4:55

3pSC11. Categorization of phonemic length contrasts in Japanese by native and non-native listeners. Hiroaki Kato (NICT/ATR, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan, kato@atr.jp), Keiichi Tajima (Hosei Univ.), Amanda Rothwell, Reiko Akahane-Yamada (ATR), and Kevin Munhall (Queen's Univ.)

Previous studies reported that exposure to Japanese and identification training allowed English listeners to improve their identification accuracy of Japanese words with phonemic length contrasts. However, this does not necessarily mean their perceptual properties approach those of native listeners in all respects. Past perceptual studies have shown that native listeners show, for example, an extremely sharp short-to-long boundary and shift of the boundary to adapt to changes in the temporal context, i.e., typically speaking rate variations. To carefully investigate the effects of training on such properties, the present study analyzed the learners' identification of stimulus continua between word pairs that minimally differed in the length of a phoneme. Overall results showed that English listeners' boundaries tended to be sharpened by the identification training, but only to a limited extent. On the other hand, the results also showed that English listeners' boundaries shifted across different speaking rates in ways that resemble native listeners' adaptation tendencies. The latter result suggests that even non-native listeners can adjust their identification boundaries according to differences in temporal context. Discussion will include both prospects and limitations of listeners' improvement by the perceptual training in this study. [Work supported by NICT and JSPS.]

THURSDAY AFTERNOON, 30 NOVEMBER 2006 MOLOKAI ROOM, 1:00 TO 5:00 P.M.

Session 3pSP

Signal Processing in Acoustics: Various Topics in Signal Processing (Poster Session)

Charles F. Gaumond, Cochair
Naval Research Lab., Code 7142, 4555 Overlook Ave., S.W., Washington, DC 20375-5320

Yoshinobu Kajikawa, Cochair
Kansai Univ., Dept. of Electronics, Faculty of Engineering, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan

Contributed Papers

All posters will be on display from 1:00 p.m. to 5:00 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 3:00 p.m. and contributors of even-numbered papers will be at their posters from 3:00 p.m. to 5:00 p.m.

3pSP1. Adaptive surround technology for stereo music signals. Yoshitaka Murayama, Akira Gotoh, Jun Ota, Hareo Hamada (DiMAGIC, Sugisho Bldg 3F, 2-4-5 Kanda-Ogawamachi, Chiyodaku, Tokyo 101-0052, Japan), Yuko Watanabe, and Hareo Hamada (Tokyo Denki Univ., Inzai-shi, Chiba, Japan)

Computation of the left and right channel instantaneous difference ($xL-xR$) is mostly used to obtain the signals for surround sound. This signal is usually referred as the antiphase components. However, there are many problems to using the antiphase components in terms of its applicability in practice as well as the sound quality. For instance, when signals xL and xR are equal, or almost equal, this variation, that is to say, the correlation coefficient value, is high; there is no way to determine the direction vector uniquely, therefore an ambiguity may appear. Instead of the antiphase component's computation, a new method is proposed by using an adaptive signal processing technique. In this paper, it is revealed that natural surround sound signals having no antiphase and uncomfortable feeling are adaptively generated by the proposed method. The other advantages from a signal processing point of view are also discussed. Finally, subjective evaluations on sound quality of a surround sound system are carried out.

3pSP2. A design of fast steering filters based on the adaptive fusion of predesigned finite impulse response filters for microphone array. Masato Nakayama (Grad. School of Sci. and Eng., Ritsumeikan Univ., 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan, gr020040@se.ritsumei.ac.jp), Takanobu Nishiura, and Yoichi Yamashita (Ritsumeikan Univ., Kusatsu, Shiga 525-8577, Japan)

