

Measuring the Voice Resemblance Extent of Identical (Monozygotic) Twins by Voiceprints Neutrosophic Domain

Yazen A. Khaleel¹, Caroline Y. Daniel¹ and Salah I. Yahya^{1,2}

¹Department of Software Engineering, Faculty of Engineering, Koya University, Koya, Kurdistan Region – F.R. Iraq

²Department of Communication and Computer Engineering, Cihan University-Erbil, Erbil, Kurdistan Region – F.R. Iraq

Abstract—The identical twins (monozygotic) are siblings (sib) created from the division of one fertilized egg (zygote), so they will be identical in their genetic characteristics and therefore in their phenotypic traits to a very large extent. The voiceprint of the twins is one of these traits. This paper proposed a method to determine the extent of the similarity and dis-similarity between the voiceprints for the brothers of identical twin and thus it is possible to distinguish between their voices. The proposed method adopts the use of the spectrogram as two-dimensional function with neutrosophic transformation. The neutrosophy theory is applied to all spectrogram elements that belong to the first sibling to determine to what extent this element belongs to the spectrogram of the second sibling for the same utterance. This study relied on the use of a number of voice clips collected from 35 identical twins. The neutrosophic transformation is used to convert the voiceprints into the neutrosophic domain represented by three membership functions (true, false, and indeterminate). The results showed that the average extent of the similarity ratio between twins' voices (true membership) is 68.15%, the difference ratio (false membership) is 31.84%, and the indeterminacy membership function ratio is 18.83%.

Index Terms—Identical Twins, Monozygotic, Neutrosophic domain, Voiceprint.

I. INTRODUCTION

Monozygotic (MZ) twins (identical twins) come from one fertilized egg (zygote). The identical twins are well-known to show the most extreme form of anatomical, physiological match among human beings. Identical twins cannot be distinguished using DNA, which has caused problems in creating evidence in many forensic cases Künzel (2011). It is recommended that the biometric identifiers, such as

fingerprints, iris, palm prints (Kong, Zhang, Lu, 2006), face, and voice (Kong, Zhang, Lu, 2006); (Kodate, et al., 2002), can still be used to discriminate them.

The physical features of the laryngeal, such as vocal fold length and structure, shape and size of the supraglottic vocal tract, and vocal mechanism, are genetically determined (Sataloff, Herman-Ackah, Hawkshaw, 2007). The excitation vocal signal which is generated by the human vocal cords is filtered by the shape of the vocal tract that includes the tongue and teeth. This shape specifies what sound comes out (Khaleel, 2020).

Although voice is unique to individuals, researches show perceptive similarity in MZ twins (Van Gysel, Vercammen, and Debryne, 2001). Many studies showed that MZ twins have similar voice characteristics, leading to perceived similarity (Whiteside and Rixon, 2000).

Fuchs, et al. (2000) presented a study to identify parameters of voice performance and acoustic features in MZ twins. They compared intrapair differences with data from a control group. Moreover, they examined the correlation of intrapair differences with the age of the MZ twins. A more significant difference in older twin pairs than in younger pairs could show the effect of an exogenous influence. They inspected seven parameters of voice performance and three acoustic features in 31 MZ twin pairs (median age 36 years, range 18–75 years) and compared them with control group pairs, which consisted of non-related persons of the same sex and age. In Cielo, et al. (2012) research, two pairs of adult twins participated, one pair of each sex. They found no differences in vocal attack, articulation, loudness, resonance, respiratory mode, and type in both pairs of MZ twins evaluated, and the results were within normal limits.

