# The Chulalongkorn Corpus of Spoken Thai (CCOST)

**Pittayawat Pittayaporn[1], Cathryn Yang[2], Sujinat Jitwiriyanont[1], James Kirby[3]**

[1]Chulalongkorn University, [2]SIL and Payap University, [3]Ludwig Maximilian University of Munich
[1]Bangkok, Thailand; [2]Chiang Mai, Thailand; [3]Munich, Germany
cathryn_yang@sil.org, jkirby@phonetik.uni-muenchen.de, {pittiyawat.p, sujinat.j}@chula.ac.th

**Abstract**

The Chulalongkorn Corpus of Spoken Thai (CCOST) is a phonetically annotated corpus of Standard Thai. The corpus comprises approximately 7 hours of interview-style spontaneous speech from 49 speakers (19 male, 30 female) ranging in age from 18 to 83 years old. Speakers represent diverse regional backgrounds across Thailand but were instructed to speak in Standard Thai. Each speaker also read a 206-item monosyllabic word list twice and a set of 25 sentences three times. The annotation pipeline combines automatic speech recognition (ASR) and forced alignment using CLARIN-D's OCTRA and Munich Automatic Segmentation System (MAUS) tools with manual correction by phonetically trained native Thai speakers. Transcriptions include orthographic, word-level, syllable-level, and phone-level annotations including toneme labels. The corpus serves as a resource in the sociophonetic investigation of segmental and tonal variation in spontaneous and controlled speech, enabling examination of individual characteristics as well as group differences across age groups, genders, and regional backgrounds. Hand-corrected annotations will additionally serve to improve alignment accuracy for Standard Thai.

**Keywords:** Standard Thai, spontaneous speech, sociophonetic variation

## 1. Introduction

The Chulalongkorn Corpus of Spoken Thai (CCOST) is a corpus of speech recordings in Standard Thai with orthographic, phonemic, and phonetic annotations. The main purpose of the corpus is to enable sociophonetic investigation of tonal and segmental variation in naturalistic speech, enabling examination of individual characteristics as well as group differences across age groups, genders, and regional backgrounds. Additionally, hand-corrected segmentation and labeling will be used to improve the accuracy of the Munich Automatic Segmentation System (MAUS: Kisler et al. 2017) and potentially of other forced alignment systems for Thai.

Variationist sociolinguistic research has traditionally focused on non-tonal languages, leaving gaps in our understanding of "sociotonetics," the subfield of sociophonetic tone variation and change (Stanford 2016a; Stanford and Yang 2023). One conspicuous gap concerns sociotonetic variation in spontaneous speech, which has remained largely unexplored due to the labor-intensive nature of transcription and annotation (Stanford 2016b).

Standard Thai, defined as the Bangkok-based form of the Central Thai language that educated Thai people consider the norm for proper speaking and writing (Smalley 1994), has five lexical tone categories, traditionally labeled as Mid (mid level), Low (low falling), Falling (high falling), High (high with an initial rise), and Rising (low rising) (Abramson 1962 ; see Figure 1). Variation in the phonetic realization of the tones has been examined across prosodic contexts (Nitisaroj 2006; Potisuk et al. 1996; Silpachai 2024), tonal contexts (Gandour et al. 1994), and speech rates (Burroni and Kirby 2025), yet sociotonetic investigation has been limited to examining generational differences in monosyllables uttered in isolation (Teeranon and Rungrojsuwan 2009; Thepboriruk 2009).

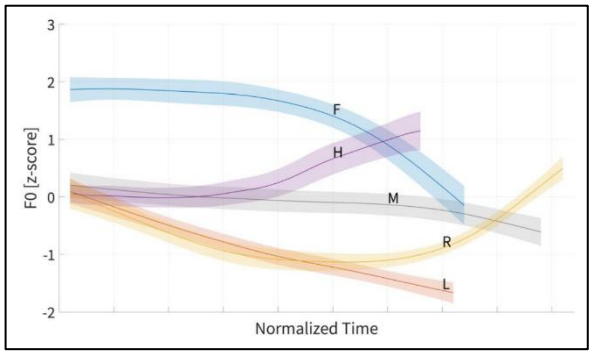


Figure 1: Thai tones produced in isolation. Tones are labeled as M (Mid), L (Low), F (Falling), H (High), and R (Rising). Reproduced with permission from Burroni (2023).

