Computational Approaches to Quantitative Analysis of Pause Duration in Taiwan Mandarin

I-Ping Wan*
Phonetics and Psycholinguistics Laboratory, Graduate Institute of Linguistics / Research Center for Mind, Brain, and Learning / Program in Teaching Chinese as a Second Language
National Chengchi University
ipwan@mail2.nccu.tw

Yu-Ju Lai
Phonetics and Psycholinguistics Laboratory
Graduate Institute of Linguistics
National Chengchi University
112555008@g.nccu.edu.tw

Pu Yu*
Phonetics and Psycholinguistics Laboratory
Graduate Institute of Linguistics
National Chengchi University
acadyupu@gmail.com

Abstract

This study presents a quantitative analysis of pause-duration patterns in a Mandarin spoken corpus to establish a baseline for prosodic and cognitive assessment. Drawing on cross-linguistic research, the distribution of pause patterns is viewed as reflecting multiple underlying factors. Longer pauses aligned with prosodic and syntactic boundaries indicate more deliberative and planned discourse rather than spontaneous speech. Such settings place higher demands on cognitive and articulatory planning, producing extended thinking time as speakers handle complex topics and specialized terminology. The spoken corpus was automatically processed and annotated using an in-house alignment and pause-tagging pipeline. Outlier detection with a 3.0×IQR threshold retained 35,474 tokens and removed extreme values exceeding 1,016 ms. Short and medium pauses remained stable across mean, median, and variability measures, while long pauses showed a moderate reduction (16,436 to 15,420 tokens), with mean duration decreasing from 535 to 426 ms and standard deviation sharply reduced from 786 to 169 ms, while the median stayed around 370–380 ms. These findings demonstrate that automatic cleaning primarily removed aberrant values while preserving linguistically meaningful long pauses. This baseline from non-impaired adult speakers underscores the need for corpus-specific frameworks and offers a reference point for cross-linguistic research on speech planning.

Keywords: Pause Duration, Speech Disfluency, Computational Approaches, Taiwan Mandarin, Spontaneous Speech, Quantitative Analysis

1 Introduction

Speech pauses and silences have been recognized as integral components of spoken interaction, reflecting cognitive processing, social norms and communicative strategy rather than mere absences of sound. Once viewed as interruptions, these pauses are now understood to serve important semantic, pragmatic and cognitive functions (Saville-Troike, 1985; Zuo, 2002). Socio-pragmatic studies have emphasized that silence, hesitations and pauses serve diverse interactional functions beyond turn-taking. For instance, Ephratt (2007) categorized pauses into four types including stillness, planned pauses, silencing and eloquent silence, highlighting their role as meaningful communicative acts. Similarly, Olaoye (2020) offered a typology of silence, including stillness, pauses, eloquent silence and judicial or commemorative silence, showing how these forms operate as communicative tools to express respect, humility, self-control, and conflict avoidance. By situating silence within sociolinguistic and pragmatic theory, these studies highlight silence as

Proceedings of the 37th Conference on Computational Linguistics and Speech Processing (ROCLING 2025), pages 116–123
November 20-22, 2025

a culturally and religiously embedded strategy with perlocutionary effects on interlocutors.

