International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:14, No:8, 2020

Minimum Data of a Speech Signal as Special Indicators of Identification in Phonoscopy

Nazaket Gazieva

***Abstract*—**Voice biometric data associated with physiological, psychological and other factors are widely used in forensic phonoscopy. There are various methods for identifying and verifying a person by voice. This article explores the minimum speech signal data as individual parameters of a speech signal. Monozygotic twins are believed to be genetically identical. Using the minimum data of the speech signal, we came to the conclusion that the voice imprint of monozygotic twins is individual. According to the conclusion of the experiment, we can conclude that the minimum indicators of the speech signal are more stable and reliable for phonoscopic examinations.

***Keywords*—**Biometric voice prints, fundamental frequency, phonogram, speech signal, temporal characteristics.

I. Introduction

THE individual parameters of the speech signal are an important argument for phonoscopic analysis. The acoustic properties of a speech signal allow us to identify a person, since it is associated with individual physiological properties [1]. The individual parameters of the speech signal are mainly controlled by spectral and formant properties. Along with parameters such as indicators of the formant properties and the pitch frequency, indicators of the intensity of instant sound are used; word duration, rhythm, pause duration and other temporal data are considered. Temporary manifestations are associated with physiological and psychological properties and coincide with signs of perception. Therefore, these data are very important for verification or identification of the speaker and are used in phonoscopy [2]-[8]. Experimental studies conducted to identify new useful parameters - biometric voice data can accurately identify a person by voice in forensic phonoscopy. The purpose of this study is to identify such data.

A. Our Approach

For the effectiveness and reliability of the experimental results, along with the speech materials of individual speakers, we also used phonograms of the voice of monozygotic twins. For this, we chose three pairs of twins. Our hypothesis is that we will observe different data both with ordinary speakers and with monozygotic twins. According to experiments [9]-[11], there are more similarities in perception between the voices of identical twins than between voices without a genetic connection. However, identification is not perfect. Since the minimum data - in this case, in particular, the average period indicators in the stationary part of the sound - are serious signs of individual signs of a speech signal, we decided to investigate the presence of stable biometric data in the speech of monozygotic twins.

Nazaket Gazieva is with the Azerbaijan National Academy of Sciences, Azerbaijan (e-mail: n.qazi@inbox.ru).

II. Materials and Method

For this experiment, we recruited 12 people (6 people 30-40 years old and 3 pairs of monozygotic twins - 2 women and 4 men, age: average ± standard; 28 years ± 2.3 years). We first recorded synchronized speech, and then asked to read the same text. They were not familiar with the text. During the experiment, we compared the phonograms of different speakers, as well as the data of the same speaker. During the experiment, the statistical method, mathematical operations, and some formulas of probability theory were used.

A. Acoustic Analysis

First Experiment

At the first stage, we analyzed the phonograms and performed a spectral analysis. All acoustic analyzes were conducted using PRAAT. Average indicators of the period of the main tone, average frequency of the main tone, standard deviation, rate of speech, sonority factor, relative frequency range of the main tone for the selected segments of the speech signal were revealed. And also using mathematical methods, we calculated the ratio of formant indicators (for example, for the same person F2/F1 = 2.2 gave an approximate result - F2/ F1 = 2.4; or F3/F2 = 2.0 and F3/F2 = 1.7). This once again proves the importance and informational content of the minimum data in a speech signal, including wavelet analysis for signal processing - to collect basic information about the signal in a small number of wavelet coefficients [12]-[15].

Based on the purpose of the experiment, we selected the minimum length from the selected materials in order to exhaust useful information for identifying a person. Three periods were distinguished in the stationary part of the sound (vowel). And for ordinary speakers, and for monozygotic twins, the average period indicator according to temporal data turned out to be different and peculiar. It took one speaker for the implementation of three periods (in the same sound in the same phonetic position) it took 0.014 seconds (average value of 0, 004 sec.), when for another it was 0, 025 sec. (average 0,008). In addition, during the experiment it turned out that for the speech of each person in a certain position for certain sounds, the average time indicators remain the same and are an indicator of individuality. In the speech of the first speaker, when pronouncing the sound in the same position, it was 0.014 sec., in another phonogram, 0.017 sec. For another

speaker, this was 0.025 seconds and 0.020 seconds, respectively.

