

More than a feeling: Expressive style influences cortical speech tracking in subjective cognitive decline

Matthew King-Hang Ma^{1,**}, Yun Feng^{1,*}, Cloris Pui-Hang Li^{1,*}, Manson Cheuk-Man Fong^{1,2}

¹ Research Centre for Language, Cognition, and Neuroscience, Department of Language Science and Technology, The Hong Kong Polytechnic University, Hong Kong

² Research Institute for Smart Ageing, The Hong Kong Polytechnic University, Hong Kong
khmma@polyu.edu.hk, yunfeng@polyu.edu.hk, ph2li@polyu.edu.hk, cmmfong@polyu.edu.hk

Abstract

Subjective cognitive decline (SCD) doubles dementia risk. This study investigates how self-perceived cognitive worsening shapes neural dynamics during naturalistic speech perception. EEG was collected from 60 cognitively normal older adults as they listened to speech varied in prosodic contexts, categorized by expressive style (scrambled, descriptive, dialogue, exciting). Encoding models mapping three speech representations—acoustic, subsyllabic segmentation and phonotactic features—to ongoing EEG signals were built. Cortical tracking strength (CTS) showed that models fitted with subsyllabic linguistic features outperformed acoustic ones. Crucially, greater SCD severity was associated with weaker CTS of (1) subsyllabic linguistic but not acoustic features, and (2) prosodically flat speech (scrambled and descriptive). Thus, the CTS of higher-level linguistic features while listening to prosodically flat speech may serve as a potential neural marker for early-stage cognitive decline.

Index Terms: subjective cognitive decline, cortical tracking, EEG, speech perception, speech prosody, phonotactics

1. Introduction

Population ageing worldwide leads to the increasing prevalence of cognitive decline, resulting in a pressing need for effective monitoring and early intervention strategies [1]. Subjective cognitive decline (SCD), a self-perceived cognitive worsening despite normal objective cognitive test performance, has emerged as an early indicator of potential impairment [2]. Evidence suggests SCD doubles the risk of progressing to mild cognitive impairment (MCI) and dementia, particularly for those reporting complaints at a younger age (around 60) [2]. While most studies focused on memory decline [2], everyday functioning relies on broader cognitive abilities. Exploring neural correlates of SCD during ecologically valid tasks, such as naturalistic speech perception, is therefore central to understanding cognitive decline comprehensively and informing early detection strategies.

Effective speech comprehension is essential for older adults' social connection and quality of life [3], yet it often declines with age. This stems from degrading auditory perception and cognitive abilities, such as inhibitory control [4]. The degradation of auditory perception, beginning as early as middle age [5], impairs both incoming phonological representations and the perception of speech prosody—the vocal modulations of suprasegmental information—that is fundamental to social interaction [6]. To compensate for degraded bottom-up

acoustic representations, speech comprehension relies increasingly on cognitively demanding top-down processing. For example, older adults rely more heavily on semantic rather than acoustic cues during prosody perception [7].

Pioneering studies have discovered that ongoing neural signals, such as electroencephalography (EEG), can track incoming auditory stimuli [8, 9]. Using encoding models like the multivariate temporal response function (mTRF), researchers have shown that cortical tracking is not merely an epiphenomenon. Instead, it facilitates general auditory processing [10, 11] alongside speech-specific processing at the phoneme [12], word [13], and prosodic levels [14, 15]. Furthermore, cortical tracking is modulated by prior context [16] and intelligibility; for instance, phoneme feature tracking is only significant when speech is understood [12]. These findings demonstrate cortical tracking's functional and predictive role in speech comprehension.

Cortical tracking studies in older adults show enhanced tracking of higher-level linguistic features [13] and reduced tracking of acoustic features [17], a pattern more pronounced with poorer inhibition abilities [18]. These results suggest that the tracking of linguistic features compensates for degraded acoustic and cognitive functions. Furthermore, tracking of the speech envelope provides unique contributions by aiding in temporal prediction [14, 19], though this ability is shown to decline with age [20]. While cortical tracking findings in cognitively declined populations remain scarce and inconsistent [11, 21], behavioral studies clearly show these populations exhibit significant impairments in recognizing emotional prosody [22, 23]. Examining how the cortical tracking of diverse speech features changes in individuals with SCD across varying prosodic contexts, could therefore reveal novel, early-stage candidate neural markers for cognitive decline and provide insight into the underlying neural changes.

