Analysis of Voice Onset Time in English Pronunciation by Teenagers: An Acoustic Phonetic Study

Prima Sholihatun[1], T Syarfina[2] and Mahriyuni[3]
[1]*Universitas Sumatera Utara, Indonesia, prima.sholihatun95@gmail.com*
[2]*Universitas Sumatera Utara, Indonesia, tengku.syarfina@usu.ac.id*
[3]*Universitas Sumatera Utara, Indonesia, mahriyuni@usu.ac.id*

DOI : https://doi.org/10.30983/mj.v4i2.10380

Submission: September 11, 2025	*Revised: Ocotober 24, 2025*	*Published: December 31, 2025*

Abstrak

Penelitian ini bertujuan untuk menginvestigasi karakteristik akustik pelafalan bahasa Inggris oleh remaja melalui parameter Voice Onset Time (VOT). Secara khusus, penelitian ini menekankan analisis durasi VOT pada konsonan letup awal [p], [b], [t], [d], [k], dan [g] pada bahasa Inggris. Penelitian menggunakan pendekatan kualitatif dengan metode eksperimental. Data diperoleh melalui rekaman ujaran kata-kata yang mengandung konsonan plosif dari beberapa remaja. Rekaman kemudian didigitalkan menjadi format sound wave dan dianalisis menggunakan perangkat lunak Praat melalui tahap digitalisasi, segmentasi data, dan pembuatan salinan kontur. Pendekatan ini memungkinkan pengukuran nilai VOT untuk konsonan plosif secara akurat. Hasil analisis VOT menunjukkan variasi yang signifikan antarindividu. Rata-rata nilai VOT untuk konsonan tak bersuara [p], [t], dan [k] masing-masing 46 ms, 82 ms, dan 49 ms, sedangkan konsonan bersuara [b], [d], dan [g] masing-masing 38 ms, -13 ms, dan 10 ms. Temuan ini menunjukkan bahwa siswa cenderung memproduksi konsonan tak bersuara dengan durasi VOT lebih pendek dibandingkan penutur asli, sehingga terdengar mendekati konsonan bersuara.

Kata Kunci: *Akustik, Konsonan Plosif, Voice Onset Time, Praat*

Abstract

This research aims to investigate the acoustic characteristics of English pronunciation by teenagers through the Voice Onset Time (VOT) parameter. Specifically, the research emphasizes the analysis of VOT duration in initial plosive consonants [p], [b], [t], [d], [k], and [g] in English. The study employs a quantitative approach with an experimental method. Data were collected through recordings of words containing plosive consonants spoken by several teenagers. The recordings were then digitized into sound wave format and analyzed using Praat software through stages of digitization, data segmentation, and contour replication. This approach enables accurate measurement of VOT values for plosive consonants. The VOT analysis results indicate significant variations between individuals. The average VOT values for voiceless consonants [p], [t], and [k] are 46 ms, 82 ms, and 49 ms, respectively, while voiced consonants [b], [d], and [g] have average VOT values of 38 ms, -13 ms, and 10 ms, respectively. These findings suggest that the students tend to produce voiceless consonants with shorter VOT durations compared to native speakers, making them sound closer to voiced consonants.

Keywords: Acoustic, Konsonan Plosive, Voice Onset Time, Praat

1. Introduction

In second language (L2) English learning, the mastery of vocabulary and grammar alone is insufficient to guarantee successful oral communication. Accurate pronunciation plays an equally critical role, as it directly affects speech intelligibility and listener comprehension. Learners who possess adequate lexical and syntactic knowledge may still experience communication breakdowns if their pronunciation deviates significantly from target language norms. Consequently, pronunciation has increasingly gained attention in applied linguistics, particularly through empirical approaches that allow researchers to objectively examine speech production.

One of the most robust approaches to investigating pronunciation is acoustic phonetics, which analyzes speech sounds based on their measurable physical properties, such as frequency, intensity, and temporal duration. Unlike impressionistic phonetic analysis, acoustic analysis provides quantifiable evidence of how sounds are produced and perceived. Within this framework, Voice Onset Time (VOT) has emerged as a central parameter for examining stop consonants, as it captures the temporal relationship between articulatory release and vocal fold vibration. VOT is especially crucial in distinguishing voiced and voiceless plosives in many languages, including English.

