Session 4pSCa

Speech Communication: Speech for Tracking Human Health State, Performance, and Emotional State I

Suzanne E. Boyce, Chair
Univ. of Cincinnati, Dept. of Communication Disorders, Cincinnati, OH 45267-0379

Chair's Introduction—1:00

Invited Papers

1:05

4pSCa1. Speech analysis in accident investigation. Malcolm Brenner (Natl. Transportation Safety Board, 490 L'Enfant Plaza SW, Washington, DC 20594, brennem@ntsb.gov)

The NTSB investigates major transportation accidents in the United States to make recommendations to prevent their recurrence. The Safety Board also examines new technologies that might assist investigations. In conducting its work, the NTSB has found speech analysis a useful new technology for providing secondary evidence on operator state for issues such as psychological stress, alcohol impairment, physical straining, and hypoxia. This talk provides examples from two investigations: the grounding/oil spill of the Exxon Valdez tanker and the crash of a Boeing 737 airliner at Pittsburgh.

1:25

4pSCa2. Common voice measures as indicators of fatigue. Cynthia M. LaJambe (201 Transportation Res. Bldg., Penn State Univ., University Park, PA 16802, cml149@psu.edu), Frederick M. Brown (Penn State Univ., University Park, PA, 16802), Rebecca M. Reichardt (Towson Univ., Towson, MD, 21252), Malcolm Brenner (Natl. Transportation Safety Board, Washington, DC, 20594), and Robert A. Prosek (Penn State Univ., University Park, PA, 16802)

Unobtrusive, economical, and readily accessible fatigue-monitoring technologies are needed especially in transportation, military operations, and security industries. Voice analysis is compatible with operational settings, given its minimal interference with hands-on work duties. Controlled laboratory studies are underway to establish the sensitivity of this fatigue-monitoring method. A recent study evaluated sleep-deprivation consequences on basic voice attributes using multiple speech tasks. Twenty-six native English-speaking 18-26 year-old subjects were screened for physical and psychological problems. Several sleep/wake cycles were monitored with actigraphy prior to laboratory participation. Vocal measures were compared between 13 speakers sleep deprived for 36 hours and 13 non-sleep-deprived controls. In the laboratory, speech was recorded during baseline sessions and on the following day. Group differences varied by speech task and vocal measure, with more sleep-deprivation sensitivity found, for example, in speaking rate as compared to fundamental frequency. Fatigue-related changes in vocal measures were associated with decrements in psychomotor reaction times and cognitive performance. Results are compared with previous studies relating psycho-physiological states to basic voice measures. Design considerations are discussed.

1:45

4pSCa3. Speaker assessment: The impact of environment on speech systems and individuals. John H. L. Hansen (Dept. of Elec. Eng., CRSS: Ctr. Robust Speech Systems, Univ. of Texas at Dallas, Richardson, TX 75083, john.hansen@utdallas.edu)

Assessing speaker variability is critical in developing a scientific understanding or model of human speech production. The environmental context plays a significant role in how this variability plays out. In this study, recent findings are presented on the variability of speech production due to environmental factors that influence man-machine interaction as well as human-to-human interaction. Speech produced under task stress, emotional stress, and background noise (resulting in Lombard effect) all cause speech production changes. This impacts both speech processing algorithms intended for speech recognition/technology and human-to-human interaction. Specifically, two speech production domains are briefly considered including Part-1: speech production under varying types/levels of background noise and how this produces flavors of Lombard effect and impacts speaker recognition systems, and Part-2: assessing the stress/emotional state of parents/care-givers in quantifying the language learning exposure of children (ages 10–36 months). In Part-1: the UT-Scope corpus is employed with speech from 30 subjects (19M,11F) for analysis of duration and spectral tilt as well as developing an automatic Lombard effect classification scheme which is incorporated into speaker recognition. Next, Part-2: considers how neutral versus stressed/emotional state of adults impacts conversational turns and adult word-count in a child language learning environment (20 child-parent interactions).

2:05

4pSCa4. Intelligibility of speech produced under sleep-deprivation conditions. Suzanne Boyce (Dept. of Comm. Sci. Disord., Univ. of Cincinnati, 3202 Eden Ave., Cincinnati, OH 45267-0379, boycese@ucmail.uc.edu), Joel MacAuslan (S.T.A.R. Corp., Bedford, MA 01730), Sandra Combs, and Alexandra Blood (Univ. of Cincinnati, Cincinnati, OH 45267-0379)

In previous work, we applied the technique of acoustic landmark detection to speech produced under rested vs sleep-deprived conditions. We found significant differences for both the number and pattern of landmarks detected. In a parallel study we found significant differences in clear vs conversational styles of speech. While it has been shown by a number of investigators that clear and conversational speech styles differ in the degree to which intelligibility to listeners is preserved in noise, it is not clear whether the articulatory changes found in sleep-deprived speech affect the ability of listeners to understand what is said, especially in noise. In this paper, we present the results of a study in which normal-hearing listeners are asked to transcribe speech presented with and without background noise. Similar levels of background noise have been shown to reduce speech intelligibility for clear vs conversational speech presented to normal listeners. The speech in this study was produced under rested and sleep-deprived conditions. Results will be compared to effects of clear vs conversational speech presented in noise and in quiet.

