ESTIMATION OF LANGUAGE COMPLEXITY FOR THE HUMAN COMPUTER INTERECTION SYSTEMS.

Objelean Nicolae

State University of Moldova, Chisinau, Mateevici 60, Moldova, MD-2009
IMB University Bordeaux1, 351 Cours de la Liberation 33405 TALENCE CEDEX, Ftance
objelean@**yahoo.com**

Abstract: We describe methods for estimation the phonemic, words and lexical ambiguity for automatic speech recognition systems. This methods are use the phonemic structure of spoken words, size of dialogue vocabulary, probability of words occurrences of spoken language and allows to predict the accuracy of word recognition and direction to increasing the accuracy on the phase of building the system. Speech recognition system is considering as the informational channel with the noise for transferring information and we estimate the information lost in this channel. The lost information can be a measure for estimation the uncertainty or complexity of language recognition for dialog systems. These methods make it possible to use linguistic knowledge and phonetic structure of words to improve the reliability of speech man-computer dialogue.

I INTRODUCTION

For automatic speech recognition systems the spoken language should seek to ensure that all phonemes are classified correctly, so we have interest to recognize the full sequence of phonetic units that make up the statement. In speech recognition systems the sound signal is the main source of uncertainty. The parametric description of the speech wave represents more uncertainty. We will consider the uncertainty of the acoustic signal and provide an assessment measure of phonetic ambiguity. Using these measures give possibility to estimate the lexical and phrase uncertainty. Together they break down in the sequence of segments on the basis of method of sound. These features are added features of places creation, which vary continuously within segments and across their borders [1,2]. With a few discrete pieces, the sounds of speech - phonemes or segments are connected in such a way that the semantic units of speech (words) were a chain of phonemes.

The majority of automatic speech recognition systems [2,3] converts the voice signal into phoneme chain, which is then compared with expected phonemes in the speech sounds. The process of converting a speech signal into a sequence of phonemes includes finding futures, segmentation and labeling of segments. Let's describe the model of phonetic ambiguity, to assess the results of improper recognition of phonemes. Then we will use errors matrix recognition of phonemes and the phonetic structure of words when assessing lexical ambiguity.

Lexical ambiguity occurs when words recognition is incorrectly because have approximately the same phonetic structure. When these two words belong to the same dictionary, their precise classification is difficult, so they can be lexically uncertain. Such situation we must exclude in real systems, if the words of task allows. We will describe the criteria of the dictionary complexity in order to be able to evaluate the degree of complexity recognition words and ambiguity of dictionaries [4].

II THE INFORMATIONAL CANAL AS THE MODEL FOR ESTIMATE OF THE PHONEMIC AMBIGUITY.

We will consider the speech recognition as the transfer of voice data through the channel with the noise and estimate the information lost in the channel. The lost information can be a measure of uncertainty or recognition complexity of phonemes. Ideally, the set phonemes on the input channel and output phonetic units obtained after segmentation of speech must be the same, but the sequence of phonemes in the output must match the input sequence. If this condition is not met, then the information channel is lost, and depending on the amount of loss can be said about the greater or lesser uncertainty classification of phonemes. For practical evaluation of phonetic ambiguity we will use the system features [4] and the algorithm for segmentation of speech into seven types of segments: V - vowel, T - transitional, M - sonority, L - low, H - high, R - a bustling, P - silence. Then the algorithm for speech signal labeling is transforming each segment into phonetic symbol, using a priory statistical information about parameters. The accuracy of the Speech Recognition System (SRS) largely depend from the reliability marking segments. Since SRS is considering as a channel for transmission information, assume that there are ***R*** possible input symbol alphabet ***A*** and the ***S*** possible outputs in the alphabet ***B***. Thus the SRS is described through channel matrix. The channel is used to describe a speech recognition system, provided by a chain of phonemes, transforming not noisy output sequence of sounds in a sequence of "machine" phonemes containing errors, badge, insert, merge and replace the sounds. Let the element of the input phonetic alphabet ***(Ai)*** appear at the entrance with some a priory probability ***p (A1), p (A2),., p (Ar),*** while elements of the alphabet ***(Bj)*** at the outlet - with the probability ***p(B), p (B2),..., p (Bs).*** As noted earlier, the work channel of the input alphabet ***(Ai)*** is the channel matrix, so

