Auditory word recognition and the locus of word frequency effects

Cynthia M. Connine and Debra Titone

Citation: The Journal of the Acoustical Society of America **87**, S108 (1990); doi: 10.1121/1.2027817
View online: https://doi.org/10.1121/1.2027817
View Table of Contents: https://asa.scitation.org/toc/jas/87/S1
Published by the Acoustical Society of America

RR25. A knowledge-based temporal alignment system involving stochastic methods. H. Kabré, J. F. Malet, J. M. Pécatte, G. Pérennou, and N. Vigouroux (IRIT/CERFIA Lab., CNRS UA824, Paul Sabatier Univ., 118 Route de Narbonne, 31062-Toulouse Cedex, France)

In this paper the principle behind an automatic signal alignment system, used for speech database labeling, is described. This system is based on both a phonetic knowledge subsystem and a stochastic training procedure. It operates in two passes. During the first one, for each frame, temporal cues are worked out, similar to those handled by the hand labeler. The frames are next recombined into phonetic events that are assigned broad labels—e.g., vocalic, fricative, liquid, sonorant, occlusive, etc. The second pass aligns phonetic events onto a phonetic transcription—provided by an expert. It resorts to a stochastic model entailing (i) a phonetic knowledge base and (ii) a channel modeling first-level decoding errors. The training of the phonetic knowledge base consists in fine tuning the assessment of probability weights. This is done on hand-labeled corpora. Results, on both French and English corpora, will be presented, attention being paid to the problem of system multilingual adaptability.

RR26. Auditory word recognition and the locus of word frequency effects. Cynthia M. Connine and Debra Titone (Psychol. Dept. and Ctr. for Cognitive and Psycholinguistic Sci., SUNY, Binghamton, NY 13901)

The locus of lexical frequency influences in auditory word recognition was investigated using a speech identification paradigm. Subjects were required to identify the initial phoneme of tokens from voicing speech continua. Here 44 continua were constructed such that the voiced and voiceless endpoint of each pair contrasted in lexical frequency. Identification responses for ambiguous stimuli tended to be the higher-frequency member. Reaction times for unambiguous stimuli showed an advantage for the high-frequency responses; ambiguous stimuli showed no such reaction time advantage. The patterns of reaction time for word frequency were similar to influences of monetary payoff [Connine and Clifton, J. Exp. Psychol.: Hum. Percept. Perform. **13**, 219–299 (1987)] and support a post-perceptual influence of word frequency. [Work supported by NIH.]

RR27. Phonetic priming in spoken word recognition: Task comparisons. Deborah A. Gagnon (Dept. of Psychology, SUNY, Buffalo, NY 14260)

In an effort to reveal the nature of the representation underlying spoken word recognition, a basic priming paradigm has been used across a series of different tasks. The paradigm involves presenting subjects with natural CVC prime-target pairs in various degrees of phonetic overlap, and asking the subjects to respond to the target as fast as possible in a task-specific manner. Previous reports on the use of this paradigm described the results from a naming task [Gagnon and Sawusch, J. Acoust. Soc. Am. Suppl. 1 **85**, S125 (1989)] and a phoneme monitoring task [Gagnon and Sawusch, J. Acoust. Soc. Am. Suppl. 1 **86**, S99 (1989)]. The pattern of results obtained in the naming task reflected a position-specific phonetic representation, while the phoneme monitoring task yielded results supportive of an allophonic representation. The methodology has been extended to a lexical decision task in which subjects determine whether the target item is a word or nonword. A comparison of the results obtained from the lexical decision task to those obtained from the naming and phoneme monitoring tasks will be made, and conclusions regarding the appropriateness of these tasks for investigating the representation used in spoken word recognition will be offered. [Work supported by NIDCD Grant DC00219 to SUNY at Buffalo.]

RR28. Phonological rule set complexity as a factor in the performance of a very large vocabulary automatic word recognition system. Philip F. Seitz, Vishwa N. Gupta, Matthew Lennig, Patrick Kenny, Li Deng, Douglas O'Shaughnessy, and Paul Mermelstein (INRS-Télécommunications, 3 place du Commerce, Ile-des-Soeurs, Verdun, Quebec H3E 1H6, Canada)

Generative phonological rules were used to derive surface forms of words for the 89 000-word lexicon of the INRS-Télécommunications speaker-dependent, isolated word recognizer. The representation of the surface forms is in terms of the 41 phonemelike subword recognition units that are trained and recognized using 25-component continuous mixture hidden Markov models. Three sets of surface forms were generated by three phonological rule sets that differ in the amount of phonetic and phonological variability they represent. The three dictionaries thus derived were of different sizes, having ratios of surface to base forms of 1.04, 1.23, and 2.40. Recognition experiments were run on the speech of one talker. Representing pronunciation variations explicitly in the dictionary affords a recognition performance advantage, but this advantage is partially offset by a higher ratio of surface to base forms, which leads to a larger and more densely populated search space and greater confusability of subword unit sequences. [Work supported by the Natural Sciences and Engineering Research Council of Canada.]

RR29. Talker variability and word recognition: A developmental perspective. Brigette R. Oliver (Speech Res. Lab., Dept. of Psychology, Indiana Univ., Bloomington, IN 47405)

The effects of talker variability on spoken word recognition were studied developmentally in 3-, 4-, and 5-yr-old children. Subjects listened to lists of words and were required to identify each word by pointing to a picture in a six alternative visual display. The words and pictures were taken from the word intelligibility by picture identification test (WIPI). Three talker conditions were examined: single talker, single talker with varying amplitude, and multiple talker. Each child heard three lists, one from each talker condition, in varying order. Results showed a main effect of age and an interaction between talker condition and list order. As expected, we found an increase in overall accuracy with age. However, the 3-yr-olds did better in the single-talker condition than the other conditions (which did not differ). This was found only when the lists were presented first during a test session. Second and third lists showed no difference due to talker condition. In most of the conditions the 4- and 5-yr-olds displayed ceiling levels of performance that may have obscured other differences. Further studies are underway with younger children to examine the nature of the differences in perception between single-talker and multi-talker lists. [Work supported by NIH Grant DC-0111-14 to Indiana Univ.]

RR30. A study of LSF parameters for speaker recognition. K. K. Paliwal[a] (Comput. Systems and Commun. Group, TIFR, Homi Bhabha Rd., Bombay 400005, India)

Linear prediction (LP) analysis of speech can lead to a number of parametric representations (such as cepstral coefficients, reflection coefficients, etc.), each of which provides equivalent information about the spectral envelope. The line spectral-pair frequency (LSF) representation is an alternate LP parametric representation. For speech coding and recognition applications, this representation has been recently found to be more useful than the other LP parametric representations. For speaker recognition application, the cepstral coefficient representation is traditionally used as an LP parametric representation. In this paper, the LSF representation for speaker identification is studied and its performance is compared with that of the cepstral coefficient representation. This representation results in speaker identificaton accuracy of 94% for the diagonal covariance matrix case and 95% for the full covariance matrix case; the corresponding results for the cepstral coefficient representation are 91.5% and 94.2%. Thus the LSF representation is better suited for speaker identification than the cepstral coefficient representation. [a] Current address: Acoustics Res. Dept., AT&T Bell Labs., Murray Hill, NJ 07974.