==========
Language in India www.languageinindia.com**ISSN 1930-2940 Vol. 19:7 July 2019**
==========

Forensic Speaker Identification: A Review of Literature and Reflection on Future

Neelu, Ph.D.
GNDU, Amritsar
2225neel@gmail.com

Mukesh Kumar, Ph.D.
IIM Amritsar
mukeshjnu@gmail.com

==========

Abstract

Forensic speaker identification is a decision-making process which determines whether a given utterance has been spoken by a particular person or not. To achieve this goal, a number of speech and voice features are evaluated. The process of Speaker Identification results in either positive identification, i.e. affirming that two voice samples belong to the same speaker or it results in the negative identification, i.e., eliminating the possibility of two voice samples coming from the same speaker. Earlier studies in speaker identification have tried to explore the components of human voice such as pitch, intensity, amplitude, intonation etc. and how they have modulated in ways that it becomes different for different individuals. A number of studies, conducted previously, show that pitch proves to be an excellent parameter in speaker identification. Pitch and intensity have been studied in various contexts such as a text-dependent vs. text-independent, single word utterance vs. continuous speech, monolingual vs. multi-lingual, reading text vs. spontaneous speech, same-sex vs. different sex etc. However, despite much research in this area, there remains uncertainty around robustness of these voice features. The present paper, therefore, besides sharing the origin and long tradition of research in speaker identification, offers a view of research gaps in this area.

Keywords: Forensic Speaker Identification; Fundamental Frequency; Speaker Identification; Language and Gender Independent Variables, Voice Cues

Forensic speaker identification is a branch of forensic phonetics which falls within the scope of applied phonetics. It is a decision-making process that uses some features of the speech signal to determine if a particular person is the speaker of a given utterance. But unlike fingerprints, the conclusion or the outcome in the process of FSI is not absolute, it is always probable. The aim of FSI is, 'to identify an unknown voice as one or none of a set of known voices' (Naik, 1994, pp. 31-8). There are unknown speech samples available from the crime scene which are matched against the known samples of suspects. The purpose is to determine whether the utterance in the unknown sample is produced by one of the suspects in the known sample or not.(Nolan, 1983).

In a number of cases of threat calls, bribery, kidnapping, terrorist activities, etc., audio tapes play a vital role in the judgment process. These audio samples help in identifying the person involved in the case concerned and can be very useful either in determining the crime of the criminal or proving the innocence of the suspect. This can be done by acoustically analyzing the recorded speech sample using the sound spectrogram. Then there is a visual comparison of graphic patterns between the question sample and suspect sample. The spectrographic analysis is the primary tool used in FSI while, the visual comparison of spectrograms helps in giving a subjective judgment about the identity of a speaker. As opinions from forensic experts are increasingly sought in courtrooms, FSI proves to be an effective tool in both conviction and acquittal of suspects.

A forensic linguist may have to identify a speaker in three different situations:

1. There are one questioned sample and one suspect.
2. There are one questioned sample and many suspects.
3. There are one questioned sample and no suspect. (Rose, 2002)

Identification of speakers by their voices may be seemingly easy under ideal conditions. This has already been manifested by automated identification systems. Humans recognize familiar voices, fairly successfully all the time. This probably entitles us to assume that different speakers of the same language do indeed have different voices. So, we do have to deal with variations between speakers, usually known as between-speaker or inter-speaker variation.

There are differences in the speech of the same speaker as well, known as within speaker or intra-speaker variation. For FSI to work effectively, these variations in speech have to be evaluated correctly. These differences are mostly audible, and always measurable and quantifiable. FSI involves being able to tell whether the inevitable differences between samples are more likely to be within-speaker differences or between-speaker differences. In this review paper, we have shared the history of speaker identification and current developments in the recent years. By presenting a comprehensive overview of research studies in the area of Forensic speaker identification, the paper helps in identifying the gaps pertaining to research in this area.

