TOWARDS DESIGNING AN INTONATION TRAINING DEVICE BASED ON SPEECH SIGNALS CLUSTERING

Leonid A. Kanter, Alexander V. Savin and Ksenia G. Guskova

Dept. of Phonetics (English), Lenin Pedagogical State University, Moscow, USSR

ABSTRACT

The paper proposes a technique for automatic intonation type recognition for the purpose of designing an intonation training device to facilitate the acquisition of prosody by foreign language learners as well as to promote intonation improvement of those who have a speech or hearing disorder. The technique is applicable to any utterance with a fixed number of syllables and was initially tested for disyllabic speech samples of Russian and English.

1. INTRODUCTION

Broadly stated, the object of this study is to provide a means for computer-assisted phonetic instruction through which sound effects produced by a spoken voice (or voices) of a strictly standardized character, free of regional accents, may be visually compared with more or less related sound effects, produced by another speaking voice of a non-standardized character, i.e. a voice of a foreign language learner or a verbally/perceptually handicapped patient. Comparing the user's pronunciation with the stored pronunciation makes it possible to measure a similarity or likelihood degree between the two types of speech patterns. More specifically, the paper proposes a technique for automatic intonation type recognition for the purpose of designing a visual display device for intonation training. The technique is applicable to any utterance with a fixed number of syllables (initially it was tested for disyllabic utterances of Russian and English).

2. CONCEPTUAL FOUNDATION

As a conceptual foundation the zonal principle of basic intonation types realization is adopted. In accordance with this principle, manifestations of a given intonation pattern are not just one-to-one reproductions of the underlying archetype. They are rather variants within the range of tolerance set by a particular invariant in the space of acoustic parameters (F_0, I, T) and in the perceptual space.

3. GENERAL DESIGN

Intonation contours clustering results, reported for Russian and English in [1-3], were used as source data for the present project.

With implementation of the training device in question visualization of intonational zones is attained, thereby enabling the learner not only to hear intonation but also "to see" it without being tied to a specific intonation curve or a set of curves. In this case the language learner has to deal with points representing the curves on a display. A cluster analysis algorithm, proposed in [4], is used to reduce every curve to a point which is mapped on a plane. In order to make the initial cluster more convenient for teaching all extraneous data (i.e. points belonging to other clusters) are eliminated. As is shown in Fig. 1, the remaining points of each cluster are linked up in straight lines. It is assumed that the interpoint distance within the cluster does not exceed an empirically found fixed value. The procedure helps the learner to assess the cluster structure for the reference samples storage zone. When a new, prosodically distorted realization is mapped on a display, a decision as to which cluster it belongs is made, judging by the two neighbouring points between which it is located. The method is known as "the nearest neighbour strategy". Some possible results are exemplified in Fig. 2. The learner can modify his intonation realization as many times as required until the point on a display reaches the right cluster, that is to say, until the learner's point is located between two reference points of the target cluster.

This sort of a technical aid for teaching intonation may be PC-based.

4. CONCLUSION

Displaying visual information for speech training purposes, as reported here, will enable the learner to make a selection from a series of options available in the reference storage zone for a particular intonation type with due account of the individual range of the learner's vocal performance.

5. REFERENCES

[1] CHIZHOV, A.P., CANTER, L.A. and SOKOLOVA, M.A. (1983), "English intonation in the space of acoustic parameters", Abstr. Xth Int. Congr. of Phonetic Sciences, Dordrecht/Cinnaminson, 360.

[2] KANTER, L.A., CHIZHOV, A.P., GUSKOVA, K.G. (1987), "A cluster-seeking technique for prosodic analysis (with special reference to Russian sentence intonation)", Proc. XIth Int. Congr. of Phonetic Sciences, Tallinn, vol.4, 59-61.

[3] KANTER, L.A. (1988), "The systems approach as a methodological principle for intonation studies" (in Russian), Funktzionalnij analiz Foneticheskikh edinitz anglijskogo jazijka, Moskva, 26-70.

[4] SAMMON, J.W., Jr., "A nonlinear mapping for data structure analysis", IEEE Transactions on Computers, C-18, 5, 401-409, (1969).