In phonetic classification, the consonants [p, b, t, d, k, g] are categorized as stop consonants because their production involves a complete closure of the vocal tract that temporarily blocks the airflow from the lungs. This closure is followed by a sudden release, producing a characteristic burst of sound. The articulation of these stops requires precise coordination between different articulators. For instance, velar stops [k] and [g] are produced when the back of the tongue makes contact with the velum, with voicing added in the case of [g]. Based on such articulatory mechanisms, English stop consonants are systematically classified according to their place of articulation (POA) and manner of articulation (MOA), as illustrated in Table 1.1.

Table 1. English Consonant Plosive

POA					
			Bilabial	Alveolar	Velar
MOA	Stop	- voice	p	t	k
		+ voice	b	d	g

Source: Koffi (2021) Relevant Acoustic Phonetic of L2 English

Note:
POA : *Place Of Articulation*
MOA : *Manner Of Articulation*

Although this articulatory classification appears straightforward, its phonetic realization particularly in terms of timing differs substantially across languages. This is where VOT becomes a decisive acoustic marker. First introduced by Lisker and Abramson (1964), VOT refers to the temporal interval between the release of a stop consonant and the onset of periodic vocal fold vibration of the following vowel. This parameter has been widely adopted to differentiate voiced and voiceless plosives in cross-linguistic phonetic research.

VOT is generally classified into three categories: negative VOT, where voicing begins before the stop release (voicing lead); zero VOT, where voicing begins almost simultaneously with the release; and positive VOT, where voicing begins after the release (voicing lag). Languages vary considerably in how these categories are distributed. In General American English (GAE), voiceless stops /p/, /t/, and /k/ are typically realized with long positive VOT due to aspiration, whereas voiced stops /b/, /d/, and /g/ are characterized by short-lag or near-zero VOT (Flege, 1995).

In contrast, Indonesian stop consonants generally exhibit shorter VOT values and lack the strong aspiration found in English voiceless stops. As a result, Indonesian learners of English frequently produce /p/, /t/, and /k/ with insufficient aspiration, causing these sounds to be perceived as voiced or unnatural by native English listeners (Misnadin, 2016). This cross-linguistic discrepancy underscores the importance of VOT as a critical source of pronunciation difficulty for Indonesian EFL learners.

Despite the growing body of research on VOT in second language acquisition, **several** significant gaps remain, particularly in the Indonesian context. Empirically, most existing studies have focused on adult learners or university students, such as Misnadin's (2016) investigation of Madurese speakers and Rahmawati and Hidayat's (2021) study involving university students in Yogyakarta. These studies provide valuable insights into adult L2 phonetic acquisition but leave younger learners underrepresented. Adolescents, especially high school students, constitute a crucial population, as their phonological systems are still developing and may exhibit different patterns of L1 interference and L2 adaptation.

Contextually, regional linguistic diversity in Indonesia has not been sufficiently explored in VOT research. Binjai, North Sumatra, represents a linguistically rich environment influenced by local dialects such as Melayu Deli and Batak. These dialects may introduce distinct phonetic timing patterns that interact with English stop production in ways that differ from learners in Java or Madura. However, to date, little empirical evidence has documented how such regional multilingual settings affect English VOT realization among Indonesian teenagers.

Pedagogically, a further gap persists between phonetic research and classroom practice. Although Indonesian learners' difficulty with aspirated voiceless stops is well documented, English teachers often lack access to objective, acoustically grounded instructional materials that could help address these specific pronunciation problems. Without empirical acoustic data from relevant learner populations, pronunciation teaching remains largely intuitive and impressionistic.

In response to these gaps, the present study investigates the production of English stop consonants through VOT analysis among high school students in Binjai, North Sumatra. By focusing on adolescent learners within a multilingual regional context, this study aims to contribute empirical evidence to L2 acoustic phonetics while simultaneously offering pedagogical implications for pronunciation instruction grounded in objective phonetic analysis.