A similar picture is observed in the materials of monozygotic twins.

Fig. 1 Data k1 and k2 reflect the speech parameters of different speakers. Data K3 and K4 refer to the same person

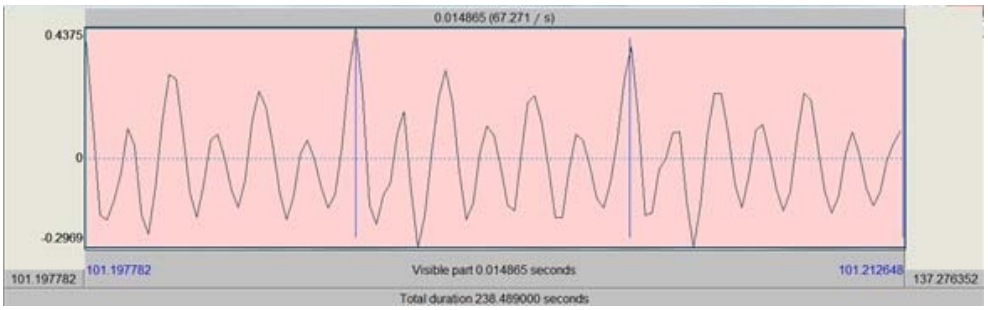

Fig. 2 First speaker

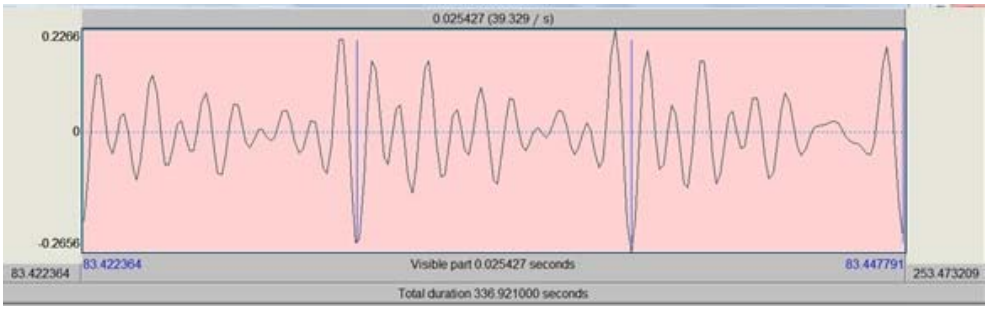

Fig. 3 Second speaker

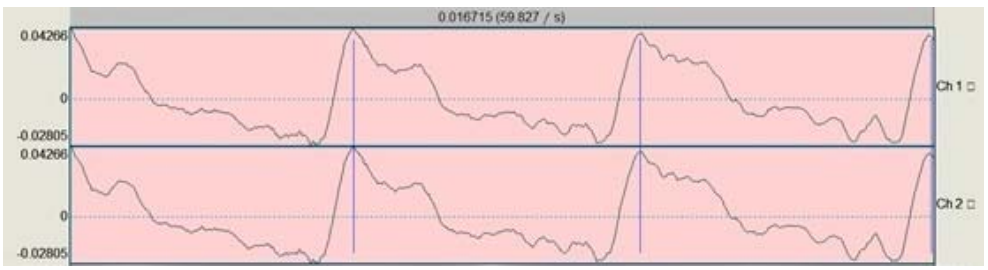

Fig. 4 The first monozygotic twin

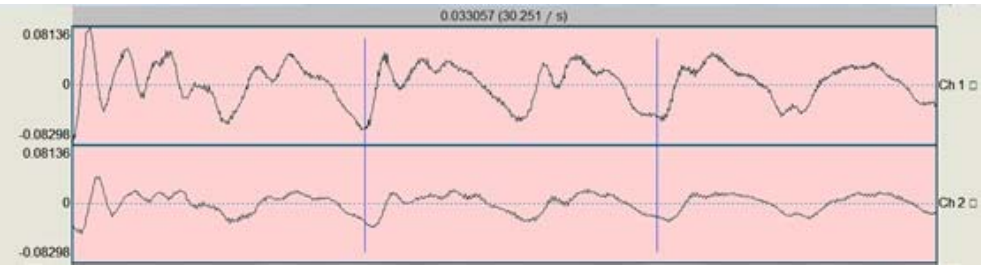

Fig. 5 The second monozygotic twin

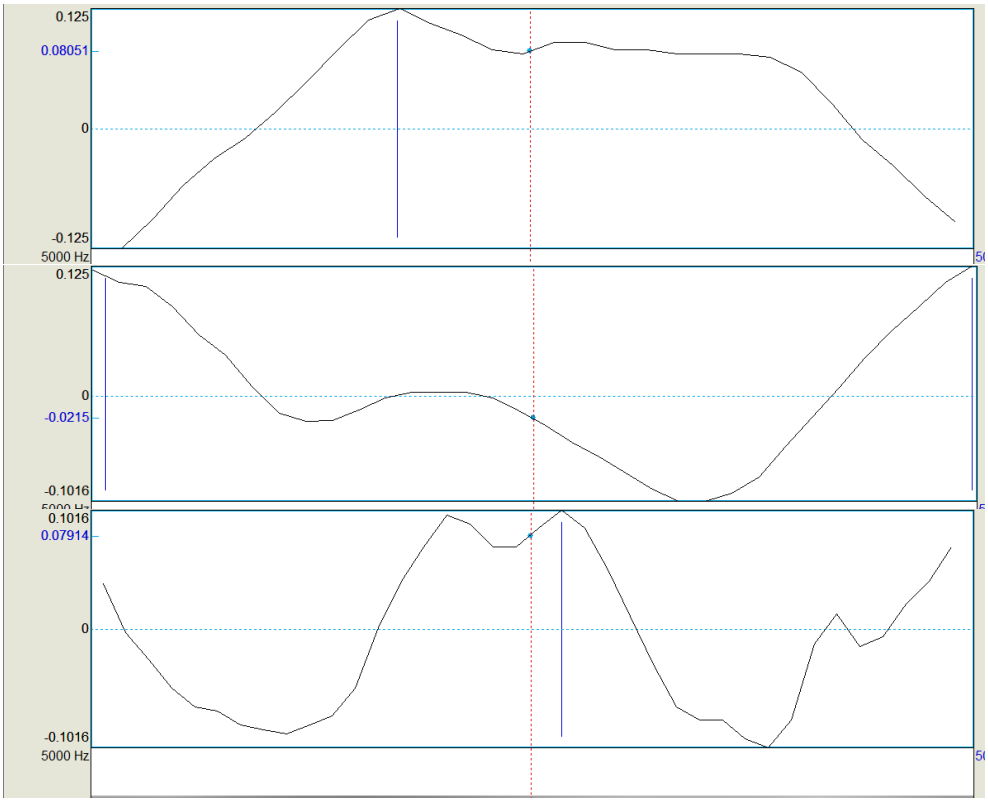

Fig. 6 One period in sound m

For one announcer from a pair, it took 0.0101 seconds to implement three periods (on average 0, 005 sec.), and another 0, 033 sec. (on average 0.011 sec.).

One of the important points in determining a person by voice is the choice of speech parameters. In order to best describe the personality of a voice, a feature system must be correctly selected. The main manifestations of individuality must be sought in spectral, formant, temporal and amplitude indicators. In the spectral picture, one can also find minimal data that contribute to determining the individuality of the voice. When pronouncing nasal sounds, the speech apparatus is in a fixed position and thus it is easier to identify individual characteristics. For this reason, we have chosen for analysis the words with the nasal consonant “m”.

