The use of context in spectrogram reading

V. W. Zue and R. A. Cole

Citation: The Journal of the Acoustical Society of America **65**, S81 (1979); doi: 10.1121/1.2017458
View online: https://doi.org/10.1121/1.2017458
View Table of Contents: https://asa.scitation.org/toc/jas/65/S1
Published by the Acoustical Society of America

10:00

FF13. Stop consonant perception: A comparison of several perceptual theories. Dennis Klatt (Massachusetts Institute of Technology, Cambridge, MA 02139)

Acoustic cues to the identity of a plosive differ depending on aspects of phonetic environment such as features of a following vowel or consonant, properties of the previous segment, and stress. A complete perceptual theory should not only describe the perceptual strategy of an adult when confronted with this variability (and variability seen across talkers, transmission mediums, and speaking rates), but also account for the acquisition of these skills. In this paper we review the relative merits of burst invariance theories, relational invariance theories, the motor theory, analysis by synthesis, the locus theory, a theory based on the equivalent second formant $F_{2'}$, and two new models called SCRIBER and LAFS [D. H. Klatt, J. Phonetics 7, (2) (1979)]. The advantage of SCRIBER and LAFS is that they can account for the differential use of various cues in different phonetic contexts by a conceptually simple decoding network structure. In its simplest form, the network represents all possible phonetic transitions of a language by distinct sequences of spectra. Perceptual abstractions such as [p]-ness or labialness are learned through association with the phonemic representations for words that are developed for speech production purposes (with the aid of a certain amount of natural acoustic similarity that exists for different realizations of a stop or of a distinctive feature). [Supported by NINCDS.]

10:10

FF14. Performance of an expert spectrogram reader. R. A. Cole, A. I. Rudnicky (Department of Psychology, Carnegie-Mellon University, Pittsburgh, PA 15213), and V. M. Zue (Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139)

The subject in our study has spent between 2000–2500 h learning to read speech spectrograms. We tested the subject's ability to identify the phonetic content of broadband speech spectrograms of unknown utterances during eight separate sessions of about 4 h each. The expert was presented with 23 spectrograms of English sentences and sequences of words and nonsense words, and 45 English words embedded in a known carrier phrase. The phonetic labels produced by the expert agreed with the phonetic labels produced by trained phoneticians (who listened to the speech) between 80% and 90% of the time, depending upon the scoring method used. When presented with words in a known carrier phrase, labeling performance improved to about 93%. A linguist who was presented with the phonetic transcriptions produced by the spectrogram reader was unable to identify all of the words in ten of 15 utterances, and missed only a single work in each of the remaining five. A film will be shown of the expert segmenting and labeling two spectrograms of unknown utterances.

10:20

FF15. The use of context in spectrogram reading. V. W. Zue (Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139) and R. A. Cole (Department of Psychology, Carnegie-Mellon University, Pittsburgh, PA 15213)

In order to assess the role of syntactic, semantic, and discourse knowledge in spectrogram reading, three short stories were recorded and speech spectrograms were made of the individual sentences of each story. The stories were presented one spectrogram at a time to an expert spectrogram reader who is instructed to read each story word-by-word without writing down segment labels. The three stories contained a total of 670 words, and 612 (91%) were correctly identified. The median reading time per sentence across the three stories was about 40 s, or about 20 times real time. However, syllable-by-syllable analysis of reading times in one story reveal that this value was inflated by a small subset of syllables and words that took a long time to decode—sometimes over a minute. The modal (most frequent) reading time per syllable fell between 1–2 s, or between 3 and 6 times real-time. Further analysis reveal that many common syllables were immediately recognized as complete patterns (e.g., "ment", "tion"), and the use of context to recognize words from partial information was evident in many cases. Implications of the results for real-time spectrogram reading will be discussed.

THURSDAY MORNING, 14 JUNE 1979
SALA DE PUERTO RICO, STUDENT CENTER, 8:00 TO 10:30 A.M.

Session GG. Physiological Acoustics III: Auditory Nerve and Central Nervous System (Poster Session)

L. S. Frishkopf, Chairman

Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Contributed Papers

All posters will be displayed from 8–10:30 A.M. To allow all contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8–9:15 A.M., and contributors of even-numbered papers will be at their posters from 9:15–10:30 A.M.

GG1. Onset of frequency specific hearing in the rabbit. M. L. Lenhardt (Biocommunications, Medical College of Virginia Commonwealth University, Richmond, VA 23298)

The onset of the first auditive startle response to frequency modulated linear ramps was studied. Upward and downward ramps (0.5–0.75, 1–2, and 2–4 kHz) served as stimuli for 50 rabbits from eight litters with five replications. Animals were tested from age 6–10 days. At 6 days there were no behavioral responses to sound. At day 7 only responses to the 0.5–0.75 kHz ramp was observed. On day