$$P\{Bj\}=\sum_{i=1}^{r}P(Ai)*P(Bj/Ai) \quad (1)$$

The information ***I(Ai, Bj),*** obtained from the channel when the input received phoneme ***Ai***, while the output is recognized as ***Bj***, is defined [1]

$$I(Ai,Bj)=LOG(P(Ai/Bj)/P(Ai)) \qquad (2)$$

The average information obtained at the outlet channel, with losses in the transmission of (recognition) input alphabet phonemes $A =(Ai)$, which is recognized as the alphabet $B = (Bj)$, will

$$I(A, B)=\sum_{A,B}P(Ai,Bj)*I(Ai,Bj)=$$
$$\sum_{A,B}P(Ai,Bj)*LOG_2 (P(Ai/Bj)/P(Ai)=$$
$$-\sum_{A,B}P(Ai,Bj)*LOG_2 P(Ai)+$$
$$\sum_{A,B}P(Ai,Bj)*LOG_2 (P(Ai/Bj);$$
$$I(A, B)=H(A)+\sum_{A,B}P(Ai,Bj)*LOG_2 (P(Ai/Bj) \quad (3\ )$$

Note that $H(A)$ - entropy, which characterizes the uncertainty of input alphabet $A = (Ai)$. From (3) we find that $H (A) - I (A, B) =- \Sigma A, BP (Ai, Bj) * LOG_2 P (Ai / Bj) =$
$=- \Sigma A, BP (Ai, Bj) * P (Bj) LOG_2 P (Ai / Bj) =$
$- \Sigma BP (Bj) \Sigma AP (Ai / Bj) LOG_2 P (Ai / Bj) =$
$H (A / B)$ (4)
where $H (A / B)$ - posterior entropy of the input alphabet of phonemes, which characterizes the measure of information lost in the system of recognition in the transfer of the input alphabet (Ai). Posterior entropy is a measure for assessing the complexity of the input dictionary for the automatic recognition with a fixed parametric description. If there are values of the entropy of the input alphabet of phonemes, you can calculate the size (volume), equal to $2^{H(A)}$, a value of $2^{H|(A/B)}$ describe the average number of possible alternative (competitive) elements of the alphabet (Ai) at the entrance of the SRS, after at the exit were a lot of (Bj), which measure the complexity of recognition of the input alphabet of phonemes. We call this measure equivalent to the size of the alphabet sounds. The value $2^{H|(A/B)}$ are the entropy criterion for assessing phonetic ambiguity, which is a generalized characteristic of the complexity of recognition of the alphabet phonemes (Ai) of the recognition system. If the SRS is working without error, then the conditional entropy $H (A / B) = 0$, and the equivalent amount of the alphabet phonemes is $2^{H|(A/B)} =1$. If $H (A / B) = 0$, then $2^{H|(A/B)} =1$, and when the SRS is not recognized by $H (A / B) = H (A)$, the equivalent size of the alphabet phonemes is equal to $2^{H|(A)}$.

The equivalent amount of the alphabet phonemes provides an opportunity to quantify the average number of possible competitive phonemes (with similar parametric descriptions), and its definition need to know a posterior probability $P(Ai/Bj)$ of input alphabet.