Building on this foundation, research on hesitation phenomena and filler use has offered valuable insights into language production processes. Modeling how speakers manage planning and execution through hesitations and disfluencies can inform both human–computer interaction and clinical applications. Grosjean and Collins (1979) long ago provided early evidence that speakers adjust breath and pause placement in read speech to match pre-planned syntactic structures, further linking silent pauses to deliberate production planning. A segment of silence exceeding 150 milliseconds in duration was classified as a speech pause (Maassen & Povel, 1984; Hammen et al., 1994). Moreover, a number of corpus studies of academic speech suggested that a relatively high proportion of long pauses align closely with prosodic or syntactic boundaries. This pattern is reminiscent of formal or highly prepared speech genres such as reading aloud and political speeches (Duez, 1982; Grosjean & Collins, 1979; Ferreira, 1993). Also, Ferreira (1993) argued that prosodic planning, rather than purely syntactic parsing, governs pause insertion, especially at sentence ends, a view echoed by Krivokapić et al. (2020), who treated grammatical pauses as anticipatory prosodic boundary events. Zellner (1994) also emphasized the close link between pauses, prosody and information packaging. In a prepared speech, pauses are more structurally aligned and semantically functional. Other work also confirmed style-sensitive variation where Gustafson-Capkova et al. (2001) observed systematic differences in pause placement, frequency and duration across spontaneous dialogue, amateur reading and professional broadcasting. In the latter, pauses were shorter, less frequent and more tightly aligned with syntactic boundaries, consistent with higher planning and rhetorical control.

In terms of pause patterning, Campione and Véronis (2002) compared pause patterns in read vs. spontaneous speech across five languages and found that read speech exhibited a more regular bimodal distribution of short and medium-length pauses. In contrast, spontaneous speech introduced a third mode, which showed rather long pauses (often >1000 ms), typically associated with hesitation, lexical search, or real-time syntactic planning. This suggests that formal and pre-planned speech tends to contain longer structurally aligned pauses, while extremely long pauses are characteristic of high planning load in spontaneous dialogue. In their studies, a methodological caveat emerges when setting the lower boundary of "long pause" at >250 ms: such a threshold may conflate two functionally distinct phenomena including boundary-aligned silences in formal registers and hesitation-induced delays in spontaneous speech. It is critical to distinguish these planned boundary-aligned pauses from extremely long pauses that more likely reflect spontaneous cognitive planning difficulties (Campione & Véronis, 2002). Šturm (2023) further compared news reading with poetry reading and demonstrated that pause patterns are shaped not only by genre but also by the underlying text structure (explicit vs. implicit cues), highlighting how increasing planning demands and formality elevate discourse-based pause control.

Computational and empirical approaches have sought to model disfluencies and pause phenomena in large-scale speech data. Aijmer (2011) and Crible (2017) utilized prosodic features as cues to indicate the presence of prosodic markers. Betz et al. (2020) investigated the form, function and modeling of disfluencies, especially hesitations, in human speech and their integration into spoken dialogue systems, providing empirical data on the frequency, distribution and acoustic characteristics of silent and filled pauses. Similarly, Wan and Allassonnière-Tang (2021) present a connectionist model of Mandarin speech production to examine how word frequency and position within an utterance influence the occurrence of speech errors, using corpus-based data and computational simulations. Zhang (2024) further applied quantitative methods to spontaneous speech corpora to uncover sociolinguistic variation linked to speech planning. These findings suggested that features such as pause duration, frequency and distribution can support speaker-state detection and automatic speech processing, extending the relevance of pause research beyond linguistics into computational and even forensic applications.

Clinical and cognitive research has increasingly begun to explore pauses and silences as sensitive markers of neurological and cognitive status. Imre et al. (2022) analyzed silent pauses, hesitations and irrelevant utterances in phonemic and semantic fluency tasks, demonstrating that silence-related parameters such as the length of pauses can effectively differentiate between individuals with mild cognitive impairment and healthy controls. In a complementary study, Sluis et al. (2020) presented an automated approach to analyzing pausing behavior in the speech of people with dementia using the Calpy open-source speech processing toolkit. They found progressive increases in pause duration and proportion of silence across groups, alongside a rise in very long pauses (≥2000 ms) and decreases in total speech duration and mean phrase length, demonstrating that automated pause detection can effectively capture speech disfluencies associated with dementia and support future diagnostic and communication research.

