional information about the character of the target is required. The current work investigates the use of frequency and aspect dependence of the target scattering for classification of canonical minelike targets, using a bistatic AUV-borne sonar receiver to interrogate spheres and cylinders in various burial conditions. Experimental evidence from GOATS98, an at-sea experiment performed off the coast of Elba Island, and supporting evidence from OASES-3D models are used to demonstrate clear distinctions between these canonical targets, and to illustrate the potential of a multi-static system for rapid and robust MCM. [Work supported by ONR and SACLANT.]

9:15

5aUW4. Impulse scattering processes of an elastic sphere. Christopher Feuillade (Naval Res. Lab., Stennis Space Center, MS 39529-5004)

The "movable rigid sphere" scattering model of Hickling and Wang [J. Acoust. Soc. Am. **39**, 276–279 (1966)] incorporates the translational "rebound" response of the scatterer to the incident field. However, as a physically unrealizable object, the rigid sphere has limited applicability. Recently, it was shown [J. Acoust. Soc. Am. **106**, 553–564 (1999)] that both the movable rigid sphere and Anderson's fluid sphere model [J. Acoust. Soc. Am. **22**, 426–431 (1950)] are limiting cases of Faran's elastic sphere scattering theory [J. Acoust. Soc. Am. **23**, 405–418 (1951)]. In this paper, the time domain impulse response is used to demonstrate the three dimensional scattering of an elastic sphere and its special cases. The critical factor is the shear rigidity modulus. For normal elastic values of the modulus, shear waves dominate the scattering response. When the rigidity is high, the shear waves travel more rapidly through the sphere, and its motion becomes increasingly characterized as a rebound translation without change of shape. Wave conversion at the boundary is also inhibited. For low or vanishing rigidity, only compressional waves are supported, and these determine the scattering response. The presentation will be illustrated by movie animations. [Work supported by ONR.]

9:30

5aUW5. Passive sonar calibration spheres. David M. Deveau (Naval Undersea Warfare Ctr. Det AUTEK, 801 Clematis St., West Palm Beach, FL 33401)

The need for calibrated sonar targets is addressed with the development and testing of a set of thin-walled spheres filled with a high-density fluid. Using historical research information as a guide, a set of thin-walled metal spheres was developed and filled with a high-density fluid. The combination of the spherical shape and the acoustic focusing effects of the fluid enhanced the acoustic scattering strength of the shape so that it was not only stable with temperature but also significantly greater in amplitude. The simple passive nature of the spheres makes them ideal acoustic

targets for long-term unattended deployments. Using a reference level measurement, each sphere was calibrated over a wide frequency range in order to provide the user with a curve of measured scattering strength vs frequency. The resulting curves showed a high degree of correlation between the individual spheres and the modeling that was used to extrapolate the theoretical values.

9:45

5aUW6. Target reconstruction from broadband acoustic backscatter. Quyen Q. Huynh, Jo E. Wilbur (Coastal Systems Station, Panama City, FL 32407), Nathan Intrator, and Nicola Neretti (Brown Univ., Providence, RI 02912)

Broadband acoustic backscatter data are examined for mine reconstruction using LFM and dolphin clicks. Broadband backscatter data from free-field and bottom mines were collected during tests conducted in ARL/UT's sand tank and sponsored by ONR. The data set includes multi-aspect broadband LFM (50–200 kHz) acoustic backscatter returns from a collection of targets. Both free-field and bottom target data were collected. Bottom target data sets include returns from two different levels of bottom roughness. Each data file is sampled over an interval that is large relative to the echo duration with sufficient sampling both prior to the echo onset and post-echo return to enable meaningful target characterization. Data collected over 360 deg in 2-deg increments are processed to generate a descriptive target pattern. The data are pulse compressed and system artifacts are removed using adjacent channels over the entire frequency band and within subbands. The target pattern is generated by applying the new adaptive wavelet filters to the scattered signals. In particular, the multi-aspect data that have several targets are tested to determine the ability of this new adaptive wavelet transform to reconstruct the external and internal structures of mines. Comparisons are made to previous results generated from broadband dolphin clicks.