History of Speaker Identification

Speaker identification has always been a part of our daily lives, in some form or the other. It begins right from the womb of a mother where a child begins to identify her/his mother's voice as a primary function of aural perception. We are under the influence of external auditory stimuli even before birth (DeCasper & Sigafoos, 1983)(Spence & DeCasper, 1987) (Ramus, Hauser, Miller, Morris, & Mehler, 2000). It seems possible that we focus more on voice recognition first and understanding a language later on (DeCasper & Fifer, 2004).In spite of that, we are able to discriminate between languages through speech rhythm at an early age (Nazzi, Bertoncini, & Mehler, 1998).

However, there is a close connection between the way speech of an individual is analyzed and the way a person's voice is analyzed. Even as newborns, we follow the same technique for analyzing speech and voice (DeCasper & Spence, 1986). Therefore, it is important to separate the inherent co-analysis of speech and voice.

The kind of voice/ speaker identification we do in our daily lives is considered naïve speaker identification process. In many experiments, it has been seen that this type of speaker identification process varies depending upon how different listeners respond to different signals in different situations (Ramos, Franco-Pedroso, & Gonzalez-Rodriguez, 2011).

But evidence from such listeners is no more accepted in courtrooms unless they are supported by an expert in speaker identification. An expert in FSI is someone who is well educated on the various parameters that describe speech and voice features and their variability in a structured manner (Schwarz et al., 2011).

In earlier times, along with many other kinds of evidence, voice and speech evidence was also considered reliable depending on upon who the witness was and how he gave the testimony. One such example where voice and speech evidence was used in a legal system is the trial of William Hulet in 1660 (Erikkson, 2005).

One of the witnesses had heard the voice of the person who executed King Charles I and declared that he recognized that speech to be of Hulet. The witness had known Hulet very well in the past so he could easily identify that the voice of the executioner and that of Hulet was same. As a result, Hulet was sentenced to death. But he was acquitted later as the real executioner, a hangman, confessed his offence. Such misidentifications were not uncommon in those times and probably happen today as well. This is just one of the examples which shows inaccuracy and unreliability of naive speaker identification. Other issues associated with naïve speaker identification arose because of the absence of recorded speech. Witnesses had to depend solely on their memory to identify the speaker. But with the delay in time, the memory of the witnesses would wear out, leading to misidentifications.

Speaker identification made by experts did not begin until speech recorders were invented. Even after the invention, it was not practical to carry recorder to every possible crime scene to record voices. But when the usage of telephones became more frequent, crimes committed over the network also became regular. It was around this time that the idea of visualization of recorded speech for its analysis floated in the world of Acoustic Phonetics.

The idea that someone could be identified through his/her voice came over hundred years ago to Alexander Melville Bell (father to Alexander Graham Bell). He developed a visual representation of how a word would look like which was based on pronunciation. He showed that there were very subtle differences among people who spoke same things. In

1941, a sound spectrograph was developed in the laboratories of Bell telephone which could map voice on a graph. It could analyze sound waves and produce a visual record of voice patterns based on frequency, intensity, and time. It was classified as a war project until the end of World War II. As a result, unfortunately not much was published on this innovative technology (Potter, 1945). The prime motive for the development of this technology was to progress research on speech and acoustic speech patterns. Another purpose was to implement this spectrographic technique in different applications for the hearing impaired. During the World War II, it was used by acoustic scientists to identify voices of the enemies on telephones and radios. However, with the end of the war, the urgency for this technology diminished and little came of it until later.

The post-war development of Speaker Identification saw the emergence of voiceprints which was followed by a huge controversy. The visual mapping of speech sounds on a spectrogram was called a voiceprint and was used as a direct analogy to fingerprints by some researchers. It was later greatly criticized, and questions were raised on the accuracy of voiceprint results.

According to extant research, voiceprints can match the accuracy of fingerprints if they are done properly. However, he agrees to the fact that fingerprints have static images that don't change unless the fingerprint ridge detail gets damaged. Voiceprints are dynamic owing to the fact that no two utterances that an individual speaks are exactly same. They may change in pitch or stress etc. Therefore, to find the range of variation in a speaker's speech, several repetitions of a speaker's voice must be recorded and analysed.

Types of Speaker Identification

The task of speaker identification can be classified into different types. One of the classifications is closed-set versus open-set speaker identification (Kekre, 2013).

