CALIBRATION OF SPEECH IN THE REAL WORLD A REQUEST FROM INDUSTRY

Robert S. Schlauch
Department of Speech-Language-Hearing Sciences, University of Minnesota
Minneapolis, Minnesota 55455

A widely used American National Standards Institute (ANSI) standard is the one titled "Method for measuring the intelligibility of speech over communication systems," ANSI S3.2 (1989). Long after this standard was published in its current form, I became the chair of its working group (WG), primarily through attrition. Shortly after becoming the WG chair, it was brought to my attention that ANSI S3.2 refers to an ANSI standard for guidelines regarding the calibration of speech. This document does not exist. Apparently, the ANSI committee working on the speech calibration standard was unable to reach a consensus on this part. The goal of this short note is to find persons in industry who are willing to work on a description of a method for speech calibration for inclusion in the standard for measuring the intelligibility of speech over communication systems. Laboratory scientists are well represented on the working group, but persons familiar with real-world applications must be included in order to develop a standard that is acceptable for industry. Below, I describe some of the issues.

"Speech is a time varying stimulus, and that raises a problem for its calibration."

Speech is a time varying stimulus, and that raises a problem for its calibration. Physical measures of the dynamic range of intensity for phonemes, the smallest segments of speech, are 50 dB or more (Boothroyd, 1994). Perceptual studies confirm that the effective dynamic range of speech is as high as 43 dB (Studebaker *et al.*, 1999).

Before the widespread use of computers, speech calibration of recorded materials was often done using a volume unit meter or the slow response setting on a sound level meter. The level of the speech was taken from the level of the peaks read on the meter, a method that is dependent on the device's damping characteristics. This method is often based on the level measured in a coupler attached to a sound level meter, and it is still used today (e.g., Warren *et al.*, 2005; Healy and Bacon, 2007; Hornsby and Ricketts, 2006). An informal survey of current issues of the *Journal of the Acoustical Society of America* revealed another popular method for speech calibration. This alternative method equates stimuli (e.g., sentences or words) for their root-mean-square (rms) level over their entire duration (Bradlow and Alexander, 2007; Erwin, *et al.*, 2007; Kramer *et al.*, 2007; Humes *et al.*, 2006; Levi *et al.*, 2007; Sommers and Barcroft, 2006; Van Engen and Bradlow, 2007). This latter approach is a trivial calculation for digitally stored stimuli. Use of a tone or other sound source with a root-mean-square (rms) amplitude equivalent to the speech stimuli is a practical method for daily calibration of such stimuli.

A frequent goal of studies involving speech stimuli is speech understanding, and these measurements are often conducted in the presence of competing sounds. The current standard (ANSI S3.2) does not specify the characteristics of a competing stimulus yet it recognizes that one may be present. In the laboratory, a typical masker for these situations is a noise filtered to have the same spectrum as the long-term spectrum of the speech stimuli used in the study. This type of shaped noise, as well as noise with other characteristics (e.g., pink noise or white noise), is easy to generate using a computer and has the benefit of being replicable. Would a standardized competing stimulus be appropriate for evaluating and comparing communication systems in industry?

The working group for the standard needs input regarding the applications in industry. We need to know whether the methods described above are appropriate for calibration for a wide variety of industrial applications. Our suspicion is that they will not be broadly accepted. At a meeting of the Acoustical Society of America, a colleague described to me a hypothetical example for the evaluation of a communication system in a tank, a tracked, armored combat vehicle. In this situation, the "talker" would produce words using live voice with background noise varying, depending on the terrain and the status of the vehicle's engine. The listener, whose responses are used to evaluate the effectiveness of the communication system, may be in a similar vehicle. So, now, the stimulus presented to the listener would be potentially affected by the noise from both vehicles. What would be an appropriate method for calibrating the speech and documenting the noise in this situation? If the speech and noise in this environment were recorded separately, using a probe-tube microphone at the listener's ear, the digital evaluations described above could be used for calibration. This method might be flexible but places additional demands on users of the standard. Also, if the noise differs across applications, one might question whether the results represent a standardized exam.

Send me an email (schla001@umn.edu) or leave me a phone message (612-624-7001), if you want to contribute to this process. I am seeking committee members, but anyone willing to share applications of the standard or other advice with our working group should also feel free to contact me. **AT**

References for further reading:

Bradlow, A. R., and Alexander, J. A. (**2007**). "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners," J. Acoust. Soc. Am. **121**, 2339–2349.

Boothroyd, A., Erickson, F., and Medwetsky, L. (**1994**). "The hearing aid input: A honemic approach to assessing the spectral distribution of speech," Ear and Hearing **15**, 432-442.

Erwin, L. J., George, E. L. J., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., and Houtgast, T. (**2007**). "Auditory and nonauditory factors affecting speech reception in noise by older listeners," J. Acoust. Soc. Am. **121**, 2362–2375.

Healy, E. W., and Bacon, S. P. (**2007**). "Effect of spectral frequency range and separation on the perception of asynchronous speech," J. Acoust. Soc. Am. **121**, 1691-1700.

Hornsby, B. W. Y., and Ricketts, T. A. (**2006**). "The effects of hearing loss on the contribution of high- and low-frequency speech information to speech understanding. II. Sloping hearing loss," J. Acoust. Soc. Am. **119**, 1752–1763.

Humes, L. E., Lee, J. H., and Coughlin, M. P. (**2006**). "Auditory measures of selective and divided attention in young and older adults using single-talker competition," J. Acoust. Soc. Am. **120**, 2926–2937

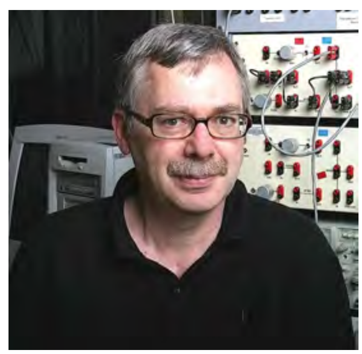
Levi, S. V., Winters, S. J., and Pisoni, D. B. (**2007**). "Speaker-independent factors affecting the perception of foreign accent in a second language," J. Acoust. Soc. Am. **121**, 2327–2338.

Sommers, M. S., and Barcroft, J. (**2006**) "Stimulus variability and the phonetic relevance hypothesis: Effects of variability in speaking style, fundamental frequency, and speaking rate on spoken word identification," J. Acoust. Soc. Am. **119**, 2406-2416.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431-2444.

Van Engen, K. J., and Bradlow, A. R. (**2007**). "Sentence recognition in native- and foreign-language multi-talker background noise," J. Acoust. Soc. Am. **121**, 519–526.

Warren, R. M., Bashford, J. A. and Lenz, P. W. (**2005**). "Intelligibilities of 1-octave rectangular bands spanning the speech spectrum when heard separately and paired," J. Acoust. Soc. Am. **118**, 3261–3266.



Robert Schlauch earned a bachelors degree from the University of Hawaii in 1980 and a PhD in Speech and Hearing Science in 1987 from the University of Washington. After a post doctoral fellowship at the University of California (Berkeley), he was hired by the University of Minnesota in 1989 where he works today as an associate professor. His research interests include intensity perception (psychoacoustics), cognitive factors influencing hearing, and diagnostic audiology. He is a fellow of the Acoustical Society of America.