Speech Prosody 2026
26-29 May 2026, Philadelphia, Pennsylvania, USA

Keep Calm and Listen: Subjective Cognitive Decline Modulates Cortical Tracking of Speech and Music in Different Expressive Styles

Matthew King-Hang Ma[1], *Manson Cheuk-Man Fong*[1,2], *Yun Feng*[1], *Cloris Pui-Hang Li*[1], *William Shiyuan Wang*[1,2]

[1]Research Centre for Language, Cognition and Neuroscience
Department of Language Science and Technology, The Hong Kong Polytechnic University
[2]Research Institute for Smart Ageing, The Hong Kong Polytechnic University
{khmma, cmmfong, yunfeng, ph2li, wsywang}@polyu.edu.hk

Abstract

Subjective cognitive decline (SCD) approximately doubles the risk of developing MCI and dementia in older adults. While existing studies focus primarily on memory deficits, we investigated how SCD modulates speech and music perception to disentangle the roles of sensory and cognitive processing at this early stage. EEG was collected from 56 Cantonese-speaking, cognitively normal older adults (aged 60–70) while they listened to speech and music across four expressive styles varying in valence, arousal, and dominance. Cortical tracking strength (CTS) was estimated by fitting encoding models between acoustic features and EEG signals filtered in delta and theta bands. We found that CTS decreased with increasing SCD severity, but this effect was specific to scrambled speech (theta band) and calm music (delta and theta bands). These findings suggest that SCD-related deficits would arise for both speech (if bottom-up acoustic cues are relied upon heavily) and music perception (if top-down sustained attention is required for tracking endogenous rhythms). By incorporating different expressive styles, our results reveal that SCD involves decline in both sensory and cognitive processing in a domain-specific manner.

Index Terms: Subjective Cognitive Decline, speech prosody, music listening, cortical tracking, EEG

1. Introduction

By the late 2070s, the number of people aged 65 or older is projected to surpass that of children under 18 [1]. The inevitable increasing prevalence of cognitive decline will impose significant challenges for the public healthcare system worldwide. Subjective cognitive decline (SCD), a self-perceived cognitive worsening despite normal objective cognitive test performance, has been proposed to indicate the early stage of cognitive impairment as well as other diseases such as cerebrovascular diseases [2]. While existing evidence suggests SCD doubles the risk of progressing to MCI and dementia [3], most of the studies focus primarily on executive functions and memory [4], with other cognitive domains receiving less attention. Deepening the understanding on how SCD manifest in other cognitive domains might offer insights on developing early detection and intervention strategies to slow down disease progression.

While memory is the primary focus of existing research, two important yet underexplored domain are speech and music perception, which are essential abilities for maintaining social connections, quality of life, and mental well-being in older adults [5, 6]. Both language and music rely on the capacity for hierarchical structure building to combine distinct elements [7]. In speech, phonemes form syllables, which are then integrated into words and sentences. Similarly, musical notes are combined into beats and form phrases [8]. Crucially, the brain tracks these structures with a corresponding hierarchy of neural oscillations [9]. While initial bottom-up sensory processing likely relies on shared auditory mechanisms, higher-level structural integration recruits distinct computational networks. This functional dissociation provides a framework to investigate if ageing and cognitive decline would affect these processes in bottom-up sensory processing, or in domain-specific top-down structural integration.

To empirically investigate how the brain constructs this hierarchy, we focus on cortical tracking, a phenomenon where neural oscillations synchronize with incoming auditory stimuli and their features. To investigate the role of cortical tracking in speech and music perception, encoding models like the multivariate temporal response function (mTRF) were used, attempting to explain the observed EEG signals with features of incoming stimuli [10]. Application of these models has shown that cortical tracking is not merely an epiphenomenon but carries functional significance [11]. Neural oscillations in the delta (1–4 Hz) and theta (4–8 Hz) bands serve as fundamental mechanisms for parsing the hierarchical temporal structure of both speech and music [12, 13]. The theta band predominantly tracks acoustic features occurring at the rate of syllables in speech and notes or beats in music, functioning largely as a segmentation mechanism driven by bottom-up acoustic edges and amplitude envelopes [13, 14, 15]. Conversely, the delta band indexes slower, higher-order structural processing. In speech, it tracks suprasegmental features such as phrasal grouping, intonation, and relative pitch [14, 16, 17]; in music, it tracks the meter and beat, with synchronization often enhanced by musical expertise [12, 18]. While some of these oscillatory mechanisms are shared across domains, recent evidence suggests that tracking in the delta band can be significantly stronger for speech than for music [19] and in different locations of the brain [20], suggesting a complex modulation of cortical tracking by different mechanisms and domains.