2. Method

2.1 Research Design

This study adopts a quantitative descriptive research design employing an acoustic-phonetic experimental approach to investigate the production of English plosive consonants by Indonesian teenage learners. While preliminary observations during data collection involved qualitative monitoring of participants' pronunciation behavior, the primary analytical framework of this study is quantitative in nature. The core data consist of numerical measurements of Voice Onset Time (VOT) expressed in milliseconds (ms), which allows for objective, replicable, and statistically interpretable analysis.

The acoustic-phonetic method is particularly appropriate for this research because it enables fine-grained measurement of temporal phonetic features that are not reliably observable through auditory perception alone. By focusing on VOT values, this study aims to provide a precise empirical description of how English plosive consonants (/p, t, k, b, d, g/) are realized by high school students in Binjai, North Sumatra, and how these realizations diverge from established General American English (GAE) norms.

2.2 Participants

The participants consisted of five high school students from Binjai, North Sumatra, who were selected using a purposive sampling technique. This sampling method was employed to ensure that all participants met specific inclusion criteria relevant to the research objectives. The criteria included: (1) being active students at the high school level, (2) having Indonesian as their primary language, (3) learning English as a second language (L2) in a formal educational setting, and (4) having no reported speech or hearing impairments.

The relatively small number of participants was intentionally chosen to facilitate a focused acoustic case analysis rather than broad generalization. In acoustic phonetics, detailed examination of individual speech tokens is crucial for capturing subtle phonetic variations, such as micro-timing differences in stop release and voicing onset. Larger sample sizes often obscure such individual deviations through averaging. Therefore, this study prioritizes depth of phonetic analysis over population-level generalizability, allowing for a high-resolution investigation of learner-specific VOT patterns.

2.3 Instruments

1. Stimuli and Word List

To ensure phonetic control and minimize extraneous variables affecting VOT, a carefully constructed word list was employed as the speech stimuli. The list consisted of six English monosyllabic words containing word-initial plosive consonants in a CVC (Consonant–Vowel–Consonant) syllable structure, namely: *pan, box, tap, dog, kids,* and *go*. These words were selected to represent all six English stop consonants (/p, t, k, b, d, g/) in initial position, where VOT contrasts are most salient and reliably measurable.

The vowel contexts following the target plosives were selected to maintain relative phonetic consistency and reduce coarticulatory effects that could influence VOT duration. Additionally, all target words carried primary lexical stress on the initial syllable, thereby eliminating variation caused by stress-related temporal differences. Each participant was instructed to produce each target word three to five times in a natural speaking rate. Multiple repetitions were necessary to enhance measurement reliability and to allow the calculation of stable average VOT values for each consonant.

2. Recording Environment and Technical Specifications

Data collection was conducted in a quiet and controlled recording environment to minimize background noise and acoustic interference. Participants were seated comfortably and instructed to speak clearly at a normal conversational volume. Speech samples were recorded using a high-sensitivity directional microphone connected to a digital recording device. A consistent distance of approximately 15–20 cm between the participant's mouth and the microphone was maintained throughout the recording session to prevent amplitude distortion and ensure uniform signal quality.

All recordings were captured at a sampling rate of 44,100 Hz with 16-bit resolution and saved in a lossless .wav format. This configuration was selected to preserve high-frequency acoustic information essential for accurate waveform and spectrographic analysis, particularly for identifying stop bursts and voicing onset. Prior to recording, participants were given brief practice trials to familiarize themselves with the task and reduce performance anxiety.

3. Acoustic Analysis Software

The acoustic analysis was carried out using PRAAT software (version 5.3.77), a widely used phonetic analysis program designed for detailed speech sound investigation (Durão, 2007). PRAAT provides integrated tools for waveform visualization, spectrogram analysis, pitch tracking, and temporal measurement, making it particularly suitable for VOT analysis.

After the recording process, each speech token was segmented and visualized in PRAAT. The software enabled the researcher to generate high-resolution spectrograms displaying both low and high frequency components of the speech signal. In addition to VOT measurement, PRAAT allows for the analysis of intensity, duration, and formant structure; however, this study focused primarily on temporal measurements relevant to stop consonant production.

2.4 Data Analysis Procedure

The digital audio recordings were analyzed following a systematic procedure. First, each token containing a target plosive was isolated and labeled. VOT was measured as the temporal interval between two acoustic landmarks: (1) the burst release, identified by a sudden spike in energy on the waveform and a vertical burst on the spectrogram, and (2) the onset of voicing, marked by the appearance of regular periodic waveforms and visible pitch striations.

