Acoustic signal processing of fundamental frequency (f0) dynamics in persuasive sentences from Indonesian presidential campaign videos

Miftahulkhairah Anwar[1*], *Gede* Satya Hermawan[2], *Sigit* Pramono[1], and *Norliza* Jamaluddin[3]

[1]Faculty of Language and Arts, Universitas Negeri Jakarta, 13220, Indonesia
[2]Faculty of Language and Arts, Universitas Pendidikan Ganesha, 81116, Indonesia
[3]Faculty of Language and Communication, Universiti Pendidikan Sultan Idris, 35900, Malaysia

Abstract. Political discourse studies have developed significantly; however, most previous research has primarily focused on textual linguistic aspects, while prosodic features remain underexplored. This study investigates the intonational contours and Fundamental Frequency (f0) dynamics of persuasive sentences in Indonesian presidential campaign discourse. A computational phonetics approach was employed using Python libraries, namely Parselmouth and NumPy. The data consisted of campaign videos presenting the visions, missions, and policy programs of the three Indonesian presidential candidates: Anies, Prabowo, and Ganjar, published by Kompas TV news portal. The findings reveal distinctive f0 patterns and pitch contour characteristics among the candidates. Anies demonstrates more varied intonational movements, with a maximum f0 of 299.23 Hz, a minimum f0 of 75.06 Hz, and a pitch range of 224.17 Hz. Prabowo exhibits a denser and more stable pitch contour, with a maximum f0 of 299.98 Hz, a minimum f0 of 74.96 Hz, and a pitch range of 225.02 Hz. Meanwhile, Ganjar shows a relatively wider and more dispersed pitch contour, with a maximum f0 of 298.41 Hz, a minimum f0 of 75.05 Hz, and a pitch range of 223.36 Hz. These findings indicate that f0 dynamics contribute significantly to political persuasion and communicative style.

1 Introduction

Speech communication extends beyond syntactic and semantic structures and also involves phonetic and discourse-related dimensions. One important aspect of spoken communication is prosody, which functions as a mechanism for expressing emotion, communicative intention and interpersonal meaning. Mozziconacci [1] argued that speech is communication not only in syntactic and semantic areas, but also in phonetic and discourse areas. Intonation can be a tool for conveying emotion in speech. She is trying to bring the intonation model and framework to show prosody and emotion, and has shown that the sender and receiver can encode and decode prosody to interpret the message. Then, the intonation model uses a

* Corresponding author: miftahulkhairah@unj.ac.id

path of notions, like pitch level or pitch range, as a reference. The data frequently uses the mean and standard deviation of f0 [1].

Linguistic prosody had two distinct threads: lexical prosody and post-lexical prosody. Lexical prosody brings words to life through features like accent, stress, tone, and the rhythm of syllable length. Meanwhile, post-lexical prosody stretches across words, spotlighting phrases, marking boundaries, and painting utterances with melodic contours [2]. This prosodic structure contributes not only to Linguistic interpretation but also to emotional and social communication.

Previous studies have demonstrated that prosodic variation influences how speech is perceived across different communicative contexts. Research on vocal profiling in animated character dubbing, for example, has shown that prosodic manipulation helps convey emotional characteristics and personality traits [3]. Similarly, studies of speech in legal and social contexts indicate that vocal characteristics influence listeners' perceptions of confidence, masculinity, and authority [4]. Furthermore, speech style itself may affect social impressions, producing perceptions of powerful or powerless communication [5]. These findings suggest that vocal properties are not merely acoustic phenomena but also function as persuasive and social resources.

Among acoustic features, f0 is particularly important because it serves as an objective indicator of vocal pitch and intonational variation. Previous research has linked f0 dynamics with charisma and persuasive effectiveness in political speech. Signorello et al. [6], [7] investigated charismatic speech across multiple communicative contexts, including monologues, interviews, and conferences, involving politicians from several linguistic backgrounds such as American, Italian, French, and Brazilian speakers. Their findings revealed systematic variations in vocal behavior across speakers and contexts, demonstrating that personal characteristics adapt according to communicative demands. The study also suggested that charismatic speech is associated with physiological vocal characteristics and context-dependent adjustments in vocal range.

