DISCOURSE INTONATION AND EXPRESSIVE SYNTHETIC SPEECH

T Brovchenko and V Veloshin

Odessa State University, USSR

ABSTRACT

Marked results have been achieved in recent years in text-to-speech synthesis.

But though intonation of discourse has received considerable attention in the past decade, prosody being one of the main factors, determining the quality of synthetic speech, surprisingly little research has been carried out into synthetic discourse intonation.

Specific intonation contours, frequency level and intervals, the rate of Fo changes, location of the melodic peak of the utterance, etc. are the characteristics that make our speech expressive. They have been found typical both of English and Russian and being applied in generating intonation contours, make synthetic speech more natural and expressive.

An experimental study of prosodic structure of natural and synthetic English and Russian speech samples makes it possible to suggest that besides the above mentioned acoustic characteristics of separate utterances, prosodic features of speech units larger than phrases may also account for the quality of synthetic speech.

The aim of this paper is to describe the acoustic cues of discourse and to determine which of them are significant for the synthesis of speech prosody.

Electroacoustic analysis and synthesis of English and Russian dialogues read by ten male native speakers revealed that the main intonation peculiarities of discourse to be taken into consideration while generating intonation on contours for synthetic speech are the following:

1. Alternation of various melodical curves, contrasting Fo and intensity values of different phrases of discourse.

2. Alternation of different allophonic variants of intonation contours within the speech samples.

3. Contrasting frequency and intensity interval values and their variability, alternation of pauses of different length at phrase junctions that constitute the text.

Typical acoustic features of expressive natural speech are presented in Figures 1, 2 (solid line – English, dashed line – Russian).

Fig 1. Frequency and temporal characteristics of phrase junctions.

Fig 2. Alternation of allophonic variants of intonation contours within speech samples.

The results of the study show that in order to obtain more natural and expressive synthetic speech, intonation contours of speech units larger than phrase are to be generated.

The set of intonation contours applied to a text are not precisely predicted. Various allophonic intonation contours of different communicative types of phrases of the language studied are to be elaborated. In generating
the prosodic portrait of the whole text
these intonation contours are to be organised
according to the above mentioned rules of
alternation of expressive discourse intonation
acoustic characteristics.