The selected period of 0.004 seconds displays the distinctive parameters of the voices. Examples of voices of three different speakers visually show their difference.

Second Experiment

In the second experiment, analysis using a contour spectrogram is used to more accurately estimate the parameters of the speech signal. This method provides the ability to highlight individual characteristics of speech based on a statistical analysis of the characteristics of the contours in the spectrogram. In the contour spectrogram, each contour contains information about the change in the instantaneous frequency and amplitude of the component of the speech signal. Contour analysis allows increasing the accuracy of the selection of contours on the spectrogram even at high noise levels. To highlight the formant structure of a speech signal, contour analysis is an effective tool.

At one time, the method of contour spectrograms was widely used in phonoscopy. Contour spectrograms allow representing the spoken words, or sounds in the form of three-dimensional images. These spectrograms depict the

parameters "time - intensity - frequency".

For the experiment, phonograms were used, where during the dialogue, the announcers conduct a conversation repeating the conversation about a particular topic. During the discourse, the interlocutors repeat the same words and phrases. This allowed us to conduct contour spectrograms of the same words, or sounds in the same position.

Contour spectrograms allow visually evaluating and comparing data. Spectrograms are shown in Fig. 7; one of them belongs to one speaker, the other two to the second. This can be determined visually. Announcers uttered the same words. The first and third spectrograms of the word “ruçka” belong to one speaker.

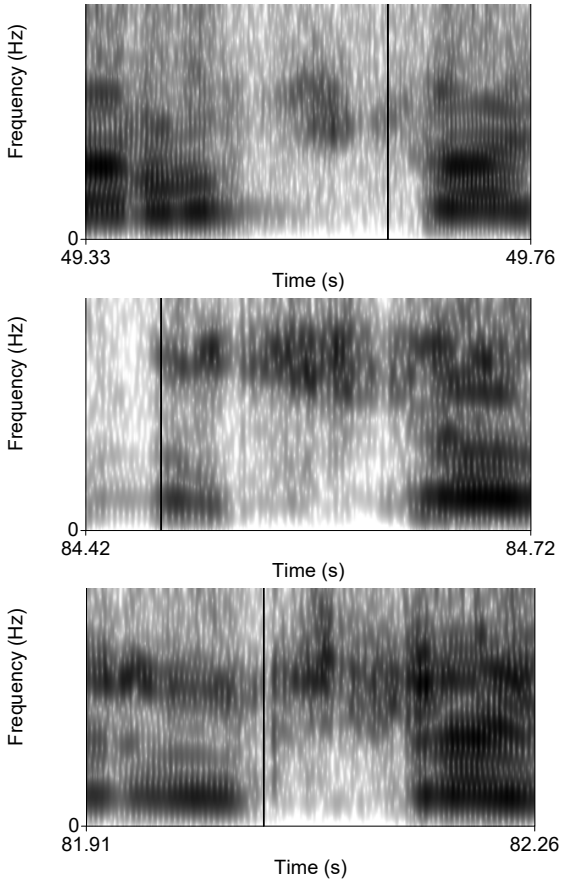

Fig. 7 Contour spectrogram of the word «ruçka»

Since we are interested in the minimum data of the speech signal, we conducted an experiment on the short part (the stationary part of the sound, or on the transition part) of the signal. To identify unique voice parameters, nasalized sounds act as one of the proven materials. Therefore, we chose the words where the sound a stands between the nasal consonants (m). In such cases, anti-formants can be observed in the contour spectrogram. As can be seen from Fig. 7, the first and third spectrogram belongs to one, the second to another speaker. Thus, the contour spectrogram method is one of the most reliable and important methods for revealing the unique properties of the voice.

