GAP INDIAN JOURNAL
OF FORENSICS AND BEHAVIOURAL SCIENCES
(ISSN – 2582-8177)
Globally peer-reviewed and open access journal.

A FORENSIC ACOUSTIC STUDY OF IDENTICAL TWIN VOICE DIFFERENTIATION USING PRAAT

[1]Prerna, [1]Kapil Verma

[1*]Dr. Manisha Kumari ([*]Corresponding Author)
[1] Department of Forensic Science, Himachal Pradesh University, Shimla.

Abstract

Audio forensics is vital in speaker identification and verifying authenticity in legal and investigative contexts. However, distinguishing between identical twins' voices remains a significant challenge due to their similar genetic and anatomical traits. Factors such as emotional state, health, ageing, speech habits, and environmental influences can cause subtle vocal variations. Current methodologies lack the precision to differentiate between identical twins' voices, particularly in overlapping acoustic features like pitch, intensity, and formants. This research addresses the forensic challenge of vocal forgery by identical twins, aiming to identify subtle acoustic variations and propose solutions to improve vocal forgery detection. WhatsApp-recorded voice samples from noise-free settings were analysed using PRAAT software, and acoustic parameters such as pitch, intensity, formant frequencies, and LPC spectra were evaluated. The study revealed significant pitch variations in vowels, minimal differences in lower formants, and variability in higher formants and LPC spectra of identical twins. These findings highlight the role of articulatory behaviour and environmental factors in shaping vocal traits.

Keywords: *Vocal forgery, Identical twins, PRAAT, Acoustic features, LPC spectrum, Pitch, Intensity, Formant frequencies, Forensic phonetics, Speaker identification.*

1. INTRODUCTION

Vocal forgery and voice-based impersonation represent an enduring and increasingly sophisticated challenge in forensic investigation and crime prevention. Criminals have long exploited the human voice as a tool of deception, employing disguise, mimicry, and deliberate vocal manipulation in the commission of offences including extortion, hoax calls, ransom demands, and impersonation (Sharma et al., 2017; San Segundo, 2013). These developments underscore the urgent need for advanced forensic methodologies capable of distinguishing authentic voices from manipulated or fraudulent ones.

One of the most complex scenarios in forensic phonetics involves identical twins, whose nearly identical genetic makeup often results in strikingly similar vocal traits. While physiological similarities in vocal cords, airflow, and vocal tract resonance complicate speaker identification, environmental factors, health conditions, and articulation styles introduce subtle variations that may be detectable through acoustic analysis. Voice production itself — regulated by genetic and neurological development — relies on phonation, articulation, and breathing, producing measurable attributes such as pitch, loudness, resonance, and articulation. These attributes manifest as acoustic features including formant frequencies (F1–F4), LPC coefficients, vowel space area, and prosodic markers such as accent, rhythm, and speech style, which together provide the foundation for forensic comparison and biometric authentication.

Accent, alongside pitch, loudness, rhythm, and speech style, plays a vital role in forensic investigations by narrowing suspects to ethnic or regional groups and aiding identification in multilingual contexts (Ilina et al., 1999; Brown & Wormald, 2017; Maher, 2009). Variations in pronunciation produce distinct acoustic and prosodic features that are crucial in forensic speaker recognition (Rafaqat Ali et al., 2014). Voice is further shaped by physiological factors — vocal cords, airflow, and vocal tract resonance — and influenced by age, health, and phonation mode (Peterson & Barney, 1952; Titze, 1994). Its measurable attributes, including pitch, formants, LPC coefficients, jitter, shimmer, and spectral features, are central to forensic comparison, speech therapy, and biometric identification (Fant, 1970; Burns, 1986).

