Session 5pAA

Architectural Acoustics, Speech Communication and Psychological and Physiological Acoustics: Psychological Aspects of Speech in Rooms II

John S. Bradley, Cochair
National Research Council, Inst. for Research in Construction, Ottawa, Ontario K1A 0R6, Canada

Hiroshi Sato, Cochair
National Inst. for Advanced Industrial Science Technology, Inst. for Human Science and Biomedical Engineering, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566 Japan

Contributed Papers

1:00

5pAA1. Calibration of consonant perception in room reverberation. Kanako Ueno (Inst. of Industrial Sci., Univ. of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505, Japan), Norbert Kopco, and Barbara Shinn-Cunningham (Boston Univ., 677 Beacon St., Boston, MA 02215)

Many studies of sound perception often assumed that our auditory sensory processes are relatively static, rather than plastic. However, in everyday environments, we naturally and fluidly compensate for interfering effects of background noise and room reverberation. In order to investigate how listeners calibrate auditory perception to such acoustic interference, a listening experiment was performed to measure the effect of sudden changes of reverberation on the identification of consonants. Test sounds were generated by convolving two types of binaural room impulse responses (BRIRs) measured in large real rooms with speech tokens. As a control condition, pseudo-anechoic BRIRs with negligible reverberation energy were used. Listeners were asked to identify the consonant in a vowel-consonant target. The target was preceded by a carrier phrase consisting of vowel-consonant pairs from the same talker. In some cases, the target and carrier phrase were processed by the same BRIRs, while in others the BRIR's processing target and carrier differed. Consistent effect of calibration was observed in one of the simulated rooms, but not in the other, suggesting that the ability to compensate for the effects of reverberation depends on the specific pattern of reverberation produced in a given room. [Work supported by AFOSR and NSF.]

1:15

5pAA2. Sentence context influences vowel perception in reverberant conditions. Janine Wotton (Dept. of Psych., Gustavus Adolphus College, 800 W. College Ave., St. Peter, MN 56082, jwotton2@gac.edu), Kristin Welsh, Crystal Smith, Rachel Elvebak, Samantha Haseltine (Gustavus Adolphus College, St. Peter, MN 56082), and Barbara Shinn-Cunningham (Boston Univ., Boston, MA 02115)

Sentences recorded with a Mid-western accent were convolved with head-related impulse responses that included different room reverberation conditions. The stimuli were presented binaurally through headphones in an echo-attenuated chamber and subjects ($n=23$) typed the sentences they heard. The target word was one of a vowel pair (cattle/kettle, jam/gem, gas/guess, past/pest) embedded as the second word in one of three sentence types. The neutral sentence provided little context for the word. Target words in sentences that provided strong contextual cues could be congruent or incongruent with the expectations of the subject, for example, ''The cattle/kettle grazed in the meadow.'' Subjects made significantly more errors in the incongruent sentences compared to the neutral (Wilcoxon$=3.572$ $p<0.05$) or congruent sentences (Wilcoxon$=3.56$ $p<0.05$). When the target word was in a congruent sentence subjects performed equally well in reverberant or pseudo-anechoic conditions (Wilcoxon$=1.298$) but they made more errors in the reverberant condition for both neutral (Wilcoxon$=3.359$, $p<0.05$) and incongruent sentences (Wilcoxon$=2.241$, $p<0.05$). Results suggest that reverberation may cause listeners to rely more heavily on linguistic context to determine word meaning. [Work supported by NOHR, AFOSR.]