Therefore, these strands of research indicate that pauses and silences are multi-layered phenomena bridging sociocultural, cognitive, and computational domains. However, despite substantial advances, most of this work has been conducted on English or other major European languages, and there remains a paucity of comparable studies in Mandarin. Chen et al.(2022) further examined how discourse functions are reflected through phonological or acoustic features. However, there is still a lack of integrated corpora that combine detailed pause-duration measurements from healthy speakers with the methodological rigor necessary for later comparison to clinical populations. Existing studies either focus on the qualitative or typological aspects of silence, or they apply automated methods primarily to clinical or task-based data without establishing a robust baseline from non-impaired speech in naturalistic settings. Therefore, this study aims to fill this gap by constructing a quantitative corpus-based resource of pause duration in Taiwan Mandarin, providing a robust baseline of silent and filled pauses in naturalistic speech. This corpus not only enables direct comparison with existing English-language studies but also lays the groundwork for future research on aging and clinical populations.

In this paper, we address this gap by constructing a speech-pause corpus that provides high-quality and time-aligned pause data from non-impaired speakers. This corpus is designed to support cross-sectional and longitudinal analyses of pause duration and distribution. By combining socio-pragmatic insights with computational modeling and corpus-based methods, our approach aims to advance both theoretical understanding and practical applications of pause analysis in naturalistic speech. Ultimately, we envision that this resource can be extended to high-risk and aging populations, enabling comparative research on pause behavior as an indicator of cognitive and communicative change in the near future. Questions to be investigated include the following:

1. How can a dedicated speech-pause corpus of non-impaired speakers be designed and annotated to capture detailed pause-duration information across spontaneous speech?
2. To what extent do pause-related parameters, such as number of pauses, average pause length or distribution, provide a reliable baseline for future comparisons with aging and clinical populations?
3. How can insights from socio-pragmatic studies of silence and computational modeling of disfluencies be integrated to improve the automatic detection and classification of pause phenomena?
4. In what ways can such a corpus support cross-linguistic or cross-task analyses, enabling the identification of sociolinguistic variation and potential early markers of cognitive decline?

2 Methodology

A subset of the corpus, totaling 16 hours, 8 minutes, and 2 seconds, drawn from a larger 202-hour multimodal Mandarin speech database, was automatically annotated using Praat (Boersma & Weenink, 2023–2025) for fine-grained analysis of features such as fillers and silent pause-related phenomena. This section outlines the participants, data collection procedures, annotation schema and analysis methods used in the study.

All participants were native speakers of Taiwan Mandarin (N = 4; 1 male, 3 females; age range = 23–25 years, M = 24.2, SD = 0.7). Although the corpus size used here is relatively limited, it was intentionally designed as a controlled case study focusing on young adult speakers with comparable linguistic and cognitive profiles. The goal of this study is not large-scale modeling, but to provide a proof-of-concept analysis demonstrating how automatic annotation can reveal pause and filler patterns in naturalistic speech.

Recordings were made in controlled environments using high-quality audio equipment. The primary content comprises graduate-level classroom settings, including instructor lectures and interactive seminar-style discussions between instructors and students. Notably, over 97% of the annotated utterances showed no statistical outliers in pause duration, indicating a high degree of internal consistency and reliability in the dataset. The combination of academic lectures, seminar discussions, free conversations, and short cognitive-linguistic exercises ensures a rich distribution of spontaneous speech, encompassing a wide range of pause types and speech planning demands.

Drawn partially from graduate classroom discussions, the corpus represents a semi-spontaneous academic register rather than a fully unplanned conversation. However, the speakers produced their utterances without any prepared script or reading material, and the recordings capture natural pauses, hesitations and fillers characteristic of spontaneous speech production. This makes the data appropriate for a case study of cognitive and prosodic pause behaviors in controlled academic discourse, which complements findings from more casual conversational corpora.

Regarding the nature of our speech data, we agree that some portions of the corpus (e.g., classroom lectures) may reflect a more deliberative and planned register. However, these data were chosen because they still involve spontaneous verbal responses, turn-taking, and hesitations typical of natural speech in academic contexts.