However, studies on political discourse have predominantly focused on linguistic content and rhetorical strategies, while acoustic dimensions remain relatively underexplored, particularly in Indonesian political communication. Moreover, computational investigations of persuasive speech using acoustic signal processing techniques remain limited. Since campaign speeches are designed specifically to persuade audiences and shape public perception, vocal characteristics such as pitch movement and f0 variation may play a significant role in constructing persuasive communication.

Therefore, this study investigates the dynamics of f0 in persuasive utterances produced by Indonesian presidential candidates during campaign speeches. Using a computational phonetic approach through Python-based acoustic signal processing with Parselmouth and NumPy libraries, this study aims to identify pitch contour patterns and vocal variation characteristics among the candidates. By examining mean f0, pitch range, and pitch contour dynamics, this study contributes to a broader understanding of how vocal characteristics function as persuasive resources in political discourse.

2 Method

This study employed a computational approach, specifically computational phonetics, to analyze speech data. Python was used as the primary programming environment, with the Parselmouth library utilized for processing the speech data of the three Indonesian presidential candidates in the 2024 election. The data consisted of campaign videos obtained from the Kompas TV Makassar Bureau news portal. The selected data focused on four-minute segments in which each candidate presented their vision, mission, and policy programs during the final presidential debate. The program 'Debat Capres-Cawapres Pemilu

2024: Debat Kelima' is running on 4th February 2024. Each candidate has four minutes to make the final remarks to the audience as voters.

The selected video sources were chosen because they contained minimal background noise and external auditory interference. To ensure the quality and reliability of the acoustic analysis, applause and other non-speech sounds occurring at the end of the recordings were removed to preserve only the candidates' speech signals. This preprocessing stage was conducted using Wondershare video editing software. After preprocessing, the audio files were labeled as follows:

pidato_video_01 for Anies
pidato_video_02 for Prabowo
pidato_video_03 for Ganjar

The first stage of data processing involved converting the original video files from MP4 format into WAV format. This conversion was necessary because WAV files are more suitable for acoustic analysis and can be processed more efficiently by the Parselmouth library. Subsequently, the WAV files were imported into the Python environment for acoustic processing. Each audio file was processed individually rather than simultaneously to ensure accurate extraction and analysis of the speech features from each candidate.

In addition to Parselmouth, other Python libraries were employed in this study. NumPy was used for numerical computation and statistical measurement, while Matplotlib was utilized to visualize the extracted pitch contours. For f0 extraction, the pitch floor was set at 75 Hz and the pitch ceiling ranged from 300 to 500 Hz. These parameters were maintained consistently throughout the analysis because all speakers in this study were male. The parameter settings followed the recommendations provided in the Praat manual [8], as Parselmouth operates as a Python interface for the Praat speech analysis framework [9].

3 Result and Discussion

3.1 Result

The f0 measurement using Parselmouth, a Python Library. The f0 parameter setup has a pitch floor at 75 Hz and a pitch ceiling at 300 Hz. All speakers are male, so this study does not compare across genders. The process calculates f0 from the vowel, then the consonant, and the pauses that are not included. The results are shown in Table 1.

Table 1 shows that the f0 Max and Min for each speaker are close to each other. They look like they had the same f0 pattern, but if you look closely at the mean f0, Ganjar had a smaller value than the others. To understand this phenomenon, we need to examine the pitch contours of each speaker.

Table 1. The Speakers F0.