The present study utilizes cortical tracking to investigate how self-perceived worsening of cognition affects speech perception. We specifically tapped into features across two levels of speech processing: acoustic and subsyllabic. While tracking of acoustic features reflects bottom-up sensory processing, tracking of subsyllabic segmentation and phonotactic features captures higher-level linguistic processing. Including both features allows us to model the tracking of subsyllabic structures, both independently and alongside their statistical predictions. Furthermore, we evaluate the impact of varying prosodic contexts using custom stimuli categorized by expressive style.

We hypothesize that (1) cortical tracking strength (CTS) for higher-level subsyllabic features will be significantly modulated by SCD severity, whereas (2) CTS for bottom-up acoustic features will remain relatively stable, reflecting a preservation of sensory processing relative to higher-level linguistic process-

*These authors contributed equally.

**indicates the corresponding author.

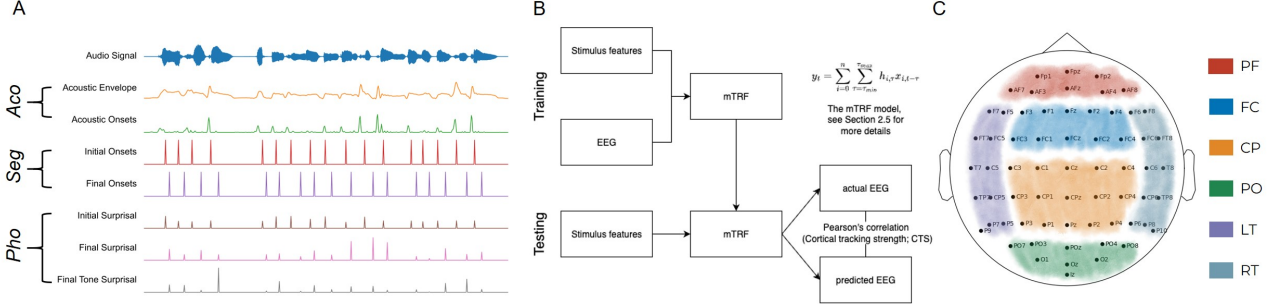


Figure 1: *Methodology. A. Three different set of stimulus features (Aco: Acoustic, Seg: Segmentation, Pho: Phonotactic); B. Training and testing of mTRF models; and the derivation of cortical tracking strength. C. The grouping of 64 electrodes into 6 sites.*

ing. Regarding the impact of prosodic contexts, we hypothesize that (3) CTS for prosodically rich stimuli will be more sensitive to SCD severity. While acoustic processing under clean speech condition may be preserved, the need to integrate rapidly changing prosodic cues and multi talker dynamics imposes higher listening demands [24].

2. Materials and Methods

2.1. Participants

A total of 62 (30F) Cantonese speakers with no known neurological disorders was recruited. These individuals were aged 60 – 70 ($M = 65.2$, $SD = 3.0$). Each participant was cognitively normal per MoCA-HK, adjusted for education years [25]. Degree of SCD was assessed using the 14-item Subjective Cognitive Decline Scale (SCDS) [26]. SCDS scores ranged from 14 to 58 out of 60 ($M = 34.6$, $SD = 11.3$). Hearing thresholds were measured by Pure-Tone Audiometry (PTA). Written informed consent was obtained from all participants.

2.2. Stimuli and experimental procedures

Speech stimuli of four expressive styles were constructed: scrambled, descriptive, dialogue, and exciting. The scrambled and descriptive style stimuli were synthesized with *Amazon Polly* from transcribed radio weather reports; the descriptive style used the original text, while the scrambled style used a word-level scrambled version. Dialogue-style stimuli were sourced from a local community radio program, and exciting-style stimuli were curated from a radio drama. While the scrambled and descriptive styles consist of single-talker speech, the dialogue and exciting styles include two or more talkers. Including multi-talker scenario was a methodological choice to enhance ecological validity, as well as to create cognitively challenging conditions that demand both sensory and cognitive resources. These complex scenarios open up the potential to reveal subtle SCD-induced neural changes, which could otherwise be masked while listening to simple, single-talker speech.

From an initial set of eight stimuli per style, we recruited an independent group of 24 (12F) raters (age = 22.0 ± 2.8) to evaluate the perceived valence, arousal, and dominance of these recordings. Then, we conducted K-means clustering ($k = 4$) on the 3D Valence-Arousal-Dominance (VAD) space. For each style, we selected four out of eight stimuli that were closest to their respective cluster center for the subsequent EEG study (average ratings in Table 1). The standard deviation (SD) of the fundamental frequency (F0) and root-mean-square (RMS) amplitude of selected stimuli were also computed. As shown in

Table 1, the dialogue and exciting styles exhibited higher F0 or intensity SD, revealing their richer prosodic nature.

