The representation of speech signals in a model of the peripheral auditory system

Richard Goldhor

Citation: The Journal of the Acoustical Society of America **73**, S4 (1983); doi: 10.1121/1.2020422
View online: https://doi.org/10.1121/1.2020422
View Table of Contents: https://asa.scitation.org/toc/jas/73/S1
Published by the Acoustical Society of America

auditory presentation of the passage. Subjects listened to three other passages without a printed version and then answered comprehension questions. Accuracy of word identification and response latency and accuracy to comprehension questions were recorded. Based on these data we will discuss the time-course of learning to perceive synthetic speech more accurately.

9:05

B5. Training away the phonemic restoration illusion. William H. Ressler and Arthur G. Samuel (Department of Psychology, Box 11A, Yale Station, New Haven, CT 06250)

Phonemic restoration is an auditory illusion which arises when a phoneme is removed from a word and replaced with noise. Listeners hear the word as intact: they perceptually restore the missing phoneneme. Recently it has been shown that the magnitude of the illusion diminishes with practice [H. C. Nusbaum, A. C. Walley, T. D. Carrell, and W. H. Resslar, *Res. on Speech and Hearing Prog. Rep. No.* 8, Indiana University (1983)]. One explanation for this result is that subjects learn to ignore their expectation of hearing an intact word by directing attention to the speech waveform itself, specifically attending to the location of the target phoneme. To test this explanation, two groups of subjects received training in the illusion. Group 1 (*Control*) saw each stimulus word printed on a CRT screen prior to hearing that word. Group 2 (*Directed Attention*) also saw each word on the screen, but with the target phoneme underlined in each. The performance of the second group relative to the first has implications both for an explanation of the illusion as a misdirector of attention, as well as for theories of the role of attention and the degree to which it may be controlled in the perception of fluent speech.

9:20

B6. Phonological short-term memory preserves phonetic detail. Ignatius G. Mattingly (University of Connecticut, Storrs, CT 06268 and Haskins Laboratories, New Haven, CT 06510), Michael Studdert-Kennedy (Queens College, City University of New York, Flushing, NY 11367 and Haskins Laboratories, New Haven, CT 06510), and Harriet Magen (Yale University, New Haven, CT 06520 and Haskins Laboratories, New Haven, CT 06510)

If phonological short-term memory preserves phonetic detail (as would seem necessary if listeners are to correct misperceptions on the basis of later information), a sequence of words in different dialects might be difficult to recall, because phonological redundancy could not be exploited. Five women (one Scottish English, one ESL with Italian interference, three GA speakers) recorded the digits *one, two···ten*. From these we compiled 20 pseudo-random, 12-word single-talker sequences, 20 three-talker same-dialect (GA) sequences, and 20 three-talker, mixed-dialect (GA, SE, ESL) sequences. Serial order recall by 18 female GA speakers was tested. Recall of single-talker sequences was not reliably better if talker and listeners shared dialect (GA), nor did recall of of later words in a sequence vary systematically with sequence-type. But earlier words were better recalled in three-talker, same-dialect than in three-talker mixed-dialect sequences, both for all words and for GA words common to both sequence-types. Implications for the role of phonetic storage in normal listening will be discussed. [Supported by NICHD and NSF.]

9:35

B7. A phonetic band? Robert E. Remez (Department of Psychology, Barnard College, New York, NY 10027) and Philip E. Rubin (Haskins Laboratories, 270 Crown Street, New Haven, CT 06510)

Explanations of phonetic perception describe the conversion of the acoustic signal to phonetic segments. By making the descriptions of phonetically relevant patterns abstract, these explanations have accomodated the great variability evident in the acoustic signals produced by different talkers and in different utterances by the same talker. In order to refine this characterization of perception, we sought to determine the precise limits of phonetically useful information over frequency variation. A frequency transposition test was performed with three sentences, employing the technique of sinusoidal replication of natural utterances. This technique emphasized time-varying acoustic information, and also controlled for timbre changes over frequency. The results show that a phonetic listening band may exist within the human range of audible frequencies. The relation of this finding to frequency coding, perception of musical pitch, and development of sensory prostheses will be discussed. [Supported by the National Institute of Child Health and Human Development.]

9:50

B8. The representation of speech signals in a model of the peripheral auditory system. Richard Goldhor (Speech Communication Group, 36-511, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139)

The effect of processing speech signals with a simple model of the peripheral auditory system will be described. Compared with a standard speech spectrogram or filter bank analysis, this model transforms the signals in each of the frequency, amplitude, and temporal dimensions. In particular, the temporal transformation models the adaptation of auditory neurons to sustained energy in a particular frequency band, and results in distinctive responses to segments of speech exhibiting rapid changes in frequency or amplitude (consonants). These transformations result in new relationships between the phonetic identity of the speech and the observable characteristics of the transformed signal. Examples of the response of this model to speech signals will be presented, and the response properties that correspond to the phonetic identity of the signals will be discussed. [Work supported by NSF.]

10:05

B9. Species differences in neural lateralization of species-specific calls. Michael R. Petersen (Department of Psychology, Indiana University, Bloomington, IN 47405)

Petersen *et al.* [Science **202**, 324–327 (1978)] reported that Japanese macaques exhibited a right ear advantage (REA) when listening to a semantically distinctive feature of their vocalizations. Several heterospecific monkeys tested in identical conditions with the Japanese monkeys' calls failed to show any reliable ear advantage. However, recent studies with humans have shown that the size and/or direction of ear advantages vary with the physical feature to which subjects attend. This raised the question of whether the species differences in lateralization obtained in our setting represented (1) a processing difference for the same signal feature or (2) a difference between species in the signal feature to which they attended. To choose between those alternatives we ascertained whether Japanese and heterospecific monkeys were in fact attending to the same acoustic feature when they showed their respective patterns of lateralization. Two Japanese and two heterospecific monkeys received extensive training on a discrimination task that required them to classify 15 vocalizations into two categories. During this time the Japanese monkeys exhibited REA's but the heterospecifics showed no ear advantage. The animals were then tested for generalization to 25 novel, natural vocalizations and six synthetic calls. These tests revealed that all four animals were attending to the same feature of the Japanese monkey calls. The animals were then returned to the original training stimuli and they exhibited their earlier patterns of lateralization. Thus this study shows that even when attending to the same physical feature, Japanese and heterospecific monkeys employ different neural processing strategies. This suggests that the communicative valence of the signals used in testing may account for the lateralization differences observed between species. [Work supported by NSF, Deafness Research Foundation and Sloan Foundation.]

10:20

B10. A pitch-synchronous, variable frame-length, variable bit-rate automatic LPC code generation system. N. Rao Vemula, Eugene H. Lee, and Philip T. McLaughlin (General Instrument Microelectronics, 600 West John Street, Hicksville, NY 11802)

There are a number of LPC-based speech synthesis integrated circuits available on the market today. For practical applications of these chips, it is advantageous to utilize an automatic code generation system to help us