Existing Thai speech corpora have focused primarily on automatic speech recognition (ASR) development. Early efforts produced text-only resources such as ORCHID by the National Electronics and Computer Technology Center (NECTEC) (Charoenporn et al. 1997). More recently developed corpora of spoken Standard Thai and other Thai dialects consist only of read speech (Liberman et al. 2022; Suwanbandit et al. 2023; Taerungruang et al. 2025; Thatphithakkul et al. 2024). While one corpus includes unscripted customer service interactions (Thatphithakkul et al. 2023), it lacks annotation at the phonetic level.

The lack of Thai spontaneous speech corpora represents a limitation for sociotonetic research that seeks insights from ecologically valid data. Spontaneous speech corpora for the Sinitic languages of Mandarin, Cantonese, and Taiwan



*Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026)*, pages 3154–3160
11–16 May 2026. 

Southern Min have advanced our understanding of phenomena such as tone merger-in-progress, codeswitching between a tonal and non-tonal language, and interaction between prosody and tone (Chen and Tseng 2019; Gao et al. 2024; Johnson et al. 2020; Nagy et al. 2024; Sun et al. 2025; Wang and Fon 2013; Winterstein et al. 2020), but there are few equivalent resources in non-Sinitic tone languages.

CCOST addresses this gap by providing a corpus of both controlled and spontaneous speech with time-aligned phonetic annotation at multiple levels. The corpus enables investigation not only of tonal variation but also of segmental phenomena such as spirantization ([x] for /$k^h$/, [ʃ] for /$c^h$/), rhotic variation ([r] ~ [l]), and consonant cluster reduction ([C-] for /Cr-/ and /Cl-/).

One challenge to building spontaneous speech corpora is the labor-intensive task of data collection. The Philadelphia Neighborhood Corpus addressed this difficulty by enlisting the efforts of undergraduate students majoring in linguistics (Labov et al. 2013). Similarly, CCOST data collection was conducted by graduate and undergraduate linguistics students who received training in interview techniques and recording protocols. This approach enabled relatively rapid corpus development while providing students with fieldwork experience, but also introduced variability in audio quality (see Section 3.4).

## 2. Corpus construction

### 2.1 Participants

49 speakers (19 who self-identified as male, 30 as female) were recorded, with ages ranging from 18-83 years, mean = 40 years, median = 29 years. Most speakers were born or grew up in the Greater Bangkok Metropolitan Area (n = 27), defined as the contiguous urban area consisting of Bangkok and three surrounding provinces—Nonthaburi, Pathum Thani, and Samut Prakan (Huebner 2006). Speakers from other regions such as non-Bangkok Central (n = 9), Northeastern (n = 7), Southern (n = 2), and Northern (n = 4) areas were also included. All speakers were requested to speak in "*Thai klang*," which literally means Central Thai but in practice refers to Standard Thai.

### 2.2 Materials

Spontaneous speech was elicited through a semi-structured interview. The interview schedule was designed to elicit monologic narratives and was adapted for the Thailand context from Tagliamonte's (2006) interview schedule. The schedule features open questions to encourage the participant to talk about personal experiences in childhood, at school, at work, in the family and in the community.

In addition to the interviews, controlled speech data were elicited through two reading tasks: a 206-item list of monosyllabic words and a set of 25 sentences. These tasks were designed to obtain targeted data on change and variation in the phonetic realization of Thai lexical tones. The wordlist crossed the five lexical tones with four initial consonant types (aspirated stops /kh, ph, th/; voiceless stops /p, k, t/; voiced stops /d, b/; sonorants /n, w, l, m/), three syllable structures (open, sonorant-final, stop-final), and vowel length (short vs. long, contrastive only in closed syllables). Thai phonotactics forbid Mid and Rising tones to occur on stop-final syllables.

The set of 25 sentences replicates Gandour et al.'s (1994) study of Thai tonal coarticulation. Each sentence embeds a sequence of two syllables (5 tones x 5 tones) between Mid tone syllables.

### 2.3 Recording Equipment and Procedure

Recordings were made in Bangkok between January 2020 and March 2023. No recordings were done in 2021 due to COVID restrictions. All interviewers were graduate or undergraduate students under 30 years of age studying at Chulalongkorn University. 43 of the 49 sessions were conducted by female interviewers. Undergraduate students completed at least one interview as part of their term project in Linguistic Analysis of Thai (course number 22309368).