Table 1: *Mean stimulus ratings and acoustic properties. Val: Valence; Aro: Arousal; Dom: Dominance. F0_{SD}: SD of F0 in Hz. RMS_{SD}: SD of RMS amplitude.*

Style	Val	Aro	Dom	F0 _{SD}	RMS _{SD}
scrambled	2.00	1.69	2.41	48.53	5.40
descriptive	2.57	1.94	2.69	46.75	5.42
dialogue	3.61	3.31	3.10	67.34	5.31
exciting	3.60	3.84	3.38	81.67	6.12

2.3. Data acquisition and preprocessing

EEG data were recorded using a 64-channel BioSemi ActiveTwo system at a 2048 Hz sampling rate. Eye movements were monitored with horizontal and vertical electrooculograms. Participants sat approximately 70 cm from a monitor while auditory stimuli were presented via two speakers. During the recording, they listened to 16 one-minute recordings of clean speech (4 per expressive style) in a random order. After each recording, they used a five-point Likert scale to rate perceived valence, arousal, and dominance with the Self-Assessment Manikin [27], as well as their overall enjoyment. Two participants with unusually low variability in their behavioral ratings ($Z \leq -1.96$ for standard deviation of valence, arousal, or dominance) were excluded, leaving 60 (30F) participants.

Prior to any preprocessing steps, the raw EEG signals were visually inspected to exclude channels contaminated by body movements. The signals were first downsampled to 512 Hz and average re-referenced. Independent component analysis (ICA) was applied to remove eye artifacts following a z-score thresholding approach [28]. Afterwards, channels that were excluded during visual inspection were interpolated. The signals were then band-pass filtered between 1 and 40 Hz using a third-order Butterworth filter. The above steps were conducted with custom Python scripts with *MNE-Python* [29].

2.4. Feature extraction

2.4.1. Acoustic features

The speech envelope and its onset were included as acoustic features following existing studies [11, 12, 14, 16, 30]. To extract the speech envelope, each audio signal was first downsampled to 128 Hz, and then filtered with 24 gammatone filters ranging

from 50 Hz to 8000 Hz. The filtered responses were then scaled to the power of 0.6 to approximate the nonlinear relationship between sound intensity and perceived loudness [30]. These 24 filtered signals were then averaged and filtered from 0.5 Hz to 25 Hz to obtain the final envelope. The envelope onset was computed as the first-order differences of the speech envelope.

2.4.2. Subsyllabic segmentation and phonotactic features

Subsyllabic segmentation was obtained by training a Cantonese Montreal Forced Aligner (MFA) on the 109-hour Common Voice Hong Kong Chinese Corpus (zh-HK 21.0), which features 3,084 speakers [31]. A pronunciation dictionary was generated with *CharsiuG2P* [32] grapheme-to-phoneme conversion tool. The resulting MFA alignments were then manually refined in Praat [33]. Following standard Chinese phonological analysis, the onset times of initials and finals rather than phonemes were extracted as segmentation features [34].

To the best of our knowledge, there are no publicly available Cantonese subsyllabic probabilistic phonotactic features. We therefore extracted them by deriving the initial, final and tone surprisal values from the Hong Kong Cantonese Corpus, which contains around 230,000 words [35]. The texts were parsed with *pycantonese* [36], and surprisal values were calculated as the negative logarithm of the probability for each component: unigram probability for initials $P(\text{initial})$, conditional probability for finals $P(\text{final}|\text{initial})$, and conditional probability for tones $P(\text{tone}|\text{initial}, \text{final})$. All subsyllabic segmentation and phonotactic features were downsampled to 128 Hz. These features are now available at [37] to facilitate future research.

2.5. Multivariate temporal response function (mTRF)

To investigate the cortical tracking of speech, we modeled the relationship between the EEG signals (dependent variable, y) and various stimulus features (predictor variables, x) using a multivariate temporal response function (mTRF). It models the EEG signals as a linear convolution of the stimulus features with a kernel h (see also Figure 1B):

$$y_t = \sum_{i=0}^n \sum_{\tau=\tau_{min}}^{\tau_{max}} h_{i,\tau} x_{i,t-\tau} \quad (1)$$

where y_t is the EEG signal at one channel at time t , n is the total number of predictor variables, and x_i is the i -th predictor. We fitted all mTRF models using the Boosting algorithm implemented in the *Eelbrain* Python toolkit [38], which can yield more accurate mTRF estimates than conventional ridge regression. Prior to fitting, the preprocessed EEG signals were downsampled to 128 Hz and low-pass filtered at 25 Hz to match the sampling rate of the stimulus features. The mTRFs were estimated over time lags ranging from -200 ms to 600 ms. To investigate the tracking of different stimulus features, we fitted three models using distinct feature sets: acoustic (*Aco*), segmentation (*Seg*), and phonotactic (*Pho*) features (Figure 1A). This procedure was repeated for each of the four expressive styles, yielding a total of 12 models.