Speaker	F0 Max (Hz)	F0 Min (HZ)	Mean F0 (Hz)
Anies	299.23	75.06	160.45
Prabowo	299.98	74.96	163.64
Ganjar	298.41	75.05	127.63

Anies' pitch contours are wave-like, with clearly visible melodic dynamics. This demonstrates a variety of intonation movements, as the speaker frequently raises and lowers his pitch to reflect the desired emphasis. The wave is clearly seen at 100-150 seconds and 150-200 seconds, and in the last second, the contour becomes denser, which means Anies rushes to wrap up all speech in the last minutes. As Alku [10] explains, an increase in f0 can decrease the pitch period according to the relationship $T = 1/f0$. This kind of physical shortening from the waveform creates a higher density of individual acoustic cycles per unit time, as seen in Figure 1. That means the speaker rushes to end the speech before the due time.

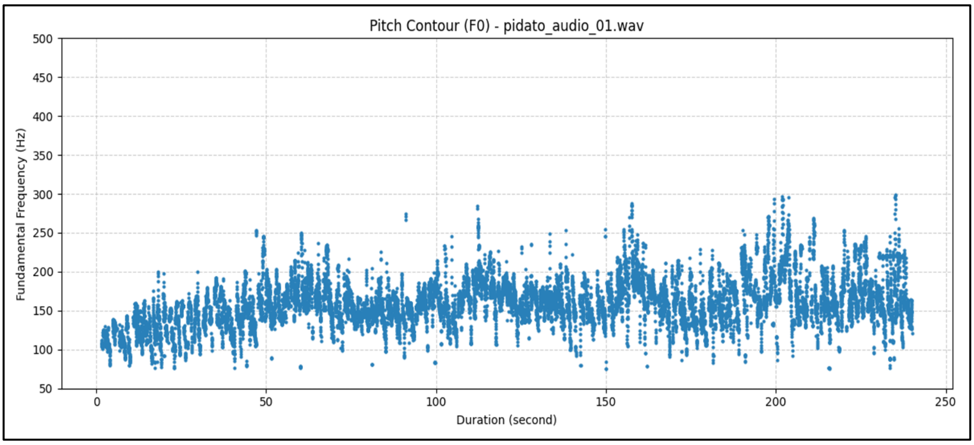

Fig. 1. Anies contours pitch

In contrast to Anies, Prabowo's pitch contour is tighter, and his melodic dynamics are less pronounced. This suggests a more formal or stiff speaking style. It sounds like someone reading a text. The emphasis remains visible, with spikes in vocal pressure at certain points. Figure 2 also shows that Prabowo starts at 100 Hz, gradually rises his f0 until 50 seconds, then remains constant within that range until the end of the speech. It means that after 50 seconds, it is Prabowo's natural speech area, his comfort zone to communicate. While the his f0 outside this baseline for expressive intentions, speech production remains bounded [10].

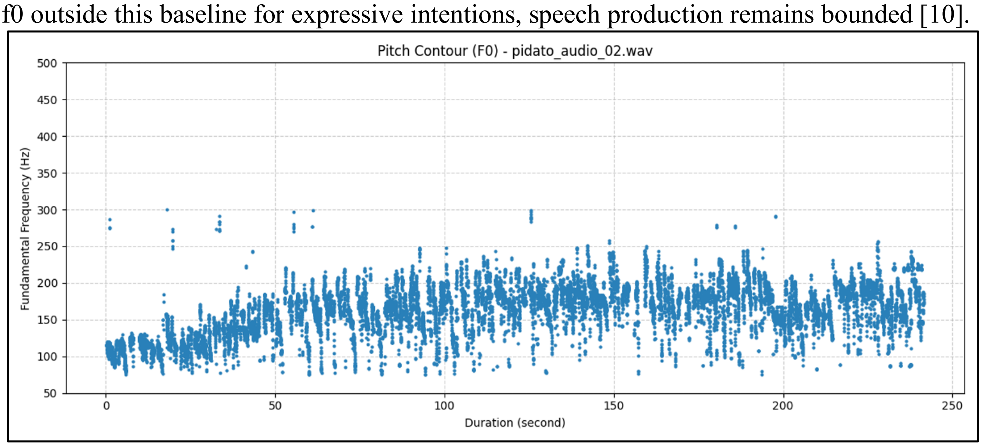

Fig. 2. Prabowo contours pitch

Ganjar displayed a hollower pitch contour, less dense than Prabowo's. This indicates that he spoke slowly and more carefully during his speech. The emphasis was placed at several points, as evidenced by the increasing pitch stress. The high-pitched sounds in Figure 3 show a dot spread from 20 to 170 seconds, with less clear evidence at the end of the speech. This is the reason the mean value is lower than the other.

