Early production of boundary tones in Mandarin declaratives:

Contour clustering of monolingual and bilingual child speech

Wangqian Zhou (The Chinese University of Hong Kong)

May 22, HISPhonCog 2026

Crosstalk between tone and intonation

- The dual function of fundamental frequency (f0): lexical and intonational.
- In the Autosegmental-Metrical (AM) framework, intonation is phonologically represented as **a composition of discrete tonal events** that associate with metrical structure (Pierrehumbert, 1980).
    - Pitch accents: associated with stressed syllables
    - **Boundary tones: associated with phrasal boundaries**
- Lexical tone languages also use intonational tones to mark phrasal boundaries. ⇒ **tonal crowding at utterance-final syllables**.
    - A challenge to acquisition: children should learn to disaggregate which aspects of the fluctuating f0 signal serve lexical or intonational information (Kager, 2018).

Tonal crowding in the AM theory

- Tonal crowding: the presence of **more tones than tone-bearing units**.
- Phonetic resolutions vary across languages (Arvaniti, et al., 2026):
  - **Compression** (adjustment in the rate of f0 change)
  - **Truncation** (target undershoot or tone delinking)
  - **Temporal realignment** of tones
  - **Segmental elongation** to accommodate the tones
- How does Mandarin resolve the crowding of lexical and boundary tones?

Boundary tones in Mandarin

Chao (1933):

- **Simultaneous addition boundary tone** (SiABT): the algebraic sum of two tonal targets.

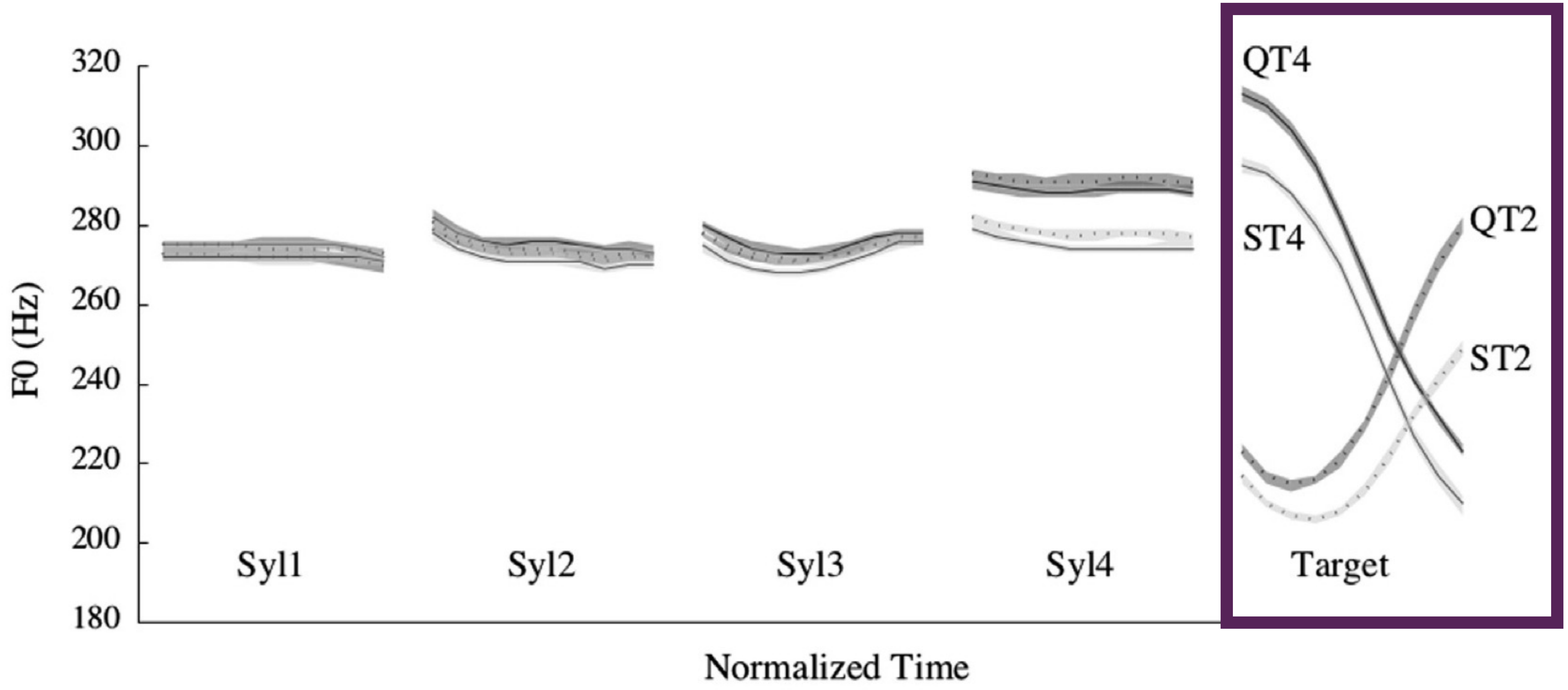

Figure 1. F0 contours of five-syllable Mandarin questions and statements ending with T2 or T4 (Liu, et al., 2016).

Boundary tones in Mandarin

Chao (1933):

- **Simultaneous addition boundary tone** (SiABT): the algebraic sum of two tonal targets.
- **Successive addition boundary tone** (SuABT): a rising or falling boundary tone appended to the complete lexical tone contour.

