Time-sequential Speaker Recognition System

Syroka Zenon[1,*] , Tomasz Zając[2], Paweł Dubiłowicz[2], Cezary Łabarzewski[2]

[1]Department of Technical Science, University of Warmia & Mazury, Poland
[2]Faculty of Mathematics and Information Technology, Oczapowskiego 11, 10-719 Olsztyn, Poland

Abstract The method presented in this publication was submitted in a patent application in May 2011, the patent was granted on December 17, 2014. It is a continuation of speech signal recognition study [1] carried out at the UWM. In 2010 a "Speech analysis and recognition method and system" [2] has been developed using sequential analysis basing on the Walsh basis functions. The present system includes time-variability of the speech signal spectrum. Basic block diagrams describing the system's mechanics have been presented.

Keywords Time-sequency Analysis, Speaker Recognition System, Walsh's Sequential Function, Wavelet Transforms, Gabor Transforms, STFT, Wigner-Ville Transforms

1. Introduction

The present paper shows a speaker recognition system based on application of join time-sequential speech signal analysis. Speech signal, after being subjected to time-sequential transformation, is stored in system memory as a template. Denoised and sequentially transformed speaker's signal is compared with the template. The comparison bases on the transforms' correlation analysis. The result of comparison with the fidelity coefficient (taking probability of error into account) is the base for a decision-making system, which resolves speaker's veracity. The foundation of the decision-making system is elimination of false positives.

Sequency can be treated as generalized frequency. In speech signals' time-frequency analysis the basic functions used to "build" transforms are trigonometric functions. These functions are periodic and zero-crossings occur in equal time intervals. In this case, the term sequency is equivalent to the term frequency. In patent application P.390450 [2] submitted on 02.19.2010, titled "Speech analysis and recognition method and system" a method of speech recognition using Walsh's sequential function basis has been shown. It didn't take speech signal's spectrum's time-variance into account.

Considering the spectrum's time-variance the system becomes more complex and the analysis time becomes longer, yet consideration of joint time-sequential features yields more precise speaker recognition results.

2. Structure and Mechanics of Time-sequential Speaker Recognition System

The time-sequential speaker recognition system has been portrayed in fig.1. It is composed of a microphone connected to a recorder (1). This device registers speech signal including ambient noise. The noise itself is also registered by the microphone and the ambient noise recorder (2). The aforementioned devices are connected to a time-sequential analysis and noise reduction system (3).

For noise elimination wavelet transforms, Gabor transforms, STFT and Wigner-Ville transforms are used. The system (3) also conducts time-sequential signal analysis.

The system's (3) first output is connected to the comparison module (4), while the second output is connected to the memory module (6). The result of speaker's signal (including the background noise) analysis obtained in the system (3) is sent to the comparison module (4). The result of speaker's signal's template analysis is sent to the memory module (6). The results take the form of time-sequential transform coefficient matrices.

The comparison module (4) is connected to the time-sequential analysis and noise reduction system (3) and the memory module (6). Speaker's spectrum template, registered in the memory (6) and denoised in the analysis system (3) is sent to this component in the form of a time-sequential transform coefficient matrix. The result of comparing the two matrices is sent to a decision-making system (5), which determines if the signal registered in the memory module (6) and the signal registered by the device (1) come from the same speaker.

The schematic of functional block connections has been shown in fig.1.

The analysis and noise reduction system (3) has been shown in more detail in fig.2. It is composed of 3 functional blocks: time-sequential signal and ambient noise analyzer (7),

noise reduction system (8), time-sequential ambient noise analyzer (9).
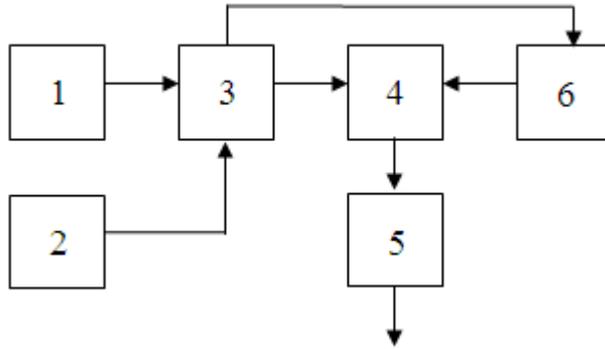
1.   - speech signal (including ambient noise) recorder;
2.   - ambient noise recorder;
3.   - time-sequential analysis and noise reduction system;
4.   - comparison module;
5.   - decision-making system;
6.   - memory module.

Figure 1. Time-sequential speaker recognition system' functional schematic.

The noise recorder's (2) output is connected to the system (9), in which the ambient noise gets analyzed. Noise coefficients (ambient noise's time-sequential transform coefficients) are calculated in the analyzer (9). The speech signal and ambient noise recorder (1) is connected to the analyzer (7). The signal along with ambient noise registered by the recorder (1) is subjected to time-sequential analysis in the analyzer (7), as a result speech signal's and ambient noise's time-sequential transform coefficients are calculated.

It is assumed, that the noise structure is the same with or without the speech signal. Otherwise, noise's time-sequential transform coefficients need to be corrected. The analyzers (7 and 9) are connected to the noise reduction system (8). In this module, transform coefficients representing noise are removed from the speech signal and ambient noise spectrum. The system's (8) first output is connected to the memory module (6) and the second output is connected to the comparison module (4). In the memory module (6) speaker's speech signal's template is stored.

The schematic of functional block connections has been shown in fig.2.