To capture and analyse these features, this study employs PRAAT, a versatile open-source platform widely used in forensic phonetics. PRAAT enables pitch tracking, formant analysis, LPC spectrum evaluation, and spectrogram visualisation — capabilities that are critical for detecting subtle phonetic differences and identifying vocal forgery techniques such as voice conversion, voice transformation, and audio copy-move forgery (Nolan & Oh, 1996; Hollien, 2002; Boersma & Weenink, 2019). By linking voice production, acoustic features, analytical instrumentation, and forensic outcomes within a unified investigative framework, the present study provides a comprehensive approach to understanding and mitigating the risks posed by vocal forgery. This analytical process ultimately supports forensic outcomes encompassing speaker identification, forgery detection, and biometric authentication, while also addressing the broader ethical and social implications of vocal misidentification.

GAP INDIAN JOURNAL
OF FORENSICS AND BEHAVIOURAL SCIENCES
(ISSN – 2582-8177)
Globally peer-reviewed and open access journal.

2. RELATED WORK

Voice production relies on three sequential physiological processes — respiration, phonation, and articulation — generating measurable acoustic parameters that underpin forensic speaker identification. The respiratory system drives airflow through the larynx, where vocal fold vibration produces sound characterised by fundamental frequency (F0), the primary correlate of perceived pitch (Titze, 1994; Peterson & Barney, 1952). This sound is subsequently shaped by the tongue, lips, and soft palate into recognisable speech, with individual differences in vocal tract morphology producing speaker-specific formant patterns (F1–F4), jitter, shimmer, harmonics-to-noise ratio, and LPC coefficients (Fant, 1970; Kent & Read, 2002). Formants F1 and F2 are particularly significant for vowel identity, while F3 and F4 contribute to individual timbre and speaker distinctiveness. Prosodic features — including accent, rhythm, and intonation — provide supplementary discriminative information, aiding speaker attribution in multilingual and regional forensic contexts (Ilina et al., 1999; Rafaqat Ali et al., 2014).

Research on vocal forgery has addressed four principal categories. Voice transformation modifies acoustic parameters such as pitch and formants to disguise a speaker's natural voice (Kinnunen & Li, 2010). Voice conversion maps the acoustic features of a source speaker onto a specific target speaker's vocal profile, producing a targeted and potentially convincing imitation (Wu et al., 2015). Audio copy-move forgery involves the excision and reinsertion of audio segments within a recording to alter its semantic content; this method is particularly resistant to detection owing to the acoustic consistency of the manipulated material with its source environment (Amerini et al., 2017; Malik & Rizvi, 2020). Human impersonation, wherein a speaker exploits natural vocal aptitude and pre-existing vocal similarity to mimic another individual, is the most forensically pertinent category in the context of identical twin research.

Studies on monozygotic (MZ) twin voices have consistently demonstrated that shared vocal tract anatomy produces closely aligned F0 and formant values that challenge both human listeners and automated speaker recognition systems. Van Gysel et al. (2001) established that the perceptual similarity between MZ twin voices substantially exceeds that observed between unrelated speakers. Debruyne et al. (2002) and Nolan and Oh (1996) confirmed that MZ twins are physiologically near-identical in vocal tract structure, making organic variation in their speech particularly informative for forensic analysis. Gonzalez Hautamaki et al. (2013) demonstrated that equal error rates in automatic speaker verification systems increase significantly when processing twin pairs, highlighting the limitations of standard biometric approaches. Despite these similarities, individual articulatory habits, health conditions, phonatory style, and environmental exposure introduce measurable inter-twin divergences detectable through detailed acoustic analysis (Hollien, 2002; Ladefoged, 2001).

Sebastian et al. (2013) conducted a perceptual and acoustic study of MZ twin voice differentiation, reporting a mean human listener identification accuracy of 80.27%. Their findings indicated that shimmer values were more discriminative than the more commonly analysed parameters of pitch and formant frequency. However, the study was conducted under controlled laboratory conditions using standardised vowel phonation tasks, and neither addressed vocal forgery nor explored the application of open-source analytical tools, limiting its applicability to naturalistic forensic scenarios.