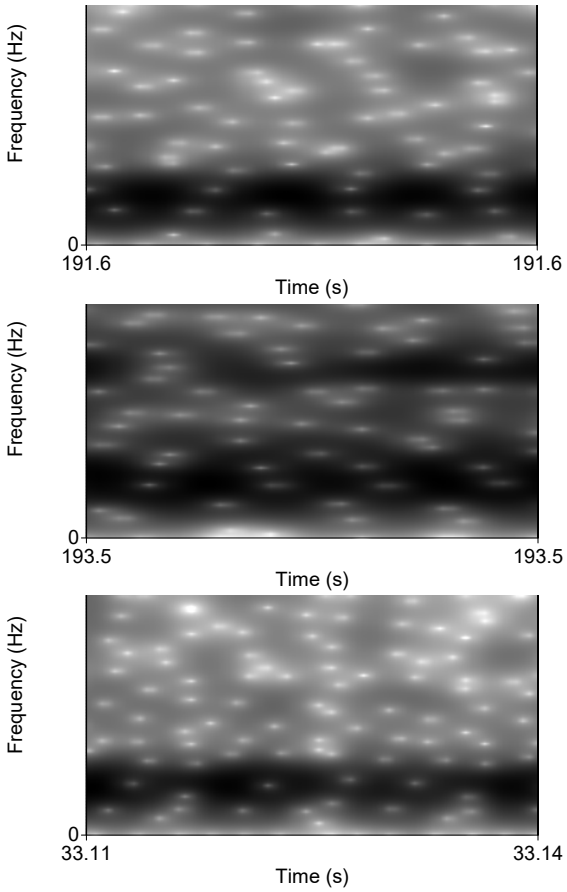

Fig. 8 Contour spectrogram of sound a

Experiments prove the feasibility of searching for personality traits in nasal sounds. Coarticulation of sound with subsequent vowels in isolated words provides excellent material for identifying individual characteristics of a voice, since coarticulation is strictly dependent on the individual. This symptom is one of those that are carried out subconsciously.

The constancy of the geometric dimensions of the nasal cavity gives an advantage in phonoscopic analysis of the speech signal. As can be seen from Fig. 9, the first and third example belongs to one, the second to another speaker.

III. Results

According to the results of experiments on the voice of monozygotic twins, it can be assumed that the temporary data of the vowel sound in the stationary zone - the pronunciation and intonation times - act as parameters for voice identification. The change in intensity from cycle to cycle was different between twins. Formant vowel indicators in the stationary zone differ mainly after the third formant - in the zone of non-lingual, but individual indicators. In addition, spectral data in the stationary zone indicate biometric fingerprints of the voice.

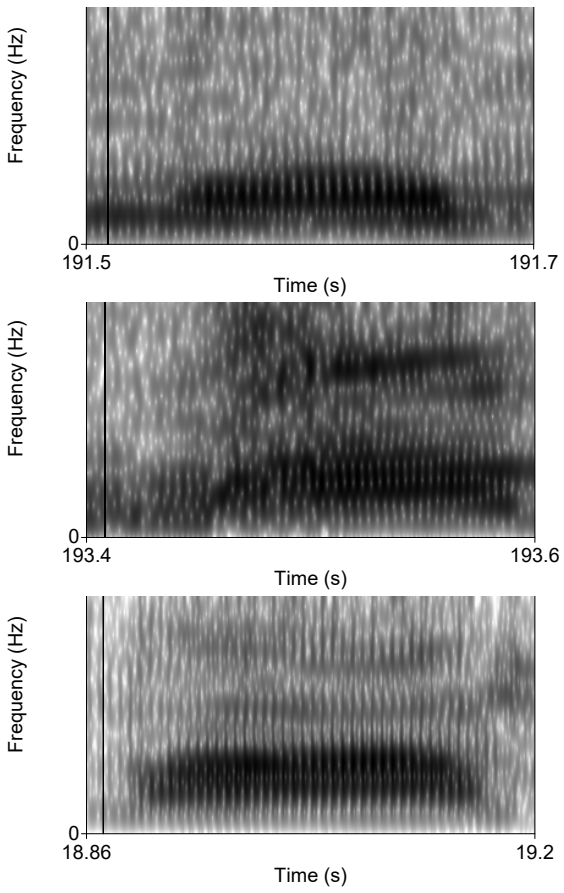

Fig. 9 Contour spectrogram of the word "mam"

Studies have confirmed the results of other studies [10], [12]. New data can be used in further studies of individual voice parameters and in judicial practice.