Recording equipment was either a Zoom H4n or Marantz PMD 661, each with either an AT831 cardioid condenser lavalier microphone or SM-10A unidirectional headset microphone. The lavalier microphone was attached to the participant's clothing within 20 centimeters from the mouth. Audio was recorded in PCM-16 WAV format at 44.1 kHz sampling rate with 16-bit resolution. Most recording sessions were conducted in quiet indoor environments with minimal background noise, such as a university meeting room or the participant's home. However, several sessions included background noise such as music, a clock ticking, or dogs barking. All recordings are graded for audio quality (see Section 3.4).

Each recording session began with a brief information session on the purposes and procedures of the research project, in order to obtain informed consent. The interview portion typically lasted 15-20 minutes. Speakers were asked to choose two general topics to discuss. Interviewers selected additional prompts from the interview schedule, choosing topics that were deemed likely to encourage extended monologues from the participants.

Following the interview, the two reading tasks (words and sentences) were interwoven to maintain participant engagement. Participants read the wordlist in two separate sections (100 items and 106 items, respectively), with the sentence set read after each wordlist section. This

procedure was repeated once more, but during the second round the sentence set was only read once. In each session, the wordlist was read twice and the sentence set was read three times. Each speaker was assigned to read the words in one of four randomized orders (labeled A, B, C, D); however, the orders were not balanced across speakers, with 25 speakers using order A. The order used by each speaker is recorded in the metadata. All materials and metadata are available in the corpus repository at https://doi.org/10.5281/zenodo.17366698.

# 3. Corpus Annotation

## 3.1 File naming

Recordings were named with the following format: sequential order of the participant (preceded by "S" = Speaker), task code, and file number. Task codes are as follows: INT = interview, WL = wordlist, and ST = sentence. Interview recordings were divided into shorter recording files for ease of processing. File number reflects the order in which the files were recorded. For example, ccost_S32_INT_1 refers to the first section of the interview of the 32nd speaker. During the reading tasks, each task section was recorded in a separate file, e.g., wordlist section 1, wordlist section 2, etc. Some of these files were later combined, and this is reflected in the file name with the sequence of relevant numbers. For example, ccost_S32_WL_12 refers to the first and second sections of the wordlist by Speaker 32.

## 3.2 Evaluation

Audio quality was assessed using a Praat script that measured signal-to-noise ratio (SNR) and signal energy. Speech and noise were distinguished using criteria relative to each recording's peak intensity, as in Nguyen et al. (2025). The threshold was defined as 25 dB less than peak intensity. Portions in labeled intervals with intensity above threshold and lasting ≥ 50 ms were classified as speech; portions below threshold and lasting ≥ 30 ms were classified as noise. SNR was calculated as the difference between average speech and noise intensities. Recordings were flagged if SNR < 10 dB or if no noise was detected (indicating background intensity level was within 25 dB of speech), or if mean intensity < 45 dB. Flagged files were assigned a grade of "C" and prompted a human evaluator to check all other files in that session for similar quality issues. Additionally, human evaluators performed spot checks (one file per task per speaker) to identify occasional background noise not caught by automatic flagging; files with such noise were graded "B" (acceptable but not ideal). The remaining non-problematic files were given an "A" grade. Grades for each file are listed in the corpus metadata.

Reading pace for wordlist and sentence tasks was evaluated by a human listener. Files where speakers read rapidly without pausing between items were assigned a pacing grade of "C," while files with somewhat fast reading were graded "B," and those with ideal pacing and consistent pauses were graded "A." Pacing grades are recorded in a separate column in the corpus metadata.

## 3.3 Transcription and segmentation

The data processing pipeline began with Google Speech-to-Text ASR in OCTRA (Pömp and Draxler 2017), which produced a TextGrid with Thai orthography transcriptions. The transcriptions were corrected manually, including marking disfluencies with label <usb> and pauses with label <p :>.