We employed an in-house automatic phonetic alignment pipeline developed and refined over several years in the laboratory, rather than relying on open-source tools.[1] This system, combined with manual verification, allows for highly accurate segmentation and annotation. Pauses are operationalized as segments of silence or silent pauses detected by our automated tagging procedure. Each pause instance is annotated with start time, end time, duration, and position relative to syntactic boundaries. From these annotations, we extracted the number of pauses, mean pause duration and distributional patterns from spontaneous speech. Metadata included various speakers, speech type and speech rate. These measures in the future hope to provide a normative baseline for future comparisons with aging or clinical populations. Quantitative analyses include descriptive statistics to identify pause-duration profiles.

Pause duration was identified by detecting segments of silence in the acoustic waveform. According to Maassen & Povel (1984) and Hammen et al. (1994), the data were categorized based on two duration thresholds, which were 150 milliseconds and 250 milliseconds, resulting in three distinct groups: pauses shorter than 150 ms, pauses between 150 and 250 ms, and pauses exceeding 250 ms.

3 Data Analysis

We first tested whether data cleaning materially altered the distribution of pause categories, as shown in Figure 1.

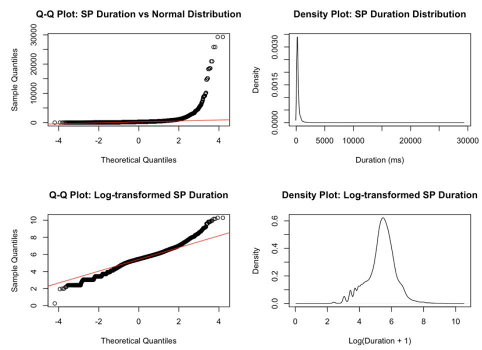

Figure 1: Distributional characteristics of speech pause duration in raw and log-transformed scales **.**

Pause durations showed a pronounced right-skewed, non-normal distribution (n = 36,490), with a peak around 230 ms and a long upper tail. Long pauses were disproportionately affected by outlier removal: 1,016 long pauses were excluded, reducing the mean from 535 to 426 ms (−20%) and compressing the standard deviation (786 to 169 ms). These changes primarily truncated extreme hesitations rather than altering the median (380 to 370 ms), suggesting that the core distribution of boundary-aligned pauses remained stable. The cleaned corpus therefore reflects a clearer

[1] The speech data were processed using a self-supervised in-house phonetic alignment pipeline developed with Praat scripting and custom Python routines, rather than relying on forced-alignment toolkits. The system performs automatic segmentation, boundary detection and iterative self-correction through acoustic feature learning, allowing cross-linguistic adaptability (see Wan et al., 2024, for how Thai preschoolers learn Mandarin). This study, however, does not address prosody or intonation, as its primary focus lies in the analysis of pause and hesitation phenomena within spontaneous speech.

distinction between short/medium pauses, which often aligned with prosodic or syntactic boundaries, and very long pauses, which tend to index planning or hesitation in spontaneous speech (cf. Campione & Véronis, 2002). Because pause durations are non-normally distributed, non-parametric methods and median/IQR statistics are used. This approach preserves linguistically meaningful contrasts between routine boundary pauses and hesitation-driven silences, while minimizing the influence of outliers.