Model performance was evaluated using a leave-one-out cross-validation scheme. Within k stimuli, a model was trained on data from $k - 1$ stimuli and then tested on the held-out stimulus. This process was repeated k times, with each stimulus serving as the test set once. CTS was quantified as the Pearson's correlation coefficient between the observed and predicted EEG signals, averaged across all k folds.

2.6. Statistical analyses

For statistical analysis, the 64 channels were grouped into 6 scalp sites (Figure 1C): prefrontal (FP); frontocentral (FC); centroparietal (CP); parieto-occipital (PO); left temporal (LT) and right temporal (RT) [28]. The maximum CTS per site served as the dependent variable to capture the strongest local tracking response while reducing dilution from weakly responsive electrodes. To examine how CTS is modulated by SCD severity, mTRF models and expressive styles, we fitted a linear mixed effect model in *R* with *lme4* [39] as in Equation 2:

$$\begin{aligned} \text{CTS} \sim & \text{SCDS} * \text{Site} * \text{Model} * \text{ExpressStyle} + \text{Age} + \text{Gender} + \\ & \text{EducationYears} + \text{MoCA} + \text{PTA} + \text{Valence} + \\ & \text{Arousal} + \text{Dominance} + \text{Enjoy} + (1|\text{Participant}) \end{aligned} \quad (2)$$

where SCDS refers to the SCDS scores, Model refers to mTRF models fitted with three distinct feature sets (*Aco*, *Seg*, *Pho*), Site refers to the six scalp sites. Covariates included age, gender, education years, MoCA-HK score, PTA, perceived valence, arousal, dominance and enjoyment, along with the participant-specific random intercepts. PTA was defined as the average of the hearing thresholds at 500, 1000, 2000 and 4000 Hz [11].

The full model in Equation 2 was then reduced to a simpler model with backward elimination. Type-III ANOVA was then performed, and all post-hoc comparisons were conducted with estimated marginal means using *emmeans* [40].

3. Results

The final reduced model is shown in Equation 3.

$$\begin{aligned} \text{CTS} \sim & \text{SCDS} + \text{Site} + \text{Model} + \text{ExpressStyle} + \text{Arousal} + \\ & \text{SCDS:Site} + \text{SCDS:Model} + \text{Site:Model} + \\ & \text{SCDS:ExpressStyle} + \text{Site:ExpressStyle} + \\ & \text{ExpressStyle:Model} + (1|\text{Participant}) \end{aligned} \quad (3)$$

As Equation 3 shows, none of the three-way or four-way interactions survived after backward elimination. Among all significant factors, we were mostly interested in the Model main effect, the interaction between SCDS scores and expressive style ($F(3, 4208.67) = 25.43, p < .001$) and that between SCDS scores and Model ($F(2, 4208.02) = 7.07, p < .001$), addressing our three predictions. Other effects fall beyond our scope.

3.1. Subsyllabic models outperform the acoustic model

A significant main effect of mTRF models was found ($F(2, 4208.02) = 21.48, p < .001$). Post-hoc comparisons (Figure 2A) revealed a hierarchy in model performance. The *Pho* model significantly outperformed the *Seg* model, which in turn demonstrated significantly higher CTS compared to the *Aco* model ($ps < .001$ for all pairs).

3.2. SCD modulates CTS in linguistic (not acoustic) models

To decompose the significant two-way interaction between SCDS scores and mTRF models, we first conducted a post-hoc simple slope analysis and found SCDS scores significantly modulated CTS negatively for the *Seg* model ($t(62.26) = -2.05, p = .044$) while the trend for *Pho* model was marginally significant ($t(62.26) = -1.76, p = .084$).

