4pSAb8. Application of vibration energy flow to evaluation of thickness. Akihiko Higashi (Dept. of Maritime Sci. and Technol., Japan Coast Guard Acad., 5-1 Wakaba-cho, Kure, Hiroshima, 737-8512, Japan)

In this study, the possibility of the useful application of the vibration energy flow is investigated. The vibration energy flow means the propagation of the vibration energy of the flexural waves. The vibration energy flow is expressed by the structural intensity. Here, it is easy to input the flexural waves in the thin plates and beam elements. Then, large structures such as ships use many of these thin plates and beam elements. But the usual methods of the evaluation and the inspection of these large structures are inefficient. Then, we investigated the possibility of the evaluation of the changed thickness of the structure by using the vibration energy flow analysis. As the result of analysis, the structural intensity suddenly changes at the position of the changed thickness. The changed quantity of the structural intensity corresponds to the change quantity of the thickness. Then, the evaluation method of the thickness of the structure is proposed. As a result, it was found that the change of the structural intensity indicates the change of the thickness. And the evaluation of the change of thickness of beams could be possible by using the proposed method.

FRIDAY AFTERNOON, 1 DECEMBER 2006　　　　　　　　　MOLOKAI ROOM, 1:00 TO 4:00 P.M.

Session 4pSC

Speech Communication: Variation in Production and Perception of Speech (Poster Session)

Heriberto Avelino, Cochair

Univ. of California—Berkeley, Dept. of Linguistics, 1203 Dwinelle Hall, Berkeley, CA 94720-2650

Haruo Kubozono, Cochair

Kobe Univ., Dept. of Linguistics, Faculty of Letters, Nada-ku, Kobe 657-8501, Japan

Contributed Papers

All posters will be on display from 1:00 p.m. to 4:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:00 p.m. to 2:30 p.m. and contributors of even-numbered papers will be at their posters from 2:30 p.m. to 4:00 p.m.

4pSC1. Cross-language perception of voice and affect. Christer Gobl, Irena Yanushevskaya, and Ailbhe N. Chasaide (Phonet. and Speech Lab., School of Linguistic, Speech and Commun. Sci., Trinity College Dublin, Dublin 2, Ireland, yanushei@tcd.ie)

The paper reports on a cross-language study of how voice quality and $f0$ combine in the signaling of affect. Speakers of Irish-English and Japanese participated in perception tests. The stimuli consisted of a short utterance where $f0$ and voice source parameters were varied using the LF-model implementation of the KLSyn88a formant synthesizer, and were of three types: (1) VQ only involving voice quality variations and a neutral $f0$ contour; (2) $f0$ only, with different affect-related $f0$ contours and modal voice; (3) VQ+$f0$ stimuli, where the voice qualities of (1) combine with specific $f0$ contours from (2). Overall, stimuli involving voice quality variation were consistently associated with affect. In (2) only stimuli with high $f0$ yielded high affective ratings. Striking differences emerge between the ratings obtained from the two language groups. The results show that not only were some affects consistently perceived by one language group and not the other, but also that specific voice qualities and pitch contours were associated with very different affects across the two groups. The results have important implications for expressive speech synthesis, indicating that language/culture-specific differences need to be considered. [This work is supported by the EU-funded Network of Excellence on Emotion, HUMAINE.]

4pSC2. An articulatory study of coronal consonants in Arrernte. Marija Tabain (LaTrobe Univ., Melbourne, Australia), Richard Beare (Monash Univ., Melbourne, Australia), Catherine Best (Univ. of Western Sydney, Sydney, Australia), and Louis Goldstein (Haskins Labs., CT)

This paper presents electro-magnetic articulography (EMA) data on the four coronal stops of Arrernte, an Australian language. The stops are: the lamino-dental "th," the apico-alveolar "t," the apico-postalveolar (or "retroflex") "rt," and the lamino-palatal "ty." Jaw, tongue tip (TT), and tongue body (TB) data were collected for two female speakers of the language. Results for the first speaker show a fronted tongue position for the laminal consonants, with the TT reflecting a similar location for both the dental and the palatal. However, for the palatal, the TB position is much higher, whereas for the dental, the TB is very low. For the apical consonants, the TT is not as far forward, and the TB is not quite as high as for the lamino-palatal. For both TT and TB, apico-postalveolar is further back than apico-alveolar. For the second speaker, the TT sensor failed, but in line with the first speaker, the TB sensor showed a higher position for the palatal. The other stops were lower and more forward, with the post-alveolar TB position higher than the laminal or alveolar stop position. For both speakers, the jaw position is lowest for the postalveolar. [Work supported by Australian Research Council and NIH: NIDCD.]