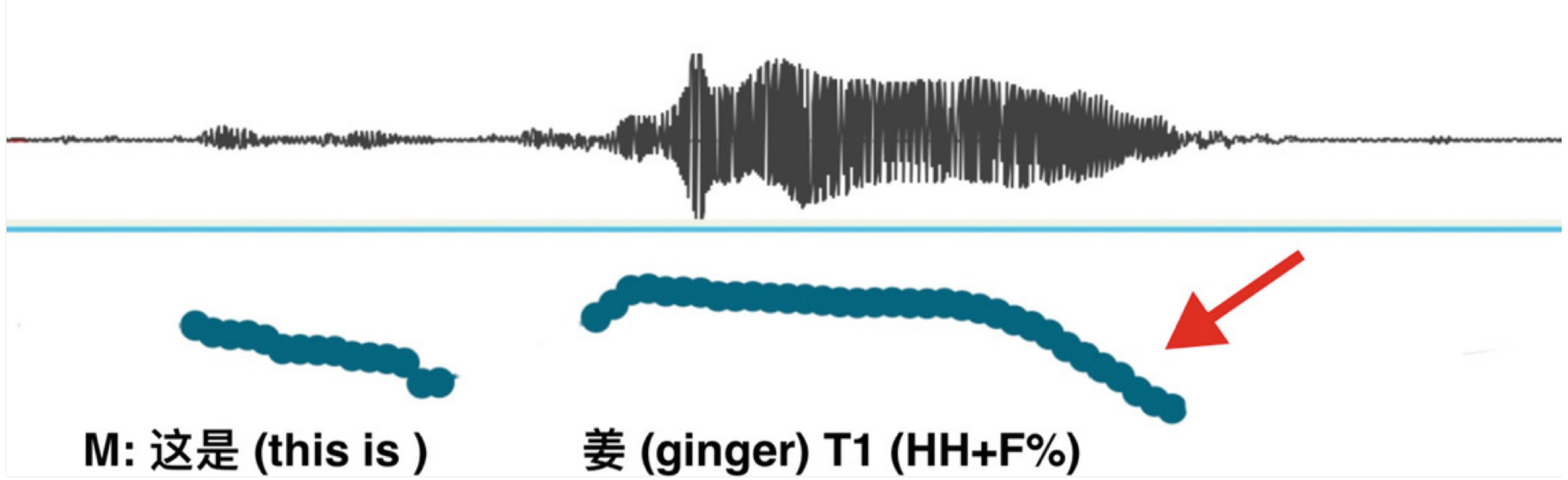

Figure 2. Waveform and Fo contours of a statement "This is ginger." produced with a falling SuABT (Li, et al., 2024).

香港中文大學
The Chinese University of Hong Kong

Mandarin children's production of SuABT

Li, et al. (2024): picture naming and imitation

- Mandarin-speaking children can produce successive addition boundary tones **(SuABT) by around age two**.
- Little is known about the **variability of early intonational productions** during tone language acquisition, nor about the **developmental trajectory** of boundary tone realisation in naturalistic speech.

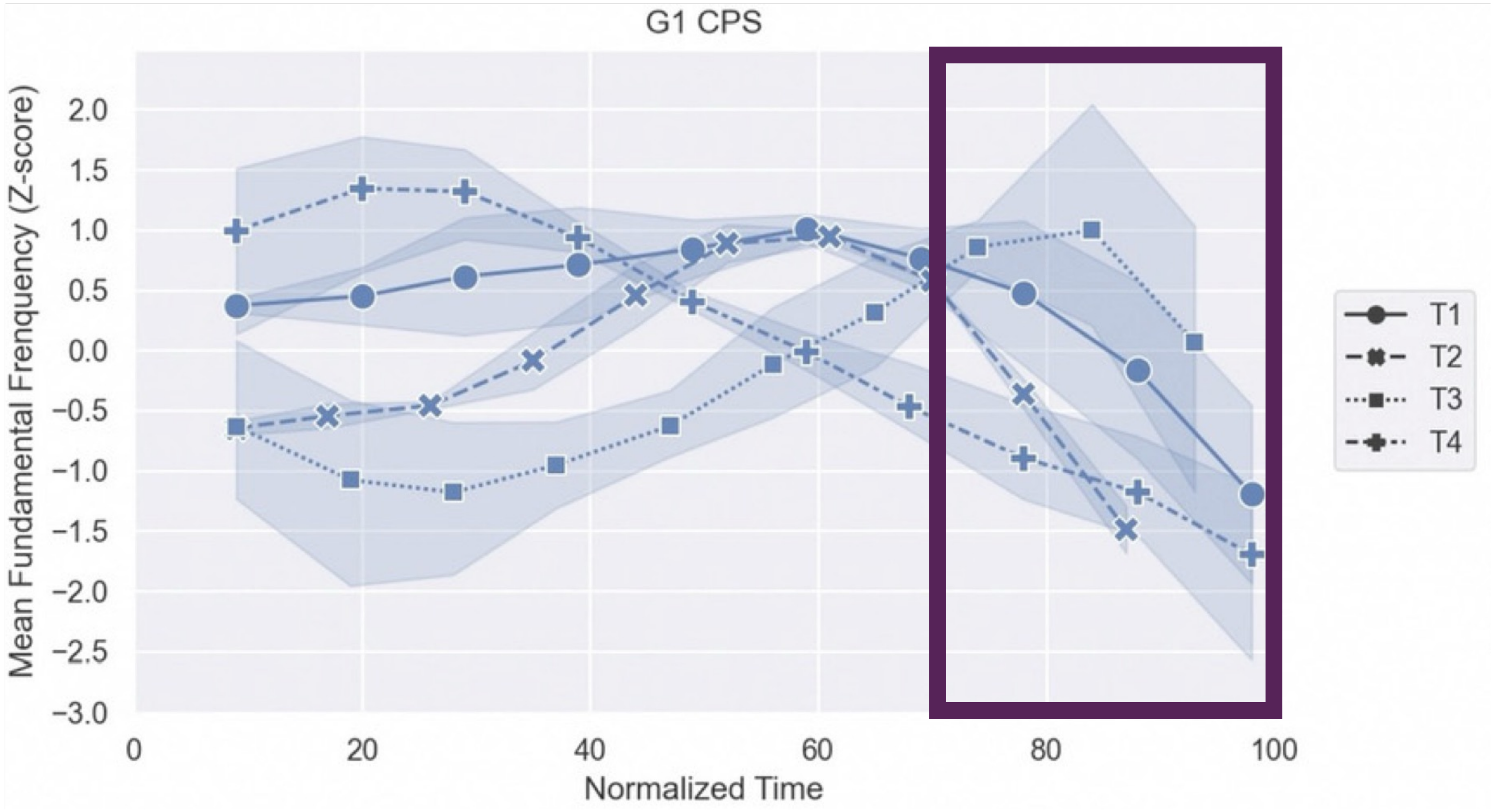

Figure 3. Average F0 contours of the falling SuABT on T1-T4 produced by five two-year-olds (Li, et al., 2024).