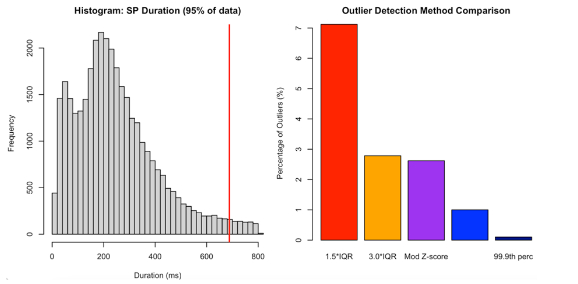

Figure 2: Detection of outliers in pause durations

In Figure 2, outlier detection analysis revealed substantial upper-tail extremes requiring data cleaning prior to modeling. Using the 3.0×IQR method (threshold ≈1,016 ms) alongside a modified Z-score approach (|z| > 3.5), we identified approximately 3% of pauses as outliers. In contrast, the standard 1.5×IQR criterion flagged over 7% of pauses, which was deemed overly restrictive for preserving natural speech variability. Frequency analysis of the central 95% of data showed a right-skewed distribution peaking around 200–300 ms, consistent with known pause distributions in formal and semi-formal speech. By selecting the 3.0×IQR criterion, we retained linguistically meaningful long pauses while trimming only extreme hesitation events, resulting in a final dataset of 35,474 observations capped at about one second. This procedure preserves the contrast between short/medium pauses—often aligned with prosodic or syntactic boundaries—and very long pauses, which tend to index planning or hesitation in spontaneous speech (Campione & Véronis, 2002).

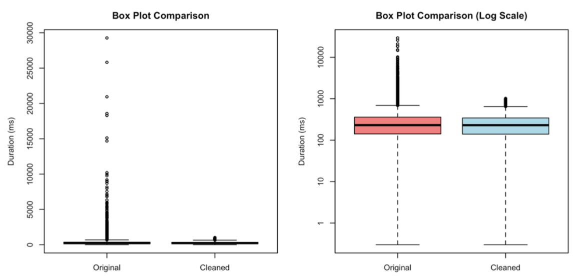

Figure 3: Impact of outlier removal on speech pause duration distributions.

As shown in Figure 3, outlier removal using the 3.0×IQR threshold produced minimal impact on the central distribution while substantially reducing extreme variability. Median pause duration remained essentially unchanged (≈230 ms), while the interquartile range decreased moderately, indicating that core pause behavior was preserved. The most pronounced effect was the elimination of extreme upper outliers without distorting the underlying distribution. Short pauses (<150 ms; n=9,767) and medium pauses (150–250 ms; n=10,287) were unaffected by data cleaning, retaining virtually identical means, medians, and standard deviations. In contrast, long pauses (>250 ms) showed the largest adjustment (n reduced from 16,436 to 15,420), with mean duration decreasing from about 535 to 426 ms and standard deviation sharply reduced, while the median shifted only slightly (380→370 ms). This selective effect confirms that the procedure primarily targeted aberrant values in the upper tail while preserving linguistically meaningful pause patterns. Short and medium pauses continue to represent routine boundary-aligned silences, whereas the cleaned long-pause category better reflects legitimate planning-related hesitations rather than measurement noise, aligning with established pause typologies (Campione & Véronis, 2002).

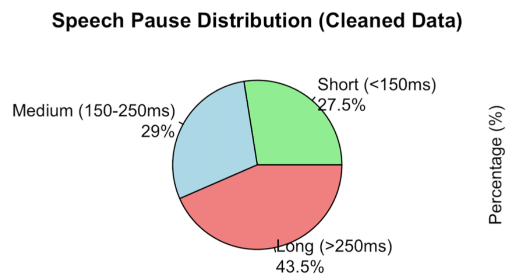

Figure 4: Speech pause distribution by duration category following data cleaning.

The present analysis revealed a distinctive pause distribution pattern that deviates substantially from typical conversational speech norms reported in the literature (Figure 4). The observed distribution—short pauses (<150ms): 27.5%, medium pauses (150-250ms): 29.0%, long pauses (>250ms): 43.5%—contrasts markedly with expected ranges where short pauses typically comprise 40-60% and long pauses 10-30% of total pause events. This inverted pattern, characterized by a predominance of long pauses and relative scarcity of brief hesitations, suggests speech production involving heightened cognitive processing demands rather than spontaneous discourse.

