TEXT-TO-SPEECH CONVERSION BY RULE AND A PRACTICAL APPLICATION

Peter B. Denes, Mark Y. Liberman and Joseph P. Olive, Bell Laboratories, Murray Hill, New Jersey 07974, USA

A system for the rule synthesis of voice answerback sentences for telephone directory-assistance purposes is described. The sentences have the form "The number for (Joe Snooks) of (518 Oaklands Avenue) is (345-6789)". Research on such a system offers the attractions of a genuine practical application for rule synthesis. It combines a non-trivial text-to-speech conversion task for the large numbers of names and addresses involved, yet avoids many of the unknowns associated with the synthesis of general English text because only a single carrier sentence is used. Also, evaluation of comprehensibility can be more realistic, using genunine users with a communication task, rather than laboratory subjects.

The task is performed in two steps. First the text is converted by rule into phonetic transcription, including stress and segment durations. The spelling-decoding involves a limited morphological analysis and a set of context-sensitive rewriting rules. Stress is assigned by a simplified version of the principles in Liberman and Prince (1977). A small dictionary of orthographically exceptional words is maintained. Durations are assigned by a set of rules which take into account the segment, its segmental context, the local stress pattern, and constituent structure.

The output of the above process serves as input to the stage of the synthesis process in which intonation is determined and the acoustic wave is calculated. The pitch contour is obtained by selecting and adjusting one of several stored contours. The acoustic wave is calculated by dyadic concatenation of vocal tract area function segments: the concatenation is based on a matrix of phoneme transitions stored as vocal tract area parameters.

The computer program implementing the first of the above two steps runs on a PDP11/45 several times faster than the associated speech time. In tests of randomly selected telephone directory entries, 91% of all entries were given a "correct" phonetic transcription and stress pattern. The second step also runs in real time, using a specially wired vocal tract area function synthesiser.

Reference

Liberman, M.Y. and A. Prince (1977): "On stress and linguistic rhythm", Linguistic Inquiry 8.2, 249-336.