A SYLLABLE BASED APPROACH TOWARDS THE INTELLIGIBILITY STUDY OF IMPAIRED SPEECH

Vasilios Keramaris
Queens College
City University of New York
Department of Computer Science
95-30 Kissena Boulevard
Flushing, NY 11367-1597
(718) 997-3566

Abstract

This paper reports the development and the preliminary results of a syllable-based research on the intelligibility study of impaired speech. In the development of this research, a syllable-based selection rule is introduced for an attempt at "binding" a set of "optimal" syllables to the set of words derived from it through the intelligibility scores. A rigorous probabilistic measure is developed to compute the intelligibility of dystarhic speech according to the result of a listening test with multiple participants. A preliminary experimental study based on a test bed of 229 syllables and 103 words derived from a vocabulary base of 3,500 words is reported.

I. Introduction

The ultimate goal of this research is to develop a causal model (Fig. 1) from which we can infer the intelligibility and the computer recognition rate of any given word based on a small set of "optimal" syllables. This paper is focused on one part of our research towards the ultimate goal — the development and testing of an experimental protocol for the assessment and prediction of the quality of dystarhic speech without exhaustively evaluating, every frequently-used word of an individual. Specifically, a syllable-based selection rule to bind a set of "optimal" syllables to a set of words, and a rigorous probabilistic measure for computing the intelligibilities of dystarhic speech are introduced. Then the experimental base and procedure for the preliminary study are detailed, followed by the analysis of the results. In the conclusion the future research work is discussed.

II. Theoretical Development

Each English word in this research is considered as a concatenation of syllables. Since many multi syllabic words share some common syllables (e.g., "tion" in "dictation", "deduction", "computation" etc.), we can represent a lot of multi syllabic words using only a small number of common syllables. By doing so, the correlation between the word intelligibility and the word recognition performance can be considered through the common parameter — the syllable intelligibility. That is: (i) the correlation between the intelligibilities of a set of test words and the intelligibilities of a corresponding set of common syllables, and (ii) the correlation between the computer recognition rate and the intelligibilities of the same set of common syllables. In this research, our focus is to develop the experimental protocol for the derivation of the correlation stated in (i).

In the development of the experimental protocol, there are three major design issues: (i) the extraction of a set of "optimal" common syllables, (ii) the compilation of a set of test words, and (iii) the figure of merit for the intelligibility test. For the extraction of a set of common syllables, we developed a syllable-clustering algorithm akin to the exchange method of K-mean clustering. This algorithm guarantees that the extracted common syllables are at least local optimum; where the optimization criterion is the number of words that can be generated out of the set of common syllables. Further details can be found in [2]. For the compilation of the set of test words, a syllable-based selection rule to relate the syllables and words is derived and can be formulated as follows:

Let L be a set of common syllables and W be a word represented as a set of syllables: {S1 S2 ... Sn}, the syllable-based selection rule is defined as a relation to bind L and W together IFF the following condition is satisfied:

\[ \frac{|O(W_i) - O(W_j)|}{|W|} \leq \frac{1}{2} \]

Remark: |X| is the number of elements in the set X.

To cope with the problem that many syllables may sound exactly the same (e.g., "tinn" as is in "dictation" and "sion" in "pension"), we developed the following probabilistic measure as the figure of merit for the Intelligibility study:

\[ \text{Intelligibility}(W_i) = \sum_j \text{C}(W_i|W_j) \text{P}(W_j|O(W_i)) \]

where \( O(W_i) = \text{listening test result of } W_i \), \( \text{P}(W_j|O(W_i)) = \text{percentage of } W_i \text{ being classified by the listeners as } W_j \text{ given the listening test result of } W_i \), and \( \text{C}(W_i|W_j) = \text{Cost function of classifying } W_i \text{ into } W_j \).

It can be easily seen that intelligibility \( (W_i) \) ranges from 0 to 1, and \( C(W_i|W_j) \) is the weight given to the class of listening test result that \( W_i \) is classified as \( W_j \). When we define \( C(W_i|W_j) = 0 \) for \( i \neq j \) and \( C(W_i|W_j) = 1 \) for \( i = j \), intelligibility \( (W_i) \) becomes a percentage measurement.