Closed set speaker identification: In this case, there is a given set of unknown speakers and a questioned sample. The questioned sample is matched with all the samples in the unknown speakers' set. The template from the unknown set which shows maximum similarity with that of the questioned sample is obtained. It is then assumed that the speaker of the questioned sample and the speaker from the unknown set who had a matching template are the same. Hence, in a closed set speaker identification system, one is forced to arrive at a decision by choosing the best matching template from the given database.

On the other hand, in an open set system, there is a questioned sample but there is no set of unknown speakers available to match the template. In the absence of a set of unknown speakers, the identification process becomes long and tedious.

A speaker identification process can also be classified into text-dependent and text-independent. The text-dependent system can also be called the constrained mode while the

text-independent system can be called as the unconstrained mode. In a system that uses text-dependent speech, individuals know the words and phrases beforehand. They just have to repeat the same utterances provided to them or as prompted by an expert. These utterances are then later on analyzed for the identification process. Text-independent identification shows a better performance with cooperative users. However, there are cases when suspects refuse to utter same phrases or disguise their voice purposely.

In a text-independent system, the speakers have no prior knowledge of what phrases they are supposed to utter. This system is more flexible in situations where a participant is unaware of the fact his/her voice sample is being obtained or in cases where suspects are unwilling to cooperate. Here, the analysis is not done on the basis of the content rather a modelling of the general underlying properties of speakers' vocal spectrum is done. (Committee on Homeland and National Security; National Science and Technology Council; Committee on Technology)

Methods and Approaches in Speaker Identification

The process of Speaker Identification results in either positive identification, i.e. affirming that two voice samples belong to the same speaker or it results in the negative identification, i.e., eliminating the possibility of two voice samples coming from the same speaker.

Some researchers believe that the reliability of Speaker Identification has been overestimated. The complexity of spoken communication makes Speaker Identification a difficult task; hence its forensic application must proceed cautiously. Since the beginning, both objective and subjective methods have been used in voice identification.

1. Objective methods relied mostly on equipment which made all decisions. These methods used automatic pattern matching of voice patterns. In one such study conducted on ten speakers, the average spectral patterns of all the speakers were obtained. These spectral patterns were stored in a computer. Later, a new pattern was obtained from each of the speakers and these spectral patterns were matched with the patterns that were already stored in the computer. The study showed almost 10 percent identification error.

2. In a subjective method, equipment such as a sound spectrograph is involved to obtain acoustic information, but the final judgement is made by an expert who carefully evaluates the available information to arrive at a decision. There are two types of subjective experiments that use spectrograms. The first type is a sorting experiment. In this experiment, the expert has sets of spectrograms of a token word spoken by different individuals at different points of time. The task of the expert is to sort those sets of spectrograms which belong to the same speaker. The second type of subjective experiment is the matching experiment. Here, the expert identifies spectrograms of

one speaker by matching them against the spectrograms in a catalogue of speakers who have uttered the same token word.

In the earlier history of FSI, primarily, two approaches were employed; the auditory approach and the acoustic approach. The differences between the two methods owed to three radically different positions. All of these three positions are encountered during the presentation of evidence in Forensic Speaker Identification.

- The auditory analysis is sufficient on its own. (Baldwin & French, 1990)
- That auditory analysis is not necessary at all. It can all be done with acoustics.
- That auditory analysis must be combined with other that is an acoustic method (Kunzel, Sprechererkennung: Grundzüge forensischer Sprachverarbeitung, 2002); (French, 'An overview of forensic phonetics with particular reference to speaker identification, 1994, pp. 173-174)

The third approach was a hybrid of the first two approaches and is more accepted today. It is known as the phonetic-acoustic approach. The strengths of the first two approaches are rather complimentary than being used as a substitute for one another in an investigation in FSI. However, the collaborative procedure depends upon the quality of the sample and the information they carry. (Nolan, Speaker Recognition and Forensic Phonetics, 1997, p. 765). These days, the joint acoustic-phonetic approach is recommended by more and more professional bodies, internationally (French, 1994). Given the indispensability of both the approaches, there is a general consensus on the usage of the hybrid approach to FSI where auditory approach must logically precede acoustic analysis.

