

## Voice Biometrical Match of Mother and Daughter for Forensic Purpose

Research Article

Kaur G<sup>1</sup>, Verma P<sup>2\*</sup>, Jain MK<sup>3</sup>

<sup>1</sup> MSC Student, University Institute of Applied and Health Sciences, Chandigarh University Gharuan Punjab, India.

<sup>2</sup> Assistant Professor, University Institute of Applied and Health Sciences, Chandigarh University, Gharuan Punjab, India.

<sup>3</sup> Senior Scientific Assistant, Department of Physics, CFSL (CBI), New Delhi, India.

### Abstract

The similarity in voices among blood relations has been always a fascinating subject in forensic phonetics. The present work is an initiatory study of investigational character that explains the resemblance of mother and her daughter phonation. This study is an attempt to report a novel approach to the speaker recognition and comparison of the voice of mother and her daughter by the use of voice biometric software. An attempt has been made to enlighten the role of voice as evidence in forensic investigation. Having noted a lack of oral corpora of mother and daughter voices, we have collected a database which consists of the 40 samples in a combination 20 each of mother & daughter respectively. Voice transcripts of mother and daughter were analyzed & compared using speech analysis software to examine the similarities & variations in sound spectrograms. The results of our study showed that the mother and daughter voice samples have higher similarities which were estimated to 88%-92%. It is due to the influence of genetic endowment, physical similarities and inheritance of vocal gesture & posture. Voice Biometrical Match cannot be 100% due to law of individuality. This suggests that the system features are to a great extent genetically conditioned and that they are hence useful and robust for comparing speech samples of known and unknown origin, as found in legal cases. This research shall enlighten an idea of getting corroborative evidence, further which can be a part of Forensic Investigation. Forensic experts can contribute to society/law with advance as well as fast crime detection solutions challenge in speaker identification.

**Keywords:** Phonetics; Automatic Speech Recognition (ASR); Computer Speech Laboratory (CSL); Spectrograms; Gold Wave.

### Introduction

Speech Recognition known as Automatic Speech Recognition (ASR) is the procedure of converting speech signals to a sequence of words through methods of algorithms accomplished as a computer programme [1]. Automatic speech recognition enables the computer to identify words that are spoken into microphone or a telephone. Speaker dependent system are those systems that identify or recognized only one person voice and those systems that recognized the voice of multiple users are known as speakers independent system [2]. Automatic Speech Recognition system (ASR), is an important branch of information and communication technology as it is used to transform spoken words into text and also has dictation, security control, and foreign language translation [3].

Voice production process begins with the production of airflow in the lungs, which is regulated by the vocal folds. Lungs, larynx, and vocal tract are the main organs for speech production. Lungs supplies air pressure signals to the larynx. The larynx regulates the airflow which is supplied by the lungs. The glottis, fluctuates with the movement of the vocal folds, and thus regulates the flow of air through them. Larynx also known as a self-vibrating acousto-mechanical oscillator. Various tiny muscles present in larynx also control the oscillation of vocal folds. Some of these tiny muscles regulate the rest position of the vocal folds while others regulate their tension [4]. The vocal cords become distressed and pulsate by the airflow when lungs expel the air. The vocal folds alternately trap and release air causing a cycle of vibrations called Glottis Cycle. When the glottis opens, the air explodes through the vocal

#### \*Corresponding Author:

Priyanka Verma,  
Assistant Professor, University Institute of Applied and health sciences, Chandigarh University, Gharuan Punjab, India.  
Tel: +919888972102  
Fax: 0160-3014402  
Email: Priyankaverma25@gmail.com

**Received:** July 20, 2019

**Accepted:** August 30, 2019

**Published:** August 31, 2019

**Citation:** Kaur G, Verma P, Jain MK. Voice Biometrical Match of Mother and Daughter for Forensic Purpose. *Int J Forensic Sci Pathol.* 2019;6(2):407-410.

**doi:** <http://dx.doi.org/10.19070/2332-287X-1900085>

**Copyright:** Verma P<sup>©</sup> 2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

folds producing a sound wave. When air pressure forces the vocal folds to open and vibrate, it produces a voiced sound. The rate of repetition of the cycles of air pressure or frequency of vocal fold vibration is called as Fundamental frequency or Frequency of oscillation and the fundamental frequency also depends on various properties of the vocal cords (length, mass and tension). There are some areas in frequency spectrum where energy is concentrated called formants. Formants are those dark bands that are observed in the vowel region. These formants are examined for the purpose of voice identification. The vocal cords that are vibrating synchronize the air stream from the lungs at a rate from 60 times per second for males to 500 times per second for children [5].