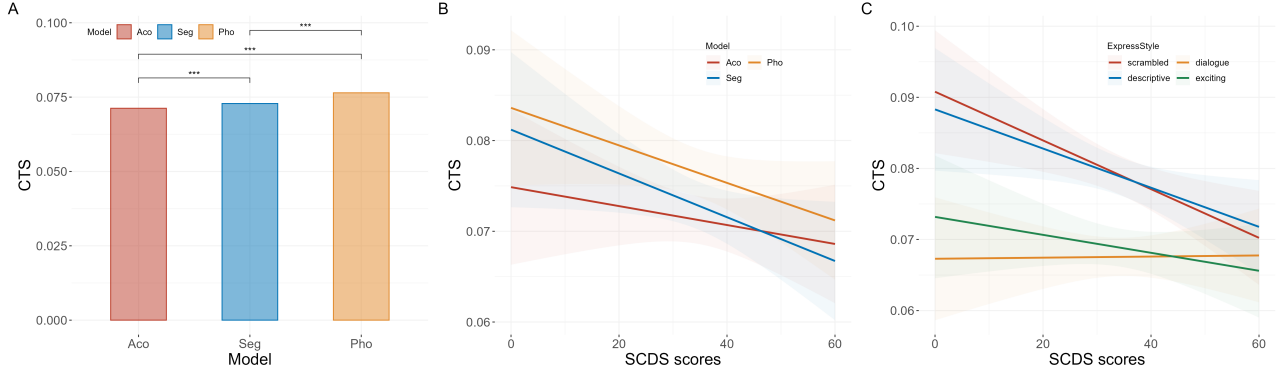


Figure 2: **A.** CTS across models; **B.** SCDS $\times$ Model interaction. **C.** SCDS $\times$ ExpressStyle interaction. Y-axis: estimated marginal means; ribbons: 95% CIs.

Subsequent pairwise comparisons were conducted with Figure 2B showing the interaction plot. The trends of both the *Seg* ($t(4208) = -3.61, p = .001$) and *Pho* ($t(4208) = -2.70, p = .019$) models were significantly more negative than that of the *Aco* model, while there were no significant differences between *Seg* and *Pho* models.

3.3. SCD modulates CTS in flat (not rich) styles

To decompose the significant two-way interaction between SCDS scores and expressive style, we first conducted a post-hoc simple slope analysis and found SCDS significantly modulated CTS negatively for scrambled- ($t(64.37) = -2.89, p = .005$) and descriptive-style speech ($t(64.44) = -2.32, p = .023$).

Subsequent pairwise comparisons were conducted with Figure 2C showing the interaction plot. The trend for scrambled-style was significantly more negative than that of the dialogue- ($t(4209.41) = -7.97, p < .001$) and exciting-style ($t(4209.53) = -4.92, p < .001$). Similarly, the trend for descriptive-style was also significantly more negative than that of the dialogue-style ($t(4208.09) = -6.47, p < .001$) and the exciting-style ($t(4208.12) = -3.41, p = .004$).

4. Discussion

The present study investigated how CTS is modulated by SCD severity, different mTRF models (trained by different acoustic or subsyllabic linguistic features), and expressive styles. We first showed that the phonotactic model outperformed both the segmentation and acoustic models in terms of CTS. Thus, higher-level linguistic information better explained EEG signals during speech perception in cognitively normal older adults. Consistent with previous findings [13, 41], this confirms that successful speech perception in aging relies more heavily on top-down linguistic processing. Establishing this baseline sets the stage to examine how SCD disrupts these processes.

Supporting our first two hypotheses, we found that SCD modulates cortical tracking differently depending on the level of speech processing. As predicted, greater SCD severity was associated with weaker cortical tracking of higher-level, subsyllabic linguistic (segmentation and phonotactics) features, but not acoustic features. It suggests that SCD may specifically affect linguistic features, reflecting early vulnerability in processing linguistic information that support speech perception.

We initially predicted that CTS for prosodically rich (di-

alogue and exciting), but not flat (scrambled and descriptive), stimuli would be most negatively impacted by SCD severity. This is due to the need to integrate rapidly changing pitch, intensity, duration, and rhythm cues, as well as multi-talker dynamics. High demands are placed on both the sensory and cognitive systems in older listeners, whether they are cognitively healthy or impaired [21]. Yet, our results showed the opposite pattern. A possible explanation is that despite the local acoustic complexity, rich prosody may be more beneficial than detrimental at the SCD stage. Existing studies have shown that prosodic cues may act as reliable acoustic and perceptual scaffolds in aiding temporal predictions [19] and syntactic parsing [42]. Thus, our results showed that, given the specific materials, the availability of prosodic cues may allow even participants with greater subjective concerns to compensate for weaker speech tracking. In contrast, flat speech provides fewer prosodic cues, requiring greater reliance on subsyllabic segmentation and phonotactic probabilities. Together with the observed SCD-related reduction in the tracking of linguistic features, our findings therefore suggest that higher-level linguistic processing, when unsupported by prosodic scaffolding, is most vulnerable to SCD.

