AUGUST 11 2005

Morph Lexicon for Speech Synthesis by Rule

Jonathan Allen

J. Acoust. Soc. Am. 51, 111 (1972)
https://doi.org/10.1121/1.1981295

thesizer. The synthesizer is excited with a pulse train, noise, or a mixture of the two. Examples of the natural and synthetic speech are presented.

3:15

P6. On Evaluating Formant Synthesizer Control Signals. W. WATHEN-DUNN, S. B. MICHAELS, L. V. KRIGER, AND H. I. SORON, *Air Force Cambridge Research Laboratories, Bedford, Massachusetts 01730.*—The problem of evaluating the inherent capability of a given synthesizer to produce good speech is inseparable from the problem of evaluating the signals that control it. Those values of control signals that give the best speech output must be accepted as optimum, and the resulting performance of the synthesizer must be accepted as maximum. Optimizing the control signals must be done, however, not in some average fashion, but with reference to the specific phonemic content of the speech. The objectives of psychoacoustic test procedures for this purpose are thus somewhat different from those used in preference testing of speech communication systems, and so are the means. Forced-choice paired comparisons have been used to evaluate different control signals for one synthesizer producing specific phonemes in specific environments. The results are used as a guide for choosing the best means for deriving the signals from real speech. Evaluations of the control signals for stop consonants from a Glace–Holmes Synthesizer will be presented.

3:30

P7. Morph Lexicon for Speech Synthesis by Rule. JONATHAN ALLEN, *Department of Electrical Engineering and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.*—This paper reports the results of representing the 1 000 000-word Brown Corpus by a set of morphs and their underlying morphemes. The detailed procedure, including practical considerations for handling such a large data base, is discussed, and the algorithm which decomposes words into morphs is presented. It was found that many words decompose into morphs ambiguously, so that a set of selectional rules is needed to indicate the correct morph sequence. Thus affixation is generally preferred over compounding in English. The data base for obtaining these rules is discussed. Some statistics for the resulting lexicon are also presented. Finally, a procedure for merging the morph lexicon (spellings) with the *Merriam Pocket Dictionary* to obtain pronunciations and parts of speech is presented, together with a description of the resulting dictionary. [This work was supported in part by the National Institutes of Health.]

3:45

P8. Digit Concatenation from Simple Rules. JOSEPH P. OLIVE, *Bell Telephone Laboratories, Speech and Communication Research Department, Murray Hill, New Jersey 07974.*—A previous paper by Olive and Nakatani ["Quality Judgments of Spliced Synthetic Telephone Numbers," J. Acoust. Soc. Amer. **50**, 145(A) (1971)] described an experiment to evaluate the relative importance of pitch, timing, and formant coarticulation rules in digit concatenation from isolated digits. That paper concluded that pitch was the most important aspect of concatenation while the other two parameters did not contribute much to the quality of the concatenated speech. This paper will describe the simple concatenation rules that were based on the recommendations of the previous paper, and evaluate the results. A digit recall test was administered to measure comprehension and its results prove that the rule-concatenated numbers are at least as recallable as the same numbers spoken naturally. A subjective preference test revealed that the rule-concatenated speech was judged to be close to speech resynthesized from natural speech. Half the subjects preferred the rule-concatenated speech to speech spliced from isolated digits even when the isolated digits were naturally spoken; the other half disliked any form of synthetic speech.

4:00

P9. Speech Frequency and Duration Measures as a Function of Chronologic Age. THOMAS SHIPP, *VA Hospital, San Francisco, California 94121,* AND HARRY HOLLIEN, *Communication Sciences Laboratory, University of Florida, Gainesville, Florida 32601.*—Fundamental frequency and duration measures were obtained on 175 normal male subjects whose ages were evenly distributed throughout the range 20–89 yr. It was found that the fundamental frequency contours by age decade show a steady but gradual lowering from the 20-yr-olds at 120 Hz to the lowest mean frequency of 107 Hz in the 40–49-yr-old group. Average frequency then raised steadily throughout the 50-, 60-, and 70-yr decades to reach a high in the 80-yr-old group of 146 Hz. Mean oral reading rate showed a steady increased duration with increased chronologic age, with the 80-yr-old group having a mean duration 50% greater than the mean for the 20-yr-olds. The increase in duration was principally a function of the increase in silent periods during the utterance or in intraphrasal pauses. For both frequency and duration measures the greatest variability was found in the 80-yr-olds with some of these older subjects having frequency and duration values similar to the young adult subjects.

4:15

P10. PARAFAC2: Extensions of a Procedure for "Explanatory" Factor-Analysis and Multidimensional Scaling. RICHARD HARSHMAN, *Phonetics Laboratory, UCLA, Los Angeles, California 90024.*—Recently developed techniques for factor analysis and multidimensional scaling (PARAFAC-INDSCAL) allow discovery of a unique orientation of axes and thus "more explanatory" factors or dimensions. A certain price is paid for these advances, however. In multidimensional scaling applications, the price is a restriction of the possible form of the solution to orthogonal "nonrelated" as opposed to oblique "perceptually related" dimensions. In factor analytic applications, the price is a limitation to those three-way data sets showing "system-variation" [Harshman, UCLA W.P.P. No. 16 (1970)]. An extension of PARAFAC to summed-cross-product matrices (e.g., covariance matrices) overcomes both of these limitations. For factor analysis with PARAFAC2 it is only necessary that the average effect of a given factor change, relative to the other factors, from one covariance matrix to the next. For multidimensional scaling, PARAFAC2 describes the perceptual space in terms of orthogonal dimensions only when this best fits the data. Otherwise, it will recover the stimulis projections on oblique dimensions and also give the angles between these dimensions. Advantages of PARAFAC2 are explored by comparing its results with those of PARAFAC. Both factor-analytic and multidimensional scaling applications are demonstrated using data from speech physiology and perception.

4:30

P11. Objective Measures of Peak Clipping and Threshold Crossings in Continuous Speech. PAUL T. BRADY, *Bell Telephone Laboratories Holmdel, New Jersey 07733.*—This study reports data on the statistics of instantaneous speech levels in continuous speech samples, with special emphasis on threshold crossings and other quantities related to peak clipping. Peak clipping is also treated as speech-correlated noise by assuming that it is caused not by a voltage limiting process, as actually occurs, but by an additive "phantom noise signal" that will cause the original signal to appear to be