Using the web services of the Bavarian Archive of Speech Signals (BAS), automatic labeling and segmentation were carried out with MAUS (Kisler et al. 2017), which provides grapheme to phoneme conversion as well as word, syllable, and phoneme and toneme segmentation and labeling. The TextGrids and recordings were run through the pipeline CHUNKPREP -> G2P -> MAUS -> PHO2SYL. This resulted in TextGrids with the following tiers (see Figure 2 for an example): orthographic word (tier label = ORT), canonical word pronunciation (KAN), canonical syllabified pronunciation KAS); phonetic segment (MAU), phonemic syllable (MAS), orthographic transcript of utterance (TRN). Phonemic and phonetic representation used the XSAMPA system (Wells, 1995). Tones were transcribed by a numeral, with _1 = Mid, _2 = Low, _3 = Falling, _4 = High, _5 = Rising, which matches the numbering system found in descriptions of Thai tone (Tingsabadh & Deeprasert 1997).

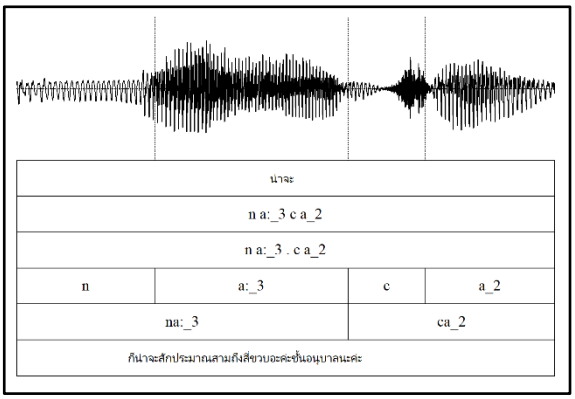


Figure 2: Part of the WAV file for the disyllabic word "น่าจะ" <na3ca2> and its corresponding TextGrid in Praat.

## 3.4 Manual correction and annotation of phonetic variation

Manual correction of the TextGrids was performed by trained native Thai research assistants to correct alignment boundaries and verify tone and segmental accuracy. If a speaker misread a word during the wordlist reading task,

that word's label was deleted from the MAS (phonemic) level. Phonetic variants, such as [l] realized for /r/ and spirantized [x] and [ʃ] for /kʰ/ and /cʰ/, were annotated on the MAU (phonetic segment) level. Deletion of the medial consonant in consonant clusters (e.g., [p] for /pl/) was annotated on the MAU level as well. English-origin items, whether loanwords or code-mixed words, were transliterated in Thai orthography but transcribed according to actual pronunciation on the MAU level.

## 4. Corpus Information

CCOST is available at https://doi.org/10.5281/zenodo.17366698 under a CC-BY-NC-SA 4.0 license. Updates will be posted to the same repository. The corpus contains 130,436 word tokens representing 5,158 word types and comprises 12 hours of annotated speech, including 7 hours of annotated free speech. Speech duration per speaker in the interviews ranged from 5:17 to 19:42 minutes (mean = 8:52, median = 8:24, std.dev. = 2.85). The majority of annotated speech received Grade A for audio quality (65% of the total duration), with 28% receiving Grade B and 7% receiving Grade C. Table 1 presents the distribution of files and speech duration by quality and task.

| Quality | Task | Files | Dur. (mins) | % Total dur. |
|---|---|---|---|---|
| A | interview | 129 | 278 | 38.8 |
| | sentence | 70 | 54 | 7.5 |
| | wordlist | 125 | 134 | 18.7 |
| B | interview | 61 | 124 | 17.3 |
| | sentence | 33 | 25 | 3.5 |
| | wordlist | 44 | 53 | 7.4 |
| C | interview | 12 | 32 | 4.5 |
| | sentence | 6 | 5 | 0.7 |
| | wordlist | 8 | 12 | 1.6 |

Table 1: Distribution of recording files and annotated speech duration (in minutes) by quality and task. Percentages represent proportion of total duration of annotated speech

## 5. Validation

To validate the suitability of CCOST recordings for sociotonetic research, we examined whether F0 measurements from Grade B recordings were comparable to Grade A recordings and consistent with previous descriptions of Standard Thai tones (e.g., Burroni 2023). We limited the analysis to monosyllables ending in a vowel or sonorant coda produced in the wordlist reading task. To control for social variation, we selected the five middle-aged female speakers (age 44-53) from the Greater Bangkok Metropolitan Area. Three of the speakers' recordings were Grade A and two were Grade B. F0 values were calculated every 5 milliseconds using Praat's pitch estimator algorithm with a pitch range of 100-500 Hz for female voices (Hirst and de Looze 2021). F0 was z-score normalized by speaker. Time was normalized within each tone category. After filtering out short tokens (< 50 ms) and tokens with pitch jump errors, 1,910 tokens were retained for analysis.