In case of the specific problems vocabulary of automatic speech recognition system have limited sets of words and then all the variety of sounds can be reduced to two or three working phonemes units (for examples, grade long noisy, calls and sounds of Blade-Tongue Occlusive) that using a simple system of speech futures and simple algorithms which give zero a posteriori entropy. However, when solving the problem of recognizing relatively complex vocabularies and go demand a reliable verification of the spoken word phonetic such number of workers affected phonemes is not enough. Work as a full set of phonemes is false because of the mistakes of their automatic recognition. Therefore is needed to find a compromise solution - to seek some optimum in the phonetic description of the working word. The conditional probability of recognizing phonemes
$P (Ai / Bj)$, defining an equivalent amount of phonetic alphabet can be estimated by several ways.

The statistical method provides the probability of recognition of phonemes using real SRS. This is done by comparing the results of the exact manual segmentation and labeling of the speech signal (or his parametrical presentation), moving at the entrance of recognition. The result is a classical matrix correct and erroneous classification of the input alphabet of phonemes.

The acoustic-parametric method describes the matrix of classification errors phonemes produced by the direct comparison of their parametric descriptions. This standard set of phonemes selected from the realization of the phonemes. The distance between the phonemes is used for the evaluation of conditional probabilities of the erroneous classification of phonemes. The accuracy of this method depends on the standard and the volume of research material.

In addition to these methods, the probability of erroneous classification of phonemes can be made by **modeling of physiological human** voice tract [5].

III. ESTIMATING THE COMPLEXITY OF RECOGNIZING WORDS BY THEIR PHONETIC STRUCTURE.

We will consider the speech recognition system as the channel for transforming information. If we will use the input vocabulary $V = (V1, V2, .., Vr, .., VR)$ that can provide words $Vr = (Ai1, Ai2, .., Ain)$ and the sequence of phonetic symbols, then the output words dictionary on channel will the chains $W = (W1, W2, .., Ws, .., WS)$, $Ws = (Bj1, Bj2, ..., Bjr)$, where $Ai\in A$, $Bj\in B$ is respectively the input and output alphabet phonemes of channel; $r = 1, R; s = 1, S; n = n (r); l = l (s)$.

For estimate of the recognition complexity of words we should comparing the input to the implementation of supply chain quasi phonetic standards can be implemented based on the analysis of the errors matrix obtained in the presentation standards of words $Wsk \in Ws , K=1, Ks$ for each output word.

In fact, the difficulty of recognizing the input vocabulary V is defined by the presence of similar reference surface of the output dictionary Wsk and the frequency of occurrence $P(Wsk)$ of the surface words forms Wsk. The main problem in constructing the errors matrix for each dictionary Ws, is to get the word pronunciation of each word and to create standards surface forms Wsk words and to get quasi phonetic graph $f (Ws)$, that taking into account all surface words forms and probabilities of their occurrence. It is difficult to describe all the of surface quasi phonetic forms Ws, that take in account not only the surface forms of the word, giving to the peculiarities of pronunciation, but also forms, which include casual segment and methods labeled quasi phonetic labels.

In the future, we will consider the impact of two factors on the formation of the reference surface of the working vocabulary of words, given that the surface forms related to the peculiarities of pronunciation and a matrix of errors quasi phonetic classification can be built manually (or automatically, using the table of acoustic-phonological rules, stored in memory, and attached to the base quasi phonetic chain), and superficial forms ***Wsk,*** due to the peculiarities of the equipment selection of informative features can be obtained by analyzing the statistics of the implementation chain quasi phonetic working vocabulary words derived from the computer. Getting these statistics is not always necessary, especially if the phrase, contrasting in their acoustic properties. Preliminary assessment of the complexity of recognition of words can make a similar assessment of the complexity of phonetic alphabet - on the phonetic structure of words, calculating a posterior verbal ambiguity and not studying statistics implementation. All models words ***Wsk*** of working vocabularies should be submitted to a sequence of phonetic labels marked segments, where the quasi phonemes must be divisible by reference, binding of the word.