Among available open-source platforms, PRAAT — developed by Boersma and Weenink at the University of Amsterdam — has emerged as the most widely adopted instrument in forensic phonetics, providing comprehensive functionality for pitch tracking, formant and LPC analysis, intensity measurement, and spectrogram visualisation (Boersma & Weenink, 2019; Hollien, 2002). Existing research on twin voice comparison has predominantly employed proprietary software and controlled recording environments, restricting the generalisability of findings to real-world forensic casework (Kawahara et al., 2005). The concurrent investigation of vocal forgery and identical twin voice differentiation within a unified forensic framework remains underexplored. The present study addresses these gaps by applying PRAAT-based pitch, intensity, and LPC spectrum analysis to naturalistic voice samples.

3. METHODOLOGY

3.1 Data Acquisition

The study cohort comprised 14 pairs of identical (monozygotic) twins (N=28), recruited through random sampling across diverse geographical regions in North India. The group included 9 male and 5 female pairs, all aged 12 years or older to ensure vocal tract maturity (Table 1). Participants recorded the standardised phrase "Hello, how are you?" in controlled, quiet environments and transmitted recordings via WhatsApp on an Android Redmi Note 9 Pro (sampling rate: 48 kHz). Informed consent was obtained from all subjects in accordance with the ethical protocols approved by Himachal Pradesh University. Computational processing was performed on an HP system equipped with a 12th Gen Intel Core i5-1235U processor (1.30 GHz) and 8 GB of RAM, operating on a Windows 11 environment.

GAP INDIAN JOURNAL
OF FORENSICS AND BEHAVIOURAL SCIENCES
(ISSN – 2582-8177)
Globally peer-reviewed and open access journal.

Table 1: Demographic Distribution of Twin Pairs (N=28)

Pair ID	Gender	Location	Pair ID	Gender	Location
01	♀	Jammu	08	♂	Patiala
02	♀	Shimla	09	♂	Baddi
03	♀	Solan	10	♂	Solan
04	♀	Delhi	11	♂	Delhi
05	♀	Baddi	12	♂	Hamirpur
06	♂	Delhi	13	♂	Solan
07	♂	Solan	14	♂	Baddi

3.2 Acoustic Analysis Pipeline

For the extraction of vocal artefacts, PRAAT (version 6.4.13) was utilised. Spectral analysis was refined using Linear Predictive Coding (LPC) with a prediction order of 16–25 poles to accurately model the vocal tract resonances. A pre-emphasis frequency cutoff of 50 Hz was applied during digital signal processing to eliminate low-frequency environmental hum. The methodology focused on identifying forensic artefacts — micro-variations in voice that persist even when twins produce the same utterance. Table 2 summarises the acoustic parameters extracted and their forensic significance.

Table 2: Acoustic Parameters and Forensic Significance

Parameter	Method of Analysis	Forensic Utility
Pitch (F0)	Autocorrelation	Measures laryngeal vibration; shows baseline similarity.
Formants (F1, F2)	Spectrogram/LPC	Reflects tongue height and front-back placement.
Formants (F3, F4)	Spectrogram/LPC	Primary Marker: Reflects unique vocal tract length.
LPC Envelope	Spectral Modeling	Identifies resonant peaks for speaker discrimination.
Intensity	Amplitude Analysis	Measures speech energy and breath control.

4. RESULTS AND DISCUSSION

4.1 Pitch and Fundamental Frequency Analysis (F0)

4.1.1 Consonantal Pitch Analysis (/h/ and /l/)

Pitch analysis conducted across 14 identical twin pairs revealed a pattern of both genetic consistency and articulatory variability. For the glottal fricative /h/, inter-twin differences were minimal; Pairs 1 and 4 exhibited divergences of only 4.00 Hz and 1.80 Hz respectively. This laryngeal similarity is consistent with observations by Ladefoged (2001) and Titze (2008), who note that genetically identical individuals tend to produce near-equivalent pitch values in sounds requiring minimal vocal tract constriction. In contrast, the lateral approximant /l/ demonstrated considerably greater inter-twin divergence, most notably in Pair 2, where a difference of 105.64 Hz was recorded (120.27 Hz versus 225.91 Hz). As Kent and Read (2002) discuss, such divergences in /l/ frequency suggest that complex tongue positioning constitutes a learned motor behaviour that may differentiate twins despite their shared anatomical substrate. These findings are summarised in Table 3.