With a microphone array, a desired distant-talking speech signal can be acquired selectively by steering the directivity in noisy environments. To form directivity, adaptive beamformers have been proposed as conventional beamformers. They can form null directivity with a small number of transducers. However, adaptive beamformers have a serious drawback: it can't realize the noise-robust adaptation in variable noisy environments, because it has necessary heavy computational costs to train multi-channel adaptive steering filters. To solve this problem, we propose to design the fast steering filters based on the adaptive fusion of predesigned FIR filters in variable noisy environments. The adaptive fusion of predesigned FIR filters is realized by adding (or subtracting) various predesigned FIR filters immediately. Therefore, the proposed method can quickly design the steering filters even in variable noisy environments. In addition, microphone calibration is indispensable for applying the proposed method. Therefore, we also propose the new microphone calibration method based on environmental noise without target signal. Evaluation experiments were con-

HRTFs are compared for each subject with the measured HRTFs in various objective criteria, such as spectral distortion (SD), signal-to-distortion ratio (SDR), and interaural time/level differences (ITD/ILD). Availability of the estimated HRTFs for the synthesis of the 3-D sound image are then evaluated via hearing experiments. In our previous study [Proceedings of WESPAC IX (2006)], the results of the experiment is insufficient since the low-frequency range of the estimated HRTFs caused the ambiguity in sound localization. In this study, the frequency range of the estimation is extended to as high frequency as possible, and the availability of the estimated HRTFs for auralization is discussed from two viewpoints: the subjective difference between the estimated and measured HRTFs, and some subjective attributes obtained from the hearing of the stimuli convoluted with the estimated HRTFs.

3pSP26. Analysis and modeling of interaction of auditory-visual information. Takahumi Asada, Yuko Watanabe, and Tatsuya Shibata (Grad. School of Information Environment, Tokyo Denki Univ., 2-1200 Muzai Gakuendai Inzai, Chiba, 270-1382 Japan, den03117@nifty.com)

We analyze and model the interaction of movement of the synchronized auditory-visual information. We construct the relationship of movement between virtually sound source and object. We suppose that cubic local contrast model (CLCM) we propose and the optical flow method are useful for analyzing moving image. CLCM is a method that describes the movement of objects from temporal and spatial difference of local contrast between pixels in moving image, and we found CLCM can measure the orientation of camera motion and the moving direction of an object in a moving image. We use the feature of head-related transfer function (HRTF) as the feature of sound movement. We have done two experiments: experiment 1 is that we analyze the data of the interaural amplitude difference (IAD) and interaural phase difference (IPD) between right and left HRTFs of a sphere by local contrast model (LCM); experiment 2 is that we analyze the relationship between the movement of an object in a moving image and a virtual sound source. We discuss the interactive features between auditory and visual information.

3pSP27. The effect of head movement on sound localization using stereo microphones: Comparison with dummy heads. Wataru Endo (Tokyo Inst. Tech., Nagatsuda, Midori-ku, Yokohama 226-8502, Japan, endo@u.ip.titech.ac.jp), Makio Kashino (Tokyo Inst. Tech., Atsugi, Kanagawa 243-0198, Japan), and Tatsuya Hirahara (Toyama Prefectural Univ., Imizu, Toyama 939-0398, Japan)

This study examined the effect of head movement on sound localization with a pair of microphones providing no head-related transfer function (HRTF) information, and with individualized, nonindividualized, and downsized dummy heads providing HRTF information with different degrees of distortion. In an anechoic room, white noise was presented for 5 s from one of 12 loudspeakers. A dummy head or a microphone pair was placed at the observation point. Listeners heard the sound via headphones outside the anechoic room and judged the direction of the source and whether the sound image was extracranial. Without head movement, localization accuracy with the microphone pair was significantly worse than that with dummy heads. When voluntary head movement was allowed and the dummy head was moved in accord with the listener's movement, accuracy with nonindividualized and downsized dummy heads improved nearly to the individualized level. Accuracy with the microphone pair also improved with head movement, but did not reach the level of dummy heads. An extracranial image was perceived with dummy heads in most cases, but not often with the microphone pair even with head movement. Head movement cannot completely compensate for the lack of HRTFs.