Several factors may account for this distributional profile. The elevated proportion of long pauses likely reflects deliberative speech planning processes, indicating that speakers engaged in more cognitively demanding language production requiring additional processing time for lexical access, syntactic formulation, or discourse organization. The reduced frequency of micropauses and brief hesitations suggests less spontaneous, more controlled speech output characteristic of formal register or task-specific contexts. This pattern is consistent with speech elicited in academic interviews, formal presentations, or complex narrative tasks where speakers prioritize accuracy and coherence over fluency.

The linguistic implications extend beyond simple temporal measurements to suggest fundamental differences in speech production mechanisms. The predominance of longer articulatory timing intervals may indicate enhanced monitoring processes, increased attention to phonetic precision, or elevated cognitive load associated with L2 speech production or specialized discourse domains. These findings underscore the importance of considering contextual factors when interpreting pause patterns and highlight the need for corpus-specific normative data in prosodic boundary analysis.

Within-category frequency analysis revealed distinct distributional characteristics across pause types in the cleaned dataset (n = 35,474), as shown in Figure 5. Short pauses (<150ms) exhibited a right-skewed distribution with modal frequency around 80-90 ms and high consistency between median (80.00 ms) and mean (83.32 ms), indicating minimal internal variability. Medium pauses (150-250ms) demonstrated the most symmetric distribution with peak frequency at 200 ms and perfect convergence of median and mean values (200.00 ms), reflecting highly standardized phrase boundary timing. Long pauses (>250ms) showed pronounced right skew with median (370.00 ms) substantially lower than mean (426.24 ms), indicating considerable internal heterogeneity despite outlier removal. The long pause category maintained an extended upper tail reaching the 1,016 ms threshold, suggesting that even within linguistically valid boundaries, substantial variation exists in processing-related articulatory timing intervals.

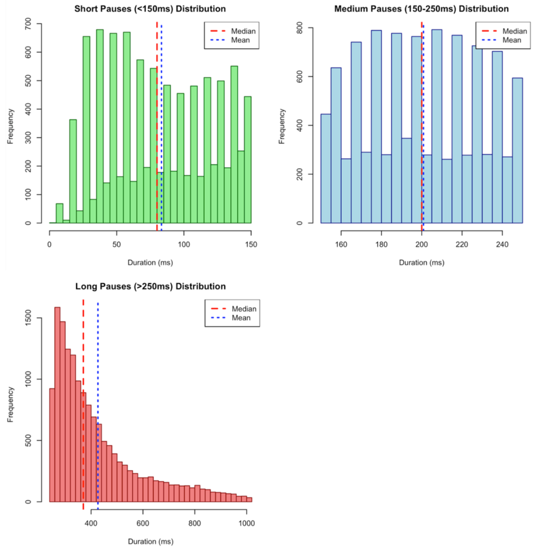

Figure 5: Frequency distributions of speech pause duration by category following outlier removal.

This figure effectively demonstrates that the three-category classification captures fundamentally different pause phenomena, with each category showing distinct statistical properties that justify separate analytical treatment.

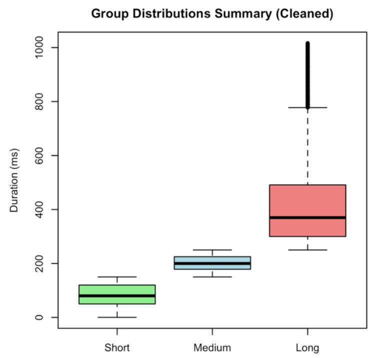

Figure 6: Comparative box plot distributions of pause duration categories in cleaned dataset.

