

Determining Concatenative Units for Speech Synthesis

Evelyne Tzoukermann* and Judith L. Klavans

Columbia University, Department of Computer Science

New York, New York 10027

klavans@cs.columbia.edu

**A.T.&T. Bell Laboratories*

600 Mountain Avenue, Murray Hill, N.J. 07974

evelyne@research.att.com

The purpose of this research is to determine the best method for deciding on concatenative units (diphones and other polyphones) for speech synthesis. Four different databases are used and a greedy algorithm (Van Santen, J, 1991) is applied in order to extract n-gram phonemic frequencies; this serves as a basis for comparison between dictionary derived frequencies and corpus derived frequencies. Two dictionaries, a large one (the Robert Encyclopedic French dictionary with 85000 headwords) and a small one (the Collins Gem containing 15000 words) and two phonetically transcribed corpora, a large one (the Hansard, of about 2.5 million words) and a smaller one (the Tubach and Boe of 80000 words) were used. The hypothesis was that a dictionary could provide essential concatenative unit data without recourse to corpora. Nevertheless, results showed that for interword phenomena, the hypothesis was correct, whereas for intraword effects, such as liaison and assimilation, the dictionary field information falls short.

This research presents two interesting aspects: on the practical side, a speech synthesis system for French (Tzoukermann, E. 1993) is being built at AT&T Bell Laboratories and the output of this research will determine the most frequent concatenative units for putting into the working system. On the theoretical side, the goal is to explore the question of what constitutes adequate data for inducing concatenative units.

Number of words: 215

Technical Area: Speech Communication, Speech synthesis

Telephone number: (908) 582-2924

Send notice to: Evelyne Tzoukermann

Special facility: overhead projector

Method of Presentation: Prefer lecture but willing to give poster