One way to explore this complex modulation is through expressive styles, as emotions can be conveyed through the speech and music domains [21]. It is well established that emotion perception declines with age [22, 23], likely contributed by the interplay between sensory and cognitive factors. Perceiving prosodic information from speech requires heightened attention to the rapidly changing acoustic characteristics. Consequently, the information degradation hypothesis posits that the age-related sensory degradation would disrupt perceptual performance, which causes an increasing demand of cognitive resources for compensation [24]. On the other hand, the inhibition

10.21437/SpeechProsody.2026-135

deficit hypothesis suggests that as older adults are less capable of suppressing task-irrelevant information [25], they prefer the more salient semantic over prosodic cues in speech. The dependency on cognition for emotion perception is supported by findings that older adults with lower overall cognition performed worse in emotion perception task [24]. The situation is more pronounced in populations with objective cognitive decline such as aMCI and AD [22, 26].

The present study leverages the cortical tracking phenomenon to examine how SCD influences speech and music perception. In particular, we created or curated a set of speech and music stimuli that vary in expressive styles in attempting to look at how neural oscillations track these stimuli.

It remains unclear whether SCD manifests primarily as a decline in sensory processing, cognitive processing, or a combination of both. If increasing SCD severity reflects progressive sensory degradation, we would expect reduced cortical tracking strength (CTS) across all expressive styles. Specifically, we predict that CTS for energetic or exciting stimuli would show the steepest decrease with SCD severity, as these higher arousal stimuli present greater challenges to the auditory system due to their rapidly modulating acoustic cues. Under this scenario, we might observe more significant patterns from theta than delta band as theta oscillations reflect more bottom-up processing. Alternatively, if cognitive processing abilities remain intact regardless of SCD severity, individuals might still be able to leverage hierarchical structure building to compensate for sensory degradation. Under this scenario, we would expect CTS to be relatively preserved for stimuli that offer rich structural cues, such as the contextual information in speech or the predictable beat in rhythmic music.

2. Materials and Methods

2.1. Participants

We recruited 62 consenting native Cantonese speakers (30F; aged 60-70, mean $= 65.2 \pm 3.03$) with no known neurological disorders. Each participant was cognitively normal per MoCA-HK, adjusted for education years [27]. SCD severity was assessed using the 14-item Subjective Cognitive Decline Scale (SCDS [28]; maximum score of 60), with participant scores ranging from 14 to 58. Hearing thresholds were assessed with Pure-Tone Audiometry (PTA). No participants reported having experience of playing musical instruments for more than 5 years and were therefore regarded as non-musicians.

2.2. Stimuli and experimental procedures

To prepare the stimuli for both speech and music domains, we first curated a set of audio recordings online across four distinct expressive styles. A subset of appropriate stimuli was selected via an independent rating procedure. For speech, the four styles were: (1) scrambled; (2) descriptive; (3) dialogue; and (4) exciting. We synthesized scrambled and descriptive speech using *Amazon Polly* based on transcribed radio weather reports; the descriptive style utilized the original text, while the scrambled style used a randomized version. Dialogue-style speech was sourced from a local community radio program, and exciting-style speech from a radio drama. For music, the four styles were: (1) neutral; (2) calm; (3) pleasant; and (4) energetic. Neutral music was generated by concatenating random piano notes, while the other styles consisted of solo and ensemble Chinese or Western instrumental pieces that were curated online.

Two independent groups of raters were recruited to evaluate the perceived valence, arousal, and dominance of the speech ($n = 24$, 12F, mean age 22.0 ± 2.82) and music ($n = 24$, 13F, mean age 21.1 ± 3.58) stimuli, respectively. Each recording was plotted in a 3D Valence-Arousal-Dominance (VAD) space to compute cluster centroids for each style. The four out of eight stimuli closest to each centroid by Euclidean distance were selected for the EEG study (average ratings in Table 1). Analysis of the VAD space revealed these closest cross-domain pairs: scrambled speech–neutral music, descriptive speech–calm music, dialogue speech–pleasant music, and exciting speech–energetic music.