Box plot comparison across pause categories confirmed distinct non-overlapping duration ranges with varying internal distributions following data cleaning. Short pauses demonstrated tight clustering with median at 80.00 ms, narrow interquartile range (51.14-110.00 ms), and minimal outliers, indicating highly consistent micropause timing. Medium pauses exhibited the most compact distribution with median at 200.00 ms and symmetrical quartile spacing (174.36-

225.64 ms), reflecting standardized phrase boundary durations. Long pauses showed the greatest variability despite outlier removal, with median at 370.00 ms, expanded interquartile range (290.91-481.09 ms), and extended upper whiskers reaching the 1,016 ms threshold. The clear separation between categories validates the literature-based classification scheme, while the progressive increase in variability from short to long pauses reflects the transition from automated articulatory timing to cognitively-mediated processing intervals.

This box plot effectively summarizes the key finding that the three categories represent genuinely distinct pause phenomena with different underlying timing mechanisms.

For word-level pauses labeled sp, cleaning produced a small but reliable shift in category composition: the proportion of long pauses decreased by 1.58 percentage points (from 45.04% to 43.47%), with corresponding increases in short (+0.77 pp, from 26.79% to 27.56%) and medium (+0.81 pp, from 28.17% to 28.98%). The association between dataset (Original vs. Cleaned) and category was significant, χ^2 (2) = 18.10, p = 1.17×10^{-4}, Cramér's V = 0.016 (small effect). Because the cleaned set is a subset of the raw set, this test quantifies a composition shift rather than independence.

The significant change is expected given the rule that removes extremely long pauses; the effect size is small (V≈0.016), indicating that cleaning mainly trims the right tail without materially altering central tendencies. Substantively, inferences about typical pause behavior should remain stable, while metrics sensitive to heavy tails (e.g., variance, mean) become less influenced by outliers. For rigor, if token-level retention flags are available, a paired/marginal-homogeneity test can confirm the finding; additionally, a sensitivity analysis across alternative cutoffs (e.g., 800–1,200 ms) can demonstrate robustness.

4 Conclusion

Based on cross-linguistic research findings, the distinctive distributional pattern observed in this Chinese corpus likely reflects several underlying factors. The speech production characteristics suggest more deliberative and planned discourse, potentially originating from formal or academic contexts with reduced spontaneous rapid speech. The cognitive processing patterns indicate increased demands for articulatory planning, with language production involving extended thinking time that may reflect topic-specific complexity or cognitive load.

The data collection context provides an interpretive framework for these findings. The corpus appears to derive from structured interactions such as interviews, presentations, or academic discussions, where speakers engage with specialized content requiring careful formulation. Notably, the speakers may represent non-native Chinese users presenting advanced academic material, a context that inherently promotes more cautious speech production with extended processing intervals. This linguistic environment naturally facilitates longer pause durations as speakers navigate complex conceptual material while managing potential language proficiency constraints.

These findings underscore the importance of contextual factors in prosodic boundary analysis and highlight how discourse demands, speaker characteristics, and communicative settings interact to shape temporal speech patterns. The results provide valuable baseline data for understanding pause distributions in formal Chinese academic discourse and demonstrate the necessity of corpus-specific normative frameworks for cross-linguistic prosodic research.

Acknowledgments

We appreciate the valuable and constructive comments from two anonymous reviewers. Our deepest appreciation goes to Dr. Chain-wu Lee for his continuous cutting-edge high-tech programming support in constructing all the corpora in the Phonetics and Psycholinguistics lab at National Chengchi University. All remaining errors in the analysis of interpretation are our own. This research was supported in part by the Ministry of Education, Taiwan, under the International Integrated Collaboration Project for the University Alliance in the Czech Republic (ICU) & University Academic Alliance in Taiwan (UAAT) to the first author (113F113085).

References

Aijmer, Karin. (2011). Well I'm not sure I think… The use of well by non-native speakers. *International Journal of Corpus Linguistics*, 16, 231–254.

Betz, Simon, et al. (2020). Hesitation Processing Analysis Using Continuous Mouse-Tracking and Gamification. In *Elektronische Sprachsignalverarbeitung 2020. Tagungsband der 31. Konferenz.* (Vol. 95).