Validation of Methods

Despite the methods being used in a structured way, validation of voice identification methods was always demanded by scientists. Speaker identification based on voice patterns was not considered reliable because even the small-scale matching experiments did not show hundred percent identification results. Also, even though the experimental methods were explicitly described, they would differ when practically applied in identifying an individual on the sole basis of his/her voice patterns.

Identification by experts was questioned because they lacked explicit knowledge and procedures. So, their opinions were not accepted as reliable. It was believed that the possibility of human eye and brain to identify a speaker on the basis of voice patterns was there, but it could not be assumed without proof. A number of suggestions followed to make the results of speaker identification valid. One of the suggestions was to develop explicit procedures based on specifications of voice features useful for identification. Another one was to develop statistically valid models for the subjective experiments. It was said that test formats should be such that they yield information about the probabilities of missed identification as well as false identification. They should also give information about the effect of various factors such as the size of the population, context of speech token, changes

in voice pattern due to noise, disguised voice etc. (Bolt, Cooper, David, Denes, Pickett, & Stevens, 1969).

Auditory and Acoustic Parameters Used in Forensic Speaker Identification

For any speaker identification study, it is important to have knowledge of more powerful dimensions and parameters because usually there is limited time and it cannot be wasted on comparing samples with respect to weak dimensions.

For comparing speech samples forensically, phonetic parameters are categorized in line with two main distinctions:

1. Whether the parameters are auditory or acoustic.
2. Whether they are linguistic or non-linguistic.

Earlier, auditory parameters were used to describe and compare voices depending upon how a voices sound to an observer. These observers were trained in recognizing and transcribing auditory features.

Acoustic parameters, on the other hand, are self-explanatory. After the invention of the spectrograph, acoustic parameters started being considered for comparing voice samples. Today, comparing voice samples with respect to their acoustic properties extracted by computer is perhaps what first comes to mind when one thinks of FSI.

As parameters for comparison the International Association for Identification (IAI) protocol lists general formant shaping and positioning, pitch variations, energy distribution, word length, coupling (how the first and the second formants are tied to each other) and a number of other features such as plosives, fricatives, and formant features (Gruber & Poza, 1995).

The FBI protocol states that examiners make spectral pattern comparison between the two voice samples by comparing beginning mean, and end formant frequencies, formant shaping, pitch timing, etc of each individual word.

Visual comparison of spectrograms involves, in general, the examination of spectrograph features of like sounds as portrayed in spectrograms in terms of time, frequency, amplitude, aural cues include resonance quality, pitch, temporal factors, inflections, dialect, articulation, syllable grouping, breath pattern disguise, pathologies, and other peculiar speech characteristics (AFTI, 2002)

In the past 50 years, speaker and speech recognition technology has made very noteworthy progress. Some of the changes that took place during this progress have been cited here. Speaker recognition began with template matching and now it has moved on to

corpus-based statistical modelling, maximum likelihood to discriminative approach, small vocabulary to large vocabulary recognition, isolated word to continuous speech recognition, from clean speech to noisy/telephone speech recognition, text-dependent to text-independent recognition, single modality (audio signal only) to multi-modal (audio/visual) speech recognition, no commercial application to many practical commercial applications. The majority of transformations have been directed towards increasing robustness of speaker and speech recognition.

It can be understood by the above review that parameters of voice play a significant role in speaker identification process. The present research tries to figure out which of them are more crucial in the process of identification.

It has been proved that in speaker identification, the accuracy of the analysis increases when both auditory and acoustic parameters are used in a combination. The auditory analysis always precedes the acoustic analysis.

Robustness of Fundamental Frequency and Intensity in Research on Speaker Identification

Robustness is a key requirement for forensic speaker-comparison parameters. F0 seems to fulfil several criteria. It is because of this reason that fundamental frequency is one of the most frequently studied parameters. But, given the variations that can occur in an individual's speech, the task for the forensic phonetician involves being able to tell whether the inevitable differences between samples are more likely to be within-speaker differences or between-speaker differences. (Rose 2002: 10). An ideal parameter is the one which shows less of within-speaker variations and more between-speaker variations. According to the criteria listed for an ideal parameter by Nolan, f0 meets most of them.