Table 1: *Averaged rating of the selected stimuli. Val: Valence; Aro: Arousal; Dom: Dominance.*

Domain	Style	Val	Aro	Dom
speech	scrambled	2.00	1.69	2.41
	descriptive	2.57	1.94	2.69
	dialogue	3.61	3.31	3.10
	exciting	3.60	3.84	3.38
music	neutral	1.66	2.20	2.23
	calm	2.89	1.83	2.53
	pleasant	3.72	3.45	3.28
	energetic	3.55	3.83	3.67

2.3. Data acquisition and preprocessing

EEG data were recorded using a 64-channel BioSemi ActiveTwo system at a 2048 Hz sampling rate. Horizontal and vertical electrooculograms (HEOG and VEOG) were collected to monitor eye movements. Participants sat around 70 cm from a monitor while auditory stimuli were presented via two speakers (see Figure 1A). They listened to 32 speech and music stimuli (16 per domain, 4 per expressive style) that were about one-minute long in a pseudorandom order. After each stimulus, they were required to rate perceived valence, arousal, and dominance with the Self-Assessment Manikin [29], as well as their overall enjoyment on a five-point Likert scale (with 3 being neutral). Six participants with extreme variability in their behavioral ratings ($|Z| \geq 1.96$ for valence, arousal, or dominance standard deviations) were excluded, leaving 56 (26F) participants.

The raw EEG signals were visually inspected to exclude channels that were contaminated by body movements before any preprocessing steps. The signals were then downsampled to 512 Hz and average re-referenced. Independent component analysis (ICA) was applied to remove eye artifacts following [30]. Afterwards, excluded channels were interpolated. The signals were then band-pass filtered between 1 and 40 Hz with a third-order Butterworth filter. All EEG-related processing was conducted with custom Python scripts using *MNE-Python* [31].

2.4. Acoustic feature extraction

Acoustic envelope and its onset were included as acoustic features. They were extracted from stimuli of both domains in the same fashion. Each audio signal was first downsampled to 128 Hz, and then filtered with 24 gammatone filters ranging from 50 Hz to 8000 Hz. Each of the 24 signals was then scaled to the power of 0.6 to approximate the nonlinear relationship between sound intensity and perceived loudness [32]. Afterwards, these signals were averaged and filtered from 0.5 Hz to 25 Hz to obtain the acoustic envelope. The envelope onset was computed

as the first-order difference of the envelope.

2.5. Multivariate temporal response function (mTRF)

To investigate the cortical tracking of speech and music, we modeled the relationship between the EEG signals (dependent variable, y) and acoustic features (predictor variables, x) using a multivariate temporal response function (mTRF). It models the EEG signals as a linear convolution of the acoustic 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. All mTRF models were fitted with the Boosting algorithm implemented in the *Eelbrain* Python toolkit [33], which was claimed to result in more accurate mTRF estimates than conventional ridge regression approach [33]. 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 how hierarchical brain oscillations track speech and music, the above procedures were repeated with EEG signals filtered in delta (1-4 Hz) or theta (4-8 Hz) band, and with acoustic features extracted from speech or music. A total of four mTRF models were fitted.

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. Encoding accuracy was quantified as the Pearson's correlation coefficient between the observed and predicted EEG signals, averaged across all k folds. This is referred to as the cortical tracking strength (CTS) hereafter.

2.6. Statistical analyses

For statistical analysis, the 64 channels were grouped into 6 scalp sites (see Figure 1C): prefrontal (FP); frontocentral (FC); centroparietal (CP); parieto-occipital (PO); left temporal (LT) and right temporal (RT) [30]. The maximum CTS per site served as the representative measure. To test how SCD severity and expressive styles hierarchically modulate CTS, four linear mixed effect models (LMMs) were fitted to the CTS resulted from the mTRF models (Section 2.5) using *lme4* [34] in *R* (Equation 2):

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

where SCDS refers to the SCDS scores and Site refers to the six scalp sites. Age, gender, education years, MoCA-HK score, PTA, perceived valence, arousal, dominance and enjoyment were included as covariates, along with the participant-specific random effects. PTA was derived as the average of the hearing thresholds at 500, 1000, 2000 and 4000 Hz [35].