Table 3: Pitch Analysis Results — Consonants

Pair No.	Consonant	Twin 1 (Hz)	Twin 2 (Hz)	Δ Difference (Hz)
01	/h/	286.69	282.69	4.00
02	/l/	120.27	225.91	105.64
03	/l/	193.33	137.13	56.20
04	/h/	295.45	293.65	1.80
05	/h/	225.95	249.82	23.87
06	/l/	153.05	167.67	14.62
07	/h/	135.23	140.80	5.57
08	/h/	132.56	135.36	2.80
09	/h/	143.87	139.72	4.15
10	/l/	121.50	119.60	1.90
11	/h/	268.36	244.11	24.25
12	/l/	141.06	166.37	25.31
13	/h/	131.18	132.31	1.13
14	/l/	168.19	162.45	5.74

4.1.2 Vowel Pitch Analysis (/a/, /o/, /u/)

Vowel pitch analysis further demonstrated that identical twins frequently occupy distinct acoustic spaces despite their shared physiology. Pair 11 (/a/) exhibited a negligible inter-twin difference of 0.56 Hz, whereas Pair 2 (/o/)

GAP INDIAN JOURNAL
OF FORENSICS AND BEHAVIOURAL SCIENCES
(ISSN – 2582-8177)
Globally peer-reviewed and open access journal.

showed a marked divergence of 95.53 Hz, reflecting substantial differences in vocal tract tension and phonatory effort. These findings are consistent with Hillenbrand et al. (1995) and Sundberg (2000), who contend that vowel production is governed not exclusively by biological structure but also by individual articulatory habits, vocal effort, and emotional state. On this basis, Pairs 1, 10, 3, and 13 emerge as the most acoustically stable candidates for intra-pair similarity analysis. Conversely, cases such as Pair 11 — where fundamental frequency is nearly indistinguishable — represent the greatest forensic challenge, and justify the subsequent application of LPC and formant analysis to extract discriminative features that F0 alone cannot reveal. Results are presented in Table 4.

Table 4: Pitch Analysis Results — Vowels

Pair No.	Vowel	Twin 1 (Hz)	Twin 2 (Hz)	Δ Difference (Hz)
01	/o/	258.65	234.72	23.93
02	/o/	178.48	274.01	95.53
03	/u/	274.93	205.55	69.38
04	/o/	350.71	373.06	22.35
05	/u/	205.30	264.43	59.13
06	/u/	150.07	133.42	16.65
07	/o/	180.53	178.03	2.50
08	/u/	114.62	121.96	7.34
09	/u/	161.61	137.68	23.93
10	/u/	131.67	118.80	12.87
11	/a/	240.11	239.55	0.56
12	/o/	177.26	155.34	21.92
13	/u/	111.46	114.17	2.71
14	/u/	136.49	150.23	13.74

4.2 Intensity Analysis

Intensity measurements yielded critical information regarding behavioural vocal variation between identical twin pairs. Mean, minimum, and maximum intensity values were examined to identify deviations in vocal effort and overall speech energy (Titze, 2000). As shown in Table 5, Pairs 4, 11, and 13 exhibited a high degree of acoustic symmetry across all intensity parameters, a pattern attributable to the shared physiological structure of the vocal folds in monozygotic twins (Titze, 2000). In contrast, several pairs demonstrated substantial inter-twin divergence.