It must be noted that the pitch shows high speaker variability when the lower range of f0 of speakers is compared. (Neelu, 2012, p. 72). It shows a high frequency of occurrence in speech samples. It is also easily extractable and measurable as we use a lot of vowels in our speech and we can easily extract and measure pitch from vowels through many software. Since f0 is a parameter of voice at the source level; it is also maximally independent of other acquired parameters.

In "An Overview of Text-Independent Speaker Recognition: from Features to Supervectors" by Tomi Kinnunen and Haizhou Li, a diagrammatic representation of the characteristics of parameters in forensic speaker identification has been presented. They restate that the choice of parameters should be based on their discrimination, robustness and practicality. It must be noted that though in the diagram the high-level features are shown as robust, they are less discriminative and easier to impersonate. It is quite possible for a mimicry artist to imitate the accent of a person. But the pitch which falls in between the

learned/acquired parameters and physiological/inherent parameters can be considered as both robust and easily extractable.

It has also been shown in an experiment that, in backward speech, the important features of voice that are retained are pitch and pitch range. (Lancker, Kreiman, & Emmorey, 1985, pp. 19-35). The results of this experiment reported that f0 and f0 contour are primary cues to familiar speaker recognition. In the backward presentation of speech, most of the articulatory and sequential characteristics get distorted. It is only f0 which is retained along with some other features such as speech rate, voice quality and vowel quality.

The intra-speaker variation in fundamental frequency is affected by paralinguistic and other factors. They can be categorized into physiological, psychological and technical factors. Physiological factors may include prolonged smoking or drinking and age of the speaker. Technical variations arise mostly due to tape speed and sample size, while emotional state of the speaker is an example of psychological factors. Fundamental frequency is a widely studied parameter in FSI and holds significant relevance in spite of these sources of variations. (Braun, 1995) for example, quotes four well-known authorities (French P., 1990)(Hollien, 1990)(Kunzel, Sprechererkennung: Grundzüge forensischer Sprachverarbeitung, 2002) and (Nolan, The Phonetic Bases of Speaker Recognition, 1983) who claim that it is one of the most reliable parameters.

Foulkes et al., in an article, 'Telephone Speaker Recognition amongst Members of a Close Social Network' mentions that the speakers with relatively higher upper range and/or low lower range of average F0 values were consistently identified with greater accuracy. This also held true for speakers who had the widest and the narrowest overall F0 range. The ones with average pitch values and pitch ranges which don't extend to either side of the extremes were rather difficult to identify. The findings of the study support the view that average F0 is a robust parameter in FSI, which also helps in naïve speaker identification.

In his article 'Speaker classification in Forensic Phonetics and Acoustics', Michael Jessen has argued if pitch (F0) or other formants can help in deciding the gender of the speaker. He concludes that average pitch is one of the strongest parameters in identifying the gender of the speaker. In most cases, auditory examination of pitch level suffices to accurately distinguish between male and female speakers. However, there are also cases of voice disguise or mimicking, where the speaker may impose a false voice creak or whisper. Pitch may either not be accessible or informative in circumstances like these. There are also situations where a speaker may have unusually high or low pitch, incongruous to the gender group he/she belongs to. This can result into a mismatch in gender identification based on pitch. In such cases, acoustic analysis plays a vital role by providing measurements of formant frequencies. It's a well-known fact that men on average have lower formant frequencies than women which can be attributed to their vocal cords which are generally longer than those of women.(Jessen)

In another study, the weight of fundamental frequency as a discriminatory parameter for sex identification has been stressed upon. The study has been conducted on transsexual voice where it becomes difficult to categorize a transsexual into male voice or female voice. Sometimes they also try to disguise or modulate their voices. The study explains that when a female vocal fundamental is modulated by a male, the vocal tract retains some of the male qualities to which listeners are perceptually sensitive. This is because the fundamental frequency can be changed but since vocal cords have fixed dimensions, it is difficult to completely wipe out the maleness in the voice quality. (Coleman, 1983) (Trollinger, 2003)

It was indicated in a speech science research concerned with the vocalizations of pre-language infants that they tend to experiment with their vocalizations via trial and error. The research suggests that initially the sounds used in vocal experimentation are reflexive, but later on the child develops vocal patterns that are appropriate to his or her culture. This happens gradually via imitation and learning (Andrews, 1999; Kuehn, 1985).