Speech Recognition systems can be divided into distinct classes by describing what types of utterances they can recognize. These classes are identified as follows:

→ Connected word systems (or more precisely ‘connected utterances’), allow apportioned utterances to be ‘run-together’ with a minimal pause between them.

→ Continuous Speech: Continuous speech recognizers enable the users to speak almost naturally, while the computer detects the content. Basically, it is a computer dictation.

→ Spontaneous Speech: This can be thought of as speech that is natural sounding and not rehearsed. An ASR system for such a speech deals with a variety of natural speech features i.e. words being run together, “ums”, “ahs”, and even slight stutters.

→ Isolated Words: Isolated word recognizers usually need each utterance to have silence on both sides of the sample window. The user speaks individual words or phrases and the recognizer accepts single words or a single utterance at a time [6].

Recent studies in voice quality are conducted towards the evaluation of phonation performance in relation to either professional voice care, or in meta-acoustic knowledge (neurological deterioration, emotion detection etc.) These fields of study are becoming more and more demanded now a day. Voice Biometrical Match on twins voice quality analysis and vocal performance of interest was done by Van Lierde et al., and Segundo in their respective studies [7, 8].

There are some behavioral phonetic characteristics that are due to learned/habitual styles and are opposed to genetic reasons. In past, various research work is done on the twin voice quality analysis and vocal performance [8-12].

Segundo and Gómez-Vilda in their study examined the voice biometrical match of twin and non-twin siblings. This is an initial analysis to explore those characteristics/features that have higher resemblance of monozygotic and dizygotic elocution/vocalization in comparison to non-twin siblings. To obtain mutual relation/similarity between monozygotic and dizygotic twins, biomechanical constants procured from phonetic fillers are used. The results of their study express the connection concerning ancestral load and environmental factors to choose vocalization manner [9].

Singh et al., emphasizes on some areas where the technique of speaker recognition can be used. The speech is one of the natural forms, the voice of an individual consists of various parameters like emotions, health, gender that carry information. Speaker rec-

ognition technique applied for authentication, surveillance and speaker recognition for identification purpose in forensics [13].

Deliyski et al., determined the influence of sampling rate on acoustic voice quality measurements, the influences of gender, intra-subject variability, microphone, environmental noise, data acquisition hardware, and analysis software considered as balancing factors. The sampling rates for acoustic voice analysis are above 26 kHz, above 19 kHz, and 12 kHz, respectively. The voice samples above 26 kHz can be taken for data examination [14].

Krajewski et al., studied the analysis and classification of a cold speech. The analysis and classification of cold speech can be done by Variation mode decomposition (VMD). The speech signals are decomposed into a number of sub-signals by VMD. Statistics, mean, variance, kurtosis and skewness are extracted from each of the decomposed sub-signals. Center frequency, energy, peak amplitude, spectral entropy, permutation entropy and Renyi's entropy are assessed, which is used as features [15].

## Materials and Methods

In our study materials and methods utilized to gather and assemble the samples, analysis of the extracted features/words of the speech samples, characteristics of the speech sample analyzed, procedure and techniques used for audio analysis and comparison in conjugation with the different stages of the automatic speaker recognition were used.

## Data Collection

In the data collection stage, mothers and her daughters were request to phonate after taking their oral consent. Having noted a lack of oral corpora of mother and daughter voice, we have collected a database which consists of the 40 samples in a combination 20 each of mother & daughter respectively from the Delhi, Haryana and Chhattisgarh population. The language used to record the audio sample was Standard Hindi as it is their mother tongue. Their speech was recorded using a mobile phone first, and then a computer microphone with sampling frequency of 8000 Hz at Wave PCM signed 16-bit per sample along with one channel (mono) WAV format used at 128kbps bit rate. Each mother & daughter read similar words respectively with variable recording time. The words tested were “*BBC News*”, “*kaha hai ki unka desh*”, “*karwai kay madey nazar*”, “*ki gyi jordar*”, “*Syria me irani*”. Each mother & daughter read every word one time to be used later during process. The approximate age of each mother varies between 45-65 years and for daughter is between 20-40 years. This particular age of mother and daughter has been selected because at this respective age, they are very able in communicating or perception as it is well developed as compared to others of a younger age. They were selected randomly from local contacts and their identity was kept in privacy. The voice recording procedure was conducted in a normal room at a noise free surrounding. Each recording was conducted with consent of individual. The subjects used in our study were not suffering from any type of auditory, speech & visual disorder. During the process of recording subjects were not suffering from cold as well as any illness. None of the subject was under the influence of any drug or alcoholic stage during the process of data collection.