Sebastian, et al. (2013) have a study aimed at perceptually and acoustically discriminating the voices of identical twins from each other. A perception test was done to find whether the voices of identical twins could be perceived as the same or different voices. Ten MZ twin pairs, five females and five males between the ages of 10 and 15 years old participated as speakers. Five native listeners were requested to judge whether each pair belonged to the same speaker or to different

ARO-The Scientific Journal of Koya University
Vol. X, No.1 (2022), Article ID: ARO.10925, 6 pages
DOI: 10.14500/aro.10925

Received: 10 January 2022; Accepted: 14 May 2022
Regular research paper: Published: 25 May 2022

Corresponding author's e-mail: yazen.adnan@koyauniversity.org
Copyright © 2022 Yazen A. Khaleel, Caroline Y. Daniel and Salah I. Yahya. This is an open access article distributed under the Creative Commons Attribution License.



speakers. The acoustic analysis exposed that shimmer values are more sensitive in discriminating twin voices among each other. This study can contribute to automatic speaker recognition.

The study of Segundo, et al. (2017), explores the forensic potential of voice features combining source and filter characteristics. A set of features have been extracted from pause fillers. Speaker similarity was measured using standardized Euclidean distances (EDs) between pairs of speakers: 54 different-speaker (DS) comparisons, 54 same-speaker (SS) comparisons, and 12 comparisons between MZ twins. Their results revealed that the differences between DS and SS comparisons were significant in both high-quality and telephone-filtered recordings, with no false rejections and limited false acceptances. Mean ED for MZ pairs lies between the average ED for SS and DS comparisons.

Cavalcanti, Eriksson, Barbosa (2021) evaluate the speaker-discriminatory potential of a set of fundamental frequency (f_0) estimates in inraidentical twin pair comparisons and cross-pair comparisons (i.e., among all speakers). Ten male identical twin pairs aged between 19 and 35 were subjected. Acoustic measurements were done on the speech samples. Although identical twins were very closely related regarding their f_0 patterns, the outcomes suggest that some pairs could still be differentiated acoustically.

The neutrosophic logic (NL) is a non-classical logic which can be used to represent a mathematical model of an uncertainty, fuzzy, vagueness, or ambiguity cases. Because of the fuzzy nature (non-crisp) in the similarity extent between the voices of the twins, so the neutrosophic theory was chosen to determine the extent of similarity (true membership), difference (false membership), or it might be the indeterminacy membership (neutrality).

The current research aims to determine whether it is possible to numerically measure the extent of similarity between the voices of identical twins using the fuzzy neutrosophic transformation (NS-T) theory. It also looks more carefully at the speech voiceprints features to determine the similarities and differences between identical twins' voices.

II. MATERIALS AND DATASET

This study used an audiovisual twins database that is collected at the Sixth Mojiang International Twins Festival held on May 1, 2010, China (Li, 2015). This database contains 39 MZ twin pairs subjects (aged 7 up to 54 years), however, only 35 pairs (16 males and 19 females) were selected and participated as speakers in the current research. The excluded samples were children twins who did not clearly perform the vocal clips as required so they are not suitable for our study. The dataset has several face images, facial motion videos, and speech records for each subject. The ages of the participants ranged from 7 to 54 years old. The used audio voice records in this dataset uttered by the participants are one speech sample. Each one for both siblings of a twin utters 10 words as he counts from 1 to 10.

The audio clip record lasts for about 8 s on average. Each uttered word is processed individually.

III. NEUTROSOPHIC TRANSFORMATION

Real-world applications provide massive information which is incomplete, imprecise, fuzzy, conflict, and inconsistent. The uncertainty may originate from acquisition errors, insufficient knowledge, or stochasticity. Biometric applications are an example of a field that deals with these kinds of information. These human biometrics types have characteristics that may or may not be completely identical to the original entity or may be indeterminate (neutrality).

Neutrosophy, a branch of philosophy that generalizes dialectics, was introduced by Smarandache in 1980. It studies neutralities' source, nature, and scope. In neutrosophy theory, every event has a substantial degree of truth (T), falsity (F), and indeterminacy (I) that have to be considered independently from each other. Therefore, A is an idea, theory, event, concept, or entity; Anti-A is the opposite of A; and the neutrality Neut-A means neither A nor anti-A, that is, the neutrality between the two extremes (Smarandache, 2001).

