SOME PRINCIPLES OF CONSTRUCTION OF SPEECH RECOGNITION SYSTEM

G.V.KRYUKOV
Dept. of Mechanics
and Mathematics
Moscow State University
Lenin Hill, Moscow
USSR, 119899

A.V.MIKHALEV
Dept. of Mechanics
and Mathematics
Moscow State University
Lenin Hill, Moscow
USSR, 119899

V.N.TRUNIN-DONSKOY
Computer Center of
Academy of Sciences of USSR
Vavilov Street, 40
Moscow, USSR, 117333

ABSTRACT

This paper is based on the theory of fuzzy sets as a mathematical means of description of speech and language. It is suggested to consider the problem of parametric representation and following analysis of speech signals with the purpose of recognizing as a problem of successive transformations from fuzzy subsets to usual ones and vice versa and an analysis of obtaining results at every step.

INTRODUCTION

The idea to use the theory of fuzzy sets for speech recognition was suggested in /1-5/. Partially it was connected with difficulties arising with attempts to use the traditional mathematical methods of description of speech and landuage, particularly when constructing nonadaptive systems of speech recognition, i.e. speaker-independent speech recognition systems for an arbitrary speaker which is a carrier of pronunciation norms for given languge. Experiments of Klatt and Stevens/6/ shows that uncertainty of speech signals in acoustic field is the main propety of speech. At the same time under phonetic decoding linguists are managing without any complicated mathematical means and even without sufficient current information. The studies on blind-spectrogram-reading experiments with various speakers which have been carried out over the years on the faculty of Philology of Moscow State University /7/ and constructions of speech understanding system based on the analysis of phonetician-expert-experience with decoding speech spectrograms confirm our thesis. Using only general lingustic knowledge on nature of formant parameters, on intensity of signals and harmonics expert-liguists make successfully phonetic decoding actually basing on

fuzzy linguistic variables (for example, high 1 formant, very low 2 formant, big total energy etc.)

FUZZY TRANSFORMATIONS IN SPEECH RECOGNITION

It has been assumed that elements of thinking of man would enable to imagine as fuzzy subsets for which the function of belonging takes not only 1 or 0 but real numbers beween 0 and 1. In practical systems of recognition it is not necesary on the parametric level to try to find some global decisins and on phonetic level to obtain unique decision on belonging to a definite class, but it is possible to accept several variants of sounds as it is doing in systems for understanding continuous speech when forming of phonetic lattices. Remark that uncertainty on the phonetic level may be setted using higher linguistic levels (lexical, syntactic and semantic). Thus extension of uncertainty of solutions for ill-definable classes of sounds on lower recognition levels may give more possibilities for right and reliable recognition of speech sounds than raising rigidness of decision-making on the same levels.

It is possible to describe speech which is by its nature rather complicated information phenomenon with the help of fuzzy linguistic variables connected with features of prime parameters which in their turn are results of some objective measurings and do not contain fuzzyness. Thus on the stage of prime data reduction of speech signal we have direct transformation from a fuzzy set to an usual one. At the same time labels indicating on belonging of learning set to suitable classes may be fuzzy. It means, that classification object must be going on with taking into account all of the probable classes. Thus on the stage of fuzzy classification we have an inverse transformation from usual subsets to fuzzy ones.

In the case of data reduction of speech signals using the notion of usual α- level subsets $A_\alpha = \{\bar{x} \mid \mu_{\underset{\sim}{A}}(\bar{x}) \geq \alpha\}$, we may formulate the general principle of construction of acoustic-phonetic processor, which consists in search of α-level which controls passage of speech signals throught some key schemes for purpose of subsequent processing. Experimentally α-levels are selected such i)basic noice of apparatus, stationary noice and room noice do not stand out against a background in pauses; ii)there is information on place of formation of weak fricative sounds (f,h) when they appear. As a result we have a direct transformation from the fuzzy subset of analogic speech signals to the usual subset of discrete figure codes corresponding to these analogic signals, i.e. $\underset{\sim}{U}(t) \to \{A_i^j\}$; where $\underset{\sim}{U}(t)$ is a fuzzy set of analogic speech signals; A_i^j is a set of discrete readings of parametres representing of a speech signal; $i = 1,2,\ldots M$ is a type of parameter;

$j = 1, 2, \ldots N$ is a number of reading. The fuzzy decision on belonging of these discrete readings to the nearest samples is described in /8/. In the process of work of fuzzy decision algorithm we have the inverse transformation from the fuzzy set of discrete readings of parameters representing of speech signal to an usual subset of hypothesis on pronounced word.

