Session 5aAA

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

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 and Technology, Inst. for Human Science and Biomedical Engineering, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566 Japan

Chair's Introduction—7:30

Invited Papers

7:35

5aAA1. Why we should design for optimum reverberation times in rooms for speech communication. J. S. Bradley (Inst. for Res. in Construction, Natl. Res. Council, Montreal Rd., Ottawa, Canada K1A 0R6)

Various texts list recommendations for optimum reverberation times as ideal goals in rooms for speech communication. However, some newer documents, such as ANSI S12.60, talk about maximum recommended reverberation times in rooms for speech communication. The source of the changed approach can be traced to interpretations of experimental results that do not consider the full implications for speech communication in real rooms, and also to the fact that reverberation time is not an ideal predictor of the quality of speech communication. In the extreme, minimizing reverberation times would lead to near anechoic rooms for speech and inadequate signal-to-noise ratios. The need for optimum reverberation times can be explained as a simple need to first achieve adequate signal-to-noise ratios. However, a more complete understanding is obtained by examining the benefits of early-arriving reflections of speech sounds on the intelligibility of the speech to listeners. Attempts to determine optimum reverberation times for normal hearing listeners, which are based on a balance between avoiding excessive reverberation and maintaining adequate signal-to-noise ratios, lead to a range of acceptable values that can vary with the ambient noise level. This discussion will be supported with examples from room acoustics measurements in classrooms and meeting rooms.

7:55

5aAA2. The relation between speech transmission index, clarity, and reverberation time and listening difficulty in the impulse response database of AIJ. Hiroshi Sato (Natl. Inst. of Adv. Industrial Sci. and Technol. (AIST), 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan), Yoshio Nishikawa (Konoike Co., Ltd., Tsukuba 305-0003, Japan), Hayato Sato, and Masayuki Morimoto (Kobe Univ., Nada, Kobe 657-8501, Japan)

The Speech Communication Research Working Group of AIJ (Architectural Institute of Japan) is collecting information on rooms including digitized impulse responses (IRs) to establish a database for evaluating and designing the speech transmission quality of rooms. This database consists of 966 measured IRs. This study presents the relationships between speech transmission index (STI), clarity (Cx) and reverberation time (T) as measures to consider for the design and evaluation of the speech transmission performance of rooms. The data show a wide range of STI and Cx values for a given T, and the minimum STI at each reverberation time can be obtained by diffused field theory. STI and Cx are seen to be better indicators than T for the design of rooms for speech and that T is not as good. Relationships between STI, Cx, and listening difficulty ratings from previous studies [Proc. of RADS (2004), Proc. of Forum Acusticum, pp. 1713–1718 (2005)] found both STI and Cx can be used as predictors of listening difficulty ratings. Finally, listening difficulty ratings of individual IR in the database are estimated and the distribution of listening difficulty ratings in a variety of rooms are presented as a solid bases to design speech transmission quality of rooms.

8:15

5aAA3. Unifying approaches for modeling and predicting speech intelligibility. Adelbert W. Bronkhorst and Sander J. van Wijngaarden (TNO Human Factors, Kampweg 5, 3769 DE Soesterberg, The Netherlands, adelbert.bronkhorst@tno.nl)

Speech intelligibility is practically always affected by the acoustic environment (reverberation, interfering speech, noise). When speech is processed electronically, distortions introduced by the transmission chain can also deteriorate intelligibility. Furthermore, intelligibility can be strongly influenced by nonacoustic factors such as contextual information, non-nativeness, and hearing impairment. Various intelligibility prediction methods have been developed (including SII, STI, SRS, and PESQ), which cover almost all factors influencing intelligibility. However, each individual method has limited applicability and no attempts have been made to unify them. Recently, the application domains of two widely used methods, the SII and the STI, have been significantly extended. They now, for example, cover interference that fluctuates in level, live speech, and digital transmission channels. These methods are actually very

9:55–10:15 Break

10:15

5aAA8. The effects of fluctuating interaural cues on the segregation of speech in rooms. Douglas Brungart and Nandini Iyer (Air Force Res. Lab., 2610 Seventh St., Bldg 441, WPAFB, OH 45433, douglas.brungart@wpafb.af.mil)

Spatial separation is known to improve the segregation of talkers in anechoic environments, but relatively little is known about the role spatial cues play in speech segregation in reverberant rooms. One might expect the random disruptions in the interaural time and level differences (ILDs and ITDs) that occur in reverberant environments to eliminate many of the intelligibility benefits that normally occur for spatially separated speech. However, the precedence effect often leads to a robust perception of spatial location even in extremely echoic environments. This can result in an apparent separation between talkers that may facilitate selective attention to the location of the target speech even in a highly reverberant room. Also, random fluctuations in ITD and ILD may lead to differences in the apparent source widths of the target and masking sounds, which might serve as a segregation cue. In this talk, we examine the effects fluctuating interaural difference cues have on the segregation of target speech from competing speech or noise. The results suggest that differences in apparent source width can be used to segregate competing speech signals even when the target and masking signals appear to originate from the same direction relative to the listener.