Neutrosophic set is a generalization of the fuzzy set theory, where each element of the universe has a degree of truth, falsity, and indeterminacy. Unlike in fuzzy sets, the neutrosophic set includes the additional domain (I), which offers a more effective method to deal with the higher degrees of ambiguity.

Let U be a universe of discourse set. A neutrosophic set A in U is described by three neutrosophic components: T, F, and I are defined to represent the truth membership degree, falsity membership degree (non-membership degree), and the indeterminacy membership degree of an element, respectively, and independently.

IV. METHODOLOGY OF TWIN'S VOICE DISCRIMINATION

The proposed method is to measure how similar and then distinguishes the voices of the identical twins from each other using the neutrosophic theory. In the beginning, it is assumed that the twins' voices are similar to the exact spoken words. Through the proposed method, the correctness or wrongness of this assumption will be proved. The technique depends on extracting some features from the speech clips recorded for the twins. One of the most important features is the spectrogram, called "*voiceprint*."

The speech signals are transformed to the spectral domain and then the spectrogram is interpreted and handled in NS domain. In NS domain, a filter is considered with the spectral and time domains.

The proposed method deals with the extracted spectrogram as a two-dimensional entity that maps the speech loudness to the frequencies and specific points of time. The neutrosophy theory is applied to all spectrogram elements that belong to the first sibling to determine to what extent this element belongs to the spectrogram of the second sibling for the same

utterance. In addition, slight differences between the siblings of the identical twins are noticed, when they are classified according to their gender and age in the examined voiceprints.

A. Preprocessing Step

The preprocessing step includes segmentation each speech records into 10 individual uttered words for the both siblings of all participated twin pairs. These segmented words are the spoken counted numbers from 1 to 10.

B. Voice Analysis in Frequency Domain

An essential part of this paper is the audio processing that involves a frequency voice analysis. The algorithms do not take the raw audio files as input; hence, extracting features from the audio files are mandatory. This helps to build the required model. An audio signal can be represented using a three-dimensional signal where these three axes represent amplitude, frequency, and time which are called spectrogram or voiceprint (vp).

MATLAB provides toolboxes for the analysis of audio. In this work, the spectrogram toolbox uses a Short-Time Fourier Transform (STFT) of the speech signal specified by a vector. The obtained spectrogram is represented in a matrix which is divided into eight segments with 50% overlap and each segment is windowed with a Hamming window. Each column of the spectrogram matrix estimates the input speech vector's short-term, time-localized frequency content. The amplitude (loudness) of an audio file can be shown in Fig. 1 that is an example of the raw speech signal uttered by both siblings of a twin.

The spectrogram is used to map different frequencies at a given point of time to their amplitudes. It is a graphical representation of the spectrum of frequencies of a voice. An example is shown in Fig. 2.

C. The Voice in the Neutrosophic Domain

A voice spectrogram is defined in the neutrosophic domain as: Let U be a universe of discourse and A is a set included in U, representing the highly valued elements. The spectrogram in the neutrosophic domain (S_{NS}) is characterized using three distinctive membership components (T), (I), and (F), where T defines the truth scale, F as the scale of false, and I characterizes the scale of intermediate.

All considered components are independent of each other. The element (S) of a spectrogram in the neutrosophic domain can be described as $S(T, I, F)$ (Smarandache, 2001). Each element belongs to set A in the following way: It is t% true membership function in the high-valued element set (similarity indicator), the i% indeterminacy membership function in the set, and the f% a falsity-membership function in the set (dissimilarity indicator). There is a valuation for each component in [0, 1].