2:25

4pSCa5. On the acoustics of emotion in speech: Desperately seeking a standard. Bjoern Schuller (Inst. for Human-Machine Commun., Technische Universitaet Muenchen, D-80333 Muenchen, Germany, schuller@tum.de)

Researchers concerned with automatic recognition of human emotion in speech have proposed a considerable variety of segmental and supra-segmental acoustic descriptors. These range from prosodic characteristics to voice quality to acoustic correlates of articulation and represent unequal degrees of perceptual elaboration. Recently, evidence has been reported from first comparisons on multiple speech databases that spectral and cepstral characteristics have the greatest potential for the task [B. Schuller *et al.*, Linguistic Insights **97**, 285–307 (2009)]. Yet, novel acoustic correlates are constantly proposed, as the question of the optimal representation remains disputed. The task of evaluating suggested correlates is non-trivial, as no agreed "standard" set and method of assessment exists, and inter-corpus substantiation is usually lacking. Such substantiation is particularly difficult owing to the divergence of models employed for the ground-truth description of emotion. To ease this challenge, using the potency-arousal-valence space as the predominant means for mapping information stemming from diverse speech resources, including acted and spontaneous speech with variable and fixed phonetic content on well-defined binary tasks is proposed. Among the various options for automatic classification, a method combining static and dynamic features representing pitch, intensity, duration, voice quality, and cepstral attributes is recommended.

2:45

4pSCa6. The detection of stress, emotion, and deception from speech: The intersection of phonetics, policy, and politics. James Harnsberger (Dept. Comm. Disord., Univ. of Florida, 68 Dauer Hall, Gainesville, FL 32611)

While prior research on the detection of stress, emotion, and deception from speech and language has shown limited progress, this has not prevented the marketing of commercial devices that purport to detect these states to a variety of customers, such as law enforcement agencies, the military, intelligence agencies, homeland security, and insurance companies. For the major products currently on the market, all independent studies to date have failed to verify their efficacy with a wide range of speech materials collected under various experimental conditions, ranging from laboratory studies with carefully controlled speech to "mock crimes" to speech produced under realistic levels of jeopardy. This literature (including two studies by the author) will be reviewed and discussed in terms of how their experimental design and results are shaped and used in policy debates by private manufacturers, elected officials and their staffers, academic researchers, and others.

3:05—3:20 Break

3:20

4pSCa7. Automatic methods to monitor the speech of Parkinson's patients with deep brain stimulators. Craig van Horne (Caritas Neurosurgery, 736 Cambridge St., CCP 8, Brighton, MA 02135), Karen Chenausky, Joel MacAuslan (STAR Corp., Bedford, MA 01730), Carla Massari, and Marianna McCormick (Caritas-St. Elizabeth's, Brighton, MA, 02135)

Parkinson's disease (PD) is a neurodegenerative disease causing hypokinetic dysarthria, associated with "blurred" or underarticulated speech, imprecise consonants, and, sometimes, irregular syllable trains. Within the past decade, deep brain stimulation (DBS) of the subthalamic nucleus (STN) has provided substantial benefit to PD patients. DBS treatment has largely been directed toward the motoric features of PD: bradykinesia, rigidity, and tremor, but its effects on speech vary. The speech of PD patients receiving DBS treatment, with or without accompanying medical therapy, was analyzed for rate (syllables per second), regularity (relative deviation of syllable length), stop consonant spirantization (a measure of stop consonant precision), vowel ratio (length of vowel to length of syllable), and other features using automatic routines written specifically for the purpose. Patients' speech is more variable on DBS stimulation than on medication or no treatment. It is possible to find a combination of DBS settings for each patient that relieves their motor symptoms and returns their speech to normal. These findings suggest that it is possible to improve speech along with the general motor symptoms of PD. Furthermore, automatic analyses show promise as sources of feedback for neurologists to use in optimizing DBS settings for speech.