A number of studies have suggested that boys' and girls' speaking voices are similar in fundamental frequency before the onset of puberty (Bennett, 1983; Bennett & Weinberg, 1979; Kahane, 1975; Kent, 1976, Wilson, 1987, Titze, 1992).

Scores of studies have emphasized that F0 is a valued parameter in speaker identification for the amount of information that it encapsulates about the speaker. F0 is influenced by a number of linguistic and non-linguistic features such as stress, tone, intonation, type of sentence, sex of speakers, properties of neighbouring vowels and consonants etc. Apart from this, speakers themselves are random effect; their normal pitch range, shape and size of vocal cords, age, health conditions, all of these interact in significant ways and are responsible for F0. These universally available effects are combined in unique ways by different speech communities, for example, Japanese women, while speaking manifest higher pitch than Dutch women (VanB ezooijen, 1995). It’s a challenge for linguists to model the way in which speakers collaborate these effects at their disposal to produce F0 in a manner that it is in congruence with their speech community (Aston, Chiou, & Evans, 2010)**.** This study points towards the language dependency of the pitch.

There have been studies which have drawn a relationship between the fundamental frequency of voice and cognitive speaker identification. It is a common knowledge that we can decode speech into language independently of who is speaking, and we can also recognize who is speaking independently of what he/she is speaking. According to the cognitive and connectionist models, this efficiency depends upon the ability of our speech perception and speaker identification systems to extract relevant features from the sensory input and to form efficient abstract representations.

However, it remains unclear how a speech form turns into a speaker's identity. Results of functional magnetic resonance adaptation suggest that there is an area specialized for voice identification in the right anterior superior temporal sulcus.

These results provide empirical support for cognitive models of speech and voice processing postulating the existence of intermediate computational entities resulting from the transformation of relevant acoustic features of vowels and F0 for speakers and the suppression of the irrelevant ones. This is an important revelation where fundamental frequency of speaker aids in cognitive voice processing and speaker identification. (Formisano, De Martino, Bonte, & Goebel, 2008)

The individual variations in F0 are often described through long-term distribution measures such as standard deviation, long-term fundamental frequency and arithmetic mean (Rose, 2002). These measures depend on the duration of an utterance. There is, however, no common consensus on minimum duration of an utterance required to yield reliable outcomes. Using traditional measures such as mean and median values of fundamental frequency have been believed to produce misleading results. This may possibly happen because the mean F0 has a roughly normal distribution across the population (Lindh, 2006). Hence, its forensic value is inherently limited and could only offer any contribution in FSI when extreme values are present.

Conclusion

It has been found that, so far, the pitch has been studied extensively and proved to be a robust parameter in speaker identification. The intensity or vocal energy, on the other hand, has been paid little attention. Studies have suggested that in different speech styles, intensity seems to make a contribution in speaker identification (Kraayeveld, 1997). Although vocal intensity has been recognized as an identification feature, it has not been extensively investigated. Therefore, we do not have a good understanding of its general nature and whether it can be termed as a speaker-specific parameter or not. The little information that we have about vocal intensity suggests that it is not a robust parameter for identifying speakers. This reason behind it is vocal intensity can fluctuate with even a little variation in the external environment. Having said that, it is nevertheless a noticeable feature that people talk at varying intensity and they also modulate it depending on the context. Therefore, it can be theorized that if the processing of vocal energy can be controlled, the evaluation of this parameter can prove to be useful in identification of speakers (Hollien F., 2002).