Then, it can be seen that Prabowo, with a tight and monotonous pitch contour, uses random voice emphasis (pitch tone) when speaking (maybe emphasizing one word, the tone rising at one point), so that the frequency variation of his voice is high in number. In contrast to Anies, whose pitch contour is dynamic but does not have high and random voice emphasis, so that the frequency variation of his voice is stable, or in the data, it does not appear high.

Meanwhile, Ganjar's pitch contour and voice variation are appropriate, even though there are many pauses (which look hollow or suggest distance), but he always emphasizes his voice when speaking, so the variation in his voice is higher.

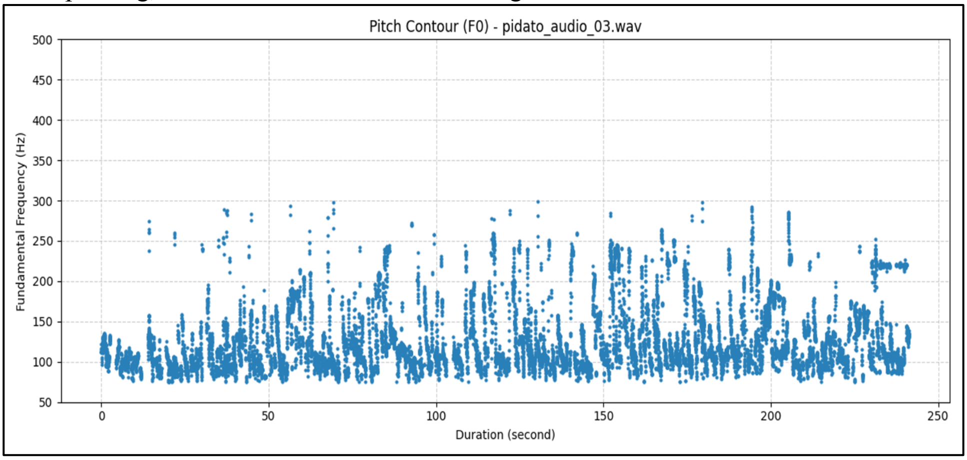

Fig. 3. Ganjar contours pitch

Figure 4 shows the average pitch (mean) of three speakers and the variation in voice (SD). The straight line in the middle bar shows a voice variation, indicating pitch movement around the mean (average pitch). The Anies deviation starts from 130-190 Hz, Prabowo deviation starts from 128-198 Hz, and Ganjar starts from 86-168Hz.