Pairs 2, 7, 9, and 12 recorded the most pronounced differences in mean and maximum intensity, reflecting individualised articulation styles and differential vocal effort (Sundberg, 2000). Of particular forensic significance is Pair 12, in which Twin 1 produced the highest mean intensity value recorded across the entire sample (80.99 dB), representing a clear and reproducible acoustic marker for speaker differentiation. On the basis of their high mean intensity values and consistent vocal dynamics, Pairs 7, 9, 12, and 14 were identified as the most discriminatively robust datasets for forensic speaker differentiation.

Table 5: Intensity Analysis Results

Pair No.	Speaker	Mean Intensity	Min Intensity	Max Intensity	Duration (s)
1	Twin 1	66.07 dB	14.63 dB	73.02 dB	0.944
	Twin 2	66.57 dB	11.47 dB	74.67 dB	0.825
2	Twin 1	69.52 dB	53.93 dB	77.14 dB	0.958
	Twin 2	75.72 dB	50.37 dB	82.71 dB	1.228
3	Twin 1	65.68 dB	8.47 dB	73.82 dB	0.913
	Twin 2	64.46 dB	3.01 dB	72.44 dB	0.993
4	Twin 1	60.81 dB	-8.42 dB	69.50 dB	0.548
	Twin 2	60.78 dB	-7.41 dB	69.50 dB	0.543
5	Twin 1	72.14 dB	-4.74 dB	78.43 dB	1.137
	Twin 2	71.64 dB	-8.48 dB	78.58 dB	0.700
6	Twin 1	69.42 dB	37.42 dB	76.89 dB	0.744
	Twin 2	60.56 dB	-7.45 dB	71.54 dB	0.606
7	Twin 1	76.97 dB	30.71 dB	84.59 dB	1.154
	Twin 2	73.82 dB	44.17 dB	82.37 dB	0.898
8	Twin 1	77.37 dB	N/A*	85.62 dB	0.821
	Twin 2	73.81 dB	-7.23 dB	83.14 dB	0.752
9	Twin 1	76.12 dB	28.95 dB	83.80 dB	1.310
	Twin 2	71.15 dB	25.95 dB	83.75 dB	0.481
10	Twin 1	75.25 dB	24.54 dB	82.47 dB	1.074

GAP INDIAN JOURNAL
OF FORENSICS AND BEHAVIOURAL SCIENCES
(ISSN – 2582-8177)
Globally peer-reviewed and open access journal.

	Twin 2	76.75 dB	31.49 dB	82.14 dB	1.337
11	Twin 1	72.15 dB	24.92 dB	79.52 dB	0.653
	Twin 2	71.18 dB	11.83 dB	79.44 dB	0.774
12	Twin 1	80.99 dB	33.68 dB	86.69 dB	1.171
	Twin 2	73.47 dB	37.12 dB	80.78 dB	0.896
13	Twin 1	58.63 dB	-0.63 dB	68.10 dB	0.499
	Twin 2	60.67 dB	5.49 dB	69.88 dB	0.915
14	Twin 1	70.55 dB	30.92 dB	76.69 dB	1.632
	Twin 2	73.53 dB	25.44 dB	82.99 dB	0.802

*Pair 8, Twin 1: minimum intensity recorded as N/A, indicating a period of complete silence at the recording boundary. This value is excluded from comparative intensity analysis for Pair 8.

4.3 Comparative Intensity Analysis of the Segment "Hello"

Intensity is fundamentally determined by the force with which air is expelled through the vocal cords, in conjunction with the size and shape of the vocal tract (Hillenbrand et al., 1995). Speech intensity varies as a function of emotional state, articulation strategy, and breath control, and therefore provides valuable diagnostic insight into individual vocal patterns (Zraick et al., 2005).

Analysis of the selected segment "hello" across 14 twin pairs revealed two distinct patterns. Pairs 4, 11, and 13 exhibited closely aligned mean, minimum, and maximum intensity values, indicating a high degree of vocal resemblance attributable to shared genetic traits and similar vocal tract anatomy (Titze, 2000). In contrast, Pairs 2, 7, 9, and 12 demonstrated notable inter-twin variation in mean and maximum intensity, likely reflecting differences in vocal effort, articulation style, or recording environment (Sundberg, 2000).