Backward elimination was applied to the full model (Eq. 2). To facilitate band comparisons, we refitted LMMs using the maximal set of retained predictors. Final models were analyzed using Type-III ANOVA, focusing on the SCDS × ExpressStyle interaction, followed by post-hoc comparisons via estimated marginal means [36].

3. Results

Table 2 summarizes the Type-III ANOVA results. While no three-way interactions were found, significant SCDS × ExpressStyle interactions were observed across all four models. These effects were decomposed with post-hoc comparisons, as detailed in Table 3 and Figure 2.

Table 2: *Summary of ANOVA results on LMMs fitted with delta and theta CTS. Significance:* $^{\dagger}: p < .1, ^{*}: p < .05, ^{**}: p < .01, ^{***}: p < .001$. *Direction of effect for covariates: (+): positive, (-): negative. S: speech, M: music.*

	Delta		Theta	
Factor	S	M	S	M
Factors				
SCDS	–	–	–	–
Site	***	***	***	***
ExpressStyle	***	***	***	***
SCDS × Expr. Style	**	**	**	***
Site × Expr. Style	***	***	***	***
Covariates				
PTA	–	†(-)	*(-)	*(-)
Valence	–	–	–	***(-)
Arousal	†(+)	†(+)	–	*(-)
Dominance	–	–	–	***(+)

Table 3: *Post-hoc analysis of SCDS × Expressive Style interaction effects on delta and theta CTS. Rows labeled "(Slope)" denote simple slope tests; all other rows denote pairwise contrasts between CTS trends.*

Comparison	*t*	*df*	*p*
Delta CTS			
Speech			
Descriptive vs. Dialogue	-3.10	1260	$.011^{*}$
Descriptive vs. Scrambled	-2.48	1270	$.064^{\dagger}$
Music			
Calm (Slope)	-2.16	80.9	$.034^{*}$
Calm vs. Neutral	-2.94	1260	$.018^{*}$
Calm vs. Pleasant	-3.08	1260	$.011^{*}$
Theta CTS			
Speech			
Scrambled (Slope)	-2.00	72.7	$.049^{*}$
Scrambled vs. Descriptive	-3.04	1270	$.013^{*}$
Scrambled vs. Dialogue	-2.68	1270	$.038^{*}$
Scrambled vs. Exciting	-3.35	1270	$.005^{**}$
Music			
Calm (Slope)	-2.51	72.8	$.014^{*}$
Calm vs. Neutral	-4.08	1260	$< .001^{***}$
Calm vs. Pleasant	-3.52	1260	$.003^{**}$
Calm vs. Energetic	-2.33	1260	$.093^{\dagger}$

In the delta band, post-hoc tests revealed that CTS for calm music significantly decreased as SCD severity increased ($p = .034$). This negative slope was significantly steeper than the trends observed for neutral and pleasant music (see Table 3). For speech, descriptive speech showed a significantly more negative trend compared to dialogue speech ($p = .011$).

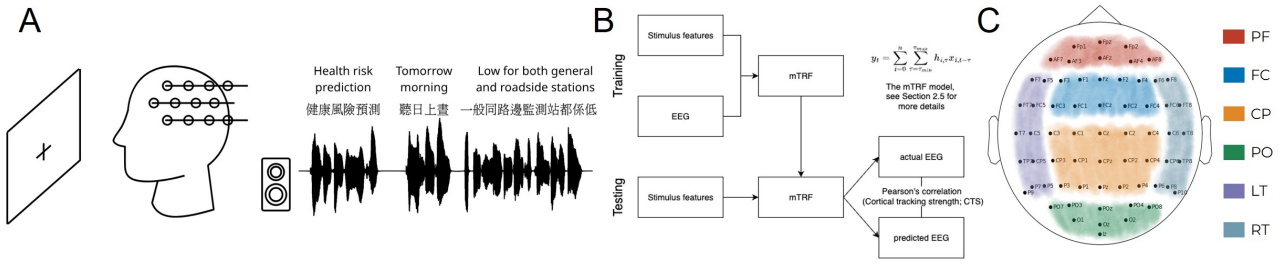

Figure 1: *Experimental paradigm. A: Schematic illustration of the experimental procedures described in Section 2.2; B: Training and testing of mTRF models and the derivation of cortical tracking strength; C: The grouping of 64 electrodes into 6 sites.*