香港中文大學
The Chinese University of Hong Kong

Research questions

This preliminary study explored

(i) Whether and how children at **three developmental time points (2;01, 2;06, and 2;11)** produce boundary tones,

(ii) How production patterns **change over time**, and

(iii) Whether **lexical tone category** or **bilingualism** influences boundary tone realisation.

Contour clustering

Corpora

The Tong corpus (Deng & Yip, 2018) & CHCC (Mai, et al., 2024)

Child	Data	MLU	No.	Child	Data	MLU	No.	Child	Data	MLU	No.
Tong	2;01.17	2.82	198	**Avia**	2;01.11	3.20	48	**Luna**	2;01.03	1.56	410
(Mandarin monolingual)	2;06.13	3.52	182	**(Mandarin-English bilingual)**	2;01.18	2.97	50	**(Mandarin-English bilingual)**	2;01.24	1.22	107
	2;11.08	3.53	106		2;01.29	2.75	23		2;06.01	3.16	22
					2;06.09	3.75	44		2;06.11	2.86	46
					2;06.26	2.62	26		2;06.29	1.92	46
					2;11.12	3.92	36		2;11.04	2.08	12
									2;11.13	2.98	49
Total No.			**486**				**227**				**692**

MLU = Mean length of utterance;
No. = Number of coded utterances.

Data extraction and coding

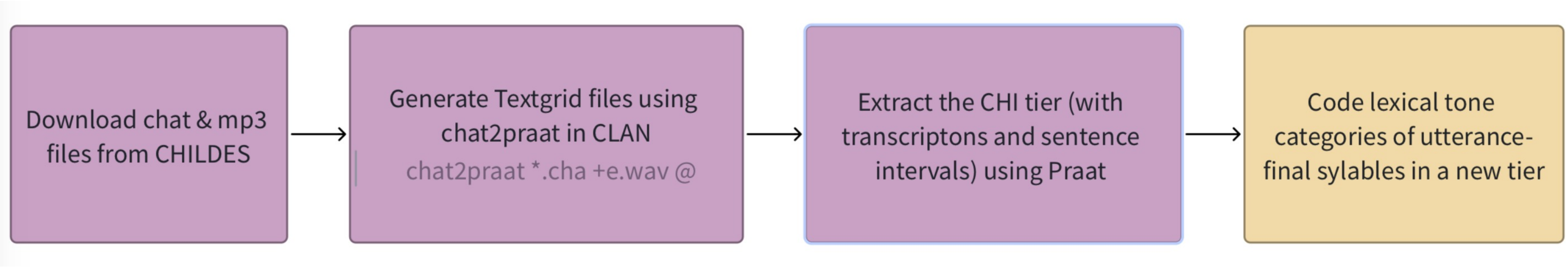

Figure 4. Diagram of the workflow in data extraction and coding.

- A total of **1405** declarative utterances ending with **full lexical tones** were extracted (no sentence-final particles or neutral tone words).
- Each utterance-final syllable was manually **coded for lexical tone category** (T1, T2, T3, or T4).

Data preparation

- Using the **Contour Clustering GUI (Kaland, 2025a)**:
    - **Time-series f0 measures** + **duration measures**
        - *F0 measurement points*: 20; *timestep*: 10 ms; *f0 fit*: 0.7; *smoothing*: 1
    - Removed missing values or extreme rates of change.
    - F0 converted to semitones, its **velocity** (1st order derivative) calculated.
    - All measures speaker-standardised.

Hierarchical agglomerative clustering

For the four lexical tones, respectively

- Complete linkage, dynamic time warping.
- **Time-series f0 velocity** and **duration** as predictors.
- Initially partitioned into 8 clusters to exclude singular outliers and clearly mispronounced tokens.

  → 300 **T1** syllables, 209 **T2** syllables, 247 **T3** syllables, and 292 **T4** syllables.
- The optimal number of clusters was determined by evaluating **within- and between-cluster variance**.