All measurements were conducted manually using PRAAT's cursor and time-display functions to ensure precision. The resulting VOT values, expressed in milliseconds, were then classified into Negative VOT, Zero VOT, or Positive VOT categories. Subsequently, the learners' VOT values were compared to established General American English (GAE) reference values reported in previous phonetic studies to identify patterns of deviation and potential first-language (L1) interference.

To enhance reliability, repeated tokens were averaged for each participant and consonant category. The analysis focused on identifying consistent tendencies in VOT realization rather than isolated production errors, thereby providing a robust acoustic profile of English stop production among teenage EFL learners in Binjai.

3. Results and Discussion

The analysis of Voice Onset Time (VOT) in six English word-initial stop consonants produced by five teenagers in Binjai reveals significant variation and instability compared to General American English (GAE) standards. The data indicates that the mean VOT values for voiceless consonants [p], [t], and [k] are 46 ms, 82 ms, and 49 ms, respectively. In contrast, the voiced consonants [b], [d], and [g] show mean values of 38 ms, -13 ms, and 10 ms, respectively.

In the voiceless consonant group, a clear tendency toward hyperaspiration was observed, particularly for the phoneme /t/, which reached 82 ms, significantly exceeding the GAE standard of 39 ms. Meanwhile, the voiced consonant group exhibited extreme instability, most notably in the phoneme /b/, with values ranging from -147 ms to +173 ms. The mean VOT for /b/ (38 ms) acoustically categorizes it as a voiceless unaspirated stop, indicating a failure to maintain consistent voicing contrast.

Table 2. Results of Voice Onset Time

Partisipant	**Sample of Data**					
	pan [p]	**box [b]**	**Tap [t]**	**dog [d]**	**kids [k]**	**go [g]**
Partisipant 1 Male	47 ms	-147 ms	82 ms	-103 ms	22 ms	23 ms
Partisipant 2 Female	0 ms	37 ms	23 ms	39 ms	28 ms	60 ms
Partisipant 3 Female	40 ms	51 ms	70 ms	19 ms	69 ms	76 ms
Partisipant 4 Female	35 ms	173 ms	112 ms	-21 ms	22 ms	-157 ms
Partisipant 5 Male	109 ms	77 ms	121 ms	0 ms	103 ms	47 ms
Rata-rata	46 ms	38	82 ms	-13 ms	49 ms	10 ms
Lisker and Abramson (1964)	**28 ms**	**7/-45 ms**	**39 ms**	**9/-56 ms**	**43 ms**	**17/-45 ms**

The acoustic analysis of Voice Onset Time (VOT) focused on six English word-initial stop consonants (/p, b, t, d, k, g/) produced by five teenage learners in Binjai, North Sumatra. The results reveal substantial variability both within and across participants, as well as notable deviations from established General American English (GAE) reference values (Lisker & Abramson, 1964).

3.1 Overall VOT Patterns

Table 3.1 summarizes the individual and mean VOT values (in milliseconds) for each target consonant. For the voiceless stops, the mean VOT values were 46 ms for /p/, 82 ms for /t/**, and** 49 ms for /k/. These values indicate that, on average, participants produced longer VOTs than the GAE norms for /p/ (28 ms) and /t/ (39 ms), while the value for /k/ (49 ms) slightly exceeded the GAE standard of 43 ms.

In contrast, the voiced stops demonstrated highly inconsistent realizations. The mean VOT values were 38 ms for /b/, –13 ms for /d/, and 10 ms for /g/**.** While /d/ and /g/ occasionally exhibited negative or near-zero VOT, /b/ showed a predominantly positive VOT, acoustically resembling a voiceless unaspirated stop rather than a voiced stop.

3.2 Voiceless Stop Consonants (/p, t, k/)

The production of voiceless stops showed a general tendency toward over-aspiration, particularly for the alveolar stop /t/. The mean VOT for /t/ reached 82 ms, more than double the GAE reference value. Individual data further illustrate this pattern, with some participants producing /t/ with VOTs as high as 112 ms and 121 ms.