Figure 3 shows F0 trajectories of individual tokens (thin lines) and loess smoothed curves (bold lines), with Grade A recordings in the top row and Grade B in the bottom row. Figure 1 from Burroni (2023) is included for reference. Tone contours in both rows qualitatively align with previous analyses of Standard Thai tones (Abramson 1962; Burroni 2023; Gandour et al. 1991), suggesting that recordings with occasional background noise are suitable for sociotonetic analysis. This finding is in line with previous study that found F0 was robust to noise (Perry et al. 2000).

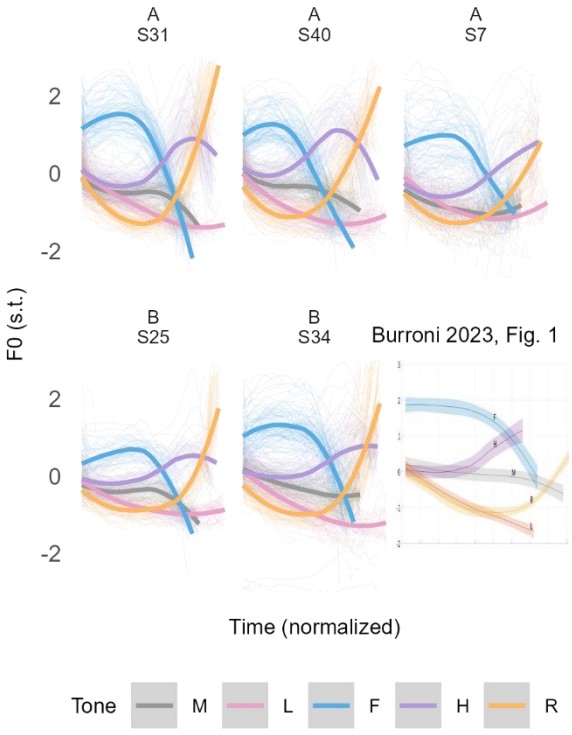


Figure 3: F0 contours for Thai tones, comparing Grade A (top row) and Grade B (bottom row) recordings.

## 6. Conclusion

CCOST is an open access resource for sociophonetic investigation of Standard Thai, with annotations at the word, phoneme, and phone level. The corpus's 7 hours of annotated spontaneous speech addresses a critical gap in sociophonetic corpora of tone languages.

## 7. Ethics statement

This research was approved by the Research Ethics Review Committee for Research Involving Human Subjects (The Second Allied Academic Group in Social Sciences, Humanities, and Fine and Applied Arts), Chulalongkorn University (COA No. 071/2563). All speakers provided informed consent prior to participation, including consent for the publication of pseudonymized original audio recordings.

## 8. Limitations

CCOST has several limitations. First, although speakers' regional backgrounds are documented, all were instructed to speak Standard Thai, making CCOST inappropriate for investigating regional dialect speech. The distribution of speakers across regions is also uneven, which may limit certain sociophonetic comparisons. Second, the wordlist presentation orders are unbalanced across speakers (25 of 49 used order A), resulting in a lack of control for fatigue or priming effects. Third, while annotators worked from transcription guidelines and corrections were spot-checked by senior team members, no formal inter-annotator agreement measures were collected.

## 9. Acknowledgments

This project has received funding from the Center of Excellence in Southeast Asian Linguistics (ChulaSEAL), supported by the Ratchadapisek Somphot Fund, Chulalongkorn University; and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No.758605).

We gratefully acknowledge the contributions of our project coordinators Apinya Singsopa, Phongbhorn Prayongdravya, and Teerawee Sukanchanon ; and the following research assistants for transcription correction and forced alignment correction: Chanakan Wittayasakpan, Chamawee Kitcharoenying, Kanisorn Tanawuttipong, Kantika Apinyovichien, Kedsaraporn Ruechai, Kumpha Sethagomut, Nanthipat Kongborrirak, Navapon Monkiatkul, Pawitsapak Akarajaradwong, Peeranut Nittayanonte, Phatharaphon Phromthong, Pimthip Kochaiyaphum, Piyapath T Spencer, Pontakorn Yuttakit, Prissana Buadum, Punyanuch Maitreenukul, Sasithon Namsiri, Sirikorn Imjongjairuk, Suphitchaya Singsahuth, Thanyanee Treewatanasuwan, Theera-anuchit Chalapinyo, Thitivut Sommool, and Uracha Tatiyanunt.