IV. Conclusion

According to the results of the experiment, it can be assumed that the minimum indicators are more reliable as a fact for personal identification. This investigation may facilitate the automatic recognition of speakers, as well as in the field of judicial phonetics, especially the judicial identification of the speaker. Identify the uniqueness of individual voices is the subject of further research.

References

[1] Sataloff, R.T, Herman- Ackah Y.D., Hawkshaw. M.J.(2007) Clinical Anatomy and physiology of the voice . In Otolaryngologic Clinics of North America. vol 40. issue 5. pp 909-929. DOI: 10.1016/j.otc.2007.05.002

[2] Koval Sergey L. Formants matching as a robust method for forensic speaker identification. SPECOM'2006, St. Petersburg, 25-29 June 2006. P.125-128

[3] Kaganov, A.Sh. (2005): Forensic expertise of sound records, Moscow: Jurlitinform (in Russian).

[4] Kaganov, A.Sh. (2008): The instrumental study of characteristics of speech path excitation for speaker identification purposes, in Forensic expertise № 3 (15), Saratov (in Russian).

[5] Olga Ilina, Serguei Koval and Michael Khitrov. Phonetic Analysis in Forensic Speaker Identification. An Example of Routine Expert Actions. ICPhS99 San Francisco. P.157-160

[6] Koval, S.L., Kaganov, A. and Khitrov, M. 1998. The chart of the standard expert actions and decision-making principles of forensic speaker identification. Proceedings of COST-250 Workshop "Speaker Recognition by Man and by Machine: Directions for Forensic Applications". Ankara, Turkey

[7] Koval, S.L., Labutin, P.V., Raev, A.N. (1995): Automatic speaker recognition using formants-based nearest-neighbor distance measure, in Proceeding EUROSPEECH'95, Madrid, vol. 2, p. 341-344.

[8] Koval, S.L., Zubova, P.I. (2007): Speaker identification by his voice and speech on the basis of complex analysis

[9] of phonograms, in Theory and practice of forensic expertise № 3 (7), Moscow: Nauka, p. 68-76 (in Russian).

[10] Decoster, W., Van Gysel, A.; Vercammen, J., Debruyne, F. (2001) Voice similarity in identical twins. Acta Oto-Rhino-Laryngologica Belgica .55, 49-55.

[11] Jayakumar, T., Savithri, S. R. (2008) Investigation into voice source of monozygotic twins using formant based inverse filtering. Journal of the All India Institute of Speech and Hearing .27, 8-14.

[12] Nolan, F. and Oh, T. Identical twins, different voices. (1996) Forensic Linguistics. 3(1):39–49.

[13] Vu Dang Hoang. Wavelet-based spectral analysis. Department of Analytical Chemistry and Toxicology, Hanoi University of Pharmacy, 13-15 Le Thanh Tong, Hanoi, Vietnam. https://doi.org/10.1016/j.trac.2014.07.010

[14] Daubechies I. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool / I. Daubechies, J. Lu, H.-T. Wu / / Journal of Applied and Computational Harmonic Analysis. – 2011. –Vol. 30, № 2. – P. 243–261. DOI: 10.1016/ j.acha.2010.08.002

[15] Gilles J. Empirical Wavelet Transform / J. Gilles // IEEE Transactions on Signal Processing. – 2013. – Vol. 61, № 16. – P. 3999–4010. DOI: 10.1109/TSP.2013.2265222