10:35

5aAA9. The intelligibility advantage of clear speech in noise and reverberation backgrounds: Effects of speaking rate. Jean C. Krause (Commun. Sci. and Disord. Dept., Univ. of South Florida, 4202 E. Fowler Ave., PCD 1017, Tampa, FL 33617, jkrause@cas.usf.edu)

Clear speech refers to a speaking style that is significantly more intelligible than conversational speech for a variety of listeners and backgrounds, including roomlike backgrounds such as noise and reverberation [Payton *et al.*, J. Acoust. Soc. Am. **95**, 1581–1592 (1994)]. Although typically spoken more slowly than conversational speech [e.g., Uchanski *et al.*, J. Speech Hear. Res. **39**, 494–509 (1996)], talkers can produce clear speech at normal rates with training [Krause and Braida, J. Acoust. Soc. Am. **112**, 2165–2172 (2002)]. Whether this form of clear speech (clear/normal speech) provides similar intelligibility benefits to clear (clear/slow) speech in all situations has not been fully characterized, though some results [Krause and Braida, Iranian Audiol. **2**, 39–47 (2003)] suggest that the amount of benefit is more dependent on talker and environment. To investigate this possibility for roomlike degradations, the intelligibility advantage provided by clear/slow and clear/normal speech was evaluated by groups of eight normal-hearing listeners at five signal-to-noise ratios (-4, -2, 0, $+2$, and $+4$ dB) and three levels of reverberation (0, 0.18, and 0.6 s). Results for four talkers will be reviewed in order to illustrate the effects of speaking rate and speaking style on speech intelligibility in rooms.

10:55

5aAA10. Separating and understanding a talker from a mixture in reverberant spaces. Barbara G. Shinn-Cunningham, Madhusudana Shashanka, and Scott Bressler (677 Beacon St., Boston, MA 02215)

Many studies show that reverberant energy can interfere with the ability to understand a listener by smearing out the temporal modulations and distorting the spectral content that convey speech meaning. However, the amount of distortion caused by the reverberation in everyday settings often is not severe enough to degrade speech understanding when there is only one talker in a room. Despite this, even modest reverberation can have an enormous impact on the ability to understand the same talker when there is another sound source in the environment. One factor that likely contributes to the problem of understanding a talker in a mixture of sounds in reverberant spaces is that it is difficult to separate the target talker from the competing sounds. This talk will explore the ways in which reverberant energy degrades the ability to separate sounds in a mixture. The effects of reverberation on different cues for source segregation, including pitch and location, will be discussed. [Work supported by AFOSR and NIH.]

11:15

5aAA11. Preprocessing speech against reverberation. Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan)

Although short reverberation in a room may help speech perception, it is known that long reverberation causes degradation in speech comprehension. This is especially true for elderly people, the hearing-impaired, and non-native listeners. In order to prevent intelligibility degradation, we can apply a signal processing technique to the public address system of a room before speech signals are radiated through the loudspeakers. So far, we have developed several such preprocessing techniques. The two main techniques are modulation filtering [Kusumoto *et al.*, Speech Commun. **45**, 101–113 (2005)] and steady-state suppression [Arai *et al.*, Acoust. Sci. Technol. **23**, 229–232 (2002)]. Both techniques essentially enhance the temporal dynamics of speech between 1 and 16 Hz [Arai *et al.*, J. Acoust. Soc. Am. **105**, 2783–2791 (1999)]. Steady-state suppression, which suppresses steady-state portions of speech, reduces overlap-masking and improves speech intelligibility for young, elderly, and non-native listeners in reverberant environments [e.g., Hodoshima *et al.*, J. Acoust. Soc. Am. **119**, 4055–4064 (2006)]. This technique is more effective when speech-rate slowing is applied in advance [Arai *et al.*, Acoust. Sci. Tech. **26**, 459–461 (2005)]. Intelligibility with this technique exceeds a simple speech-rate slowing approach. [Work partially supported by JSPS.KAKENHI (16203041).]