The automatic recognition of the choice of standards (standards of word) to be primarily due to the presence of input received on the implementation of strong, binding marked segments, taking into account that due to not perfect segmentation of the total number of segments of the input implementation might not coincide with the possible number of segments of the reference graph, at the expense of not supporting segments generated or falling accidentally.

The errors of classification give the appearance of "mess" of surface form for different vocabulary words [6]. Let us assume that the errors matrix of the recognition words, a priori, is formed in such a way that (with the surface similarities of various forms of dictionary words), more frequent surface forms of the words of one class be deemed to apply only to the words of this class, but rarely seen similar surface forms for other words the dictionary gives the error of recognition. However, we can use synonyms or semantic-syntactic restrictions for increasing the rate words recognition. Always make sure that such incidents do not occur (the difficulties are words belonging to the same a semantic-syntax group that can not be replaced by using synonyms, for example, the names of numbers). Thus, for a final decision on membership of the input word ***Vx*** to the class ***Ws*** must select two of the most likely candidate ***Ws1*** and ***Ws2***, which correspond to the probability ***P (Vx/Ws1)*** and ***P (Vx/Ws2),*** and to check whether the conditions are met:

$$P(Vx/Ws1)>\Delta s1;$$

$$P(Vx/Ws1)-P(Vx/Ws1))>\Delta s1s2,$$

where ***Δs1*** - threshold probability that the input corresponds to the realization of the word ***Ws1,***
Δs1s2 - threshold the difference of conditional probabilities of belonging input ***Vx*** implementation classes ***Ws1*** and ***Ws2***, in which a decision on the classification of ***Vx***.

For a given speech recognition system the value of thresholds ***Δs1***, ***Δs1s2*** should choose experimentally by using of phonetic signs, as well as the required system accuracy and the probability of failures recognition. If the selection of thresholds, raised requirements for the system of recognition can not be a more detailed analysis than the reference segments, go to attempt to improve the system of signs. In some cases, to meet the requirements specified in the system should be used synonymously. We consider a further how to assess the lexical ambiguity of dictionary ***V*** for language speech communication system. Just as the estimated uncertainty of the alphabet sounds, you can determine the complexity of recognition of the input vocabulary ***V***, consisting of the ***R*** word, and to calculate the equivalent amount of the input dictionary. It is necessary to obtain the probability ***P (Vr/Ws)*** describing words near areas of attributes $Vr\in V$, $Ws\in W$, $r=1\div R$, $s=1\div S$, which is presented as a sequence of phonetic units (phonetic transcription of words). Next, we evaluate the probability ***P (Vr / Ws).***

As already noted, on the basis of linguistic knowledge, patterns of words $Ws\in W$ are presented in the type of phonetic (or rather, quasi-phonetic) chains, the totality of which is described by a graph with a finite number of states, and each phoneme - signs of the method and place of formation. The word Ws corresponds to one or more paths (chains of surface forms) on the graph (number of trajectories depends on the method of pronunciation and features narration). Directed graph of ***f (Ws)*** represents all the phonemes of $Ws\in W$, which is surface form ***Wsk,*** and ***K = 1,2,3 .., Ks;*** each speech standard $Wsk\in Ws$ contains ***L = L (s, k)*** reference quasi phonemes. It should be noted that the number of reference segments in the surface forms of output words depend of vocabulary and is limit, it changes in the index ***L*** depends on the number of words, and on its surface form ***L = L (s, k).***

In order to assess the incorrect classification of words in the dictionary stage of lexical recognition of phonetic structure of words, the operation of splitting all the surface forms of words in the standards of ***M*** phonetic groups with the same amount of reference segments of ***L = L (s).*** In this speech, the surface forms which belong to different groups, will not mess with each other because they are easily classified by the number of anchor phonemes that make up words. In generally it is possible to imagine a group of reference phonetic surface forms that differ not only in the number of reference phonemes, but also their character, as well as the procedure is followed. If we take into account all three factors to break the standards to significantly larger number of phonetic groups, another methods pronunciation of words can be attributed to each of these groups. For simplicity, however, we assume that we have ***M*** phonetic groups, each with the same number of reference segments. In practical tasks when words is to break into the group we should take into account all these factors, however, it is necessary to strictly limited the number of reference segments, choosing only those that are not confused with each other and are characterized by the place of group formation signs - drums vowels, occlusive, and fricative phonemes [4].