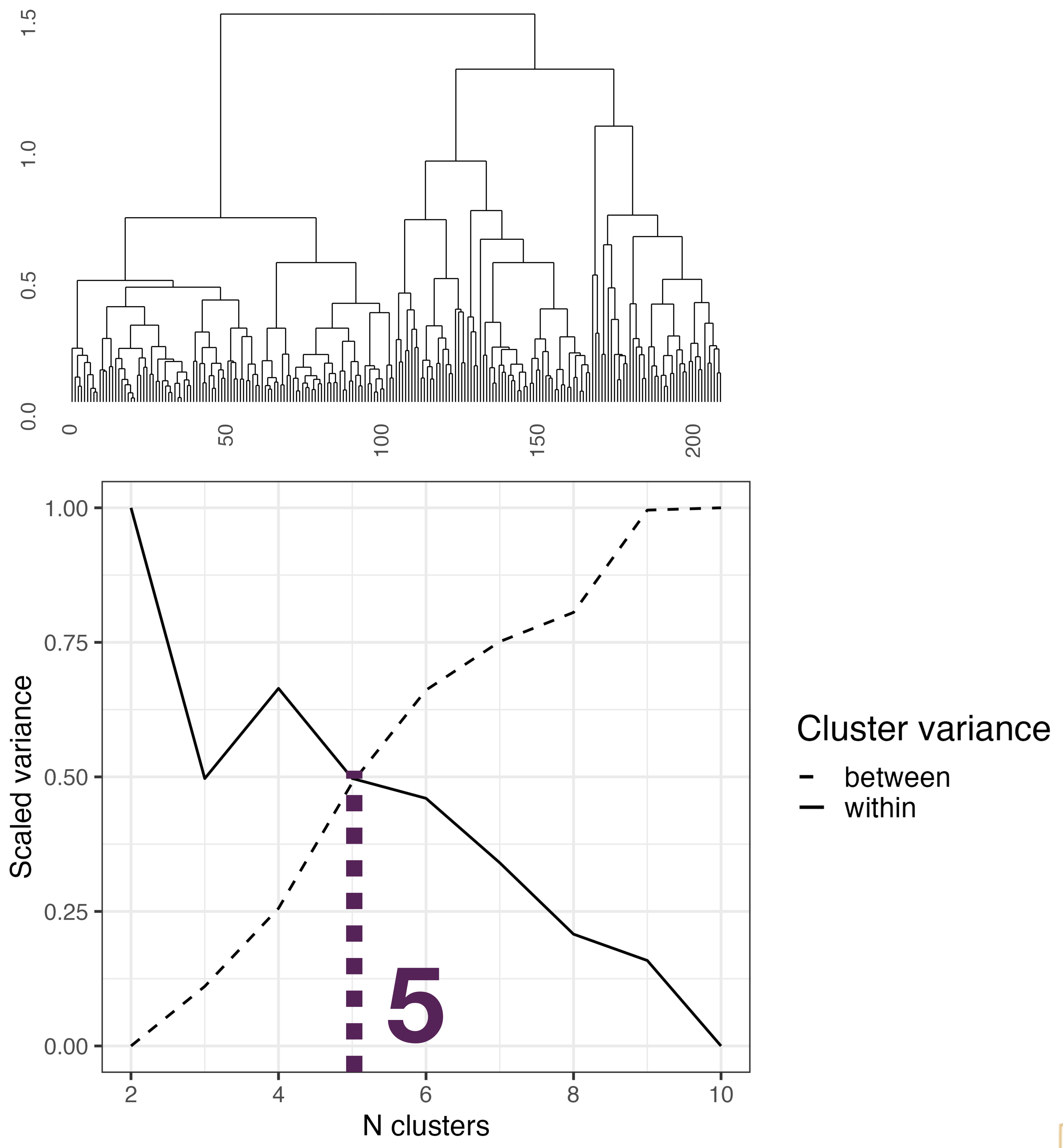

Figure 5. The cluster dendrogram and the evaluation of clustering solutions of T2.

Clustering results

T1 clusters

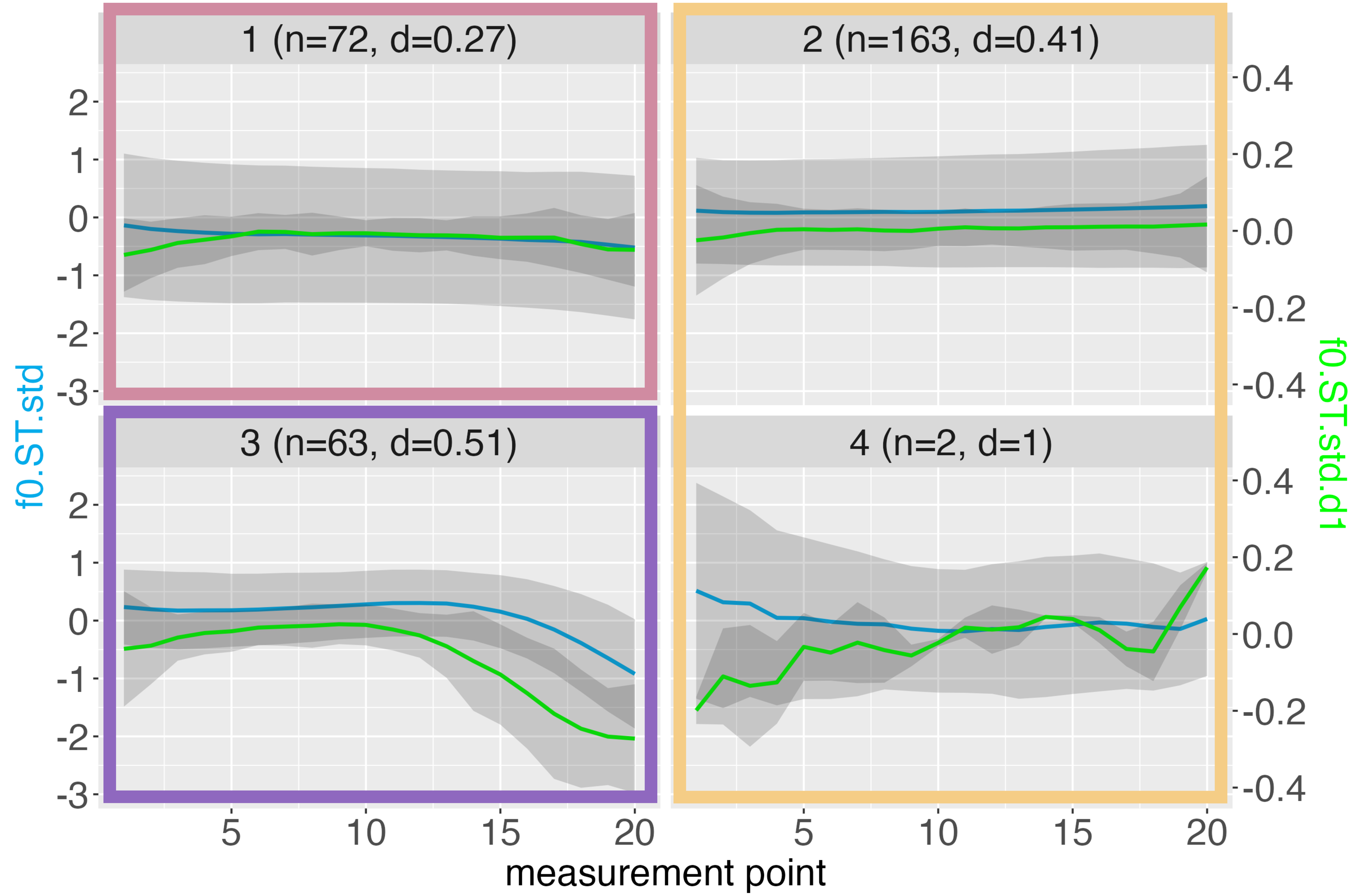

Figure 6. Clustering results of T1.