Among all pairs, Pairs 7, 9, 12, and 14 were identified as the most analytically robust for further forensic examination, on account of their high mean intensity levels, consistent temporal patterns, and stable vocal dynamics. Pair 12 was particularly notable, with Twin 1 producing the highest mean intensity in the entire sample (80.99 dB). Collectively, the intensity analysis confirms that although identical twins share significant vocal similarities arising from their common genetic background, individual differences in speech behaviour are reliably detectable through intensity measurement.

4.4 Pitch Analysis — Summary Across Samples

Pitch analysis across 14 twin pairs revealed that Pairs 1 (/h/), 10 (/l/), 7, and 13 exhibited minimal inter-twin pitch variation, rendering them well-suited for forensic comparison. Among the vowels examined, /u/ displayed the greatest pitch discrepancy, with a difference of 192.37 Hz recorded in Pair 2, while Pair 13 exhibited the smallest divergence, suggesting a degree of vowel-dependent pitch similarity. Moderate inter-twin differences were observed for /e/ and /a/, particularly in Pairs 3, 5, and 12. These variations likely reflect subtle anatomical and physiological differences arising from factors including vocal tract morphology, environmental exposure, habitual posture, and oral behaviours, all of which influence resonance and formant characteristics (Kent & Read, 2002; Ladefoged, 2001).

4.5 LPC Spectrum Analysis

Linear Predictive Coding (LPC) provides a compact parametric representation of the speech signal, modelling the resonance characteristics of the vocal tract through the estimation of spectral envelope peaks corresponding to formant frequencies (Rabiner & Schafer, 2011). In forensic phonetics, LPC analysis is particularly valuable for speaker identification in demanding scenarios, including the differentiation of identical twins, owing to its sensitivity to fine-grained vocal tract differences (Kunzel, 2001).

LPC spectrum analysis was performed on voice samples drawn from eight selected twin pairs (Pairs 1, 2, 3, 4, 7, 9, 12, and 13), encompassing the vowels /o/, /u/, /a/, and /e/. Four formant peaks (F1–F4) were extracted for each sample, as presented in Table 6.

Table 6: LPC Spectrum Analysis — Formant Peaks (F1–F4)

Pair No.	Speaker	Vowel	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
1	Twin 1	/o/	903.29	3738.29	7053.14	9382.76
	Twin 2	/o/	859.25	3804.20	7395.65	9038.90
2	Twin 1	/u/	421.81	2081.20	2995.41	4062.67
	Twin 2	/u/	438.92	1678.57	3069.77	3392.72
3	Twin 1	/e/	2890.55	4677.91	7202.87	8472.46
	Twin 2	/e/	3355.24	4804.37	6796.81	8907.81
4	Twin 1	/o/	670.47	3326.74	4868.23	5920.09
	Twin 2	/o/	676.79	3349.63	4768.23	5990.24
7	Twin 1	/o/	747.57	3155.04	6640.71	8851.89
	Twin 2	/o/	804.81	3381.65	6784.95	9149.07

9	Twin 1	/u/	429.47	2196.12	2964.07	5700.53
	Twin 2	/u/	336.76	1761.90	2706.59	5479.99
12	Twin 1	/a/	717.23	5116.31	6316.19	9746.40
	Twin 2	/a/	688.07	4269.42	6676.48	9204.29
13	Twin 1	/u/	290.95	2072.83	3526.76	5742.11
	Twin 2	/u/	316.33	1929.99	4604.10	6623.35

Across the majority of pairs, F1 and F2 exhibited minimal inter-twin variation, reflecting the shared vocal tract dimensions and configuration characteristic of monozygotic twins. Pair 4 (/o/) demonstrated the least divergence, with F1 and F2 differing by only 6.32 Hz and 22.89 Hz respectively. The higher formants, F3 and F4, showed considerably greater inter-twin variability, attributable to more complex articulatory factors including tongue placement, degree of lip rounding, and vocal fold elasticity. Pair 2 (/u/) recorded a 74.36 Hz difference in F3, while Pair 9 exhibited a substantially larger divergence of 257.48 Hz at the same formant.