1:30

5pAA3. Perceptual compensation for reverberation: Effects of noise-context bandwidth. Simon J. Makin, Anthony J. Watkins, and Andrew P. Raimond (School of Psych., The Univ. of Reading, Earley Gate, Reading RG6 6AL, UK, s.j.makin@reading.ac.uk)

Perceptual compensation for reverberation is observed when the reverberation is applied to a test word (from a ''sir''-to-''stir'' continuum) and its context (e.g., ''next you'll get_to click on'') are varied independently. Increasing reverberation in test words decreases listeners' ''stir'' responses, as reverberation ''fills the gap'' that cues the [t]. Compensation occurs when the context's reverberation is commensurately increased, and ''stir'' responses increase back to the level found with minimal test-word reverberation. Compensation is strongest with speech contexts but also occurs with some noise-like contexts, including ''signal-correlated noise'' that has the wideband temporal envelope of the original speech. Also effective is a wideband noise that is given the temporal envelope seen at the output of a single auditory filter in response to speech. A narrow-band version of this ''auditory-filtered'' noise is not effective, but when contexts are made by summing of three or five of these bands, their effectiveness increases correspondingly. Compensation appears to be informed by the ''tails'' that reverberation adds at offsets, so it merely requires contexts with suitable temporal-envelope fluctuations. However, effects seem confined to the context's frequency region, so the crucial offsets need to be in a wide range of frequency bands. [Work supported by EPSRC.]

1:45

5pAA4. Aural localization of speech stimuli. Evelyn Way (Talaske, 105 N. Oak Park Ave, Oak Park, IL 60301, evelyn@talaske.com)

Localization error was tested for a variety of signals to answer the question: do humans aurally localize different speech stimuli with a different localization blur? A series of tests was conducted comparing the effect of the sentence length, gender of the talker, and frequency content of the signals on localization. Results were applied to ongoing research into constructing an aurally accurate telepresence system.

the presentation room for a simple spoken speech used with a relatively excessive absorption at a middle-sized multiuse conference room. The space occupies 1100 ft^2 with a ceiling height of 9 ft. It is designed for demonstration of sound equipment as well as lecture without audio equipment. Since the acoustics in the room are highly absorptive (average absorption coefficient, 0.3), considering a primary use for product demonstration, it causes a flutter echo between longitudinal walls due to less sound energy in a late field. We demonstrate here a system adjustment of the AFC system emphasizing voice amplification as well as an acoustical solution to a defect of a fluttering echo here. Acoustical descriptors such as RT, echo diagram, and SPL are discussed for a tuned room acoustics with the AFC system.

Contributed Poster Papers

Poster papers 5pAA11 to 5pAA14 will be on display from 1:00 p.m. to 4:45 p.m. Authors will be at their posters from 3:45 p.m. to 4:45 p.m.

5pAA11. Slowed speech spreading into reverberant environments; steady-state suppression improves speech intelligibility. Yuki Nakata, Yoshiaki Murakami, Nao Hodoshima, and Takayuki Arai (Dept. of Elec. and Electrons Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan)

Speech intelligibility is often decreased in reverberant environments. For improving speech intelligibility under reverberant environments, Arai *et al.* [Acoust. Sci. Technol. **23**, 229–232 (2002)] suggested steady-state suppression, which suppresses steady-state portions of speech to reduce overlap-masking (causing degradation intelligibility) and improves speech intelligibility as preprocessing. Previous studies showed that speech intelligibility was improved through this processing under some reverberant environments [Hodoshima *et al.*, J. Acoust. Soc. Am. **119**(6), 4055–4064 (2006)]. On the other hand, speech intelligibility is known to be increased greatly by speaking slowly [Bolt and MacDonald, J. Acoust. Soc. Am. **21**(6), 577–580 (1949)]. However, it is not optimal for improving speech intelligibility in terms of reducing overlap masking because vowels causing major parts of overlap masking are expanded by speaking slowly. Therefore, in the current study, we investigated the effects of steady-state suppression on speech perception with a decreased speech rate from 6 to 4 and 5 morae/s. Results showed that the slowest speech (4 morae/s) with steady-state suppression was the most intelligible. Also, steady-state suppression improved speech intelligibility at a speech rate of 4 morae/s in each reverberant condition (reverberation times of 1.5, 2.0, and 2.5 s). [Work partially supported by JSPS.KAKENHI (16203041).]