- f0 contour in blue,
  f0 velocity in green,
  d = mean duration,
  grey band = SD.
- **Canonical T1 (cluster 2 & 4)**: a full level tone, relatively long duration;
- **Undershot T1 (cluster 1)**: a lower level tone with shorter duration;
- **SuABT T1 (cluster 3)**: a negative slope appended to a level tone.

Clustering results

T2 clusters

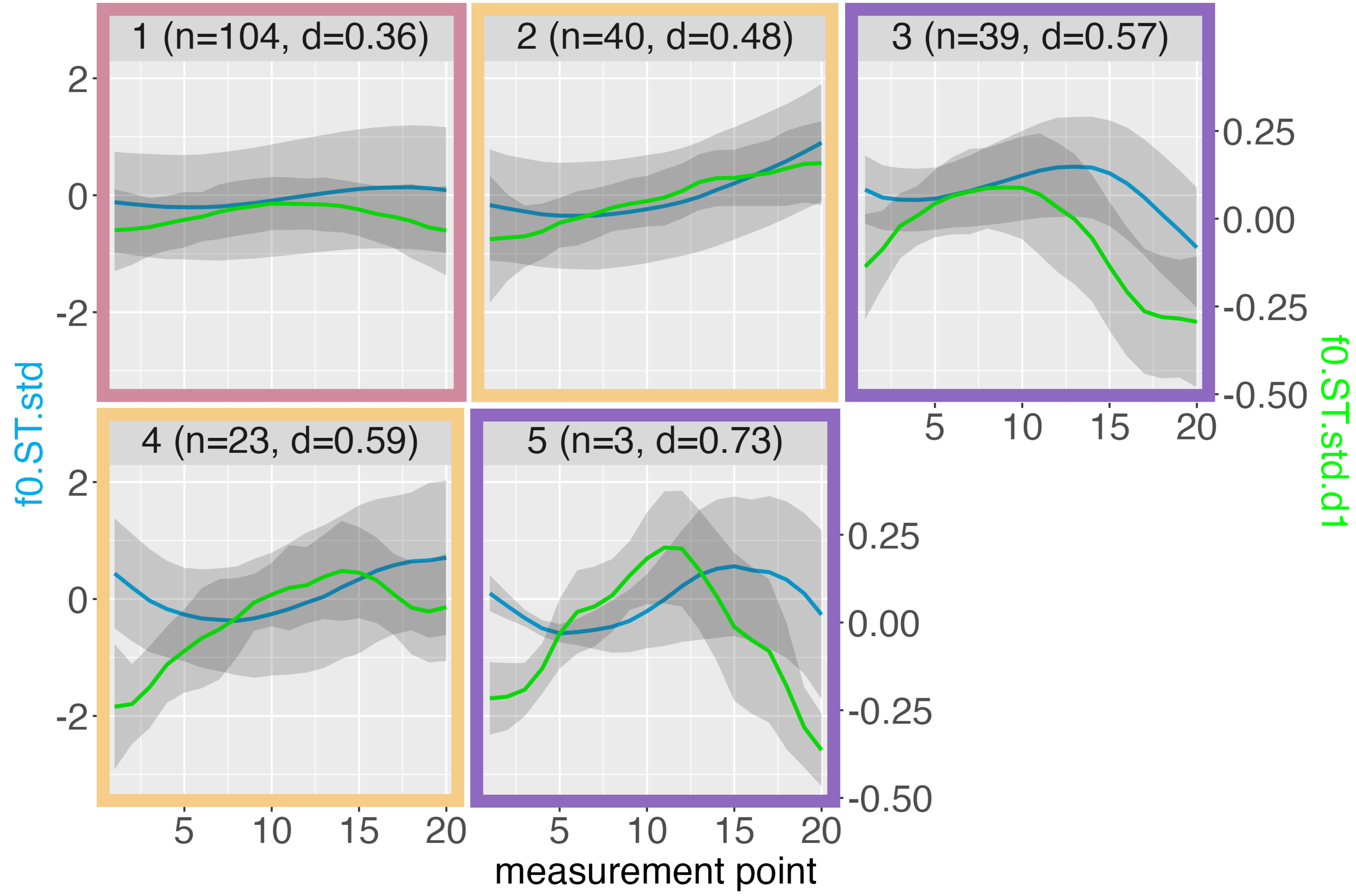

Figure 7. Clustering results of T2.

- **Canonical T2 (cluster 2 & 4)**: a full rising contour;
- **Undershot T2 (cluster 1)**: scaled lower with a compressed f0 range;
- **SuABT T2 (cluster 3 & 5)**: a late decline following the lexical rise.