4pSC3. Symbolic phonetic features for pronunciation modeling.
Rebecca A. Bates,[a)] , Mari Ostendorf (Dept. of Elec. Eng., Univ. of Washington, Box 352500, Seattle, WA 98195), and Richard A. Wright (Univ. of Washington, Seattle, WA 98195)

A significant source of variation in spontaneous speech is due to intraspeaker pronunciation changes, often realized as small feature changes, e.g., nasalized vowels or affricated stops, rather than full phone transformations. Previous computational modeling of pronunciation variation has typically involved transformations from one phone to another, partly because most speech processing systems use phone-based units. Here, a phonetic-feature-based prediction model is presented where phones are represented by a vector of symbolic features that can be on, off, unspecified, or unused. Feature interaction is examined using different groupings of possibly dependent features, and a hierarchical grouping with conditional dependencies led to the best results. Feature-based models are shown to be more efficient than phone-based models, in the sense of requiring fewer parameters to predict variation while giving smaller distance and perplexity values when comparing predictions to the hand-labeled reference. A parsimonious model is better suited to incorporating new conditioning factors, and this work investigates high-level information sources, including both text (syntax, discourse) and prosody cues. Detailed results are under review with *Speech Communication*. [This research was supported in part by the NSF, Award No. IIS-9618926, an Intel Ph.D. Fellowship, and by a faculty improvement grant from Minnesota State University Mankato.] [a)] Currently at Minnesota State University, Mankato.

4pSC4. Acoustic phonetic variability and auditory word recognition by dyslexic and nondyslexic children.
Patricia Keating, Kuniko Nielsen (Phonet. Lab., Linguist., UCLA, Los Angeles, CA 90095-1543, keating@humnet.ucla.edu), Frank Manis, and Jennifer Bruno (USC, Los Angeles, CA 90089)

The hypothesis that dyslexia involves degraded phonological representations predicts impairments in behaviors that rely on these representations, such as auditory word recognition. Normal adult listeners recognize different pronunciations of a word as instances of the same lexical item, but more slowly and less accurately; dylexics should be even more impaired by acoustic phonetic variability. Children with and without dyslexia performed a word recognition task: on each trial, a child hears a target word, then eight probes (matching the target or not), responding yes/no to each probe. On some trials, probes are spoken by multiple talkers who differ in age, sex, speech style, etc.; on some trials the match probes also differ from the target in final stop consonant allophone. Responses are scored for accuracy and speed. Research questions include: Do all children demonstrate less accurate/slower recognition of words spoken by multiple talkers versus by one talker? Do all children demonstrate less accurate/slower recognition of words spoken with different allophones? Do dyslexic children demonstrate less accurate/slower recognition than nondyslexic children and, if so, for all trials, only for multiple talker trials, and/or only for different allophone trials; for all dyslexic children, or only those with particular phonological impairments? [Work supported by NIH.]

4pSC5. Intertalker differences in intelligibility of cochlear-implant simulated speech.
Tessa Bent, Adam B. Buchwald, and David B. Pisoni (Indiana Univ., Dept. of Psychol. and Brain Sci., 1101 E. 10th St., Bloomington, IN 47405, tbent@indiana.edu)

Are the acoustic-phonetic factors that promote highly intelligible speech invariant across different listener populations? Two approaches have been taken to investigate intelligibility variation for a variety of listener populations including hearing-impaired listeners, second language learners, and listeners with cochlear implants: studies on how speaking style affects intelligibility and other research on how inherent differences among talkers influence intelligibility. Taking the second approach, we compared intertalker differences in intelligibility for normal-hearing listeners under cochlear implant (CI) simulation ($n = 120$) and in quiet (n =200). Stimuli consisted of 20 native English talkers' productions of 100 sentences. These recordings were processed to simulate listening with an eight-channel CI. Both clear and CI-processed tokens were presented to listeners in a sentence transcription task. Results showed that the most intelligible talkers in quiet were not the most intelligible talkers under CI simulation. Furthermore, listeners demonstrated perceptual learning with the CI-simulated speech but showed little learning in the quiet. Some of the acoustic-phonetic properties that were correlated with intelligibility differed between the CI-simulated speech and the speech in the quiet. These results suggest that the intertalker variations that result in highly intelligible speech observed in earlier studies are dependent on listener characteristics. [Work supported by NIDCD.]