Our study is limited in that only two levels of speech processing were modeled. Future studies should examine features including but not limited to word-level (e.g., word surprisal [43]) and explicit prosodic-level (e.g., pitch contour [44]) features to gain a more comprehensive understanding of declining linguistic processing abilities and to test our speculated scaffolding role of prosody. Additionally, leveraging the EEG oscillatory hierarchy [10] could help localize specific linguistic deficits and corresponding brain rhythms.

5. Conclusion

Utilizing speech stimuli across varying prosodic contexts, we investigated how subjective cognitive decline (SCD) impacts the cortical tracking of acoustic, subsyllabic segmentation and phonotactic features. We found that greater subjective cognitive concern is associated with weaker cortical tracking of (1) higher-level, subsyllabic linguistic features, and (2) prosodically flat speech. These results suggest that individuals with greater subjective cognitive concerns may make use of rich prosodic cues to compensate for their diminished higher-level linguistic processing. Consequently, the cortical tracking of prosodically flat speech could serve as a candidate neural marker for early-stage cognitive decline.

6. Acknowledgements

This work was supported by the HKRGC Postdoctoral Fellowship Scheme awarded to M.K.-H.M.. We thank the University Research Facility for Big Data Analytics (UBDA), HKPolyU, for making the GPU virtual machines available for the computations involved.

7. Generative AI Use Disclosure

Gemini 3.1 Pro was used to improve the grammar and syntax after the author had written the draft. All AI-generated suggestions were manually reviewed.