Clustering results

T3 clusters

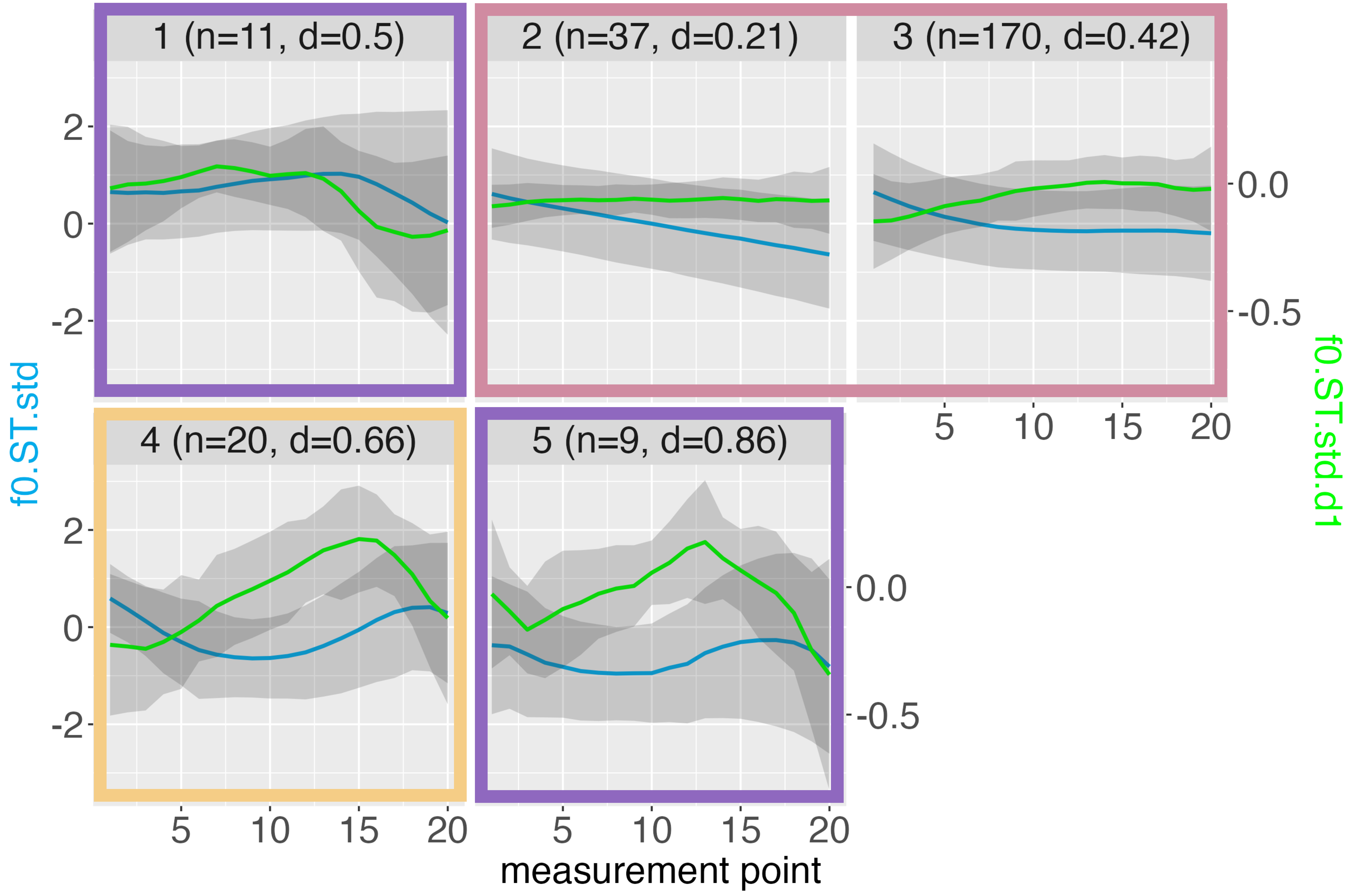

Figure 8. Clustering results of T3.

- **Canonical T3 (cluster 4)**: the canonical fall-rise;
- **Undershot T3 (cluster 2 & 3)**: a half T3 sandhi pattern, viz. the canonical dipping contour was truncated into a low fall.
- **SuABT T3 (cluster 1 & 5)**: a (fall-) rise that occupies most of the contour, followed by a shorter but non-negligible negative-slope section.