The bilabial stop /p/ also demonstrated variability, ranging from 0 ms (Participant 2) to 109 ms (Participant 5). Although some productions fell within or near the GAE range, others showed exaggerated aspiration. Similarly, /k/ exhibited VOT values ranging from 22 ms to 103 ms, indicating inconsistent control of aspiration duration.

These findings suggest that while learners recognize voiceless stops as requiring aspiration, they lack stable temporal boundaries, resulting in excessive and uneven aspiration across tokens.

3.3 Voiced Stop Consonants (/b, d, g/)

The voiced stop consonants exhibited greater instability than the voiceless stops**.** The most striking case was the phoneme /b/, which showed an extreme range of VOT values from –147 ms to +173 ms**.** Such a wide range indicates the absence of a consistent phonetic target. The mean VOT of 38 ms places /b/ firmly in the positive VOT range, which acoustically aligns it with voiceless stops rather than voiced ones.

The phoneme /d/ demonstrated relatively more negative VOT values, with a mean of –13 ms, though several productions still fell into zero or positive VOT territory. Similarly, /g/ showed a mean VOT of 10 ms, reflecting short-lag or marginally positive VOT, with individual values ranging from –157 ms to 76 ms.

Overall, the voiced consonants failed to maintain a stable voicing lead, resulting in **overlap** between voiced and voiceless categories. This overlap indicates a breakdown in the phonetic contrast that is essential for intelligibility in English.

The findings of this study demonstrate that teenage EFL learners in Binjai experience persistent difficulty in producing stable Voice Onset Time (VOT) contrasts in English stop consonants. Two dominant patterns emerge from the data: excessive aspiration in voiceless stops and substantial instability in voiced stop production. These patterns indicate that learners have not yet internalized the temporal phonetic boundaries that differentiate English voiced and voiceless plosives, resulting in acoustic overlap that compromises phonemic contrast.

From a theoretical standpoint, these results strongly support Flege's Speech Learning Model (SLM) (1995), which posits that L2 learners struggle to form new phonetic categories when L2 sounds are perceived as similar to L1 categories. Indonesian stop consonants do not employ aspiration as a phonemic feature, and voicing distinctions are not realized through systematic VOT contrasts. Consequently, English voiceless stops /p, t, k/ are often perceived as equivalent to their Indonesian counterparts, leading learners to rely on familiar L1 timing patterns. The hyperaspiration observed in this study particularly in the production of /t/ suggests a form of phonetic overcompensation, where learners are aware of aspiration as a feature of English but lack precise control over its duration.

The instability observed in the production of voiced stops, especially /b/, further illustrates the challenge of phonetic category formation. The wide range of VOT values for /b/, spanning from substantial negative to long positive values, indicates that learners do not consistently associate this phoneme with pre-voicing or short-lag VOT. As a result, many productions acoustically resemble voiceless stops, effectively neutralizing the voicing contrast. This neutralization is particularly problematic in English, where VOT serves as a primary acoustic cue for distinguishing minimal pairs. When voiced and voiceless categories overlap, the listener's ability to accurately decode lexical meaning is compromised.

These findings are consistent with previous studies on Indonesian learners of English, which report similar difficulties in maintaining aspiration and voicing contrasts. However, the present study extends existing research by focusing on teenage learners in a multilingual regional context. The linguistic environment of Binjai, which is influenced by local dialects such as Melayu Deli and Batak, may introduce additional phonetic timing patterns that interact with English production.

This multilingual exposure may contribute to the heightened variability observed in the data, as learners negotiate multiple phonological systems simultaneously.

From a pedagogical perspective, the results highlight the limitations of traditional pronunciation instruction that emphasizes auditory repetition without explicit phonetic guidance. Learners may be able to imitate surface-level pronunciation models, yet still fail to acquire the underlying temporal properties of speech sounds. The instability of VOT boundaries observed in this study suggests a need for instructional approaches that raise learners' awareness of duration and timing as critical components of pronunciation. Incorporating acoustic visualization tools such as PRAAT into pronunciation teaching could provide learners with concrete, objective feedback on their speech production. By observing the visual representation of aspiration and voicing onset, students can better understand how their pronunciation deviates from target norms and gradually refine their articulatory timing.