Besides this, the present review of literature, aimed at exploring several aspects of pitch and intensity parameters has helped in finding research gaps with respect to the use of parameters that remain stable even with the change in the linguistic and non-linguistic environment and provide useful insights when speakers of different genders are involved. Even though a number of studies, conducted previously have shown that pitch proves to be the most robust parameter in speaker identification, we could further investigate the components of the pitch, including F0, concerning how robust their results remain as

speakers switch languages. Using existing voice analysis tools like Praat and MDVP, various aspects of pitch in combination with amplitude may be investigated to obtain the properties of f0 and intensity which are minimally influenced by internal and external variations but are maximally discriminatory in nature. Future studies can, therefore, focus on various aspects of pitch and intensity such as jitter, shimmer, mean fundamental frequency along with duration etc. Insights from these studies will make the exercise of speaker identification far more reliable.

==========

References

AFTI. (2002). Voice Print Identification. In P. Rose, *Forensic Speaker Identification.* New York: Taylor & Francis. Retrieved from Applied Forensic Technologies International, Inc.

Andrews, M. L. (1999). Manual of voice treatment: Pediatrics through geriatrics. Cengage Learning.

Aston, J. A., Chiou, J. M., & Evans, J. P. (2010). Linguistic pitch analysis using functional principal component mixed effect models. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 59*(2), 297-317.

Baldwin, J., & French. (1990). *Forensic Phonetics.* London: Pinter.

Bennett, S., & Weinberg, B. (1979). Acoustic correlates of perceived sexual identity in preadolescent children's voices. Journal of the Acoustical Society of America, 66, 989-1000.

Bennett, S. (1983). A three-year longitudinal study of school-aged children's fundamental frequencies. Journal of Speech and Hearing Research, 26, 137-141.

Bolt, R. H., Cooper, F. S., David, E. E., Denes, P. B., Pickett, J. M., & Stevens, K. N. (1969). Identification of a speaker by speech spectrograms. *Science, 166*(3903), 338-343.

Braun, A. (1995). Fundamental frequency – how speaker-specific is it? In A. Braun, & J. P. Koster, *Studies in Forensic Phonetics: Beiträge zur Phonetik und Linguistik 64* (pp. 9-23). Buske.

Coleman, R. O. (1983). *Acoustic Correlates of Speaker Sex Identification: Implications for the Transsexual Voice.*

Committee on Homeland and National Security; National Science and Technology Council; Committee on Technology. (n.d.). *Speaker Recognition.* Retrieved November 2017, from www.fbi.gov: https://www.fbi.gov/file-repository/about-us-cjis-fingerprints_biometrics-biometric-center-of-excellences-speaker-recognition.pdf/view

DeCasper, A. J., & Fifer, W. P. (2004). On Human Bonding: Newborns Prefer Their Mothers' Voices. *Readings on the Development of Children*, 1174-1176.

DeCasper, A. J., & Sigafoos, A. D. (1983). The Intrauterine Heartbeat: A Potent Reinforcer for Newborns. *Infant Behaviour and Development*, 19-25.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal Maternal Speech Influences Newborns' Perception of Speech Sounds. *Infant Behaviour and Development, 9*(2), 133-150.

Erikkson, A. (2005). Tutorial on forensic speech science. *European Conference on Speech Communication and Technology*, (pp. 4-8).

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). " Who" is saying" what"? brain-based decoding of human voice and speech. *Science, 322*(5903), 970-973.

French, P. (1990). Acoustic Phonetics. In J. Baldwin, & P. (. French, *Forensic Phonetics* (pp. 42-63). London: Pinter.

French, P. (1994). 'An overview of forensic phonetics with particular reference to speaker identification. *Forensic Linguistics*, 169-184.

Gruber, J. S., & Poza, F. T. (1995). Voicegram identification evidence. In *American Jurisprudence Trials 54.* Lawyers Cooperative Publishing.

Hollien, F. (2002). *Forensic voice identification* . Academic Press.

Hollien, H. (1990). *The Acoustics of Crime.* New York: Plenum.