3pSP28. A biologically inspired pitch determination algorithm. Arturo Camacho and John G. Harris (Computational Neuro-Eng. Lab, Elec. and Comput. Eng., Univ. of Florida, Gainesville, FL 32611, arturo@cnel.ufl.edu)

A biologically inspired pitch determination algorithm is presented. This algorithm combines existing models of the cochlea and inner-hair-cell based spike generation [Lopez *et al.*, J. Acoust. Soc. Am. **110**, 3107 (2001); Sumner *et al.*, *ibid.* **113**, 893–901 (2003)] to model spike trains in the auditory nerve. The pitch and its salience are then estimated using a method proposed by Cariani [Cariani *et al.*, J. Neurophysiol. **76**, 1698–1716], which computes a summary autocorrelation function over the spike trains. Unlike Cariani's work, where spike trains are obtained experimentally, we simulate the spike trains from biologically inspired models. The main contribution of our approach is to combine models of the auditory system and a pitch estimation method based on neural spike trains. The proposed algorithm was tested using standard synthesized sounds, speech, and singing voices. Results show that this algorithm better matches human performance as compared to traditional pitch detection algorithms used in automatic speech processing.

3pSP29. End-point detection of speech using spectral transition for captioning system. Ayako Koga, Yuki Fujikashi, Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554), Noboru Kanedera (Ishikawa Natl. College of Technol.), and Junko Yoshii (Fujiyama, Inc.)

In recent years, captioning video contents with text translations is increasingly necessary because of the burgeoning use of media internationally, resulting from the rapid development of communication technology. In addition, within one language, video captioning is very important for hearing-impaired people. However, the process of captioning videos, including speech and nonspeech decisions, is often done manually by translators at present. Therefore, an efficient automatic end-point detection of speech for captioning video contents has been proposed. We attempted to detect speech end-points based on acoustic landmarks that identify times when acoustic changes are prominent in the speech signals [K. N. Stevens, Acoust. Phonetics (1998)]. In this study, landmarks were obtained by combining the mean square for the regression coefficients of logarithmic envelopes of $\frac{1}{3}$-oct bands in time, which resembles the parameter proposed by Furui to measure spectral transition [S. Furui, J. Acoust. Soc. Am. **80**(4), 1016–1025 (1986)], with other ones such as the logarithmic power of speech signals. An experiment was carried out using the proposed technique for speech detection. Results showed a high correct rate and introduced the possibility of its application to an efficient video captioning system. [Work partially supported by Open Research Center Project from MEXT.]

3pSP30. Inverse filter analysis of common harmonic structure on Specmurt using Riemann's ζ function. Nobutaka Ono, Shoichiro Saito, Hirokazu Kameoka, and Shigeki Sagayama (Dept. Information Phys. and Computing, Grad. School of Information Sci. and Technol., Univ. Tokyo, 7-3-1 Hongo Bunkyo-ku Tokyo, 113-8656 Japan)

In this paper, based on the interesting relationship between the log-spaced δ function sequence and Riemann's ζ function, the analytical properties of the inverse filter of the common harmonic structure on *specmurt analysis* are discussed. Specmurt [S. Sagayama *et. al.*, in Proc. SAPA (2004)] is a simple and efficient technique for the multi-pitch analysis of polyphonic music signals. If all tones have the same harmonic pattern, the power spectrum on the log-scaled frequency can be regarded as the convolution of the common harmonic structure and the distribution of fundamental frequencies. Based on the model, overtones are effectively suppressed by the inverse filtering of the common harmonic structure in specmurt. Thus, for the stable processing, analytic properties of the inverse filter are significant. Our new finding is that when the common harmonic structure is expressed as a log-spaced δ function sequence with a particular kind of decay, the Fourier transform is exactly equal to Riemann's ζ