4pSC6. The effect of phonological neighborhood density and word frequency on vowel production and perception in clear speech.
Rajka Smiljanic, Josh Viau, and Ann Bradlow (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208)

Previous research showed that phonological neighborhood density and word frequency influence word recognition (Luce and Pisoni, 1998) and vowel production (Wright, 2002; Munson and Solomon, 2004; Munson, to appear), suggesting an interaction of lexical and phonetic factors in speech production and perception Here, we explore whether hyperarticulated, intelligibility-enhancing clear speech shows similar sensitivity to lexical-level structure. Nine American English talkers (five females, four males) produced 40 monosyllabic easy (frequent words with few lexical neighbors) and hard (infrequent words with many lexical neighbors) words in conversational and clear speech. Twenty-four subjects participated in a word-in-noise listening test. Results revealed a large effect of style on intelligibility and vowel production: words were more intelligible and vowels were longer and more dispersed in clear compared to conversational speech. Moreover, the female talkers produced larger vowel spaces than male talkers in both speaking styles. Vowels in hard words were marginally more dispersed than vowels in easy words in both speaking styles. However, within both speaking styles, easy and hard words were equally intelligible and of approximately equal duration. These results showed that phonetic properties of vowels were enhanced equally in clear speech regardless of their lexical properties.

4pSC7. Phoneme dependency of accuracy rates in familiar and unknown speaker identification.
Kanae Amino, Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan, amino-k@sophia.ac.jp), and Tsutomu Sugawara (Sophia Univ., Chiyoda-ku, Tokyo, 102-8554 Japan)

For perceptual speaker identification, the identification accuracy depends on the speech contents presented to subjects. Our previous studies have shown that stimuli containing nasals are effective for identifying familiar speakers [Amino *et al.*, Acoust. Sci. Tech. **27**(4) (2006)]. We have also presented the possibility that the interspeaker spectral distances reflect perceptual speaker similarities. In the present study, we conducted an experiment in which four unknown speakers were identified by 15 subjects. The stimuli were identical to those used in the previous study, in which ten speakers were identified by familiar listeners, although the speakers were fewer this time. Nine consonants in the CV structure were used as stimuli. The consonants were /d/, /t/, /z/, /s/, /r/, /j/, /m/, /n/, and /nj/; the vowel was restricted to /a/ for all CV syllables to simplify the experiment. The results showed that the nasals /n/ and /nj/ obtained higher scores. Tendencies in the differences among consonants were on the same order as those of the

previous experiment, but the average scores were lower than those for familiar listeners. [Work supported by Grant-in-Aid for JSPS Fellows 17-6901.]

4pSC8. Speech style and stereotypical character in Japanese. Akiko Nakagawa (Grad. School of Cultural Studies and Human Sci., Kobe Univ., 1-2-1 Tsurukabuto, Nada-ku, Kobe 657-8501, Japan, akiko.nakagawa@atr.jp) and Hiroko Sawada (Kyoto Univ., Kyoto 606-8501, Japan)

This study shows that ''stereotypical character'' is necessary to understand Japanese speech communication in addition to existing conceptions such as emotion, communicative strategy, register, and so on. Stereotypical character is here defined as a complex entity, consisting of information about gender, age, social status, physical features, characteristics, and speech style. The necessity of stereotypical character was shown through an auditory experiment involving a total of 70 speech sounds comprised of 15–19 short phrases (mean duration 1.4) selected from recordings of spontaneous speech of four adult female speakers of Japanese. Ten participants were asked to listen to these speech sounds randomly, and to classify them into four speakers. Each of the resulting auditory-perceptual categories was found to contain speech sounds from more than one speaker. Further analyses of these results suggested that the participants classified the speech sounds not according to invariant speaker characteristics but according to virtual stereotypical characters that are common in Japanese society. Therefore, such changeable speaker characteristics as ''busybody'' ''thoughtful,'' ''high-handed,'' and so on, can be elicited through speech sounds by Japanese speakers. [This work was partially supported by the Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research (A), 16202006.]

4pSC9. Perceived vocal age and its acoustic correlates. Hiroshi Kido (Dept. of Commun. Eng., Tohoku Inst. of Technol., Taihaku-ku, Sendai, Japan 989-8577, kidoh@tohtech.ac.jp) and Hideki Kasuya (Intl. Univ. of Health and Welfare, Otawara, Japan 324-8501)

This study investigates relationships between perceived and chronological age of talkers and acoustic correlates of the perceived age. Most of the past studies were primarily concerned with the instability of the vocal-fold vibration extracted from sustained vowels. This study focuses on the dynamic nature of sentence utterances. Talkers included 115 healthy men, aged 20–60 years, who read a short sentence in Japanese. Listeners consisted of 70 men and women, aged 20–40 years, who made direct estimations of age. The results showed a strong correlation ($r = 0.66$) between the perceived and chronological age as well as the tendency toward overestimating the ages of younger talkers and underestimating those of older talkers, supporting past investigations [e.g., R. Huntley *et al.*, J. Voice **1**, 49–52 (1987)]. Acoustic parameters considered were median of the fundamental frequency ($F0$) contour, $F0$ range, declination of $F0$ contour, spectral tilt, median of the boundary frequencies above which irregularities dominate, and speaking rate. From both statistical graphical modeling and regression tree analysis, the speaking rate, $F0$ declination, and spectral tilt were found to be dominant acoustic correlates to the perceived age. [Work supported partly by a Grant-in-Aid for Scientific Research, JSPS (16300061).]