So, let us assume that there are M groups of phonetic words ***W1, W2, W3, ..., Wn, ..., Wm***, each with an equal number supporting a quasi phonemes, then the total number of benchmarks $W= U_{n=1}^{m} Wm$, and the number of phonemes that comprise: a word (phonetic chain length) of each group, denote by ***Lm; m = 1, M***.

The erroneous classification phonemes of words dictionary is presenting as matrix:

$$P(a/b)=[Pij], \qquad (5)$$

Then can estimate the probability confusing recognition ***Pm (Vr/Ws)*** of surface forms of words within each group of words as follows:

$$\mathbf{Pm\ (Vr/Ws)}=\Pi \prod_{T=1}^{Tm} P(Art/Bst) \quad (6)$$

where ***T = 1,2, .., Tm*** phonetic length of the chain words group ***Wm , Art ∈ Vr , Bst ∈ Ws.*** . In general, the same word ***Ws*** can have ***Ks,*** surface form, with varying number of phonetic elements and are in different groups of words ***Wm***. Therefore, the overall conditional probability erroneous words classification of the dictionary define as:

$$P(Vr/Ws)=\sum_{N} P(Wsk)Pm(Vr/Ws) \quad (7)$$

To determine the loss of information in the SRS, which is seen as a channel of information, in the case of recognition of words using the expression

$$\mathbf{I(V/W)}=-\sum_{N} P(Ws)\sum_{V} P(Vr/Ws)LOG_{2}P(Vr/Ws) \quad \mathbf{(8)}$$

Then $2^{I(V/W)}$ defines as the equivalent amount of vocabulary - the number of alternative input words of the system of recognition, and $2^{I(V)}$- is the estimate value or volume of the dictionary, where:

$$\mathbf{I(V)}=-\sum_{\kappa=1}^{K} P(Vr)LOG_{2}P(Vr) \quad (9)$$

These expressions are evaluating the lexical ambiguity and define the recognition complexity of vocabulary.

IV CONCLUSION

When the automatic marking, along with errors of incorrect classification of phonemes, there are, the error of incorrect segmentation, leading to a merger of sections of the related phonemes in one segment, or dismember a segment corresponding to single phonemes, in several different classes of related phonemes. When choosing alternative vocabulary words should be taken to ensure that this kind of trouble did not cause the similarity of sequences of phonetic units that correspond to different words. To do this we must use the matrix, reflecting a possible segmentation of the dictionary words and the frequency of occurrence of different options for segmentation of the differences of surface forms of words. Since the information about the words contained in the phoneme, the excess, it is often in the evaluation of legibility of the words the dictionary is quite enough to use the basic phonemes, allowing a minimum of errors and the dismemberment of the merger. Therefore for approximation of estimation rates recognition words must to use the probability of erroneous phonemes recognition, which in speech signals does not give the error of merging and dismemberment.

REFERENCES.

[1] Fano R. Transmission Information. *The statistical Theory of Communication,* M. Mir, 1969.

[2] Speech Recognition. *The future now!* Prentice Hall PTr., 1997.

[3] Newell A. at all . *Speech understanding System,* Final Rapport. Amsterdam, 1973.

[4] Trunin-Donskoi V. N. *Speech dialogue in man-computer systems.* Chisinau, Stiinta, 1986

[5] The CMU Sphinx Group. *Open Source Speech Recognition Engines*. www.cs.cmu.edu.

[6] Method for error correction in strings with applications in speech recognition. .P*apers for proceedings of the 27ARA Congress. Oradea, Romania. 233-236. 2003*