## Analysis

**Feature extraction:** After the completion of data collection, voice samples are repeatedly listened carefully and extract the features/words that occur frequently in the transcripts of mother and daughter that are further analyzed by Voice Biometric Software such as 'Gold Wave'. The words tested were "BBC News", "kaha hai ki unka desh", "karwai kay madey nazar", "ki gyi jordan", "Syria me irani. These extracted words of each mother and daughter were tested and further analyzed.

**Feature analysis:** After the completion of Feature extraction, the samples (Sample1-20) mentioned here are labeled as **S1, S2, S3,....., S20** respectively. The extracted phonate "BBC News" marked as "**M1**" for mothers, "**D1**" for daughters. The phonate/feature "**Syria me irani**" marked as "**M2**" for mothers and "**D2**" for daughters. For mothers, the extracted word/ sample "ki gyi jordan" marked as "**M3**" and for daughters as "**D3**". Extracted word "karwai kay madey nazar" marked as "**M4**" for mothers and "**D4**" for daughters and word "kaha hai ki unka desh" marked as "**M5**" for mothers and "**D5**" for daughters of all samples. The extracted features of each mother and daughter samples were analyzed by using Voice Biometric software such as Gold Wave.

**Feature Comparison: Computer Speech Laboratory (CSL) or Speech Science Laboratory (SSL)** is a personal computer based instruments used for this purpose which records, edits and quickly analysis the speech signal. The sound energy converted to electrical energy when the speech is subjected to spectrographic analysis which operates the stylus and creates a trace in form of graph called **Voiceprint or spectrogram**.

On horizontal left to right plane, it shows sound duration (in milliseconds) and on vertical bottom to top shows speech sound characteristics like frequency. After obtaining spectrograms, spectrographic analysis can be done by analyzing the formants formed in the spectrograms. The analyzed phonates/words were compared by spectrographic analysis. Further, spectrographic analysis can be done using SSL software. Samples are compared through Spectrogram analysis by using Speech Science Laboratory. The Spectrogram formants of both the voice samples (mother and her daughter) are then compared visually. In the spectrographic analysis, the number of formants formed, the intonation patterns of formants and the formant frequency distribution of formants in the mother's spectrograms are calculated and compared with the number of formants formed in the spectrogram of her daughter's voice.

## Results and Discussion

Results of our study showed that Visual comparison work of spectrograms of mother voice samples marked exhibit "S1(M1-M5) and daughter voice sample marked exhibit "S1(D1-D5) have been rendered. According to [Figure 1-100](#), it is observed that the number of formants and their positions are 88%-92% similar in both mother and her daughter's spectrograms which shows that they have a high level of similarity and due to law of individuality, similarities are not 100%. The intonation pattern of formants and the formant frequency distribution of formants of mother are observed similar/ homogenous to her daughter.

Similarly, visual comparison work of spectrograms of mother voice samples marked exhibit "S2-20(M1-M5)" and daughter voice sample marked exhibit "S2-20(D1-D5)" have been rendered. According to [Figure 1-100](#), it is observed that the number of formants and their positions are 88%-92% similar in both mother and her daughter's spectrograms which shows that they have a high level of similarity and due to law of individuality, similarities are not 100%. The intonation pattern of formants and the formant frequency distribution of formants of mother are observed similar/ homogenous to her daughter. This is the result that we are expected taking into account the degree of shared genes and shared environmental factors by pairs. A voice is driven by anatomical structure as well as non-biological (behavioral factors), dialectal aspects, environmental effects, mimicking characters. Qualities of individual's voice are driven from corporeal inheritance. The results allowing insightful discussion concerning the influence of genetic endowment and environmental factors in the type of speakers analyzed. The results of comparing each mother speaker with her pair (daughter) support or validate the hypothesis that the higher similarity values would be found in mother and her daughter's voice. The influence of genetic endowment and environmental factors are the most common factors of such higher similarity. Finally, in view of the results obtained, it is concluded that the mother and daughter voice samples have higher similarities which is estimated to 88%-92%. It is due to the influence of genetic endowment, physical similarities and inheritance of vocal gesture & posture. Also, similarities cannot be 100% efficient due to law of individuality.