The logic of the neutrosophic concept is an obvious frame trying to calculate the truth, indeterminacy, and falsity. Smarandache notes the dissimilarity of fuzzy logic and NL. NL could differentiate absolute truth (AT) and relative truth (RT) by assigning 1+ for AT and 1 for RT and is also applied in the field of philosophy. Hence, the standard interval [0,1] used in IFS is extended to non-standard $[-0, 1^+]$ in NL. There is no condition on truth, indeterminacy, and falsity, which are all the subsets of a non-standard unitary interval (Broumi, et al., 2019).

In the spectral domain, element $S(j, k)$ is transformed into a neutrosophic domain by calculating $S_{NS}(j, k) = \{T(j, k), I(j, k), F(j, k)\}$ in Eqs. (1–6), where $T(j, k)$, $I(j, k)$, and $F(j, k)$ are considered as a probability of that element $S(j, k)$ belongs to a highly valued set, indeterminate-valued set, and low-valued set, respectively. This is the primary benefit of neutrosophic in colormap processing, and it can be taken at the same time when the decision is made for each element in the spectrogram. The following basic equations were used for transforming voiceprint from a spectral domain into the neutrosophic domain (Ali, 2018):

$$S_{NS}(j, k) = \{T(j, k), I(j, k), F(j, k)\} \tag{1}$$

$$T(j, k) = \frac{\bar{s}(j, k) - \bar{s}_{min}}{\bar{s}_{max} - \bar{s}_{min}} \tag{2}$$

$$\bar{s}(j, k) = \frac{1}{L^2} \sum_{m=j-\frac{L}{2}}^{j+\frac{L}{2}} \sum_{n=k-\frac{L}{2}}^{k+\frac{L}{2}} s(m, n) \tag{3}$$

$$I(j, k) = \frac{h(j, k) - h_{min}}{h_{max} - h_{min}} \tag{4}$$

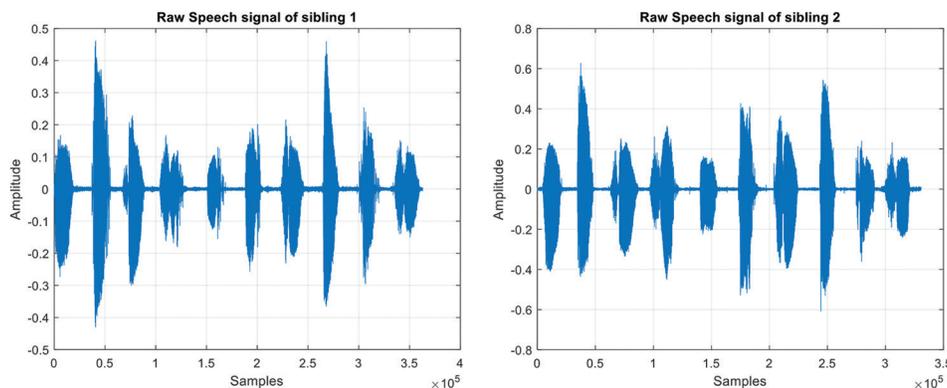


Fig. 1. Example of the raw audio signal record of an identical twin uttering the same words sample. Each column is from one sibling of the twin.

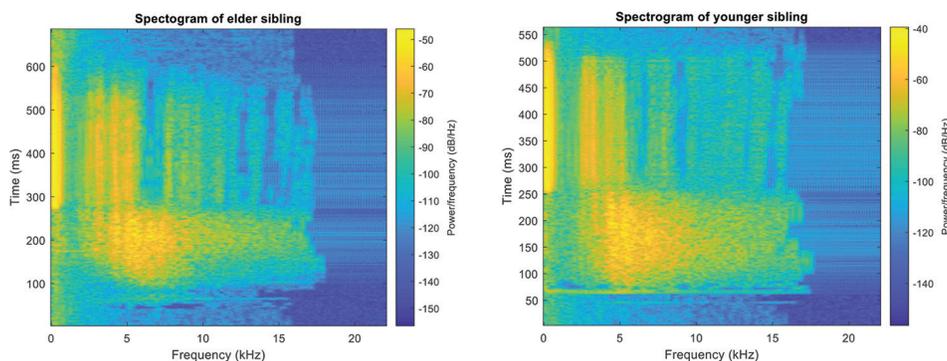


Fig. 2. The corresponding spectrograms of identical twins for a same uttered word shown in Fig. 1. Each column is from one sibling of the twin.