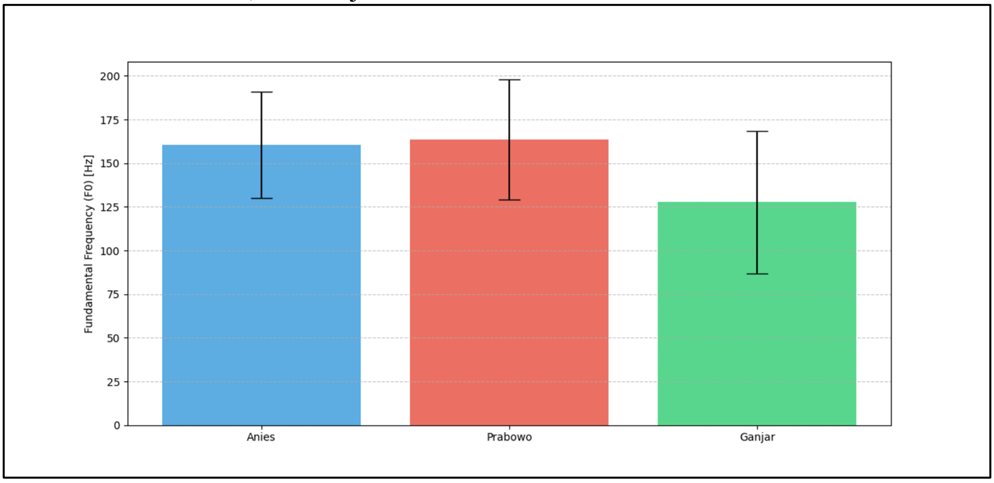

Fig. 4. Average pitch and voice variation

For comparison, Anies has an average pitch of 160.45 Hz when giving a speech, Prabowo has 163.64 Hz, and Ganjar has 127.63 Hz. If we look at the voice variation, Ganjar appears dynamic, using low and high tones well, with an SD f0 of 40.93. Because the range of his vocal variation is wider, he is more dynamic when speaking. Even Prabowo looks more dynamic, not monotonous, with an SD f0 of 34.39 Hz, and Anies, who, in the pitch contour, looks dynamic, has an SD f0 of 30.29. This means that dynamics play a meaningful role in political persuasion and communicative style.

The data show that, even though the Probowo pitch contour was dense, the voice variation indicates otherwise: it is more melodic than Figure 2 suggests. It happened to Ganjar too, he has a more dynamic range than two speakers. Anies data only show a significant correlation among pitch contour, mean pitch, and voice variation. Variation in mean and SD reflects not only physiological voice characteristics but also strategic vocal modulation that shapes audience perception. Previous research suggests that vocal attributes influence judgments of credibility, authority, emotional engagement, and leadership effectiveness [11].

3.2 Discussion

In this paper, we study the f0 dynamic in a persuasive sentence. The speakers are three Indonesia president candidates in 2024. The result shows f0 dynamic use to persuade the audience to choose the right leader. Anies had a dynamic flow, using high f0 and low f0 precisely to convey the message to the audience.

Based on a systematic review of research over the past two decades, Barkat-Defradas et al. [11] showed that charisma is shaped by the use of high f0 and the variability of f0 perceived by listeners. This finding also aligns with a study by Signorello et al [6], [7].

Prabowo's tight and stagnant pitch contour may be due to his military background. Studies show that military training produces changes, particularly in the intensity of the vowel /a/ and the fricative consonants /s/and /z/. The study also showed that these changes are more likely to occur if military training begins at a younger age (17-22 years old) [12].

Furthermore, Ganjar's seemingly hollow pitch contour is likely attributable to his Central Javanese origins. Previous studies have shown that the Javanese sound has hissing vowels following heavy stops [13]. They also have vowels with a lower first formant (F1) and a higher second formant (F2) in the same context [14]. Javanese male speakers tend to have lower frequencies across all words compared to Javanese female speakers. Javanese declarative sentences begin with a high tone, with female Javanese speakers exhibiting greater prominence [14], [13].

From this research, we can also see that pitch contour does not directly indicate dynamic f0. Ganjar's case shows that melody can be indicated by variations in sound and average f0. A wide range of sound variations indicates effective use of high and low f0 to convince listeners [15]. Three speakers used the dynamic f0 during speech to bold the message.

The use of monologue data in this study is also appropriate. Using interview data would yield narrow findings, and the persuasive purpose might even be obscured. Consequently, there would be little variation, with a narrow f0. In the future, this research can also be expanded by examining listener perceptions to clarify the relationship between the use of high and low f0 and persuasive power.