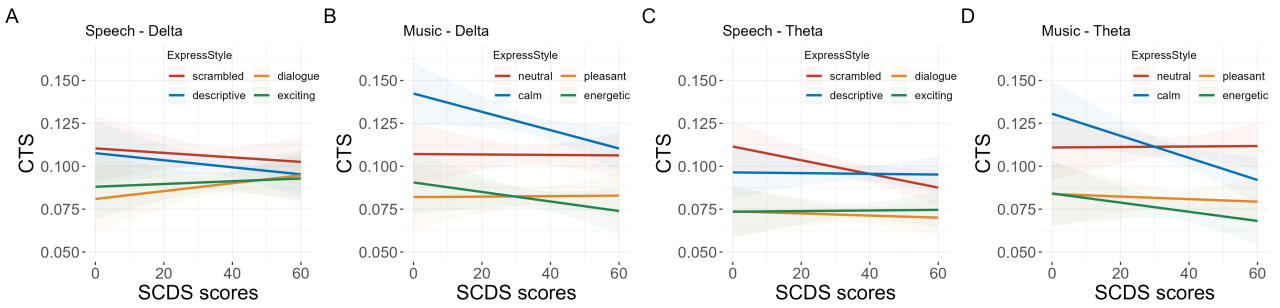

Figure 2: *Interaction plots of the SCDS × Expressive Style interaction on delta and theta CTS. A: delta CTS of speech; B: delta CTS of music; C: theta CTS of speech; D: theta CTS of music. The shaded ribbons correspond to the 95% CI of the estimated marginal means.*

In the theta band, significant negative trends were found for both calm music ($p = .034$) and scrambled speech ($p = .049$) as SCD severity increased. Pairwise comparisons indicated that the slope for calm music was significantly more negative than neutral and pleasant music. Similarly, the trend for scrambled speech was distinct from other speech types, showing a significantly steeper decline compared to descriptive, dialogue, and exciting speech (all $ps < .05$; see Table 3).

4. Discussion

This study investigated whether SCD modulates the cortical tracking of speech and music, and how these effects are manifested with distinct neural oscillations and various expressive styles. Contrary to our prediction based on the sensory degradation account, we only observed decreased CTS in scrambled speech and calm music but not all styles. Yet, our results suggest a complex interplay between sensory and cognitive deficits, differentially reflected by speech and music perception.

In the speech domain, the observed negative association was restricted to the theta band and was specific to scrambled speech. Scrambled speech preserves the acoustic structure of the signal but removes its context, forcing individuals to rely solely on acoustic cues without the aid of higher-level linguistic knowledge. Consequently, the decreasing CTS in this condition likely reflects a decreased fidelity in the auditory representation of the incoming stimuli. The lack of significant findings in other styles suggests that participants with higher SCD severity may still be able to compensate for the increasing sensory difficulties by leveraging contextual information in natural speech, consistent with the information degradation hypothesis.

In contrast to the speech domain, the negative association was observed in both delta and theta bands in the music domain. Interestingly, this effect was observed specifically for calm music; it contradicts our prediction that the stimuli that are rich in rapidly modulating acoustic cues would lead to reduced tracking. One possible explanation is that, calm music often lacks a clear rhythm, requiring the listener to recruit top-down predictive mechanisms in both slower delta and faster theta band to track musical phrases and notes, respectively [8]. This is contrasted by more energetic music, which is often characterized by strong rhythms. Such music might elicit more robust, stimulus-driven evoked responses, which are posited as a possible generative mechanism of cortical tracking [10]. Therefore, the reduced tracking with increasing SCD severity in calm music points to a subtle impairment in maintaining the attention to track endogenous rhythms.

Collectively, these findings indicate that cortical tracking in speech and music reveals distinct declining abilities in SCD. The results from scrambled speech suggest sensory deficits, while those from calm music point to cognitive deficits. To further refine this distinction, future studies should include younger adults that would allow for a clearer dissociation between normal ageing and the specific SCD-related decline.

5. Conclusion

To conclude, our study demonstrates that cortical tracking of speech and music serves as a sensitive marker for SCD. Crucially, we found that the decreased tracking with increasing subjective cognitive concerns are style-specific, emerging only when the stimuli lack salient rhythmic features (calm music) or contextual information (scrambled speech). These findings highlight that SCD involves decline in both bottom-up sensory and top-down cognitive processing.