Vowel-specific articulatory dynamics exerted a consistent influence on the pattern of formant variation. The back vowel /o/, characterised by posterior tongue positioning and lip rounding, produced slight but consistent F3 and F4 differences. The high back vowel /u/ exhibited the largest inter-twin discrepancies, while /a/ maintained stable lower formants with variation concentrated in the higher spectral range. The mid-front vowel /e/ produced moderate variation across all four formant peaks. These findings demonstrate that while lower formants reflect the shared anatomical substrate of identical twins, the higher formants encode individualised articulatory behaviour and may therefore serve as critical discriminative parameters in forensic speaker differentiation.

5. CONCLUSION

This study demonstrates that acoustic analysis using PRAAT can reliably differentiate the voices of identical twins despite their shared genetic makeup. Pitch analysis confirmed that phoneme complexity governs the degree of inter-twin divergence, while intensity measurements identified Pairs 7, 9, 12, and 14 as the most forensically discriminative. LPC spectrum analysis established that higher formants F3 and F4, reflecting individualised articulatory behaviour, are the most diagnostically valuable parameters for twin voice differentiation. These findings affirm that a convergent multi-parametric approach — integrating fundamental frequency, intensity, and LPC-derived formant data — provides the most reliable basis for forensic speaker identification in identical twin cases, with direct implications for phonetic casework and the judicial use of voice evidence. Future research should expand sample sizes and work towards standardised forensic protocols for twin voice comparison.

REFERENCES

[1] Amerini, I., Caldelli, R., Del Bimbo, A., & Lombardi, L. (2017). Splicing forgeries localization through the use of first digits features. IEEE Workshop on Information Forensics and Security, 1–6.
[2] Boersma, P., & Weenink, D. (2001). PRAAT, a system for doing phonetics by computer. Glot International, 5(9/10), 341–345.
[3] Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer (Version 6.4.13) [Software]. http://www.praat.org/
[4] Boone, D. R., McFarlane, S. C., Von Berg, S. L., & Zraick, R. I. (2014). The voice and voice therapy (9th ed.). Pearson.
[5] Brown, G., & Wormald, L. (2017). Dialect accent features for accent classification in forensic speaker recognition. Proceedings of the International Conference on Language and Technology.
[6] Burns, E. M. (1986). Variations in vocal pitch and quality in classical and country singing. Journal of the Acoustical Society of America, 80(S1), S97. https://doi.org/10.1121/1.2023633
[7] Davis, S. B., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), 357–366.
[8] Debruyne, F., Decoster, W., Van Gijsel, A., & Vercammen, J. (2002). Speaking fundamental frequency in monozygotic and dizygotic twins. Journal of Voice, 16(4), 466–471. https://doi.org/10.1016/S0892-1997(02)00118-2
[9] Fant, G. (1970). Acoustic theory of speech production (2nd ed.). Mouton.
[10] Gonzalez Hautamaki, R., Kinnunen, T., Hautamaki, V., Leino, T., & Laukkanen, A.-M. (2013). I-vectors meet imitators: On vulnerability of speaker verification systems against voice mimicry. Proceedings of Interspeech 2013, 930–934. https://doi.org/10.21437/Interspeech.2013-289
[11] Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America, 97(5), 3099–3111. https://doi.org/10.1121/1.411872
[12] Hollien, H. (2002). Forensic voice identification. Academic Press.
[13] Ilina, N., Koval, S., & Varaksin, V. (1999). Phonetic analysis in forensic speaker identification. In Proceedings