III. Experimental base and procedure

In this research, 6500 commonly-used words are selected to serve as the vocabulary base to compile the set of syllables and test words for the experiment. All syllables are first clustered according to their occurrence position in a word. Then all position-dependent syllables are divided into two groups, one containing those with frequency 10 or higher, and the other the rest. A local optimal set of 229 syllables is found and this set constitutes the syllable list of the experiment for the first stage.
The subject of the experiment is a 30+ years old male, with Cerebral Palsy, who lost most of the motor control of his right hand. In preparing the dysarthric speech for the syllable listening lest, the following procedure is adopted: the subject is asked to utter each syllable, in the entire (syllable) list twice. This procedure is repeated so that we get a total of 4 utterances for each syllable. One out of the four utterances is randomly selected for the listening test in order to avoid subjective selection of particularly good or bad dysarthric speech, which might bias the result. The syllables are then randomized for a context-based intelligibility.

In the listening test session, each syllable in the randomized set is played twice to 10 normal listeners (*). Each listener then has to choose the best matching syllable out of the given list of 229 syllables, with the added option — "syllable not in list".

Finally, the intelligibility of each syllable, with the anticipation of two syllables, with identical phonetic transcription such as "TE" or "TY", is computed based on the probabilistic measure discussed in II. All the syllables are grouped into 11 ranges. Range 1 is assigned for 100% recognition, and the rest ten are the classes with 10% interval, going from 100 down to 0.

Then the syllable-based selection rule is applied to generate the set of test words for the second phase of the experiment. This way we come up with 103 words, out of the total 11 ranges. To prepare the data for the word listening lest, the subject speaks all the words in a pre-randomized order once, with a time interval of 2 sec. Then, we-perform the intelligibility study of the words, following the same procedure as is in the syllable-phase. The intelligibility of the words are computed, and the words are divided accordingly into 11 ranges, as the same way the syllable listening test does.

IV. Results
The experimental results are studied in terms of a margin of acceptance — defined as the rate difference between the ranges assigned and the ranges classified. In our study the margin of acceptance is set to 3. For example a word from range 3, is considered acceptable if it is classified within the ranges 1 to 6. This margin seems quite big, but it has to be so, because the number of listeners is only 10. This means that a random guess made by one listener can contribute as much as 10% of error on one range difference! With acceptance margin of 3, the results are: 48 classified correctly and 55 incorrectly. This, looks discouraging at a first look, but with respect to ranges much can be said. Out of the 54 incorrect, only 4 came from ranges 1 - 6. And, from these incorrect four, three started with the same syllable "SUB", which tells us that "SUB" is assigned to an incorrect range on the syllable test. Therefore, the results were conclusive towards the high intelligibility ranges, specifically, ranges 1 - 6. In other words, whenever we get syllables that were well understood by the listeners, the majority of the corresponding words are also well understood. On the other hand, the listening test results of the 50 words, which are originated from a total of 75 words in the ranges 7 - 11, do not fall within the margin of acceptance. As a result, no conclusions could be drawn at these ranges.

After a careful study we found that these extreme errors, were mainly caused by three factors:
1. The 10 % range division of this experiment is too sensitive to the errors due to the random guessing. This problem can be alleviated by either increasing the number of listeners with a future research experiment, while widening the percentage of the ranges.
2. C(Wi|Wj), is not optimal yet even though it has already taken the phonetic similarity into consideration. The cost function needs to be further optimized to reduce the misclassifications due to the similar phonetic transcription of two syllables.
3. Consideration has to be given on how syllables should be presented to the listeners on the intelligibility test. (Problem of too many similar syllables on the list).

V. Conclusion
In this paper, a syllable-based experimental protocol towards the intelligibility study of impaired speech is discussed and the result of the preliminary study is reported. Our future work will focus on two major issues. One is the refinement of the cost faction and the range for the classification such that the experimental result will not be sensitive to a random guess made by the listeners. Second is the development of the experimental protocol for the evaluation of the speech recognition performance to complete the causal model shown in Fig. 1.

Acknowledgment: The author is indebted to our human subject, to Dr. Bon K. Sy for working along on all aspects of this research and on providing valuable help on writing this article, the listeners for their participation in the experiment of this research, the students who prepared the data entry for the experiment, and David Horowitz of New England Medical Center for the valuable comments of the early draft of this paper.

References

Fig. 1. CAUSAL MODEL.