Boersma, Paul & Weenink, D. (2023-2025). *Praat: Doing phonetics by computer* [Computer program] (Version 6.4.27). Retrieved January 27, 2025, from http://www.praat.org/

Campione, Estelle, & Jean Véronis. (2002). A large-scale multilingual study of silent pause duration. In *Speech prosody* (Vol. 2002, pp. 199-202).

Chen, Pin-Er, et al. (2022). Analyzing discourse functions with acoustic features and phone embeddings: Non-lexical items in Taiwan Mandarin. *International Journal of Computational Linguistics & Chinese Language Processing*, 27(2).

Crible, L. (2017). Discourse markers and (dis) fluency across registers. *A Contrastive Usage-Based.*

Duez, Danielle. (1982). Silent and non-silent pauses in three speech styles. *Language and speech*, 25(1), 11-28.

Ephratt, Michal. (2007). On silence—introduction. *In Silence in Culture and in Interpersonal Relations* (pp. 7–25). Resling: Tel Aviv.

Ferreira, Fernanda. (1993). Creation of prosody during sentence production. *Psychological review*, 100(2), 233.

Grosjean, François, & Maryann Collins. (1979). Breathing, pausing, and reading. *Phonetica*, 36, 98–114.

Gustafson-Capkova, Sofia, & Beata Megyesi. (2001). A comparative study of pauses in dialogues and read speech. In *INTERSPEECH* (pp. 931-934).

Hammen, Vicki L., Kathryn M. Yorkston, and Fred D. Minifie. (1994). The effect of temporal alterations on speech intelligibility in Parkinsonian dysarthria. *Journal of Speech and Hearing Research*, 37(2), 244–253. https://doi.org/10.1044/jshr.3702.244

Imre, Nóra, et al. (2022). Temporal speech parameters indicate early cognitive decline in elderly patients with type 2 diabetes mellitus. *Alzheimer Disease & Associated Disorders*, 36(2), 148-155.

Krivokapić, Jelena, Will Styler, & Benjamin Parrell. (2020). Pause postures: The relationship between articulation and cognitive processes during pauses. *Journal of Phonetics*, 79, 100953.

Maassen, Ben, & Dirk-Jan Povel. (1984). The effect of correcting temporal structure on the intelligibility of the deaf. *Speech Communication*, 3(2), 123–135. https://doi.org/10.1016/0167-6393(84)90034-7

Olaoye, Anthony Ayodele. (2020). A socio-pragmatic analysis of silence in communication: An ethnographic review. *Veritas Journal of Humanities*, 2(1), 108–113.

Saville-Troike, M. (1985). The place of silence in an integrated theory of communication. In D. Tannen & M. Saville-Troike (Eds.), *Perspectives on Silence* (pp. 3–18). Norwood, NJ: Ablex.

Sluis, Rachel A., et al. (2020). An automated approach to examining pausing in the speech of people with dementia. *American Journal of Alzheimer's Disease & Other Dementias*, 35, 1533317520939773.

Šturm, Pavel, & Jan Volín. (2023). Occurrence and duration of pauses in relation to speech tempo and structural organization in two speech genres. *Languages*, 8(1), 23.

Wan, I-Ping & Allassonnière-Tang, Marc. (2021). The effect of word frequency and position-in-utterance in Mandarin speech errors: A connectionist model of speech production. *The Post-conference proceedings of CLSW2020 in the LNAI series*. Springer Singapore. 491-500.

Wan, I-Ping and Marc Allassonnière-Tang & Pu Yu. (2024). Early segmental production in Thai preschool children learning Mandarin. *International Journal of Asian Language Processing, 34*(2), 2450005.

Zhang, Jing. (2024). Variational pragmatics in Chinese discourse markers zhege and nage: The influence of region and gender. *Journal of Pragmatics,* 230, 76–88.

Zuo, Yan. (2002). The golden silence: A pragmatic study on silence in dyadic English conversation. *München: Lincom Europa*.