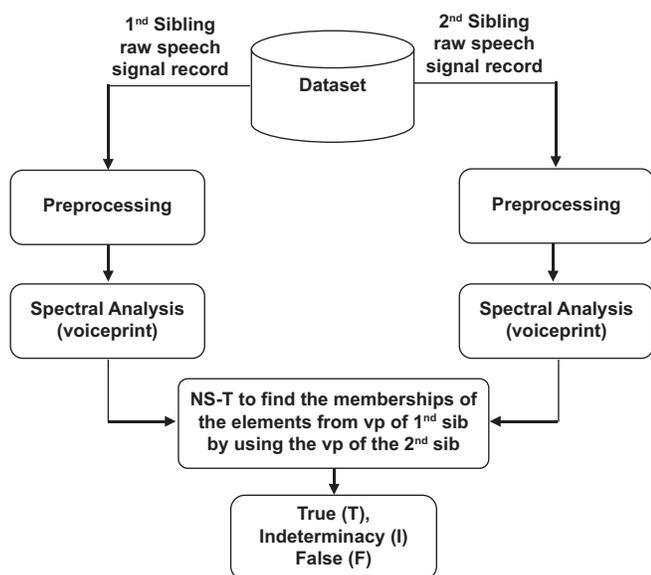


Fig. 3. General flow block diagram of the proposed method.

$$h(j, k) = |s(j, k) - \bar{s}(j, k)| \quad (5)$$

$$F(j, k) = 1 - T(j, k) \quad (6)$$

Where:

- $s(j, k)$ represents the amplitude value of the element in the domain
- T, I, and F are true, indeterminacy, and false sets, respectively, in the neutrosophic domain
- $\bar{s}(j, k)$ can be defined as the local average (mean) value of $s(j, k)$
- L is the length of the averaging window, for the current research, $L=5$
- $h(j, k)$ is the homogeneity value of T at (j, k) , which is described by the absolute value of the difference between the amplitude value of the spectrogram $s(j, k)$ and its local average value $\bar{s}(j, k)$.

The output of the proposed process is the true, indeterminacy, and false membership functions ($t\%$, $i\%$, and $f\%$, respectively). They are calculated for each element in the spectrogram function of the uttered word (1–10). This

TABLE I
TRUE, INDETERMINACY, AND FALSE MEMBERSHIP VALUES FOR ONE TWIN FOR EACH UTTERED WORD

Uttered Word	T	I	F
One	0.749	0.154	0.251
Two	0.642	0.156	0.358
Three	0.622	0.152	0.378
Four	0.643	0.142	0.357
Five	0.739	0.156	0.261
Six	0.739	0.220	0.261
Seven	0.696	0.205	0.304
Eight	0.789	0.195	0.211
Nine	0.702	0.168	0.298
Ten	0.699	0.208	0.301
Average	0.702	0.176	0.298

process is repeated for all the participated twins. Finally, the average for each membership is calculated. Fig. 3 shows a general block diagram of the proposed method.

V. RESULTS AND DISCUSSION

The output of this process is the calculated probability of the true membership function in the set $t\%$, indeterminacy membership function in the set $i\%$, and a falsity-membership function in the set $f\%$ for each uttered word, as shown in Table I. This process is repeated for all the participated twins.

Then, the average of each membership value is calculated as in the last row in Table I. The mean values of T, I, and F memberships are computed for each twin, as shown in Fig. 4. Finally, the total average is calculated, showing the mean values of each probability value of the three membership functions for all the 35 twins that participated in the process, as shown in Fig. 5.