## 10. Bibliographical References

Abramson, A. S. (1962). *The vowels and tones of Standard Thai: Acoustical measurements and experiments*. Bloomington: Indiana U. Research Center in Anthropology, Folklore, and Linguistics.

Burroni, F. (2023). *Dynamics of F0 planning and production: Contextual and rate effects on Thai tone gestures* (Ph.D. dissertation). Cornell University, Ithaca, N.Y.

Burroni, F., and Kirby, J. (2025). Effects of speaking rate on f0 trajectories: Evidence from Thai lexical tones. *The Journal of the Acoustical Society of America*, *158*(2), 1204–1226. https://doi.org/10.1121/10.0038965

Charoenporn, T., Sornlertlamvanich, V., and Isahara, H. (1997). Building a large Thai text corpus - part-of-speech tagged corpus: ORCHID. In *Proceedings of the Natural Language Processing Pacific Rim Symposium (NLPRS '97, Incorporating SNLP '97)* (pp. 509–512). Phuket, Thailand.

Chen, A. C.-H., and Tseng, S.-C. (2019). Prosodic encoding in Mandarin spontaneous speech: Evidence for clause-based advanced planning in language production. *Journal of Phonetics*, *76*, 100912. https://doi.org/10.1016/j.wocn.2019.100912

Gandour, J., Potisuk, S., and Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics*, *22*(4), 477–492. https://doi.org/10.1016/S0095-4470(19)30296-7

Gandour, J., Potisuk, S., Ponglorpisit, S., and Dechongkit, S. (1991). Inter- and intraspeaker variability in fundamental frequency of Thai tones. *Speech Communication*, *10*(4), 355–372. https://doi.org/10.1016/0167-6393(91)90003-C

Gao, X., Voeten, C., and Liberman, M. (2024). The impact of prosodic boundary and information structure on tonal coarticulation in spontaneous Cantonese. In *Proceedings of Speech Prosody 2024* (pp. 393–397). Leiden, The Netherlands: ISCA. https://doi.org/10.21437/SpeechProsody.2024-80

Hirst, D., and de Looze, C. (2021). Fundamental frequency and pitch. In R.-A. Knight & J. Setter (Eds.), *The Cambridge Handbook of Phonetics* (pp. 336–361). Cambridge: Cambridge University Press.

Huebner, T. (2006). Bangkok's linguistic landscapes: Environmental print, codemixing and language change. *International Journal of Multilingualism*, *3*(1), 31–51. https://doi.org/10.1080/14790710608668384

Johnson, K. A., Babel, M., Fong, I. and Yiu, N. (2020). SpiCE: A new open-access corpus of conversational bilingual speech in Cantonese and English. In *Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 4089–4095). Marseille, France: European Language Resources Association. https://www.aclweb.org/anthology/2020.lrec-1.503

Kisler, T., Reichel, U. and Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347. https://doi.org/10.1016/j.csl.2017.01.005

Labov, W., Rosenfelder, I. and Fruehwald, J. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, *89*(1), 30–65.

Nagy, N., Tse, H. and Stanford, J. N. (2024). Have Cantonese tones merged in spontaneous speech? In R. Rao (Ed.), *The Phonetics and Phonology of Heritage Languages* (pp. 302–320). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108966986.015

Nguyen, D. D., Gupta, R., Gunjawate, D. R., Holik, J., Jin, C., & Madill, C. (2025). Speech-to-noise ratio and voice-to-noise ratio of voice databases with implications for acoustic voice analysis. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2025.05.029

Nitisaroj, R. (2006). Thai tonal contrast under changes in speech rate and stress. In *Proceedings of Speech Prosody 2006* (p. 096). Dresden, Germany: ISCA.

Perry, C. K., Ingrisano, D. R.-S., Palmer, M. A. and McDonald, E. J. (2000). Effects of environmental noise on computer-derived voice estimates from female speakers. *Journal of Voice*, *14*(2), 146–153. https://doi.org/10.1016/S0892-1997(00)80021-1

Pömp, J. and Draxler, C. (2017). OCTRA—a configurable browser-based editor for orthographic transcription. In *Proceedings of Phonetik und Phonologie im deutschsprachigen Raum (P&P)* (pp. 145–148). Berlin.