Comparison module (4) has been shown in more detail in fig.3. It consists of 4 functional blocks: speech signal's spectrum and speaker's signal's template's spectrum correlator (10), speaker's signal's template's spectrum autocorrelator (11), correlator's and autocorrelator's result comparison system (12), fidelity coefficient bank (13).

Memory module (6) and noise reduction system (8) are connected to the correlator (10). The correlator (10) calculates correlation between speaker's signal's template's spectrum stored in the memory module (6) and speech signal's spectrum sent from the noise reduction system (8), the result is saved in matrix format. The autocorrelator (11) is connected to the memory module (6) and the comparison

system (12). The autocorrelator (11) conducts speaker's signal's template's autocorrelation and the result is saved in matrix format. The comparison system (12) is connected to the correlator (10), the autocorrelator (11) and the fidelity coefficient bank (13). The results sent from the correlator (10) and the autocorrelator (11) are compared in the comparison system (12). Both matrices' elements are compared one-by-one. They may vary by a percentage of a predefined fidelity coefficient, sent from the bank (13). Comparison system's (12) output is connected to the decision-making system (5), where the speaker's veracity is determined.

The schematic of functional block connections has been shown in fig.3.
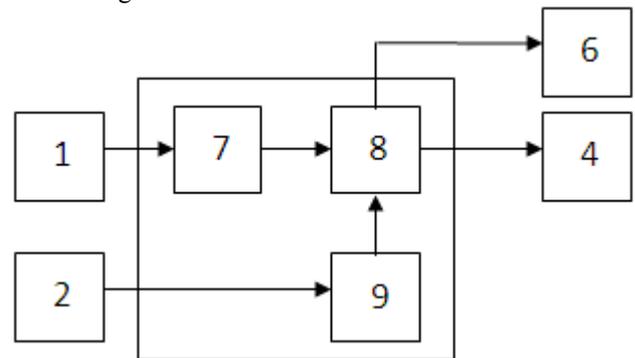
(1) - speech signal (including ambient noise) recorder;
(2) - ambient noise recorder;
(4) - comparison module;
(6) - memory module;
(7)- time-sequential signal and ambient noise analyzer;
(8)- noise reduction system;
(9)- time-sequential ambient noise analyzer.

Figure 2. Analysis and noise reduction system's functional schematic (block 3 from fig.1.)
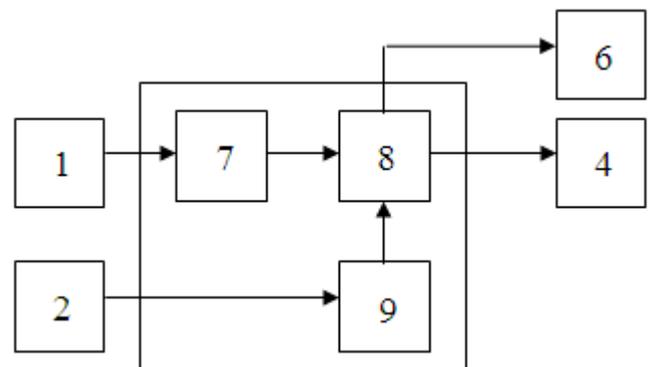
(3) - time-sequential analysis and noise reduction system;
(5) - decision-making system;
(6) - memory module;
(10) - speech signal's spectrum and speaker's signal's template's spectrum correlator;
(11) - speaker's signal's template's spectrum autocorrelator;
(12) - correlator's and autocorrelator's result comparison system;
(13) - fidelity coefficient bank.

Figure 3. Comparison module functional schematic (block 4 from fig.1.).

3. Conclusions

1. Speaker recognition does not require precise knowledge of frequencies present in speech signal; it only requires precise knowledge of signal's expansion coefficients in a suitable orthogonal basis.

2. The basis should be selected with regard to avoiding complex number calculations. The authors prefer Walsh and Rademacher functions.

3. Analysis using Walsh functions is much faster than when using classic trigonometric polynomials. Fast Walsh Transform is faster than Fast Fourier Transform when applied in computing "windowed" transforms, such as STFT.

4. Using wavelet transforms, Gabor transform and Wigner-Ville transform allowed taking speech signal's spectrum's time-variance into account.

REFERENCES

[1] Z. Syroka, T. Zając, P. Dubiłowicz, C. Łabaszewski *Czasowo sekwencyjny system rozpoznawania mówców* (Time-sequential speaker recognition system), patent P.395038 granted on December 17, 2014.

[2] Z. Syroka, D. Słomian, K. Kusyk (2010) *Sposób analizy i rozpoznawania mowy oraz układ do analizy i rozpoznawania mowy* (The method of analysis and speech recognition, and system analysis and speech recognition), patent application P.390450, 02.19.2010

[3] H. Harmuth, Sequency (1977) *Theory, Foundations and Applications,* Academic Press, New York

[4] H. Harmuth (1972) *Transmission of Information by Orthogonal Functions,* Springer-Verlag, Berlin

[5] P. Flandrin (1998) *Time-Frequency / Time-Scale Analysis*, Academic Press, San Diego

[6] V. Chen, H. Ling (2002) *Time-Frequency Transform for Radar Imaging and Signal Analysis,* Artech House, London.

[7] D. Elliott (1982) *Fast Transform, Algorithms, Analyses, Applications,* Academic press, New York

[8] H. Nussbaumer (1982) *Fast Fourier Transform and Convolution Algorithns,* Springer-Verlag, Berlin

[9] R. Carmona, Wen-Liang Hwang, B. Torresani (1998) *Practical Time-Frequency Analysis,* Academic Press, San Diego.

[10] S. Furui, M. Sandhi (1992) *Advances in Speech Signal Processing,* Marcel Dekker. Inc. New York.