LIKESNESS FUNCTIONS OF THE ACOUSTIC PATTERNS AS AN INDEX FOR OBJECTIVE ESTIMATION OF SPEECH TRANSMISSION QUALITY

Czeslaw Basztura
Institute of Telecommunication and Acoustics Technical University of Wroclaw ul.B.Pruna 53,55
50-317 Wroclaw POLAND

ABSTRACT

The purpose of the paper is a presentation of a new objective measure for estimation of speech transmission quality and to perform a preliminary evaluation of conformity between the results obtained by means of the proposed method and subjectively measured speech intelligibility. The new method uses likeness functions of the acoustic patterns as an index for the evaluation of speech transmission quality. Eight likeness functions as distance or proximity were examined, i.e. Hamming, Euclidean, Minkowski, Chebyshev, Camber, Chi-square, Tanimoto and directional cos were examined.

The likeness function have a form:

\[ d^{m}(X,Y) = \max \left( \frac{|X^p - Y^p|}{|X^p + Y^p|} \right) \]

where: \( p=1,2,... \) \( p \) - dimensionality of vector parameters from speech signal.

The likeness function have a form:

\[ d^{m}(X,Y) = \max \left( \frac{|X^p - Y^p|}{|X^p + Y^p|} \right) \]

was calculated for the telephone channel as an investigation object.

INTRODUCTION

Speech intelligibility as a measure of speech transmission quality may be classified as either subjective or objective. The subjective measurement is a procedure for determining the communication channel's intelligibility using a predetermined vocabulary and selected speakers and listeners panel. Subjective measurement techniques generally attempt to determine intelligibility for an information presented in one of the following forms:

a) nonsense syllables (logatoms) list
b) limited list of words
c) list of sentences

In subjective measurement methods the intelligibility is determined by the ability of the listeners to identify system (or recorded) syllables, logatoms, words or sentences. A number of subjective methods have been devised with the desirable results. However, the requirements for listeners panels greatly restricts the utility of these methods, and a long-sought goal is to replace the subjective scoring with objective measurement. An objective measure for the fidelity of a speech communication system is a measure that is computed from data which contain no human subjective response.

There is a hypothesis that it is possible to design a relatively compact objective measures which are in a good correlation with subjective results over a subset of distortions and disturbances introduced by speech transmission channels (12.7.8.). Over the years some number of papers contained informations about evaluation of speech transmission systems by objective measures [13.7.8.].

The task of comparing and contrasting the validity of such measures is immense. To verify the likeness functions which are in a good correlation with subjective results over a subset of distortions and disturbances introduced by speech transmission channels (12.7.8.).

The likeness function have a form:

\[ d^{m}(X,Y) = \max \left( \frac{|X^p - Y^p|}{|X^p + Y^p|} \right) \]

between objective and subjective quality measure.

a) The typical telephone channels, represented by an adjustable model of the telephone system under investigation, are used.

b) The criterion reference for objective measure is subjectively measured logatoms speech intelligibility.

c) The measurement conditions for both subjective and objective procedures are the same.

d) As a test signal for objective measurement three following Polish key phrases were used:

1. ALO (part of word “Hello”)
2. JUTRO BEDZIE LADNY (Tomorrow will be a fine day)
3. PRZEPIS ZE JESIEŃ NASZEGO POLSKIEGO KOSZAROWISKOWA POLSKIEGO KOSZAROWISKA ZIMOWEGO KOSZAROWISKA ZIMOWEGO KOSZAROWISKA

EXPERIMENT AND RESULTS

First step in the experiment is a choice of an adequate test signal. Second problem relies on finding an effective set of parameters representing the test signal and presence of distortions and disturbances in these signals.

Fig. 1 The computation of a relationship between objective and subjective quality measure.

- Hamming distance:

\[ d_{\text{Hamming}}^{m}(X,Y) = \frac{1}{p} \sum_{p=1}^{P} \left( \frac{|X^p - Y^p|}{|X^p + Y^p|} \right)^2 \]

- Camber distance:

\[ d_{\text{Camber}}^{m}(X,Y) = \frac{1}{p} \sum_{p=1}^{P} \left( \frac{|X^p - Y^p|}{|X^p + Y^p|} \right)^3 \]

- Chi-square distance:

\[ d_{\text{Chi}}^{m}(X,Y) = \frac{1}{p} \sum_{p=1}^{P} \left( \frac{|X^p - Y^p|}{|X^p + Y^p|} \right)^4 \]

- Directional cos proximity:

\[ d_{\text{cos}}^{m}(X,Y) = \left( \frac{X^p \cdot Y^p}{|X^p| \cdot |Y^p|} \right)^2 \]

- Tanimoto proximity:

\[ d_{\text{Tanimoto}}^{m}(X,Y) = \frac{X^p \cdot Y^p}{2 \cdot |X^p| \cdot |Y^p|} \]

- Likinson proximity:

\[ d_{\text{Likinson}}^{m}(X,Y) = \frac{X^p \cdot Y^p}{|X^p| \cdot |Y^p| + \epsilon} \]

EXPERIMENT AND RESULTS

First step in the experiment is a choice of an adequate test signal. Second problem relies on finding an effective set of parameters representing the test signal and presence of distortions and disturbances in these signals.
Next problems depend on the assumed method (Compare "METHOD").

The analysis of the previous investigations [3, 4] shows a fairly large effectiveness of representing the voice and speech features by parameter set with being the distribution of the time intervals between the zero-crossing of a speech signal (Fig. 2).

An especial computer program counted likeness functions (as objective measures) and made statistical correlation analysis of figure of merit L1-function of LF.

Some of the results are shown in Fig. 3.

Comparison of the results of the experiments on the 40 different telephone channels implemented by a physical model of telephone channel permits the following observations and conclusions to be given:

a) Hamming distance gives a good correlation L1 with LF for all of the test signals (especially for 3 degrees of polynomial regression).

For example:

\[ L_1 = 92.91 - 0.0421 d^{10} + 0.151 \times 10^7 (d^{10})^2 - 0.631 \times 10^{10} (d^{10})^3 \]

b) Euclidean and Minkowski (r=3) distances (likewise the Hamming distance) give a good correlation for third key phrase. For example:

\[ L_1 = 90.2 - 0.0184 d^{12} + 0.364 \times 10^5 (d^{12})^2 - 0.24 \times 10^9 (d^{12})^3 \]

c) Chebyshev distance gives a good correlation (especially for third key phrase and for higher degree of polynomial regression).

d) Camber distance and Tanimoto and directional cos proximity did not give satisfactory results.

The future investigations will concentrate mainly on the selection of test signals and other parameters of speech signal.

LITERATURE