Potisuk, S., Gandour, J. and Harper, M. P. (1996). Acoustic correlates of stress in Thai. *Phonetica*, *53*(4), 200–220. https://doi.org/10.1159/000262201

Silpachai, A. (2024). The boundary-induced modulation of obstruents and tones in Thai. *Journal of Phonetics*, *102*, 101291. https://doi.org/10.1016/j.wocn.2023.101291

Smalley, W. A. (1994). *Linguistic diversity and national unity: Language ecology in Thailand*. Chicago: University of Chicago Press.

Stanford, J. N. (2016a). A call for more diverse sources of data: Variationist approaches in non-English contexts. *Journal of Sociolinguistics*, *20*(4), 525–541. https://doi.org/10.1111/josl.12190

Stanford, J. N. (2016b). Sociotonetics using connected speech. *Asia-Pacific Language Variation*, *2*(1), 48–82. https://doi.org/10.1075/aplv.2.1.02sta

Stanford, J. N. and Yang, C. (2023). Sociophonetics and tone: The world of sociotonetics. In C. Strelluf (Ed.), *The Routledge Handbook of Sociophonetics* (pp. 76–92). London: Routledge.

Sun, J., Audibert, N., Wu, Y. and Adda-Decker, M. (2025). Corpus-based insights into Mandarin neutral tone: effects of tonal context and structural patterns in spontaneous speech. In *Proceedings of Interspeech 2025* (p. Paper ID: 960). Rotterdam, Netherlands: International Speech Communication Association. https://doi.org/10.21437/Interspeech.2025-960

Suwanbandit, A., Naowarat, B., Sangpetch, O. and Chuangsuwanich, E. (2023). Thai dialect corpus and transfer-based curriculum learning investigation for dialect automatic speech recognition. In *Proceedings of Interspeech 2023* (pp. 4069–4073). Dublin, Ireland: ISCA. https://doi.org/10.21437/Interspeech.2023-1828

Taerungruang, S., Taninpong, P., Chunwijitra, V., Thatphithakkul, S., Kasuriya, S., Inthanon, V., et al. (2025). PaSCoNT - parallel speech corpus of northern-central Thai for automatic speech recognition. *Computer Speech & Language*, *89*, 101692. https://doi.org/10.1016/j.csl.2024.101692

Tagliamonte, S. A. (2006). *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511801624

Teeranon, P. and Rungrojsuwan, R. (2009). Change in the Standard Thai High tone: An acoustic study. *Manusya*, *12*(3), 34–44. https://doi.org/10.1163/26659077-01203003

Thatphithakkul, S., Thangthai, K. and Chunwijitra, V. (2024). The development of LOTUS-TRD: A Thai regional dialect speech corpus. In *Proceedings of the 27th Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)* (pp. 1–6). Hsinchu City, Taiwan: IEEE. https://doi.org/10.1109/O-COCOSDA64382.2024.10800335

Thatphithakkul, S., Thangthai, K., Sriphol, S. and Chunwijitra, V. (2023). The development of a Thai telephone conversational speech corpus. In *Proceedings of the 26th Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)* (pp. 1–6). Delhi, India: IEEE. https://doi.org/10.1109/O-COCOSDA60357.2023.10482925

Thepboriruk, K. (2009). Bangkok Thai tones revisited. *Journal of the Southeast Asian Linguistics Society (JSEALS)*, *3*(1), 86–105.

Wang, S.-F. and Fon, J. (2013). A Taiwan Southern Min spontaneous speech corpus for discourse prosody. In B. Bigi & D. Hirst (Eds.), *Proceedings of TRASP 2013: Tools and Resources for the Analysis of Speech Prosody. An Interspeech Satellite Event.* (pp. 20–23). Aix-en-Provence, France: ISCA.

## 11. Language Resource References

Liberman, M., Yuan, J., Cieri, C., Wright, J. and

Chanchaochai, N. 2022. *Global TIMIT Thai.* distributed via Linguistic Data Consortium: LDC2022S13, ISLRN 471-761-626-780-5. PID https://doi.org/10.35111/vr5f-6m39.

Winterstein, G., Tang, C. and Lai, R. 2020. CantoMap: a Hong Kong Cantonese MapTask Corpus. ISLRN 167-857-138-471-9. PID https://github.com/gwinterstein/CantoMap.