Jessen, M. (n.d.). *Speaker Classification in Forensic Phonetics and Acoustics.* Retrieved December 21, 2011, from http://www.springerlink.com: http://www.springerlink.com/content/978-3-540-74186-2/#section=373799&page=1&locus=4

Kahane, J. (1975). The developmental anatomy of the human prepubertal and pubertal larynx (Doctoral dissertation, University of Pittsburgh). Dissertation Abstracts International, B36-10, 4964. (UMI No. ATT7608806).

Kekre, H. B. (2013). Closed set and open set Speaker Identification using amplitude distribution of different Transforms. *International Conference on Advances and Technology in Engineering*, (pp. 1-8).

Kent, R. D. (1976). Anatomic and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. Journal of Speech and Hearing Research, 19, 421-447

Kinnunen, T., & Li, H. (2010). An overview of text-independent speaker recognition: From features to supervectors. *Speech Communication*, 12-40.

Kraayeveld, J. (1997). Idiosyncrasy in Prosody. *PhD Dissertation*. Nijmegen.

Kunzel, H. J. (2002). Sprechererkennung: Grundzüge forensischer Sprachverarbeitung. In P. Rose, *Forensic Speaker Identification* (p. 1). New York: Taylor & Francis.

Lancker, V., Kreiman, J., & Emmorey, K. (1985). Familiar Voice Recognition: Patterns and Parameters. *Journal of Phonetics*, 19-38.

Lindh, J. (2006). Preliminary Descriptive F0-statistics for Young Male Speakers. Lund University.

Naik, J. (1994). Speaker Verification over the Telephone Network: Databases, Algorithms and Performance Assessment. *ESCA Workshop on Automatic Speaker Recognition, Identification, and Verification*, (pp. 31-38). Martigny, Switzerland.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language Discrimination by Newborns: Toward an Understanding of the Role of Rythm. *Journal of Experimental Psychology - Human Perception and Performance*, 756-766.

Neelu. (2012). *Unpublished M.Phil dissertation.*

Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition.* Cambridge: Cambridge University Press.

Nolan, F. (1997). Speaker Recognition and Forensic Phonetics. In W. Hardcastle, & J. (. Laver, *A Handbook of Phonetic Science* (p. 765). Oxford: Blackwell.

Ohala, J., Bronstein, A., Busa, G., Lewis, J., & Weigel, W. (1999, January 1). *A Guide to the History of the Phonetic Sciences in the United States*. Retrieved December 20, 2017, from eScholarship.org: https://escholarship.org/uc/item/6mr8317x

Potter, R. P. (1945, November). Visible Patterns of Speech. *Science*, 463-470.

Ramos, D., Franco-Pedroso, J., & Gonzalez-Rodriguez, J. (2011). Calibration and Weight of the Evidence by Human Listeners. The ATVS-UAM Submission to NIST Human-aided Speaker Recognition 2010. *In Acoustics, Speech and Signal Processing (ICASSP), International Conference on IEEE*, 5908-5911.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language Discrimination by Human Newborns and by Cotton-top Tamarin Monkeys. *Science*, 349-351.

Rose, P. (2002). *Forensic Speaker Identification.* New York: Taylor & Francis.

Schwartz, R., Campbell, J. P., Shen, W., Sturim, D. E., Campbell, W. M., Richardson, F. S., ... & Granville, R. (2011, May). USSS-MITLL 2010 human assisted speaker recognition. In 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5904-5907). IEEE.

Spence, M. J., & DeCasper, A. J. (1987). Prenatal Experience with Low-Frequency Maternal Voice Samples 102 Voice Sounds Influences Neo-natal Perception of Maternal Voice Samples. *Infant Behaviour and Development*, 133-142.

Titze, I. R. (1992). Critical periods of vocal change: Early childhood. The Journal of Singing, 48 (6), 16-18

Trollinger, V. L. (2003). Relationships between pitch-matching accuracy, speech fundamental frequency, speech range, age, and gender in American English-speaking preschool children. *Journal of Research in Music Education, 51*(1), 78-94.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. Language and speech, 38(3), 253-265.

Wilson, D. K. (1987). Voice problems of children (3rd ed.). Baltimore, MD: Williams & Wilkins Co.

==========