Fig. 4 shows that the average true membership value (similarity indicator) equals 0.6815, the average indeterminacy membership value equals 0.1883, and the false membership value (dissimilarity indicator) equals 0.3184. These results indicate that the extent of similarity between the voiceprints of the two siblings of a twin is about 68.15% and the dissimilarity is about 31.84%. At the same time, the indeterminacy probability is 18.83%. These results were obtained for a text-dependent utterance. The maximum

TABLE II
COMPARISON WITH THE PREVIOUS LITERATURE WORKS

Ref.	No. of MZ twin pairs	Technique	Conclusion
Fuchs, et al., 2000	31	Identify 10 parameters of voice performance and acoustic features	They found significant differences in seven of 10 parameters
Cielo, et al., 2012	2	Auditory, auditory-perceptual vocal, and glottic source acoustic evaluations	No differences in vocal attack, articulation, loudness, resonance, respiratory mode, and type in both pairs of monozygotic twins evaluated, and the results were within normal limits
Sebastian, et al., 2013	10	Acoustic analysis and five native listeners were requested to judge	The measured values are more sensitive in discriminating twin voices among each other
Segundo, et al., 2016	12	Extraction a set of features. Speaker similarity was measured using standardized Euclidean distances	Mean ED for MZ pairs lies between the average ED for SS comparisons and DS comparisons
Cavalcanti, Eriksson, Barbosa, 2021	10	Assess the speaker-discriminatory potential of a set of fundamental frequency (f_0) estimates in intra identical twin pair comparisons and cross-pair comparisons	Although identical twins were very closely related regarding their f_0 patterns, some pairs could still be differentiated acoustically
Current research	35	Neutrosophic domain transformation of the voiceprints	The extent of similarity between the voiceprints of the two siblings of a twin is about 68.15%, and the dissimilarity is about 31.84%. At the same time, the indeterminacy probability is 18.83%

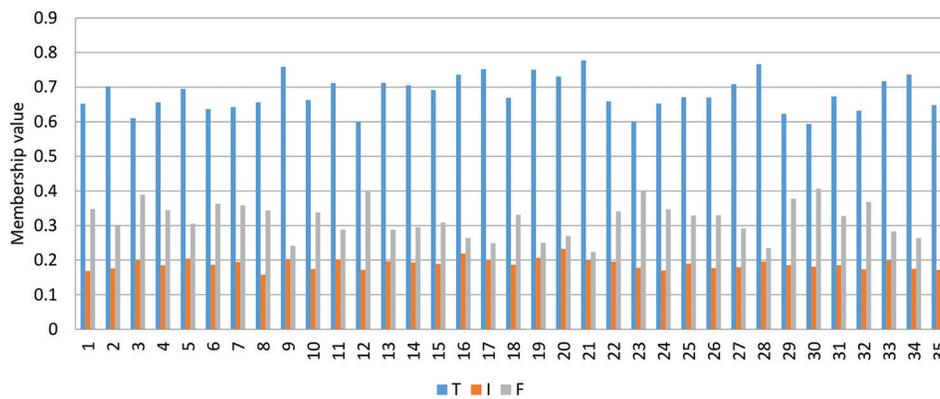


Fig. 4. True (T), indeterminacy (I), and false (F) membership values for each participating twin.

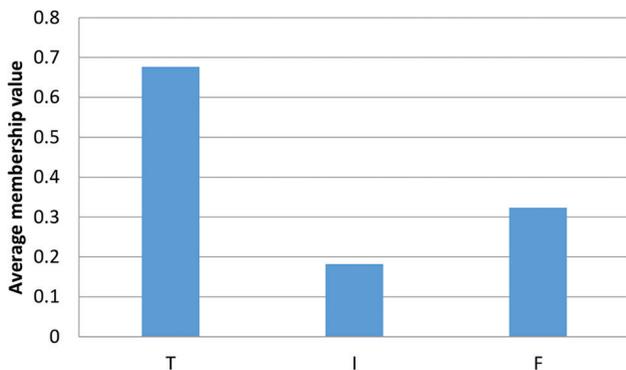


Fig. 5. The total average of true (T), indeterminacy (I), and false (F) membership values membership value for all 35 twins.