Such transformation may be realized using decomposition theorem /9/, which from the chain of usual subsets $A_0 \subset A_1 \subset A_2 \subset \ldots \subset A_{n-1} \subset A_n$ of acoustically similar families of segments of words with phonetic labels defined by parametric matrices give us a fuzzy subset of hypothesis on pronounced word (here $A_0 \to a_1^{(0)} \ldots a_{N_0}^{(0)}$, $A_n \to a_{1n}^{(n)} \ldots \ldots a_{N_n}^{(n)}$ are phonetic transcription of acoustically similar words; $a_i^{(j)}$ is a phonetic label of i-th segment and j-th word; n is the word number; N_n is length of phonetic transcription of n-th word. The inclusion relation of acoustically similar words is defined by relative embedding coefficient.

DEFINITION 1. By a relative inclusion coefficient K_e from A_0 in A_e is meant the maximal number of coincidence of phonemes in transcription arranged on the order of appearence: $K_e = K(A_0, A_e) = max(m) | \exists a_{i_m}^{(e)} = a_{j_k}^{(0)}, a_{i_2}^{(e)} = a_{j_2}^{(0)}, \ldots, a_{i_m}^{(e)} = a_{j_m}^{(0)}$.

$i_1 \leq i_2 \leq \ldots \leq i_m;\ j_1 \leq j_2 \leq \ldots \leq j_m$

DEFINITION 2. Let $A_{e_1} \subset A_{e_2}$ be meant that $K_{e_1} \geq K_{e_2}$.

The fuzzy subset $\underset{\sim}{A}$ of hypotheses on pronounced word defined as follows: $\underset{\sim}{A} = \underset{\alpha_i}{MAX}[\alpha_i \cdot A_{\alpha_1}, \alpha_2 \cdot A_{\alpha_2}, \ldots, \alpha_n \cdot A_{\alpha_n}]$ (here the meanings of α_i for A_i are such that $\alpha_1 > \alpha_2 > \ldots > \alpha_n$). Meanings $\alpha_1, \ldots, \alpha_n$ have significance of related between speech waveform and nearests samples.

The following transformation (from a fuzzy subsets of hypotheses on pronounced word to a precise belonging of introducted realization to a definite standard) may be realized, for example, with regard for syntax and semantics of languge.

On these grounds we may consider the problem of parametric reprezentation and subsequent analysis of speech recognition as a problem of subsequent transformations from fuzzy subsets to usual ones and vice versa and analysis of results on each step. Such transformations may be described using Galois relations for fuzzy sets /10/.

These theses were used when working out of acoustic-phonetic and phonologic processor and hardware-software speech recognition systems.

REFERENSE

/1/R. DE MORI, «Computer Models of Speech Using Fuzzy Algoritms», New York, Plenum Press, 1983.

/2/P. DEMICHELIS, R. DE MORI, P. LAFACE, M. O KANE, «Computer Recognition of Plosive Sounds Using Contextual Information», IEEE transactions on acoustics, speech and signal processing, vol. ASSP-31, p.p. 359-377, April 1983.

/3/R. DE MORI and G. GIORDANO, «A parser for segmenting continuous speech into pseudo

syllabic nucleir», Proc. IEEE-ICASSP, Denver, Colo., p.p. 876-879, 1980.

/4/Э.Б.Донбаев, В.Н.Трунин-Донской, Моделирование систем понимания киргизской речи на ЭВМ, Изд."ИЛИМ", Фрунзе, 1977.

/5/А.Ж.Чымбаев, Предварительная сегментация и маркировка слитной речи, М.ВЦ АН СССР, 1979.

/6/D.KLATT, K.STEVENS, «Sentens Recognition from Visial Examination of Spectrograms and Macineaided Lexical Searching», IEEE Proc. of 1972 Conference on Speech Communication and Processing.

/7/Н.В.Зиновьева, Особенности интерпретации "слепых" сонограмм при работе с временным окном, В кн."Тезисы докладов и сообщений 13-й Всесоюзной школы-семинара "Автоматическое распознавание слуховых образов (АРСО-13)", ч.2, Новосибирск, 1984, ИМ СО АН СССР, с.30-32.

/8/Г.В.Крюков, А.В.Михалев, В.Н.Трунин-Донской, Относительная глобальная проекция в автоматическом распознавании образов, В кн. VII Всесоюзная конференция "Проблемы теоретической кибернетики", Тезисы докладов, I часть, Иркутск, 1985.

/9/А.Кофман, Введение в теорию нечетких множеств, М.,"Радио и связь", 1982.

/10/A.ACHACHE, «Galois connection of fuzzy Subset», Fuzzy Sets and System, 8(1982), p.p. 215-218.