Clustering results

T4 clusters

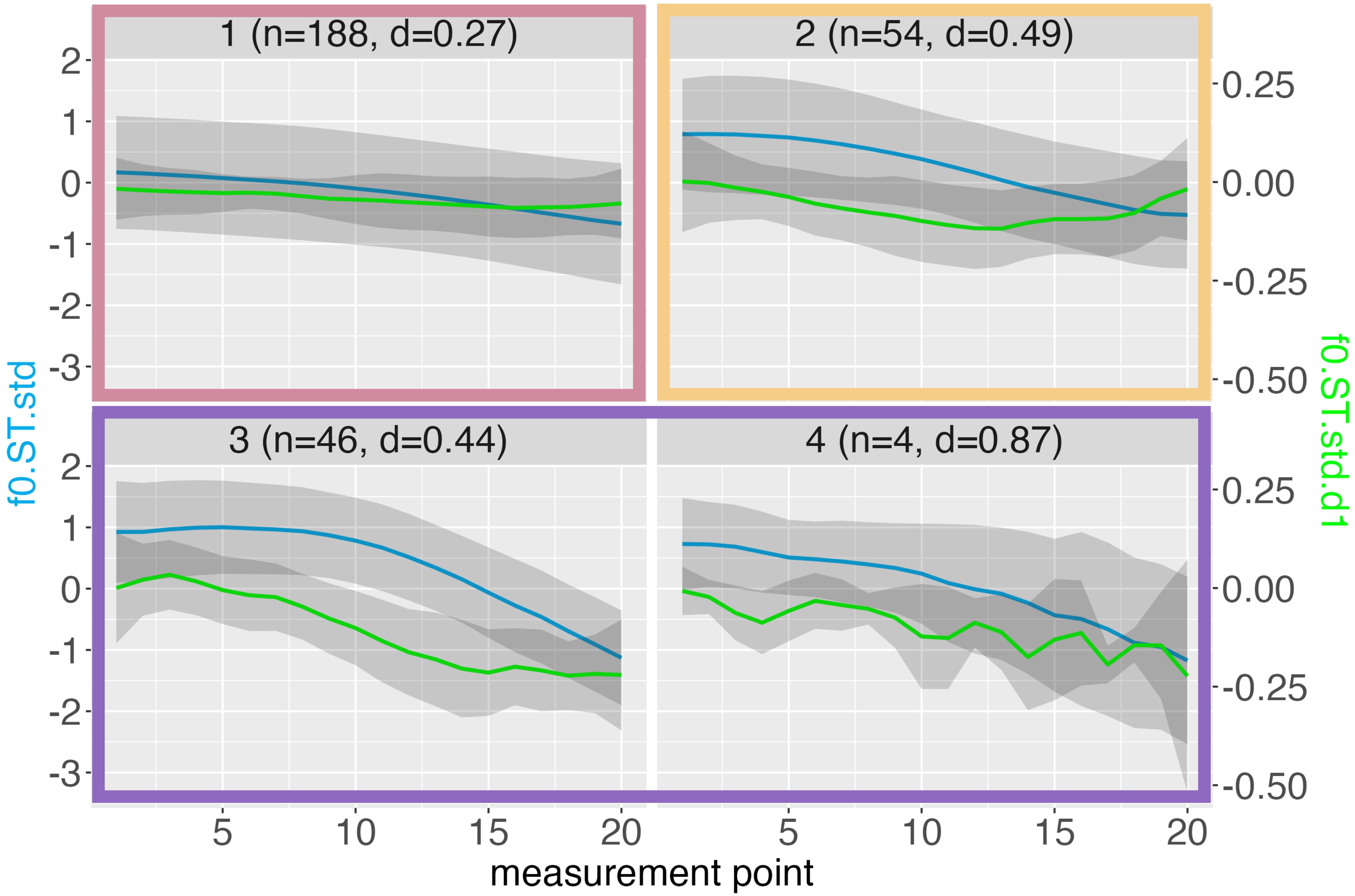

Figure 9. Clustering results of T4.

- **Canonical T4 (cluster 2)**: a negative slope that gradually levels off toward zero in the end;
- **Undershot T4 (cluster 1)**: a compressed f0 range;
- **SuABT T4 (cluster 3 & 4)**: a sustained negative slope.

Statistical analysis

Dirichlet regression

- Response variable: percentages of canonical, undershot, and SuABT contours per Speaker × Age × Tone.
- Predictor variables: speaker, age, and tone.

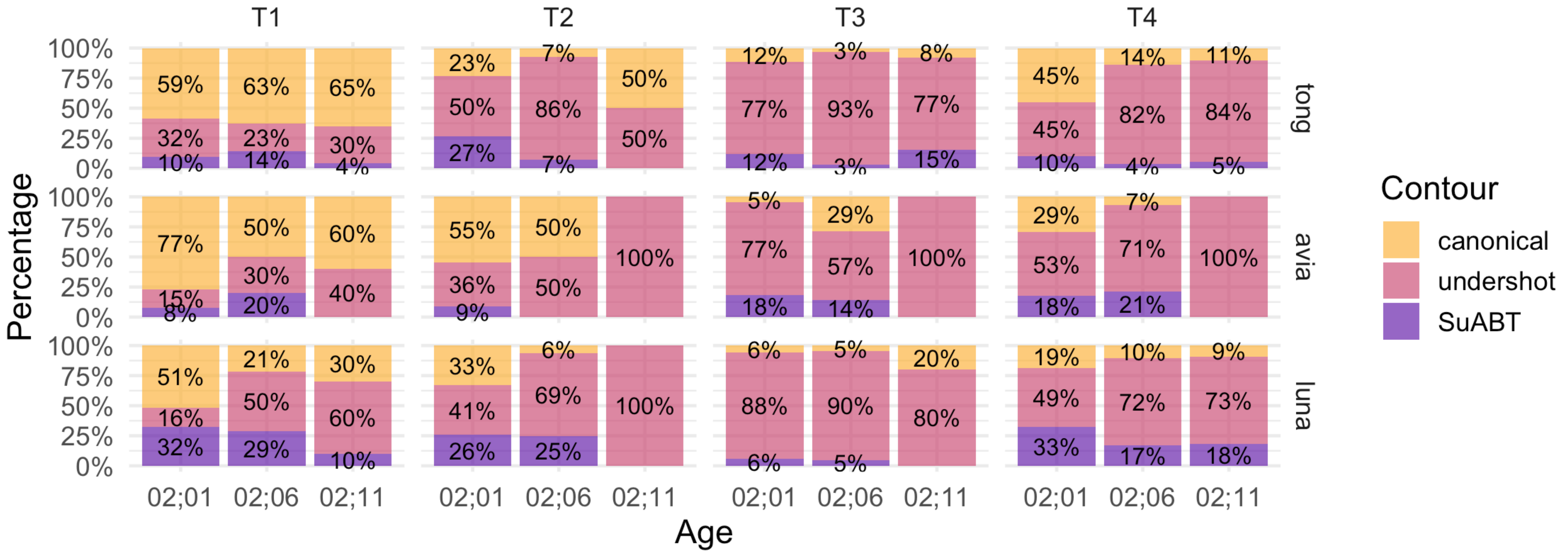

Figure 10. Proportion of contour categories (Speaker × Age × Tone) for different declarative-final lexical tones produced by the three children.

Dirichlet regression

An exploratory analysis

- Dirichlet regression models **compositional data**, where several proportions are modelled together because they are interdependent and must sum to one (Maier, 2014).
- The ***DirichletReg*** Package in *R* (Maier, 2025).
    - Extreme values of 0 & 1 were adjusted using the transformation implemented in *DR_data*() before model fitting.
    - Interaction models failed to converge due to data sparsity, so a main-effects model was reported.
    - Given the current sample size, the model was interpreted as an exploratory analysis rather than a strong inferential model.

Dirichlet regression: results

- **Age effects**
  - 02;11 vs. 02;01: **SuABT** contours significantly **decreased** ($\beta = -2.12$, $p < .001$); **canonical** contours significantly **decreased** ($\beta = -1.97$, $p < .001$).
- **Tone effects**
  - **T2** vs. T1: **SuABT** significantly **less frequent** ($\beta = -1.20$, $p = .015$); **canonical** contours significantly **less frequent** ($\beta = -1.91$, $p < .001$).
  - **T3** vs. T1: **SuABT** significantly **less frequent** ($\beta = -1.34$, $p = .013$); **canonical** contours significantly **less frequent** ($\beta = -2.10$, $p < .001$).
- **Speaker effects**
  - **Luna** vs. Tong: **Higher** proportion of **SuABT** ($\beta = 1.29$, $p = .017$).
  - **Avia** vs. Tong: **Lower** proportion of **undershot** contours ($\beta = -0.99$, $p = .047$).

Discussion

General discussion

- Overall, the results suggest that children at all three time points produced a range of boundary tone realisations, but the patterns were not stable across development.