3:40

4pSCa8. Impact of cognitive load and frustration on drivers' speech. Hynek Bořil (Erik Jonsson School of Eng. and Comp. Sci., The Univ. of Texas at Dallas, 2601 N. Floyd Rd. 75080, Richardson, TX 75083-0688, hynek@utdallas.edu), Tristan Kleinschmidt (Speech and Audio Res. Lab., Queensland Univ. of Technol., GPO Box 2434, Brisbane, Queensland 4001, Australia), Pinar Boyraz, and John H. L. Hansen (The Univ. of Texas at Dallas, Richardson, TX 75083-0688)

Secondary tasks such as cell phone calls or interaction with automated speech dialog systems (SDSs) increase the driver's cognitive load as well as the probability of driving errors. This study analyzes speech production variations due to cognitive load and emotional state of drivers in real driving conditions. Speech samples were acquired from 24 female and 17 male subjects (approximately 8.5 h of data) while talking to a co-driver and communicating with two automated call centers, with emotional states (neutral, negative) and the number of necessary SDS query repetitions also labeled. A consistent shift in a number of speech production parameters (pitch, first format center frequency, spectral center of gravity, spectral energy spread, and duration of voiced segments) was observed when comparing SDS interaction against co-driver interaction; further increases were observed when considering negative emotion segments and the number of requested SDS query repetitions. A mel frequency cepstral coefficient based Gaussian mixture classifier trained on 10 male and 10 female sessions provided 91% accuracy in the open test set task of distinguishing co-driver interactions from SDS interactions, suggesting—together with the acoustic analysis—that it is possible to monitor the level of driver distraction directly from their speech.

THURSDAY AFTERNOON, 22 APRIL 2010 GRAND BALLROOM V, 4:05 TO 5:05 P.M.

Session 4pSCb

Speech Communication: Speech for Tracking Human Health State, Performance, and Emotional State II (Poster Session)

Suzanne E. Boyce, Chair
Univ. of Cincinnati, Dept. of Communication Disorders, Cincinnati, OH 45267-0379

Contributed Papers

All posters will be on display and all authors will be at their posters from 4:05 p.m. to 5:05 p.m.

4pSCb1. Using temporal cycles in spontaneous speech to quantify linguistic impairments in patients with neurodegenerative disorders. Eden Kaiser (Prog. in Linguist., Univ. of Minnesota, 214 Nolte Ctr., 315 Pillsbury Dr. SE, Minneapolis, MN 55455, kaise113@umn.edu), Serguei V. S. Pakhomov, Angela K. Birnbaum, Daniel Boley (Univ. of Minnesota, Minneapolis, MN 55455), and David S. Knopman, (Mayo Clinic, Rochester, MN 55905)

Fronto-temporal lobar degeneration (FTLD) is a form of dementia which may manifest through symptoms similar to Alzheimer's, including language-specific impairment. Linguistic manifestations of the disorder are often described in subjective assessments of the disorder, but not always easily quantifiable using objective tests. This paper investigates a speech characteristic not yet assessed in FTLD patients, that of "temporal cycles" [Henderson *et al.* (1966); Butterworth and Goldman-Eisler (1979)]. Temporal cycles in the speech of healthy adults consist of alternating and roughly equal periods of fluent and hesitant speech [Roberts and Kirsner (2000)]. A time series analysis of temporal cycles was conducted using spontaneous speech from 45 adults diagnosed with FTLD. Patients' cognitive functioning was assessed using the language-specific clinical dementia rating (CDR) scale by board-certified neurologists. Periodicity of temporal cycles was quantified using the proportion of the energy in the highest peak to the total energy in the power spectrum. It was found that this measure was correlated with independent CDR assessments. The results of this study indicate that temporal cycles may be used to characterize the effects of neurodegenerative disorders on speech communication. [Work supported by US NIA Grants Nos. R01-AG023195, P50-AG16574, P30-AG19610, and NIH NIA-1R01AG026390, and Univ. of MN Academic Health Center.]

4pSCb2. Transition characteristics in speakers with dysarthria and in healthy controls: Part IV: Additional data on vital capacity transitions and stroke patients. Gary Weismer, Christina Kuo, and Phoebe Allen (Dept. of Communicative Disord., Univ. of Wisconsin-Madison, 1975 Willow Dr., Madison, WI 53706, gweismer@wisc.edu)

Formant transitions are known to provide important cues for speech perception, sound identification, and inferences to articulatory behavior. This study describes and examines three types of formant transitions [consonant-vowel (CV), vowel-consonant (VC), and diphthong transitions] in four groups of speakers: healthy, ALS, Parkinson's disease, and stroke. This is an extension from previous work by Weismer *et al.* [J. Acoust. Soc. Am. **121**, 3135 (2007)] and Weismer *et al.* [J. Acoust. Soc. Am. **124**, 2558 (2008)], who showed shallower slopes for CV and diphthong transitions in persons with dysarthria (ALS and PD). To better understand the characteristics of the different transition types in healthy and disordered populations, two questions are addressed here. First, are CV transitions associated with dysarthria different from those in healthy speakers in a way comparable to the observed differences in diphthong transitions? Second, do VC transitions show the same normal characteristics, and are the differences between healthy speakers and speakers with dysarthria the same as in CV transitions? Distributional analyses for 50-ms CV, 50-ms VC, and diphthong (highest-lowest/lowest-highest) F2 transition measures will be presented for 18 healthy speakers, 4 speakers with ALS, 4 speakers with PD, and 20 stroke patients. [Work supported by NIDCD R01 DC003723.]