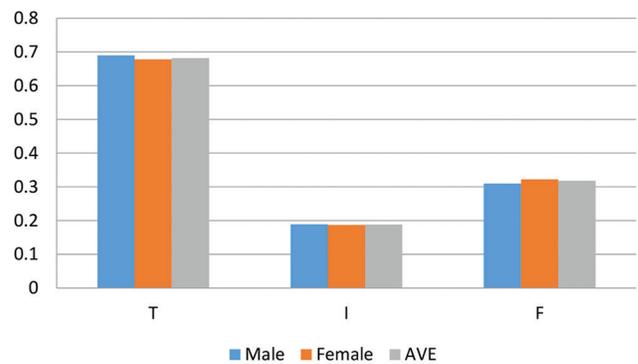


Fig. 6. True (T), indeterminacy (I), and false (F) membership values membership value for 16 males versus 19 females.

true membership value is 93.13% for the individual spoken words, whereas the minimum is 47.79%. At the average level for each twin, the maximum true membership value is 77.7%, whereas the minimum is 59.3%.

A comparison of T, I, and F membership values was made, but this time the data were classified by gender (male/female). Fig. 6 shows the average true (T), indeterminacy (I), and

false (F) membership values membership value for 16 males versus 19 females. The obtained values are compared with total average of all 35 pairs results. It is noticed clearly, that is, no significant effect between the results relating the gender difference when compared with total average. However, for the male, it is slightly bigger in the true membership value.

Another comparison was carried out after classifying the participants according to their age groups (children,

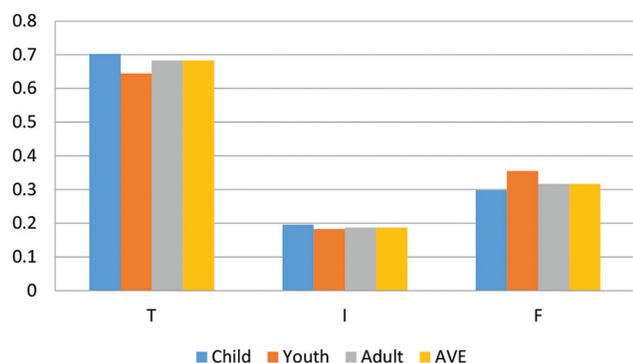


Fig. 7. True (T), indeterminacy (I), and false (F) membership values according to the age categories of the participants twin pairs.

youth, and adults). The T, I, and F membership values were compared for these three age categories and with the total average as well. Fig. 7 shows the results for the true (T), indeterminacy (I), and false (F) membership values according to the age of the participants twin pairs.

There is a small impact relating to the ages of the twin pairs. The average true membership value for the children group is slightly greater than the other groups, and in the youth category, it takes the least true membership value among them. Table 2 shows a comparison between the suggested method in the current research and some of the previous literatures. The comparison is focused on the used technique and the with the main obtained conclusions.

VI. CONCLUSION

The identical twins come from one fertilized egg (zygote), and therefore, their genetic personalities are expected to be very similar. The voice is one of these traits. This paper proposed a method to measure the degree of similarity or difference between the voiceprint between twin siblings. The audio clips of 35 twins were collected during the twins' festival held in 2010. During these audio clips, the participants count from 1 to 10. The neutrosophic transformation is applied to the voice spectrograms "voiceprints" of the twin brothers. The outputs of this process are three functions: $t\%$ true membership function (similarity indicator), the $i\%$ indeterminacy membership function, and the $f\%$ a falsity-membership function (dissimilarity indicator). These results indicate that the extent of similarity between the voiceprints of the two siblings of a twin is about 68.15%, and the dissimilarity is about 31.84%. At the same time, the indeterminacy probability is 18.83%. These results were obtained for a text-dependent utterance. In addition, slight differences between the siblings of the identical twins are noticed, when they are

classified according to their gender and age in the examined voiceprints.

