Appropriateness of different rule types in speech synthesis

Sue Hertz

Citation: The Journal of the Acoustical Society of America **65**, S130 (1979); doi: 10.1121/1.2017052
View online: https://doi.org/10.1121/1.2017052
View Table of Contents: https://asa.scitation.org/toc/jas/65/S1
Published by the Acoustical Society of America

son, Vasundara V. Varadan, and Vijay K. Varadan (Wave Propagation Group, Department of Engineering Mechanics, The Ohio State University Columbus, OH 43210)

A new scattering matrix theory is developed for investigating the scattering of acoustic waves by elastic or viscoelastic obstacles of arbitrary shape immersed in a fluid. The problem is difficult since the Q matrix obtained in the usual way is not square and hence cannot be inverted. In this paper, a T-matrix formalism is presented by considering additional representations of the scattered and refracted fields so that one arrives at matrix equations that are invertible. Numerical results for the scattering cross sections of prolate, oblate, and spherical obstacles immersed in water are presented as a function of the dimensionless wave number.

FRIDAY AFTERNOON, 15 JUNE 1979 KRESGE AUDITORIUM, 2:00 TO 4:40 P.M.

Session YY. Speech Communication X: Synthesis and Recognition

R. M. Schwartz, Chairman

Bolt Beranek and Newman Incorporated, 50 Moulton Street,
Cambridge, Massachusetts 02138

Contributed Papers

2:00

YY1. MITalk-79: The 1979 MIT text-to-speech system. Jonathan Allen, Sharon Hunnicutt, Rolf Carlson, and Bjorn Granstrom (Room 36-575, Massachusetts Institute of Technology, Cambridge, MA 02139)

To mark the completion of a ten-year effort to develop a high performance text-to-speech algorithm, we have established a benchmark system called "MITalk." Components of the computer-simulated bench mark include: (1) conversion of abbreviations and special text symbols, (2) a lexicon consisting of about 11 000 morphs with pronunciation and parts of speech, (3) morpheme analysis, (4) letter-to-sound rules, (5) syntactic analysis, (6) rules for stress assignment, boundary placement and phonological recoding, (7) fundamental frequency and segmental duration prediction, (8) phonetic-to-parametric conversion, and (9) digital formant synthesis. The MITalk-79 system is being extensively documented and its performance is being evaluated. The presentation will summarize aspects of system organization and performance. (A more complete description will be given in a one-week course to be offered June 25–29, 1979.) The oral presentation will include a five-minute demonstration of synthetic speech generated from English text with absolutely no human intervention. While currently simulated on a large digital computer, MITalk-79 is amenable to practical IC technology. Implementation issues will be briefly discussed. [We gratefully acknowledge the synthesis-by-rule programs and advice provided by Dennis Klatt.]

2:10

YY2. Appropriateness of different rule types in speech synthesis. Sue Hertz (Department of Modern Languages and Linguistics, Morrill Hall, Cornell University, Ithaca, NY 14853)

The Cornell Speech Research System has been augmented to support a third kind of rule. In addition to feature and parameter rules, the user can now express letter-to-sound rules. Feature rules modify the feature composition of an utterance, and parameter rules transform the output of the feature rules into a file of synthesizer parameters. Letter-to-sound rules convert standard orthography into a phonetic transcription. A powerful notation has been developed for expressing these rules. By using the system, we have found that the three kinds of rules can be written to interact in linguistically meaningful ways to yield speech.

2:20

YY3. A set of concatenative units for speech synthesis. Joseph Olive and Mark Liberman (Bell Laboratories, Murray Hill, NJ 07974)

In a previous paper [J. P. Olive, "Rule Synthesis of Speech From Dyadic Units", IEEE International Conference on Acoustics, Speech and Signal Processing, 568–570, (1977)], we discussed a speech synthesis by rule scheme where the segments used for synthesis were obtained from natural speech. These segments included the consonants and the transitions from consonants to vowels, vowels to vowels, and vowels to consonants. Each synthesis parameter was defined by two sets of LPC area parameters, and in the concatenative process, straight line interpolation was used to obtain the complete set of area parameters. Informal listening and some formal intelligibility testing revealed that this simplified description of the synthesis segments was not sufficient to produce the kind of speech that would satisfy us. Consequently, it was decided to expand the definition of the set of concatenative units. Specifically: (1) the number of phonetic segments defining a synthesis unit can vary from one to three; (2) arbitrary number of area parameter vectors can be stored for a given synthesis unit; and (3) a facility is provided for expressing transition durations and amplitude rules for synthesis units or classes of units. This paper will discuss the details of a more complete and accurate set of concatenative units.

2:30

YY4. A demisyllable inventory for speech synthesis. J. B. Lovins, M. J. Macchi, and O. Fujimura (Bell Laboratories, Murray Hill, NJ 07974)

English syllables can be treated as a syllable core, decomposable into initial and final demisyllables, plus optional word-final phonetic affixes (e.g., /s,z,t,d,*θ*/) [Lovins and Fujimura, J. Acoust. Soc. Am. **60**, S75 (A) (1976)]. An inventory of 835 demisyllables, five phonetic affixes, and 100 reduced-vowel units (CV or VC syllables) has been prepared in the form of LPC parameters. This inventory allows us to generate any of the ten thousand possible English syllables. An initial demisyllable is composed of an initial consonant or cluster, plus a fixed-length segment which includes the CV transition; a final demisyllable contains the major portion of the syllable nucleus, plus postnuclear consonant(s). We have found it generally appropriate to produce "tense" or "long" syllabic nuclei by concatenations such as /Cɪ−/ + /−iyC/. For stressed syllables, this allows a