8. References

- [1] Y. J. Hong, S. Ho, J. H. Jeong, K. H. Park, S. Kim, M. J. Wang, S. H. Choi, D. W. Yang, and CoSCo study group, "Impacts of baseline biomarkers on cognitive trajectories in subjective cognitive decline: The CoSCo prospective cohort study," *Alzheimer's Research & Therapy*, vol. 15, no. 1, p. 132, Aug. 2023.
- [2] F. Jessen, R. E. Amariglio, R. F. Buckley, W. M. van der Flier, Y. Han, J. L. Molinuevo, L. Rabin, D. M. Rentz, O. Rodriguez-Gomez, A. J. Saykin, S. A. M. Sikkes, C. M. Smart, S. Wolfgruber, and M. Wagner, "The characterisation of subjective cognitive decline," *Lancet Neurology*, vol. 19, no. 3, pp. 271–278, Mar. 2020.
- [3] A. Heinrich, J.-P. Gagne, A. Viljanen, D. A. Levy, B. Ben-David, and B. A. Schneider, "Effective communication as a fundamental aspect of active aging and well-being: Paying attention to the challenges older adults face in noisy environments," *Social Inquiry into Well-Being*, vol. 2, no. 1, Sep. 2016.
- [4] M. C.-M. Fong, T. S.-T. Law, M. K.-H. Ma, N. Y. Hui, and W. S. Wang, "Can inhibition deficit hypothesis account for age-related differences in semantic fluency? Converging evidence from Stroop color and word test and an ERP flanker task," *Brain and Language*, vol. 218, p. 104952, Jul. 2021.
- [5] K. S. Helfer, G. R. Merchant, and P. A. Wasiuk, "Age-Related Changes in Objective and Subjective Speech Perception in Complex Listening Environments," *Journal of Speech, Language, and Hearing Research*, vol. 60, no. 10, pp. 3009–3018, Oct. 2017.
- [6] H. Mixdorff, A. Hönemann, A. Rilliard, T. Lee, and M. K. Ma, "Audio-visual expressions of attitude: How many different attitudes can perceivers decode?" *Speech Communication*, vol. 95, pp. 114–126, Dec. 2017.
- [7] B. M. Ben-David, S. Gal-Rosenblum, P. H. H. M. van Lieshout, and V. Shakuf, "Age-Related Differences in the Perception of Emotion in Spoken Language: The Relative Roles of Prosody and Semantics," *Journal of Speech, Language and Hearing Research (Online)*, vol. 62, no. 4S, pp. 1188–1202, Apr. 2019.
- [8] E. Ahissar, S. Nagarajan, M. Ahissar, A. Protopapas, H. Mahncke, and M. M. Merzenich, "Speech comprehension is correlated with temporal response patterns recorded from auditory cortex," *Proceedings of the National Academy of Sciences*, vol. 98, no. 23, pp. 13 367–13 372, Nov. 2001.
- [9] N. Ding and J. Z. Simon, "Neural coding of continuous speech in auditory cortex during monaural and dichotic listening," *Journal of Neurophysiology*, vol. 107, no. 1, pp. 78–89, Jan. 2012.
- [10] E. C. Lalor and A. R. Nidiffer, "On the generative mechanisms underlying the cortical tracking of natural speech: A position paper," 2025.
- [11] E. Bolt and N. Giroud, "Neural encoding of linguistic speech cues is unaffected by cognitive decline, but decreases with increasing hearing impairment," *Scientific Reports*, vol. 14, no. 1, p. 19105, Aug. 2024.
- [12] K. D. Prinsloo and E. C. Lalor, "General Auditory and Speech-Specific Contributions to Cortical Envelope Tracking Revealed Using Auditory Chimeras," *Journal of Neuroscience*, vol. 42, no. 41, pp. 7782–7798, Oct. 2022.
- [13] J. Mesik, L. Ray, and M. Wojtczak, "Effects of Age on Cortical Tracking of Word-Level Features of Continuous Competing Speech," *Frontiers in Neuroscience*, vol. 15, p. 635126, 2021.
- [14] A. Bernardo, P. Correia, M. Vigário, R. Vigário, and S. Frota, "Neural tracking of prosodic structure in delexicalized speech," in *Speech Prosody 2024*. ISCA, Jul. 2024, pp. 1140–1144.
- [15] G. Antonicelli, N. Molinaro, P. D. L. Riva, R. Laspiur, A. L. D. Turiso, and S. Mancini, "Cortical tracking of prosody after stroke and in aging: Evidence from magnetoencephalography," in *Speech Prosody 2024*. ISCA, Jul. 2024, pp. 21–26.
- [16] G. M. Di Liberto, M. J. Crosse, and E. C. Lalor, "Cortical Measures of Phoneme-Level Speech Encoding Correlate with the Perceived Clarity of Natural Speech," *eneuro*, vol. 5, no. 2, pp. ENEURO.0084–18.2018, Mar. 2018.
- [17] B. Herrmann, B. Maess, and I. S. Johnsrude, "A neural signature of regularity in sound is reduced in older adults," *Neurobiology of Aging*, vol. 109, pp. 1–10, Jan. 2022.
- [18] L. Decruy, J. Vanthornhout, and T. Francart, "Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties," *J. Neurophysiol.*, vol. 122, no. 2, pp. 601–615, Aug. 2019.
- [19] Y. Lamekina, L. Titone, B. Maess, and L. Meyer, "Speech Prosody Serves Temporal Prediction of Language via Contextual Entrainment," *Journal of Neuroscience*, vol. 44, no. 28, Jul. 2024.
- [20] M. Gillis, J. Kries, M. Vandermosten, and T. Francart, "Neural tracking of linguistic and acoustic speech representations decreases with advancing age," *NeuroImage*, vol. 267, p. 119841, Feb. 2023.
- [21] G. M. Bidelman, J. E. Lowther, S. H. Tak, and C. Alain, "Mild Cognitive Impairment Is Characterized by Deficient Brainstem and Cortical Representations of Speech," *J. Neurosci.*, vol. 37, no. 13, pp. 3610–3620, Mar. 2017.
- [22] J. Amlerova, J. Laczó, Z. Nedelska, M. Laczó, M. Vyhánek, B. Zhang, K. Sheardova, F. Angelucci, R. Andel, and J. Hort, "Emotional prosody recognition is impaired in Alzheimer's disease," *Alzheimer's Research & Therapy*, vol. 14, no. 1, p. 50, Apr. 2022.
- [23] J. Jiang, J. C. S. Johnson, M.-C. Requena-Komuro, E. Benhamou, H. Sivasathiaselan, A. Chokesuwattanaskul, A. Nelson, R. Nortley, R. S. Weil, A. Volkmer, C. R. Marshall, D.-E. Bamio, J. D. Warren, and C. J. D. Hardy, "Comprehension of acoustically degraded emotional prosody in Alzheimer's disease and primary progressive aphasia," *Scientific Reports*, vol. 14, no. 1, p. 31332, Dec. 2024.
- [24] P. Kovács, B. Tóth, F. Honbolygó, O. Szalárdy, A. Kohári, K. Mády, L. Magyar, and I. Winkler, "Speech prosody supports speaker selection and auditory stream segregation in a multi-talker situation," *Brain Research*, vol. 1805, p. 148246, Apr. 2023.
- [25] A. Wong, L. S. N. Law, W. Liu, Z. Wang, E. S. K. Lo, A. Lau, L. K. S. Wong, and V. C. T. Mok, "Montreal Cognitive Assessment: One Cutoff Never Fits All," *Stroke*, vol. 46, no. 12, pp. 3547–3550, Dec. 2015.
- [26] H.-F. Tsai, C.-H. Wu, C.-C. Hsu, C.-L. Liu, and Y.-H. Hsu, "Development of the Subjective Cognitive Decline Scale for Mandarin-Speaking Population," *Am. J. Alzheimers. Dis. Other Dement.*, vol. 36, p. 15333175211038237, 2021.
- [27] M. M. Bradley and P. J. Lang, "Measuring emotion: The Self-Assessment Manikin and the Semantic Differential," *J. Behav. Ther. Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994.
- [28] M. K.-H. Ma, M. C.-M. Fong, C. Xie, T. Lee, G. Chen, and W. S. Wang, "Regularity and randomness in ageing: Differences in resting-state EEG complexity measured by largest Lyapunov exponent," *Neuroimage: Reports*, vol. 1, no. 4, p. 100054, Dec. 2021.
- [29] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. Hämäläinen, "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroscience*, vol. 7, p. 267, Dec. 2013.