Further in continuation to the research, it would be important in order to identify the role of some specific parameters formed in the results from each comparison in samples of mother and daughter. A reanalysis with more subjects shall be performed in future for better analysis using world class workbench and voice analysis tools. Also, as it is followed from the past research papers' review carried out in [\[16\]](#) and which is summarized in the introduction format to this research. This heterogeneity mostly appears as a main factor in various studies on twins' voices.

Past various analyses obtained from the same collection of Spanish twins showed the same results, that different pairs of twins exhibit different variation in results when an IP differentiation is carried out, regardless of phonetic speech examination, be it a formant ambit or other characteristics as likely as glottal [\[9\]](#). Though, these are not needed to be a characteristic apparently pointed for twins but are common in speaker/speech recognition. Have gone through the previous researches [\[8, 9, 11, 12, 16\]](#).

Künzel in his study showed that there is an important gender related difference in the presentation of the automatic system, this being superior for male as related to female voices [\[17\]](#). In a study made by Kim showed that analyzed twins' voices is focused only on female voices and these studies are almost comparable with our study [\[18\]](#).

In this study, an automatic speech recognition system are used to determine the similarity or variation between mother and her daughter's voice, which consist of combination of automatic speech recognition software as a feature extraction technique combined with formants by spectrographic analysis. The voice identification includes the feature extraction & feature comparison in which the features/words that occurs frequently can be

extracted, analyzed & compared with the specimen using voice biometric software such as Gold wave. Speech was recorded using a mobile phone first, and then a computer microphone with sampling frequency of 8000 Hz at Wave PCM signed 16-bit per sample along with one channel (mono) WAV format is used at 128 kbps bit rate. There are many different aspects which are discussed in the results obtained with the automatic speech recognition system. The intonation patterns of formants of mother spectrogram of speech exhibit are compared with the intonation pattern of formants of daughter spectrogram of speech exhibit to determine the similarity or variation. The formant frequency distribution of formants in mother spectrogram of speech exhibit is compared with the formant frequency distribution of formants in daughter spectrogram of speech exhibit.

The results obtained from spectrographic analysis using SSL have a higher similarity between mother and their daughter's voices. In spectrographic analysis it is found that the number of formants formed and their positions are similar in the spectrograms of both mothers and her daughter's voice samples. The intonation patterns of formants of mother spectrogram of speech exhibit are compared with the intonation pattern of formants of daughter spectrogram of speech exhibit to determine the similarity or variation. The formant frequency distribution of formants in mother spectrogram of speech exhibit is compared with the formant frequency distribution of formants in daughter spectrogram of speech exhibit.

## Conclusions

This chapter summarizes the results of voice biometrical match of mother and daughter. Speech provides a vital conversation media between people. Vocal tract consists of oral, nasal and pharyngeal cavities. Each of which has a plangent profile, which is typical and idiosyncratic for each speaker/person, at least similarly to what happens with other parts of the human anatomy, which are more or less individual. Automatic speech recognition system extricates features depicts the vibration profile of the speaker's voice cavities which generates a multifaceted direction. These are some of criteria (low-level features) that are useful in the analysis and with high-level features, which attributes to other phonological aspects that also helps to characterize a speaker, such as intonation patterns, pausing behavior, vernacular etc. The present study is an attempt to report a novel approach to the speaker recognition and comparison of the voice of mother and her daughter by the use of voice biometric software. An attempt has been made to enlighten the role of voice as evidence in forensic investigation. A voice is driven by anatomical structure as well as non-biological (behavioral factors), dialectal aspects, environmental effects, mimicking characters. Qualities of individual's voice are driven from corporeal inheritance. All the facts or features are typically mentioned in the literature to explain the similarities in mother and her daughter's voices, namely, genetic and environmental factors. In future research, we would recommend to incorporate the factor of epigenetic (which explains the alteration in the expression of specific genes caused by mechanisms other than changes in the underlying DNA sequence) which is often neglected in this

research because it may play an important role in our understanding of the striking dissimilarities found for some pairs. Finally, in view of the results obtained, it is concluded that the mother and daughter voice samples have higher similarities which is estimated to 88%-92%. It is due to the influence of genetic endowment, physical similarities and inheritance of vocal gesture & posture. Also, similarities cannot be 100% efficient due to law of individuality. This research shall enlighten an idea of getting corroborative evidence, further which can be a part of Forensic Investigation. Forensic experts can contribute to society/law with advance as well as fast crime detection solutions.