Overall, this study underscores the importance of integrating acoustic phonetic insights into English pronunciation pedagogy, particularly at the high school level. Addressing VOT-related difficulties early may help learners establish more stable phonetic categories, thereby enhancing intelligibility and reducing long-term fossilization of pronunciation errors

The implications of these findings are critical for communicative intelligibility. The inability to maintain stable VOT boundaries can trigger semantic ambiguity during communication with native speakers. Therefore, pedagogically, English instruction at the high school level should integrate objective acoustic awareness training. Utilizing software such as PRAAT can serve as a visual tool for teachers to provide feedback on students' pronunciation duration, helping them reduce negative L1 interference and improve phonetic accuracy.

4. Conclusion

This study concludes that the production of Voice Onset Time (VOT) in English word-initial plosive consonants by teenagers in Binjai exhibits significant deviations from the standards of native speakers (General American English). The primary findings indicate two dominant articulatory patterns: hyperaspiration in voiceless plosives (/p, t, k/) and voicing instability in voiced plosives (/b, d, g/). The exceptionally high average VOT for the phoneme /t/ (82 ms) reflects a hypercorrection strategy, where teenage learners over-emphasize aspiration a phonetic feature that is absent in the Indonesian (L1) phonological system to distinguish it from their native speech patterns.

Conversely, the failure to produce consistent pre-voicing (negative VOT) in voiced plosives leads to a neutralization of phonetic contrasts. This results in an acoustic overlap where voiced sounds (such as /b/ with a mean of 38 ms) are perceived as voiceless by native speakers. This phenomenon reinforces the Speech Learning Model (SLM), suggesting that L1 interference profoundly influences the mastery of temporal duration in a second language, which ultimately reduces communicative intelligibility and can lead to semantic ambiguity.

Practically, this research recommends a paradigm shift in the pedagogy of pronunciation instruction at the secondary school level. Rather than relying solely on auditory imitation, educators should integrate acoustic phonetic awareness training using specialized software such as PRAAT. By providing visual feedback on aspiration and voicing durations, teenage learners can more objectively correct their articulatory errors. Such an approach is expected to enhance phonetic accuracy and oral fluency, ensuring better intelligibility in global communication.

Acknowledgments

"The author is grateful to the Master's Program in Linguistics at Universitas Sumatera Utara for their invaluable assistance and support in fulfilling the requirements for this master's thesis."

Declarations

Author Contribution Statement

"The first author was responsible for study conceptualization, design, data collection, and drafting the initial manuscript. The second and third authors acted as supervisors, contributing to the data analysis and interpretation, as well as providing critical revisions for important intellectual content. All authors gave their final approval for the manuscript to be submitted."

Funding Statement

The authors received no specific funding for this work.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request

Declaration of Interests Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AI Use Statement

During the preparation of this manuscript, the authors used Gemini 3 Flash solely for language editing (grammar, clarity, and readability). The authors reviewed, revised, and verified the final text and take full responsibility for the content of the publication.

Additional Information

Correspondence and requests for materials should be addressed to: prima.sholihtaun95@gmail.com

References

Arikunto, S. (2006) Metode Penelitian Kualitatif. Jakarta: Bumi Aksara

Arfan, E.F., Nofida, R.D., Tetti, N.M, (2011) "Identitas suara vokal suku banjar berdasarkan frekuensi formantt" Universitas Lambung Mangkurat, Banjarbaru.

Bhattacharya, P., & Ray, J. (2019). Analysis of voice quality using Praat: A study on Bengali vowels. International Journal of Speech, Language and the Law, 26(2), 185-198.

Boersma, P., Weenink, D. (2014). PRAAT: doing phonetics by computer (Version 5.3.77). Amsterdam.

Chen, S. H., & Xu, Y. (2006). Measurement of pitch and intensity perturbation using Praat. Journal of the Acoustical Society of America, 119(5), 3639-3642

.Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. Journal of Phonetics, 27(2), 207–229. https://doi.org/10.1006/jpho.1999.0094

Choi, H., Kim, M., & Broersma, M. (2017). Voice onset time in second language acquisition: Korean learners of English. Journal of Phonetics, 64(2), 1–14.

Dardjowidjojo, S. (2009). English Phonetics and Phonology for Indonesians. Jakarta: Yayasan Obor Indonesia.

Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. Journal of Phonetics, 29(2), 217–230.

Durao (Richards & Weber). (2007). Contrastive approach in phonological disorders therapy: theoretical considerations. Rio de Janeiro, Brasil.

Escudero, P., Boersma, P., & Rauber, A. (2021). Vowel space and formant distribution in L2 speech. Applied Linguistics, 42(5), 923–940.

Ferrand, C. T. (2013). Speech science: An integrated approach to theory and clinical practice. Pearson Higher Ed.

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. Language and Speech, 24(2), 125–146.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. The Journal of the Acoustical Society of America, 89(1), 395–411. https://doi.org/10.1121/1.400473

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp. 233–277). Timonium, MD: York Press.

García-Lázaro, M., Llisterri, J., & Machuca, M. J. (2020). The use of Praat in the analysis of vowel formantts in Spanish. Journal of Phonetics, 83, 101010.

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.

Hartini. L (2011) *"Penerapan Fonetik Akustik Dan Teori Grice Pada Rekaman Penyadapan Telepon Sebagai Alat Bukti Hukum: Kajian Linguistik Forensik Terhadap Percakapan Antara Artalyta Suryani Dengan Jaksa Urip Tri Gunawan"* Tesis. Sekolah Tinggi Hukum Bandung.

Inge Genee, Fangfang Li., (2023) "Voice Onset Time in a language without voicing contrast: An acoustic analysis of Blackfoot oral stops" University of Lethbridge.

Irawan, Y. (2007) Fonetik Akustik. Bandung: Angkasa.

Kent, R. D., & Read, C. (2002). The Acoustic Analysis of Speech (2nd ed.). San Diego: Singular/Thomson Learning.

Kingston, J., & Diehl, R. (2019). Phonetic categorization of voicing: Cross-linguistic VOT patterns. Phonetica, 76(4), 205–230.

Koffi, E. (2021). Relevant acoustic phonetics of L2 English: Focus on intelligibility. CRC Press.

Kreidler, C. W. (2002). Phonology: Critical Concepts (Vol. 4). London: Routledge.

Labov, W., Ash, S., & Boberg, C. (2006). The Atlas of North American English: Phonetics, Phonology, and Sound Change. Berlin: Mouton de Gruyter.

Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M., & Nagy, N. (2013). Sociophonetics: A Handbook. Cambridge: Cambridge University Press.

Ladefoged, P., & Johnson, K. (2014). A course in phonetics (7th ed.). Cengage Learning.

Laver, J. (1994). Principles of Phonetics. Cambridge University Press.

Lestari, D. (2019). Voicing contrasts in Indonesian stops: An acoustic study. Indonesian Journal of Applied Linguistics, 9(3), 656–667. https://doi.org/10.17509/ijal.v9i3.23220

Levy, R., & Law, S. (2022). Acoustic correlates of intelligibility: The role of formants in L2 English learners. Language and Speech, 65(3), 445–465.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Li, S., Zhang, L., & Yang, Y. (2021). Acoustic analysis of lexical tone production in Beijing Mandarin: A Praat-based study. Journal of the Acoustical Society of America, 149(4), 2883-2883.

Liu, C., Chen, L., & Wang, W. (2019). Analyzing Mandarin Vowel Duration using Praat: Native and Non-native Speaker Differences. Journal of the Chinese Language Teachers Association, 54(3), 61-76.

Lubis, R. (2022). Formant vowel distribution in English pronunciation by Indonesian EFL learners. Jurnal Bahasa dan Sastra, 20(3), 201–214.

Major, R. C. (2001). Foreign accent: The ontogeny and phylogeny of second language phonology. Mahwah, NJ: Lawrence Erlbaum Associates.

Miller, A., & Cho, Y. (2020). VOT and second language perception: A cross-linguistic study. Journal of the Acoustical Society of America, 148(2), 789–801.

Mi-Ryoung Kim,(2021) "Voice Onset Time In English And Korean Stops With Respect To A Sound Change" Korea

Misnadin. (2016). Comparative study of VOT in Madurese and English initial plosives. Prosodi: Jurnal Ilmu Bahasa dan Sastra, 10(1), 1–12. https://journal.trunojoyo.ac.id/prosodi/article/view/48

M. N. Al-Azhar. (2001). Audio Forensic: Theory and Analysis. Pusat Laboratorium Forensik Polri Bidang Fisika dan Komputer Forensik.

