Mimic Resistance of Speaker Verification Using Phoneme Spectra

T. W. Rekieta and G. D. Hair

Citation: The Journal of the Acoustical Society of America **51**, 131 (1972); doi: 10.1121/1.1981414
View online: https://doi.org/10.1121/1.1981414
View Table of Contents: https://asa.scitation.org/toc/jas/51/1A
Published by the Acoustical Society of America

2:15

Z2. *Abstract withdrawn.*

2:30

Z3. Vowel and Speaker Identification in Natural and Synthetic Speech. DAVID MELTZER, *IBM Corporation, Poughkeepsie, New York and Ohio State University,* AND ILSE LEHISTE, *Department of Linguistics, Ohio State University, Columbus, Ohio 43210.*—The purpose of this study was to develop a simple means for evaluating the relative quality of synthesizers. A set of 10 monophthongal English vowels was produced by a man, woman, and child. These vowels were synthesized on a Glace–Holmes synthesizer, using values measured from spectrograms. In addition, synthetic stimuli were generated on the basis of averages published by Peterson and Barney [J. Acoust. Soc. Amer. **24**, 175–184 (1951)]. In the latter set, formant values for men, women, and children were combined with the respective fundamental frequencies, resulting in 9 different combinations for each of the 10 vowels. The 150 stimuli were presented, in random order, to 60 trained listeners for both vowel and speaker identification. The over-all vowel identification score for the normal set (all three speakers combined) was 79.46%; the over-all speaker identification score (all 10 vowels combined) was 90.03%. The corresponding scores for the set synthesized from measured spectrograms were 50.87 and 69.73%, respectively. The differences from the normal set (−28.59 and −20.30%) constitute an evaluation measure for the performance of the synthesizer (and the synthesizers). Results of the listeners' responses to the set of vowels synthesized from averages will also be discussed.

2:45

Z4. Aural Identification of Children's Voices. S. CORT AND T. MURRY, *Department of Speech, Central Connecticut State College, New Britain, Connecticut 06050.*—Recent studies have emphasized the importance of aural identification of speakers; this study sought to assess children's ability for aural-identification purposes. Specifically, the purpose of this study was to determine if children can make aural identification of their peers and, secondly, to determine if three repetitions of the sample increased identification scores. Fourth-grade children were divided into two groups of 10 subjects, 5 male and 5 female in each group. Children recorded a paragraph, sentence, and sustained vowel which were randomized and later played back to each group. Group 1 heard the samples once; group 2 heard them three times. The children were instructed to write the name of the speaker. The results indicate that children can identify peers from a group of 10. Identification increased for both groups as the repertoire of the sample increased from vowel to sentence to paragraph; however, three repetitions of the same sample-type did not increase identification significantly. The results suggest that repertoire rather than actual duration of the sample provided cues to the children.

3:00

Z5. Automatic Speaker Verification Using Phoneme Spectra. G. D. HAIR AND T. W. REKIETA, *Texas Instruments Incorporated, P. O. Box 5621, M/S 939, Dallas, Texas 75222.*—A technique, amenable to completely automatic real-time application, for verifying the identity of a cooperative speaker has been demonstrated and evaluated. Wide-band digital recordings of 10 isolated words were made on each of 9 weeks for 230 speakers. Segments of selected phonemes were automatically edited and power-density spectra computed over a 7.5-kHz bandwidth. A preliminary analysis of 40 randomly selected speakers has been performed. Phoneme spectra for five repetitions were averaged to produce speaker standard pattern vectors with the remaining four repetitions used for error-rate estimation. Pattern vectors of dimension 180 (smoothed spectra of six phonemes), 26 (features derived from the spectra), and 206 (combination of spectra and features) were used for technique evaluation. Combined error rates of 2, 5, and 1%, respectively, were obtained using a simple hypersphere decision rule. It is concluded that reliable real-time speaker verification can be achieved with completely automatic digital processing techniques and modest pattern storage requirements.

3:15

Z6. Mimic Resistance of Speaker Verification Using Phoneme Spectra. T. W. REKIETA AND G. D. HAIR, *Texas Instruments Incorporated, P. O. Box 5621, M/S 939, Dallas, Texas 75222.*—The susceptibility to mimicry of the speaker-verification technique discussed in the paper "Automatic Speaker Verification Using Phoneme Spectra" was evaluated with the assistance of a professional performer specializing in impersonations. Subjects to be mimicked were six speakers selected from the previously recorded population. After becoming familiar with each speaker, the impersonator was permitted to mimic the speaker immediately after his utterance of each selected word. The subjects' utterances and the impersonator's mimic attempts were then processed for speaker verification on a single-phoneme basis and using combinations of several phonemes. Spectral analysis of individual phoneme segments revealed that some increase in similarity was accomplished by the mimic for certain speakers and phonemes. However, when the verification procedure was applied using features from five phonemes, the impersonator was unsuccessful in all mimic attempts.

3:30

Z7. Test of an Automatic Speaker Verification Method with Intensively Trained Professional Mimics. R. C. LUMMIS AND A. E. ROSENBERG, *Bell Telephone Laboratories, Murray Hill, New Jersey 07974.*—Several professional mimics selected by audition from a large group, were given intensive training on utterances of the eight "customers" in Doddington's 40-speaker population [J. Acoust. Soc. Amer. **49**, 139(A) (1971)]. The training method uses an interactive DDP-516 computer to provide the mimic with immediate playback of his practice utterances as well as immediate $A-B$ comparison between his own and customer utterances. Recordings of the *best* utterances from the *best* four mimics were processed by the computer verification system described previously [Doddington, op cit., and R. C. Lummis, J. Acoust. Soc. Amer. **50**, 106(A) (1971)]. The system uses five features for verification: pitch, level, first, second, and third formant frequencies. If the acceptance–rejection criterion that yields equal-error performance in Doddington's speaker population is used for the mimics, 27% of the best utterances by the best mimics