  - By 02;11, SuABT and canonical contours both decreased relative to 02;01, suggesting **a potential shift toward phonologically less transparent undershot realisations**.

  - T2 and T3 showed lower proportions of SuABT and canonical contours than T1. **Contour complexity may constrain boundary tone production**.

  - Luna's preference for SuABT and Avia's relative dispreference for undershot contours may reflect **bilingual effects**, but **should be interpreted cautiously** given the current sample size.

Future directions

- **The input question**:
  - How variable is boundary tone production in adult (esp. child-directed) speech?
  - How might this variability influence children's acquisition patterns?
- **The form-meaning mapping question**:
  - Can the hypothesised categories be attested in perceptual judgements?
  - Is the variability in production associated with a more detailed classification of pragmatic functions of declaratives?
- **Other factors**:
  - What roles do utterance length and sentence-final particles play?
  - How might these factors further differentiate bilingual from monolingual developmental trajectories?

香港中文大學
The Chinese University of Hong Kong

Conclusion

- This study provides the first corpus-based evidence that Mandarin-speaking children, both monolingual and bilingual, acquire boundary tones early in development.

- Theoretically, it shows that children's boundary tone production and its developmental trajectory may be shaped by lexical tone category and children's linguistic profile.

- Methodologically, it showcases how contour clustering can be used to probe variability in naturalistic child speech data and identify recurrent patterns of tone–intonation interaction.

香港中文大學
The Chinese University of Hong Kong

Special thanks to

Ziya Liao,
for assistance with data coding;

Constantijn Kaland,
for Contour Clustering GUI support;

Virginia Yip,
whose course inspired the initial idea for this study;

Peggy Mok,
for suggestions on conceptual framing and presentation;

Aijun Li,
for suggestions on discussion;

Aoju Chen,
for suggestions on discussion.

References

Arvaniti, A., Grice, M., & D'Imperio, M. (2026). Advancements of phonetics in the 21st century: Intonation. *Journal of Phonetics, 114*, Article 101459.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. University of California Press.

Cole, J. & Steffman, J. (2021). The primacy of the rising/non-rising dichotomy in American English intonational tunes. *Proceedings of the 1st International Conference on Tone and Intonation*(TAI), Sonderborg, 122–26.

Deng, X., & Yip, V. (2018). A multimedia corpus of child Mandarin: The Tong corpus. *Journal of Chinese Linguistics*, *46*(1), 69-92.

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in R: The dtw package. *Journal of Statistical Software, 31*(7).

Hyman, L. M., & Monaka, K. C. (2011). Tonal and non-tonal intonation in Shekgalagari. In G. Elordieta, S. Frota, & P. Prieto (Eds.), *Prosodic Categories: Production, Perception and Comprehension* (pp. 267–289). Springer Netherlands.

Hulk, A., & Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition, 3*(3), 22-–255.

Kager, R. (2018). How do infants disaggregate referential and affective pitch? *Frontiers in Psychology, 9*, 2093.

Kaland, C. (2023a). Intonation contour similarity: f0 representations and distance measures compared to human perception in two languages. *The Journal of the Acoustical Society of America, 154*(1), 95–107.

Kaland, C. (2023b). Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association, 53*(1), 159–188.

Kaland, C. (2025a). *Contour Clustering* (Version 2025-03) [Computer software]. https://constantijnkaland.github.io/contourclustering/

Kaland, C. (2025b). F0 derivatives in the classification of meaningful tonal movements. *Journal of Phonetics, 113*, Article 101454.

Kaland, C., Steffman, J., & Cole, J. (2024). K-means and hierarchical clustering of f0 contours. *Proceedings of Interspeech 2024*, 1520–1524.

Kung, C., Chwilla, D. J., & Schriefers, H. (2014). The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: An ERP study on Cantonese Chinese. *Neuropsychologia, 53*(Jan), 293–309.

Li, A. (2025). 普通话儿童互动话语中后续叠加边界调初探 [Exploring the successive addition boundary tone in Mandarin child-parent interactional discourse]. In *互动语言学与汉语研究* [*Interactional Linguistics and Chinese Studies*] (Vol. 5). Social Sciences Academic Press.

Li, A., Gao, J., & Wang, Z. (2024). The production of successive addition boundary tone in Mandarin preschoolers. In *Man-Machine Speech Communication* (Vol. 2006, pp. 138-150). Springer.

Liu, M., Chen, Y., & Schiller, N. O. (2016). Online processing of tone and intonation in Mandarin: Evidence from ERPs. *Neuropsychologia, 91*, 307–317.

Liu, F., & Xu, Y. (2005). Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation. *Phonetica, 62*(2–4), 70–87.

Mai, Z., Shang, M., Liu, J., Yan, S., Matthews, S., & Yip, V. (2024, July 15-19). Acquiring Chinese in US, Hong Kong and Beijing: Three new corpora and three verbal structures [Conference presentation]. *XVIth International Congress for the Study of Child Language* (IASCL-2024), Prague, Czech Republic.

Maier, M. J. (2025). *DirichletReg: Dirichlet Regression in R*. R package version 0.7-2.

Peng, S.-H., Chan, M. K. M., Tseng, C.-Y., Huang, T., Lee, O. J., & Beckman, M. E. (2005). Towards a Pan-Mandarin system for prosodic transcription. In S.-A. Jun (Ed.), *Prosodic typology: The Phonology of Intonation and Phrasing* (pp. 230–270). Oxford University Press.

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation* [Doctoral dissertation]. Massachusetts Institute of Technology.

Sorace, A., & Filiaci, F. (200,). Anaphora resolution in near-native speakers of Italian. *Second Language Research, 22(3*), 339–368.

Sybesma, R., & Li, B. (2007). The dissection and structural mapping of Cantonese sentence final particles. *Lingua, 117*(10), 1739–1783.

Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America, 130*(6), 4063–4069.

香港中文大學
The Chinese University of Hong Kong

Thank you for listening!
All comments are welcome.

Wangqian Zhou (Email: zhouwangqian@link.cuhk.edu.hk)

香港中文大學
The Chinese University of Hong Kong