Putri, A. (2019). A comparative study of vowel formants among Javanese and Minangkabau speakers. Humaniora, 31(2), 156–166.

Peterson, Stacey. (2016). Vowel Dispersion in English Diphthongs: Evidence from Adult Production. Proceedings of AMP 2015 Completed March 15, 2016

Rabiner, L., Bing-Hang Juang,(1993) "Digital Processing of speech signals", Prentice-Hall International Inc,.

Raees, N., Gustianingsih, Alemina Br P., (2024) Phonological Disorder in Speaking Among Stroke Patients: A Neurolinguistics Study. Modality: International Journal of Linguistics and Literature https://ejournal.uinbukittinggi.ac.id/modality/article/view/8891

Rahmawati, F., & Hidayat, T. (2021). Acoustic analysis of English plosives: VOT study on Indonesian high school students. Journal of English Language Studies, 6(1), 55–67.

Ren, Y., & Zhang, L. (2022). Acoustic analysis of creaky voice in Mandarin: A study using Praat. Proceedings of the International Symposium on Chinese Spoken Language Processing, 1-5

Rasya Risyafa P., (2024) A Language Production In Children With Dysarthria Language Disorder: A Psycholinguistic Study.
https://ejournal.uinbukittinggi.ac.id/index.php/modality/article/view/8374

Sadegi, V.S., Yaghmaie K., "Vowel Recognition Using Neural Network", International Journal of Computer Science and Network Society, VOL.6 No.12, December 2006.

Sanjaya, D. (2016). Voice onset time in English plosives produced by Indonesian learners. Lingua Cultura, 10(1), 12–21.

Smith, J., Johnson, A., & Davis, K. (2017). Using Praat for Acoustic Analysis in English Phonetics Courses. Journal of English Linguistics, 45(2), 159-185.

Stevens, K. N. (2000). Acoustic phonetics. MIT Press.

Sugiyono. 2003. Pedoman Penelitian Bahasa Lisan: Fonetik. Pusat Bahasa Departemen Pendidikan Nasional, Jakarta.

Sukardi. 2011. Metodologi Penelitian Pendidikan Kompetensi dan Praktiknya. Jakarta: PT Bumi Aksara.

Sukmawijaya, A., Rahmawati, R., & Yuliana, L. (2020). Analisis akustik bunyi /t/ dalam bahasa Sunda, bahasa Indonesia, dan bahasa Inggris oleh penutur Sunda. Metahumaniora, 10(1), 45–60. https://doi.org/10.24198/metahumaniora.v10i1.26597

Suyudi, Ichwan., Saptano, Debyo. (2016). Analisis Vocal Tract Pada Kajian Akustik Vokal Bahasa Indonesia. Prosiding PESAT (Psikologi, Ekonomi, Sastra, Arsitektur & Teknik Sipil) Universitas Gunadarma - Depok - 20-21 Oktober 2015

Syarfina, Tengku., (2009) "Ciri akustik bahasa melayu deli" Medan

Tilman, Rosa da Costa., Nurhayani, Ika. (2015). Kesalahan Pengucapan Vokal Depan Bahasa Inggris Oleh Mahasiswa Semester I Universitas Nasional Timor Lorosa'e. Jurnal LITE Vol. 11, No. 2. Universitas Dian Nuswantoro: Semarang.

Van Lieshout, Pascal. (2003). PRAAT Short Tutorial A basic introduction. University of Toronto, Graduate Department of Speech Language Pathology, Faculty of Medicine, Oral Dynamics Lab V. 4.2.1, October 7, 2003

Vogel, A. P., Maruff, P., Snyder, P. J., & Mundt, J. C. (2009). Standardization of pitch-range settings in voice acoustic analysis. Behavior Research Methods, 41(2),

Wilson, Ian. (2008). Using Praat and Moodle for Teaching Segmental and Suprasegmental Pronunciation. 2008 World CALL Proceeding.

Zheng, H., & Johnson, K. (2018). Formant analysis of Mandarin speakers learning English vowels. Second Language Research, 34(3), 347–370.