- [30] M. F. Issa, I. Khan, M. Ruzzoli, N. Molinaro, and M. Lizarazu, "On the speech envelope in the cortical tracking of speech," *NeuroImage*, vol. 297, p. 120675, Aug. 2024.
- [31] K. K. So, C. Xu, G. W. Cao, and P. Mok, "Performance of Montreal Forced Aligner on Cantonese Spontaneous Speech," in *Proc. Interspeech 2025*, 2025, pp. 5398–5402.
- [32] J. Zhu, C. Zhang, and D. Jurgens, "ByT5 model for massively multilingual grapheme-to-phoneme conversion," in *Proc. Interspeech 2022*, 2022, pp. 446–450.
- [33] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer," Jun. 2024.
- [34] W. S.-Y. Wang, "The Chinese Language," *Scientific American*, vol. 228, no. 2, pp. 50–60, Feb. 1973.
- [35] LYM. Wong and KK. Luke, "The Hong Kong Cantonese Corpus: Design and uses," in *Linguistic Corpus and Corpus Linguistics in the Chinese Context*. Berkeley: The University of California, 2019, vol. 43, p. 312.
- [36] J. L. Lee, L. Chen, C. Lam, C. M. Lau, and T.-H. Tsui, "PyCantonese: Cantonese Linguistics and NLP in Python," in *Proceedings of The 13th Language Resources and Evaluation Conference*. European Language Resources Association, Jun. 2022.
- [37] neurothew, "neurothew/cantonese-phonotactic-features: v1.0.0," Jun. 2026. [Online]. Available: <https://doi.org/10.5281/zenodo.20748010>
- [38] C. Brodbeck, P. Das, M. Gillis, J. P. Kulasingham, S. Bhattachali, P. Gaston, P. Resnik, and J. Z. Simon, "Eelbrain, a Python toolkit for time-continuous analysis with temporal response functions," *Elife*, vol. 12, Nov. 2023.
- [39] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models using lme4," Jun. 2014.
- [40] R. V. Lenth, *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*, 2025.
- [41] E. Bolt, K. Kliestenec, and N. Giroud, "Hearing and cognitive decline in aging differentially impact neural tracking of context-supported versus random speech across linguistic timescales," *PLOS ONE*, vol. 19, no. 12, p. e0313854, Dec. 2024.
- [42] G. Degano, P. W. Donhauser, L. Gwilliams, P. Merlo, and N. Golestani, "Speech prosody enhances the neural processing of syntax," *Communications Biology*, vol. 7, no. 1, pp. 1–10, Jun. 2024.
- [43] A. Goldstein, Z. Zada, E. Buchnik, M. Schain, A. Price, B. Aubrey, S. A. Nastase, A. Feder, D. Emanuel, A. Cohen, A. Jansen, H. Gazula, G. Choe, A. Rao, C. Kim, C. Casto, L. Fanda, W. Doyle, D. Friedman, P. Dugan, L. Melloni, R. Reichart, S. Devore, A. Flinker, L. Hasenfratz, O. Levy, A. Hasidim, M. Brenner, Y. Matias, K. A. Norman, O. Devinsky, and U. Hasson, "Shared computational principles for language processing in humans and deep language models," *Nature Neuroscience*, vol. 25, no. 3, pp. 369–380, Mar. 2022.
- [44] G. N. Gnanateja, K. Rupp, F. Llanos, J. Hect, J. S. German, T. Teichert, T. J. Abel, and B. Chandrasekaran, "Cortical processing of discrete prosodic patterns in continuous speech," *Nature Communications*, vol. 16, no. 1, p. 1947, Mar. 2025.