4pSC10. A cross-linguistic study of informational masking: English versus Chinese. Bruce A. Schneider, Liang Li, Meredyth Daneman (Dept. of Psych., Univ. of Toronto at Mississauga, Mississauga, ON, L5L 1C6 Canada, bschneid@utm.utoronto.ca), Xihong Wu, Zhigang Yang, Jing Chen, and Ying Huang (Peking Univ., Beijing, China 10087)

The amount of release from informational masking in monolingual English (Toronto, Canada), and Chinese (Beijing, China) listeners was measured using the paradigm developed by Freyman *et al.* [J. Acoust. Soc. Am. **106**, 3578–3588]. Specifically, psychometric functions relating percent-correct word recognition to signal-to-noise ratio were determined under two conditions: (1) masker and target perceived as originating from the same position in space; (2) masker and target perceived as originating from different locations. The amount of release from masking due to spatial separation was the same for English and Chinese listeners when the masker was speech-spectrum noise or cross linguistic (Chinese speech masking English target sentences for English listeners or English speech masking Chinese target sentences for Chinese listeners). However, there was a greater release from masking for same-language masking of English (English speech masking English target sentences) than for same-language masking of Chinese (Chinese speech masking Chinese target sentences). It will be argued that the differences in same-language masking between English and Chinese listeners reflect structural differences between English and Mandarin Chinese. [Work supported by China NSF and CIHR.]

4pSC11. Cross-linguistic differences in speech perception. Keith Johnson and Molly Babel (UC Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650)

This research explores language-specific perception of speech sounds. This paper discusses two experiments: experiment 1 is a speeded forced-choice AX discrimination task and experiment 2 is a similarity rating task. Experiment 1 was intended to investigate the basic auditory perception of the listeners. It was predicted that listeners' native languages would not influence responses in experiment 1. Experiment 2 asked subjects to rate the similarity between two tokens on a five-point equal interval scale; the purpose of this experiment was to explore listeners' subjective impression of speech sounds. In experiment 2 it was predicted that listeners' language would affect their responses. The same stimuli were used in both experiments. The stimuli consisted of vowel-fricative-vowel sequences produced by a trained phonetician. Six fricatives were used: /f, th, s, sh, x, h/. These fricatives were embedded in three vowel environments: /a_a/, /i_i/, and /u_u/. Tokens were presented to listeners over headphones with a 100-ms interval. Independent groups of 15 native Dutch and English listeners participated in each of the two experiments. Results suggest that listeners' language influenced responses in both experiments, albeit the result was larger in experiment 2. [Work supported by NIH.]

4pSC12. Neural coding of perceptual interference at the preattentive level. Yang Zhang (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, Minneapolis, MN 55455), Patricia Kuhl, Toshiaki Imada (Univ. of Washington, Seattle, WA 98195), Toshiaki Imada, and Masaki Kawakatsu (Tokyo Denki Univ., Inzai-shi, Chiba 270-1382, Japan)

Language acquisition involves neural commitment to language-specific auditory patterns, which may interfere with second language learning. This magnetoencephalography study tested whether perceptual interference could occur at the preattentive level. Auditory mismatch field (MMF) responses were recorded from ten American and ten Japanese adult subjects in the passive oddball paradigm. The subjects read self-chosen books and ignored the sounds. Three pairs of synthetic /ra-la/ syllables were used: one cross-category pair varied only in the third formant (F3), and the other two within-category pairs varied only in the second formant (F2). ANOVA results showed a main effect of acoustic dimension with significant interaction with subject groups ($p < 0.01$). As reported earlier, American listeners showed larger but later MMF responses for the F3 change. By contrast, Japanese listeners showed larger and earlier MMFs than Americans for changes in F2. Moreover, Japanese listeners had larger and earlier MMF responses for the changes in F2 as against changes in F3, which was more prominent in the right hemisphere than in the left. These results provided further support for the hypothesis that language experience produces neural networks dedicated to the statistical properties of incoming speech experienced in infancy, which later interfere with second language acquisition.