Automatic Speaker Authentication Using Speech Recognition Techniques

W. F. Meeker, T. B Martin, M. B. Herscher, et al.

Citation: The Journal of the Acoustical Society of America **42**, 1182 (1967); doi: 10.1121/1.2144065
View online: https://doi.org/10.1121/1.2144065
View Table of Contents: https://asa.scitation.org/toc/jas/42/5
Published by the Acoustical Society of America

boundary for speakers of the language. Optimal binary features based exclusively on this criterion have been derived, using the 23 consonant confusion matrix data of Wickelgren [J. Acoust. Soc. Am. 39, 388 (1966)]. These optimal features are closely related to the binary distinctive features of phonology.

S5. Phonemic Confusion in Whistled Turkish. JARVIS BASTIAN AND CAROL WALL, *Department of Psychology, University of California, Davis, California.*—A method of whistling the local Turkish dialect has been developed in Northeastern Turkey that involves whistling while executing approximations to the standard articulator maneuvers of the spoken dialect. The signals produced are sufficiently intense and intelligible to those who have learned to use this signaling system to provide an auxiliary means of linguistic communication over long distances in the rugged terrain they inhabit. This system of "whistled speech" is very similar to those that have been recorded in much the same ecological conditions in one of the Canary Islands and in the Pyrenees. However, the Turkish system is based on a far richer phonemic inventory than the others, which are based on Spanish dialects. We therefore analyzed the effectiveness of the Turkish system in transmitting the phonemic contrasts maintained in the spoken language. Confusion test results indicate a severe reduction in the phonemic contrasts transmitted in the whistled version. The nature of these confusions and of their acoustical basis is discussed together with their consequences for sentence intelligibility. [Field work supported by Wenner–Gren Foundation.]

S6. Machine Recognition of Speech. T. B. MARTIN, H. J. ZADELL, R. B. COX, AND M. B. HERSCHER, *Radio Corporation of America.*—This paper summarizes the status of a continuing research program concerned with the recognition of phonemes in both isolated and continuous speech. The acoustic analysis technique employs analog-threshold logic to abstract features that provide the basis for the recognition decisions. These feature-abstraction networks have been organized in a hierarchy of processing levels, ranging from the recognition of broad classes of sounds to decisions on a phoneme-by-phoneme level. As part of this program, a detailed comparison has been made of the characteristics of phonemes in continuous speech as contrasted with well-articulated discrete speech. The development of a continuous speech recognition capability makes possible the realization of several specific applications. Among these are spoken ZIP-code recognition systems that will improve the efficiency of parcel-sorting operation. In this application, it is necessary to recognize continuous strings of digits in a very high-noise background. A speech communications system is also being developed utilizing a speech recognition system at one end of a communications link, a narrow bandwidth channel transmitting the recognition results, and a speech synthesizer at the receiver.

S7. Automatic Speaker Authentication Using Speech Recognition Techniques. W. F. MEEKER, T. B. MARTIN, AND M. B. HERSCHER, *Radio Corporation of America, Camden, New Jersey,* AND DOUGLAS PHYFE AND MARTIN WEINSTOCK, *U. S. Army Electronics Command.*—While some degree of speaker recognition can be achieved using gross measures such as average glottal frequency or average spectrum, it seems certain that, for highly accurate speaker recognition or authentication, it will be necessary to know what is spoken. In applications where there is no control over what the speaker utters, automatic speech recognition becomes a prerequisite for automatic speaker authentication. Considerable success in automatic speech recognition has been achieved using analog-threshold logic and this equipment has been used to select automatically portions of continuous speech for speaker authentication. Characteristics of continuous speech deviate substantially from those predicted by classical phonetics and many of these deviations can be used for speaker authentication. A number of different measures have been studied and automatically extracted. For vowels, the basic measure is the slope of the spectrum. For consonants, spectral characteristics, duration, and sequence of occurrences are useful measures. The over-all accuracy of authentication can be quite high when results for many individual measurements are combined. Results for single and multiple measurements will be discussed. [Work sponsored by the U. S. Army Electronics Command.]

S8. Properties of the Instantaneous Frequency and its Application to Speech Classification. K. HIRAMATSU, D. K. RALEY, R. WACKERBARTH, AND C. L. COATES, *Laboratories of Electronics and Related Science Research, Electrical Engineering Department, The University of Texas, Austin, Texas 78712.*—The instantaneous frequency, corresponding to the derivative of the phase function of the analytic signal, expressed in terms of the zeros of the function and its mean value, is formulated, and the properties of the instantaneous frequency will be illustrated using some typical examples. An expression for the instantaneous frequency in the complex frequency domain has been derived, and has the following properties: (1) When the spectral distribution is symmetric, the average value of the instantaneous frequency corresponds to the mean value of the distribution. (2) When asymmetric, the average value corresponds to the predominating region of the spectrum. Emphasizing or weighting the spectral distribution results in the predomination of frequencies that would not predominate under ordinary circumstances. The average value of the instantaneous frequency then corresponds to that region of the spectrum. The properties of these expressions for the instantaneous frequency will be applied to the classification of a limited number of connected speech sounds; i.e. zero, one, . . ., nine.

S9. Rules for Word Stress Analysis: for Conversion of Print to Synthetic Speech. JANE H. GAITENBY, *Haskins Laboratories, New York, New York.*—Successful conversion of graphemes to English allophones depends on correct assignment of syllable stress (word accent). In turn, stress reflects morphological and syntactic patterns that are intrinsic to the language. Although written English lacks obvious prosodic markers on the word level, its spelling generally represents phonetic realities if stress is not overlooked. (Printed words are also conveniently framed by space.) An important truism is that normal printed texts closely follow spoken grammatical routines. With these facts made operational, and coupled to a modest storage of letter sequences (with appropriate syllables treated as stable stress elements), a fair degree of prosodic contour prediction is attainable. In this paper, preliminary rules for deriving stress information from a printed input are exemplified. The usefulness of these rules in a program for basic intonation in a synthetic speech output is demonstrated. [This research on reading machines for the blind was made under contract to the Research and Development Division of the Prosthetic and Sensory Aids Service, Veterans Administration, New York.]

S10. Speaker-Listener Dialect Error in Discrimination Testing. CORNELIS W. KOUTSTAAL AND MARTHA HOSACK, *Department of Speech, Bowling Green University, Bowling Green, Ohio 43402.*—Articulation curves were obtained from a group of listeners who listened to speech discrimination-test words in their own dialect and a dialect different from their own. The speakers were from the General American dialect and the Southern American dialect regions. A male and a female speaker represented each dialect. The listeners were all from