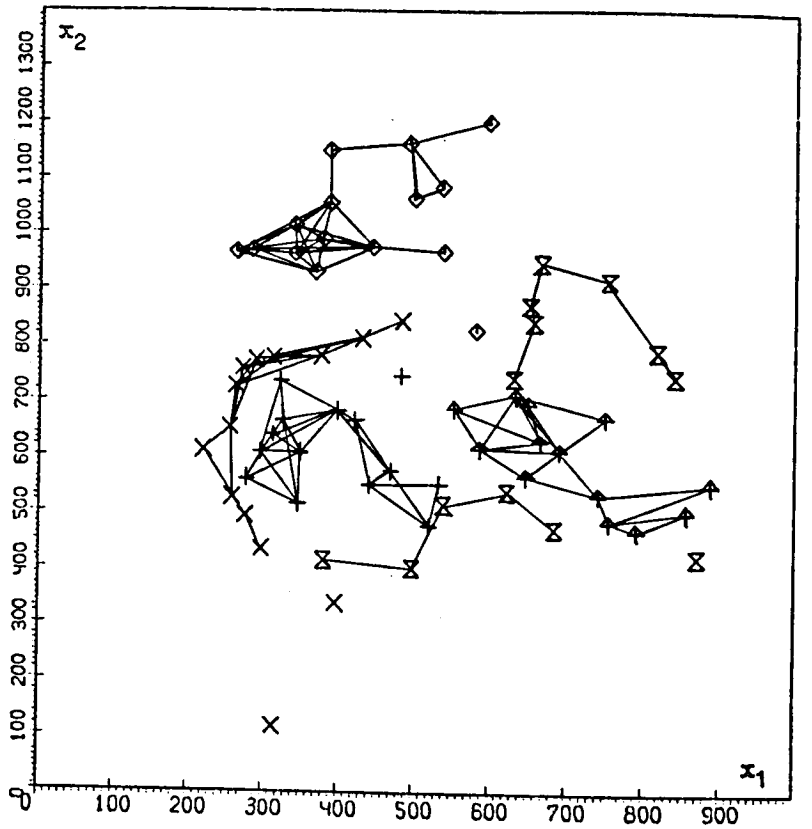

Fig. 1. Internal cluster structure for the reference samples storage zones (interpoint distance ⩽170). The test phrase ОН ЗНАЛ , pronounced by 16 male native speakers of Standard Russian was considered.

Symbols for intonation types read as follows:

+ - final statement
× - reply statement
◇ - general question
↑ - exclamation
⧖ - non-final statement

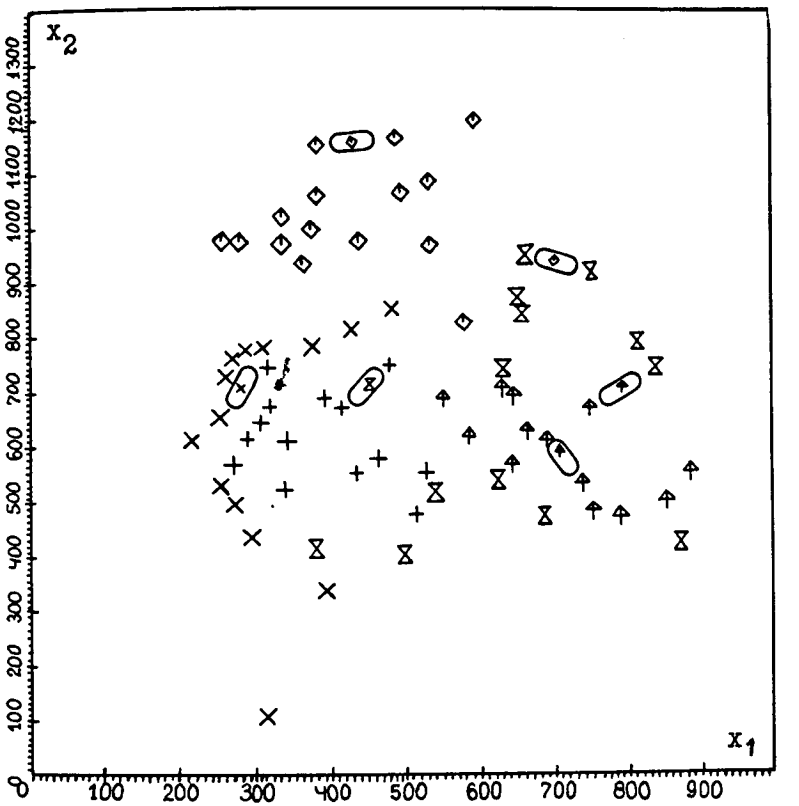

Fig. 2. Reference samples storage zones and non-native imitations of the Russian test phrase ОН ЗНАЛ , produced by English learners of Russian. Non-native imitations are marked in smaller symbols within an ellipse.