REFERENCES

- Ali, M.L., Son, H., Khan, M. and Tung, N.T., 2018. Segmentation of dental X-ray images in medical imaging using neutrosophic orthogonal matrices. *Expert Systems With Applications*, 91, pp.434-441.
- Ariyaeinia, A., Morrison, C., Malegaonkar, A. and Black, S., 2008. A test of the effectiveness of speaker verification for differentiating between identical twins. *Science and Justice*, 48, pp.182-186.
- Broumi, S., Nagarajan, D., Bakali, A. and Talea, M., 2019. Implementation of neutrosophic function memberships using MATLAB program. *Neutrosophic Sets and Systems*, 27(1), pp.44-52.
- Cavalcanti, J.C., Eriksson, A. and Barbosa, P.A., 2021. Multiparametric analysis of speaking fundamental frequency in genetically related speakers using different speech materials: Some forensic implications, *Journal of Voice*, In Press.
- Cielo, A.C., Agustini, R. and Finger, L.S., 2012. *Vocal Characteristics of Monozygotic Twins*. Revista CEFAC, pp.1234-1243.
- Fuchs, M., Oeken, J., Hotopp, T., Täschner, R., Hentschel, B. and Behrendt, W., 2000. Similarity of monozygotic twins regarding vocal performance and acoustic markers and possible clinical significance. *HNO*, 48(6), p.462.
- Khaleel, Y.A., 2020. High security and capacity of image steganography for hiding human speech based on spatial and cepstral domains. *ARO Journal*, 8(1), pp.95-106.
- Kodate, K., Inaba, R., Watanabe, E. and Kamiya, T., 2002. Facial recognition by a compact parallel optical correlator. *Measurement Science and Technology*, 13, pp.1756-1766.
- Kong, A.W.K., Zhang, D. and Lu, G., 2006. A study of identical twins' palmprints for personal verification. *Pattern Recognition*, 39, pp.2149-2156.
- Künzel, H.J., 2011. Automatic speaker recognition of identical twins. *The International Journal of Speech, Language and the Law*, 17(2), pp.251-277.
- Li, J., Zhang, L., Guo, D., Zhuo, S. and Sim, T., 2015. Audio-visual twins database. International Conference on Biometrics, 2015, pp.493-500.
- Sataloff, R.T., Herman-Ackah, Y.D. and Hawkshaw, M.J., 2007. Clinical Anatomy and physiology of the voice. *Otolaryngologic Clinics of North America*, 40(5), pp.909-929.
- Sebastian, S., Benedict, A.S., Sunny, G.K. and Balraj, A., 2013. An investigation into the voice of identical twins. *Otolaryngology Online Journal*, 3(2), pp.1-7.
- Segundo, E.S., Tsanas, A. and Vilda, P.G., 2017. Euclidean Distances as measures of speaker similarity including identical twin pairs: A forensic investigation using source and filter voice characteristics. *Forensic Science International*, 270, pp.25-38.
- Smarandache, F., 2001. An introduction to neutrosophy, neutrosophic logic, neutrosophic set, and neutrosophic probability and statistics. *1st International Conference on Neutrosophy, Neutrosophic Logic, Neutrosophic Set, Neutrosophic Probability and Statistics*, University of New Mexico-Gallup 1-3 December, pp.1-17.
- Van Gysel, W.D., Vercammen, J. and Debruyne, F., 2001. Voice similarity in identical twins. *Acta Otorhinolaryngologica Belg*, 55(1), pp.49-55.
- Whiteside, S.P. and Rixon, E., 2000. The identification of twins from pure (single speaker) syllables and hybrid (fused) syllables: An acoustic and perceptual case study. *Perceptual and Motor Skills*, 91, pp.933-947.