of the 14th International Congress of Phonetic Sciences (ICPhS 1999), San Francisco (pp. 157–160). International Phonetic Association.
[14] Kawahara, H., Masuda-Katsuse, I., & de Cheveigne, A. (2005). Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. Speech Communication, 27(3–4), 187–207. https://doi.org/10.1016/S0167-6393(98)00085-5
[15] Kent, R. D., & Read, C. (2002). The acoustic analysis of speech (2nd ed.). Singular/Thomson Learning.
[16] Kinnunen, T., & Li, H. (2010). An overview of text-independent speaker recognition: From features to supervectors. Speech Communication, 52(1), 12–40. https://doi.org/10.1016/j.specom.2009.08.009
[17] Kunzel, H. J. (2001). Beware of the telephone effect: The influence of telephone transmission on the measurement of formant frequencies. Forensic Linguistics, 8(1), 80–99. https://doi.org/10.1558/sll.2001.8.1.80
[18] Ladefoged, P. (2001). A course in phonetics (4th ed.). Harcourt College Publishers.
[19] Maher, R. C. (2009). Audio forensic examination. IEEE Signal Processing Magazine, 26(2), 84–94.
[20] Malik, H., & Rizvi, S. A. (2020). Audio splicing detection using acoustic environment consistency. EURASIP Journal on Audio, Speech, and Music Processing, 2020, Article 3. https://doi.org/10.1186/s13636-020-0171-8
[21] Nolan, F., & Oh, T. (1996). Identical twins, different voices. International Journal of Speech, Language and the Law, 3(1), 39–49. https://doi.org/10.1558/ijsll.v3i1.39
[22] Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 24(2), 175–184. https://doi.org/10.1121/1.1906875
[23] Rabiner, L. R., & Schafer, R. W. (2011). Theory and applications of digital speech processing. Pearson.
[24] Rafaqat, A., Irtza, S., Farooq, M., & Hussain, S. (2014). Accent classification among Punjabi, Urdu, Pashto, Saraiki and Sindhi accents of Urdu language. In Proceedings of the 2014 Conference on Language and Technology, Lahore, Pakistan (pp. 1–7).
[25] San Segundo, E. (2013). A phonetic corpus of Spanish male twins and siblings: Corpus design and forensic application. Procedia — Social and Behavioral Sciences, 95, 59–67. https://doi.org/10.1016/j.sbspro.2013.10.622
[26] Sebastian, S., Benadict, A. S., Sunny, G. K., & Balraj, A. (2013). An investigation into the voice of identical twins. Otolaryngology Online Journal, 3(2), 9–15.
[27] Sharma, S., Jain, A., & Kumar, S. (2017). Characterisation of temporal and acoustic parameters for speaker identification in disguised speech. Journal of Forensic Science & Criminal Investigation, 5(1), 001–009. https://doi.org/10.19080/JFSCI.2017.05.555655
[28] Sundberg, J. (2000). The science of the singing voice. Northern Illinois University Press.
[29] Titze, I. R. (1994). Principles of voice production. Prentice Hall.
[30] Titze, I. R. (2000). Principles of voice production (2nd ed.). National Center for Voice and Speech.
[31] Titze, I. R. (2008). Nonlinear source–filter coupling in phonation: Theory. Journal of the Acoustical Society of America, 123(5), 2733–2749. https://doi.org/10.1121/1.2832337
[32] Van Gysel, W., Vercammen, J., & Debruyne, F. (2001). Voice similarity in identical twins. Acta Oto-Rhino-Laryngologica Belgica, 55(1), 49–55.
[33] Wu, Z., Evans, N., Kinnunen, T., Yamagishi, J., Alegre, F., & Li, H. (2015). Spoofing and countermeasures for speaker verification: A survey. Speech Communication, 66, 130–153. https://doi.org/10.1016/j.specom.2014.10.005
[34] Zraick, R. I., Wendel, K., & Smith-Olinde, L. (2005). The effect of speaking task on perceptual judgment of the severity of dysphonic voice. Journal of Voice, 19(4), 574–581. https://doi.org/10.1016/j.jvoice.2004.07.009