10:00–10:15 Break

10:15

5aUW7. Sonar backscattering enhancements for small targets caused by caustics resulting from reflecting ocean surfaces. Ben Dzikowicz and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

The reflection of sound by curved top or bottom ocean surfaces can produce caustics that enhance the sound level incident on a small target. Similarly, the scattering by the target reflected off a curved surface can be focused back onto the source transducer. These effects are simulated in a laboratory water tank experiment using a concave, cylindrical, Styrofoam surface and a small spherical target. The echo structure is simplest when the target is located at the cusp of the reflection caustic. A pulse can travel directly to the target and return or it can reflect off the surface on its way there, its way back, or both. For each reflection off the concave surface there is an observed enhancement so that the latest arriving pulse has a magnitude much larger than the earlier returns because it is doubly focused. The second return (though weaker than the doubly focused return) is also stronger than the direct echo. Application of a model for caustic wave fields [K. L. Williams *et al.*, *J. Acoust. Soc. Am.* **96**, 1687–1702 (1994)] enables simple predictions for the scaling of the relative amplitudes of the echoes with frequency. [Work supported by ONR.]

10:30

5aUW8. Broadband acoustic scattering from a bubbly liquid-filled compliant cylinder with known bubble population statistics. Preston S. Wilson, Ryan D. McCormick, Eun-Joo Park, Ronald A. Roy, and William M. Carey (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215)

In previous experiments, acoustic scattering from a bubbly liquid-filled latex tube was investigated in the 5–20 kHz frequency range. The mean bubble resonance frequency was 113 kHz, and a Wood limit effective medium theory was found to quantitatively describe the scattering throughout the measurement range, using a void fraction as the model's only fit parameter [Wilson *et al.*, *Acoust. Res. Lett. Online* (in press)]. New measurements will be presented that extend the frequency range to 150 kHz and include *in-situ* void fraction measurement. A preliminary analysis of the new data shows qualitative agreement with the effective medium theory up to about 35–40 kHz. Above this range, the qualitative trends predicted by the model are no longer apparent, and an increased ping-to-ping variability in the received echo is observed. As the frequency approaches individual bubble resonance frequency, one proceeds from a continuum response where the bubble distribution is not important to a region where individual bubbles and groups of bubbles play a role. Moreover, as the frequency goes up, losses in the latex rubber tube wall should become increasingly important. An analysis of the data, which includes bubble resonance effects and tube wall losses, will be presented. [Work supported by ONR.]

10:45

5aUW9. The three-dimensional Fredholm equation of the first kind and its application to the inverse problem of submerged objects in a waveguide. M. F. Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., Catholic Univ. of America, Washington, DC 20064)

Assume one has knowledge of the Green's function in some ocean waveguide and a measurable signal scattered from a source. This may be treated by solving the Fredholm equation of the first kind (FEFK). The task is to determine any features from the measured field by solving the FEFK. We use synthetic data and produce measurements and determine what information may be recovered by solving this equation.

11:00

5aUW10. Do submerged objects twinkle from gated pulse signals in turbulent wave guides? Michael Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., Catholic Univ. of America, Washington, DC 20064)

If one illuminates the ocean with a time-gated acoustical signal will an inclusion such as an elastic shell insonified by the pulsed twinkle or, in the jargon of some researchers, will it scintillate? Twinkling from stars owes its origin to a mildly turbulent atmosphere. The ocean is often much more turbulent and we examine the distortions that one gets from scattered pulse-gated signals.

11:15

5aUW11. More on the nonhomogeneous Sturm–Louville problem and its application to inclusions in a waveguide and the extraction of target properties. M. F. Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., Catholic Univ. of America, Washington, DC 20064)

The second order solution of the wave equation is useful in adopting a method for solving the related inhomogeneous version in which the inhomogeneous term comes from a known target or an unknown object in the waveguide. Formulations are presented for these cases with examples.