4 Conclusion

This study demonstrates that f0 dynamics and intonational contours play a significant role in constructing political persuasion in Indonesian presidential campaign videos. Each presidential candidate exhibits distinctive prosodic characteristics in realizing persuasive sentences. Anies Baswedan tends to employ more varied and dynamic intonational patterns, Prabowo Subianto demonstrates denser and more stable pitch contours, while Ganjar Pranowo displays wider and more dispersed pitch patterns. These differences in f0 patterns indicate that prosodic strategies contribute to the construction of each candidate's political communication style, emotional emphasis, and persuasive image. The findings confirm that political communication is not solely constructed through textual aspects, such as lexical choice and sentence structure, but also through the acoustic realization of speech. Therefore, prosodic features, particularly f0 dynamics, can be understood as strategic rhetorical resources in contemporary political discourse. This study also advances the field of political linguistics by integrating pragmatic analysis with computational phonetics via Python-based acoustic signal processing. Theoretically, this study contributes to the development of acoustic phonetics, pragmatics, and political discourse analysis. Methodologically, the findings demonstrate that a computational phonetics approach can effectively reveal prosodic patterns in digital political communication. Future studies are recommended to investigate additional acoustic features, such as intensity, speech tempo, and pauses, and to expand the dataset across various political communication contexts and other digital media platforms.

This publication was funded by the Indonesia Endowment Fund for Education (LPDP) on behalf of the Indonesian Ministry of Higher Education, Science and Technology, and managed under the EQUITY Program (Contract Number 4308/B3/DT.03.08/2025 and B/284/UN39/HK.07.00/2025).

References

1. S. Mozziconacci, Proc. Speech Prosody 2002, 1 (2002). https://doi.org/10.21437/SpeechProsody.2002-1
2. C. Themistocleous, Neurosci. Biobehav. Rev. 175, 106210 (2025). https://doi.org/10.1016/j.neubiorev.2025.106210
3. A. Crochiquia, A. Eriksson, M.A.S. Fontes, S. Madureira, Delta 36(3) (2020). https://doi.org/10.1590/1678-460X2020360311
4. D. Chen, Y. Halberstam, A.C.L. Yu, PLoS One 11, e0164324 (2016). https://doi.org/10.1371/journal.pone.0164324
5. B. Erickson, E.A. Lind, B.C. Johnson, W.M. O'Barr, J. Exp. Soc. Psychol. 14, 266 (1978). https://doi.org/10.1016/0022-1031(78)90015-X
6. R. Signorello, D. Demolin, N.H. Bernardoni, et al., J. Voice 34, 808.e1 (2020). https://doi.org/10.1016/j.jvoice.2019.04.007
7. R. Signorello, F. D'Errico, I. Poggi, D. Demolin, Proc. ASE/IEEE Int. Conf. Social Computing, 435 (2012). https://doi.org/10.1109/SocialCom-PASSAT.2012.68
8. P. Boersma, D. Weenink, Praat: Doing Phonetics by Computer, University of Amsterdam, Amsterdam (2024). Available at: https://www.praat.org
9. Y. Jadoul, B. Thompson, B. de Boer, J. Phonetics 71, 1 (2018). https://doi.org/10.1016/j.wocn.2018.07.001
10. T. Bäckström, O. Räsänen, A. Zewoudie, P. Perez Zarazaga, L. Koivusalo, S. Das, E. Gómez Mellado, M. Bouafif, D. Ramos, S. Kadiri, P. Alku, M. Vali, Introduction to Speech Processing, 2nd edn., Aalto University, Fundamental frequency (F0) (2022). https://doi.org/10.5281/zenodo.6821774
11. M. Barkat-Defradas, M. Raymond, A. Suire, in Voice Attractiveness, ed. by B. Weiss, Springer Nature Singapore, Singapore, 55 (2021). https://doi.org/10.1007/978-981-15-6627-1_4
12. C.L. Nascimento, A.C. Constantini, L.F. Mourão, J. Voice 30, 61.e1 (2016). https://doi.org/10.1016/j.jvoice.2015.03.005
13. N. Mawarni, T. Syarfina, P.A. Zulhantiar, R. Rangkuti, Fonologi: J. Ilmuan Bahasa Sastra Inggris 2, 63 (2024). https://doi.org/10.61132/fonologi.v2i4.1136
14. S.C.A. Xu, C. Wilson, Proc. 20th Int. Congress of Phonetic Sciences, 843 (2023).
15. C.A. Klofstad, R.C. Anderson, S. Nowicki, PLoS One 10, e0133779 (2015). https://doi.org/10.1371/journal.pone.0133779