## References

- [1]. M Kalamani, S Valarmathy, C Poonkuzhali, Catherine JN . Featuuer selection algorithms for automatic speech recognition. International Conference on Computer Communication and Informatics. (2014). IEEE, Coimbatore, India.
- [2]. Chao Wang , Ruifei Zhu , Hongguang Jia , Qun Wei ,Huhai Jiang,Tianyi Zhang ,et al., Design of speech recognition speech. 2013 IEEE Third International Conference on Information Science and Technology (ICIST), 2013. IEEE, Yangzhou, China. 152.
- [3]. BR Sharma.Forensic Science in Crime Investigation and Trials. 5th ed. Universal Law Publishing Co. Pvt. Ltd;2014.
- [4]. Sondhi MM, Schroeter J. Speech production models and their digital implementations. The Digital Signal Processing Handbook, VK Madiseti, DB Williams (Eds.), CRC Press, Boca Raton, Florida. 1997:44-1.
- [5]. Emeeshat JS. Isolated Word Speech Recognition System for Children with Down Syndrome (Doctoral dissertation, Youngstown State University). 2017.
- [6]. Pires JN. Industrial robots programming: building applications for the factories of the future. Springer Science & Business Media; 2007.
- [7]. Van Lierde, KM, Vinck, B, De Ley S, Clement Z, and Van Cauwenberge P. Genetics of vocal quality characteristics in monozygotic twins: a multiparameter approach. *J Voice*. 2005 Dec;19(4):511-8. PubMed PMID: 16301097.
- [8]. Fernández ES. A phonetic corpus of Spanish male twins and siblings: Corpus design and forensic application. *Procedia-Social and Behavioral Sciences*. 2013 Oct 25;95:59-67.
- [9]. Segundo ES, Gómez-Vilda P. Voice biometrical match of twin and non-twin siblings. In *Proceedings of the 8th International Workshop Models and analysis of vocal emissions for biomedical applications*. 2013; 253-256.
- [10]. Segal N. The importance of twin studies for individual differences research. *J. Couns. Dev*.1990; 68(6):612-622.
- [11]. San Segundo. Glottal source parameters for forensic voice comparison: An approach to voice quality in twins' voices .Paper presented at the 21<sup>st</sup> Conference of the International Association for Forensic Phonetics and Acoustics. 2012; Santander, Spain .
- [12]. San Segundo E, Künzel H. Automatic speaker recognition of Spanish siblings:(monozygotic and dizygotic) twins and non-twin brothers. *Loquens*. 2015 Dec 30; 2(2):021.
- [13]. Singh N, Khan RA, Shree R. Applications of speaker recognition. *Procedia engineering*. 2012 Jan 1;38:3122-6.
- [14]. Deliyiski DD, Shaw HS, Evans MK. Influence of sampling rate on accuracy and reliability of acoustic voice analysis. *Logopedics Phoniatrics Vocology*. 2005 Jan 1;30(2):55-62.
- [15]. Deb S, Dandapat S, Krajewski J. Analysis and classification of cold speech using variational mode decomposition. *IEEE Transactions on Affective Computing*. 2017 Oct 10. 99:1-1.
- [16]. San Segundo Fernández E. Forensic speaker comparison of Spanish twins and non-twin siblings: A phonetic-acoustic analysis of formant trajectories in vocalic sequences, glottal source parameters and cepstral characteristics (Doctoral dissertation, Doctoral Thesis, UIMP).2014.
- [17]. Künzel HJ. Automatic speaker recognition of identical twins. *International Journal of Speech, Language & the Law*. 2010 Dec 1;17(2):251-277.
- [18]. Kim K (2010) Automatic speaker identification of Korean male twins. In *Proc. 19th Ann. Conf. of the Int. Assoc. for Forensic Phon. and Acoust. (IAFPA)* .2010 Jul ; 21.