5pAA12. Suppression of speech intelligibility loss through a modulation transfer function-based speech dereverberation method. Masashi Unoki, Masato Toi, Yohei Shibano, and Masato Akagi (School of Information and Sci., JAIST, 1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan, unoki@jaist.ac.jp)

The concept of modulation transfer function (MTF) can successfully be applied to evaluate the quality of speech transmission in the room acoustics [Houtgast and Steeneken, J. Acoust. Soc. Am. **77**, 1069–1077 (1985)]. We previously proposed a speech dereverberation method based on the MTF concept, which consisted of MTF-based power envelope inverse filtering and the carrier regeneration in the filterbank [Unoki *et al.*, EuroSpeech2003 (2003)]. This paper evaluates how the proposed method can suppress the loss of speech intelligibility caused by reverberation, by comparing various methods. We have carried out massive simulations of dereverberation for reverberant speech signals to objectively evaluate these methods. We also subjectively evaluated the methods via listening tests. In these simulations, artificial reverberations were convolved with a clean speech signal in which the impulse responses in the room acoustics can be approximated from the exponential decay as a function of reverberation time with a white-noise carrier. The results of both evaluations show that, in addition to reducing the averaged log-spectrum distortion by about 1 dB, the proposed method reduces the loss of speech intelligibility by about 30%. [Work supported by a Grant-in-Aid for Science Research from the Japanese Ministry of Education (No. 18680017).]

5pAA13. Word and mora intelligibility in "Familiarity-controlled word-lists 2003 (FW03)." Tadahisa Kondo, Shigeaki Amano (NTT Commun. Sci. Labs., NTT Corp., 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa, 2430198, Japan, tkondo@brl.ntt.co.jp), Shuichi Sakamoto, and Yôiti Suzuki (Tohoku Univ., Aoba-ku, Sendai, Miyagi, 9808577, Japan)

"Familiarity-controlled Word-lists (FW03)" was developed to make it possible to perform intelligibility tests on the same person repeatedly and/or under several different conditions. FW03 consists of 20 lists of 50 words in four word familiarity ranks. These words were selected taking phonetic balance into consideration to maximize their variety of initial moras and vowel-consonant sequences. To confirm that all FW03 word lists from a particular familiarity rank present the same hearing difficulty, we measured the word intelligibility of all the words in FW03. Sixteen subjects listened to 4000 words spoken by four speakers with seven signal-to-noise ratios. The mean intelligibility scores for the lists were significantly different even for the same familiarity rank. The intelligibility was apparently influenced by some specific moras and their position in words. These results suggest that it is difficult to equalize the word intelligibility of the lists even when the word familiarity is controlled more precisely. We therefore examined several methods for equalizing the intelligibility among the lists such as controlling the sound pressure level for each word when operating an actual intelligibility test. The effectiveness of these methods is discussed with detailed analyses of word and mora intelligibility in FW03.

5pAA14. Influence of Deutlichkeit value and reverberation time on improved speech intelligibility in reverberant environments because of steady-state suppression. Nahoko Hayashi, Nao Hodoshima, Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo 102-8554 Japan, nhayash@sophia.ac.jp), and Kiyohiro Kurisu (TOA Corp., Takarazuka, Hyogo 665-0043 Japan.)

To improve speech intelligibility in reverberant environments, Arai *et al.* proposed "steady-state suppression (SSS)" as preprocessing [Arai *et al.*, Acoust. Sci. Technol. **23**, 229–232 (2002)]. In this study, a perceptual experiment under artificial reverberant conditions with simulated impulse responses was conducted to elucidate the effect of the Deutlichkeit (D) value and reverberation time (RT) on improvements of speech intelligibility because of SSS. Artificial impulse responses were simulated with white noise multiplied by a decay curve. The advantage of this method is that the simulated impulse responses have mutually similar frequency characteristics; consequently, we can evaluate them using only the D value and RT regardless of their different frequency characteristics. Two parameters, the energy of the impulse response 50 ms from the direct sound and the attenuation rate of the decay curve, were controlled to obtain several impulse responses having certain D value and RT. Results show that SSS improved speech intelligibility in the conditions of low D value, even if RT was long or short. We could also interpret these results as indicating that processing is effective when the original speech intelligibility is less than 60%. [Work supported by JSPS.KAKENHI (16203041).]