6. Acknowledgements

This work was supported by the Inter-Faculty Collaboration Scheme for FH, FHSS and FENG by Hong Kong Polytechnic University awarded to W.S.W. (PI), and the HKRGC Postdoctoral Fellowship Scheme awarded to M.K-H.M..

7. References

[1] United Nations, *World Population Prospects 2024: Summary of Results*, ser. UN DESA/POP/2024/TR/NO. 9. New York: United Nations, 2024.

[2] H. Pitti *et al.*, "Cerebrovascular damage in subjective cognitive decline: A systematic review and meta-analysis," *Ageing Research Reviews*, vol. 82, p. 101757, Dec. 2022.

[3] F. Jessen *et al.*, "The characterisation of subjective cognitive decline," *Lancet Neurol.*, vol. 19, no. 3, pp. 271–278, Mar. 2020.

[4] C. E. Munro *et al.*, "Recent contributions to the field of subjective cognitive decline in aging: A literature review," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 15, no. 4, p. e12475, 2023.

[5] A. Heinrich, J.-P. Gagn, 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.

[6] P. Laukka, "Uses of music and psychological well-being among the elderly," *Journal of Happiness Studies*, vol. 8, no. 2, p. 215, Jun. 2007.

[7] R. Asano, "The evolution of hierarchical structure building capacity for language and music: A bottom-up perspective," *Primates*, vol. 63, no. 5, pp. 417–428, Sep. 2022.

[8] M. F. Assaneo and J. Orpella, "Rhythms in Speech," in *Neurobiology of Interval Timing*, H. Merchant and V. De Lafuente, Eds. Cham: Springer International Publishing, 2024, vol. 1455, pp. 257–274.

[9] G. Buzsaki, *The Brain from Inside Out*. Oxford University Press, May 2019.

[10] E. C. Lalor and A. R. Nidiffer, "On the generative mechanisms underlying the cortical tracking of natural speech: A position paper," 2025.

[11] G. N. Gnanateja *et al.*, "On the Role of Neural Oscillations Across Timescales in Speech and Music Processing," *Frontiers in Computational Neuroscience*, vol. 16, p. 872093, Jun. 2022.

[12] E. E. Harding, D. Sammler, M. J. Henry, E. W. Large, and S. A. Kotz, "Cortical tracking of rhythm in music and speech," *NeuroImage*, vol. 185, pp. 96–101, Jan. 2019.

[13] A. Fiveash and B. Tillmann, "Shared Mechanisms for the Processing of Rhythm in Music and Speech," Jul. 2024.

[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] J. Rimmele and A. Keitel, "Region-specific endogenous brain rhythms and their role for speech and language," Sep. 2023.

[16] 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.

[17] E. S. Teoh, M. S. Cappelloni, and E. C. Lalor, "Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features," *Eur. J. Neurosci.*, vol. 50, no. 11, pp. 3831–3842, Dec. 2019.

[18] K. B. Doelling and D. Poeppel, "Cortical entrainment to music and its modulation by expertise," *Proceedings of the National Academy of Sciences*, vol. 112, no. 45, pp. E6233–E6242, Nov. 2015.

[19] N. J. Zuk, J. W. Murphy, R. B. Reilly, and E. C. Lalor, "Envelope reconstruction of speech and music highlights stronger tracking of speech at low frequencies," *PLOS Computational Biology*, vol. 17, no. 9, p. e1009358, Sep. 2021.

[20] S. Osorio and M. F. Assaneo, "Anatomically distinct cortical tracking of music and speech by slow (1–8Hz) and fast (70–120Hz) oscillatory activity," *PLOS ONE*, vol. 20, no. 5, p. e0320519, May 2025.

[21] E. Coutinho and N. Dibben, "Psychoacoustic cues to emotion in speech prosody and music," *Cognition and Emotion*, vol. 27, no. 4, pp. 658–684, Jun. 2013.

[22] C. Oh, R. J. Morris, and X. Wang, "A Systematic Review of Expressive and Receptive Prosody in People With Dementia," *Journal of Speech, Language, and Hearing Research*, vol. 64, no. 10, pp. 3803–3825, Oct. 2021.

[23] X. Fan, E. Tang, M. Zhang, Y. Lin, H. Ding, and Y. Zhang, "Decline of Affective Prosody Recognition With a Positivity Bias Among Older Adults: A Systematic Review and Meta-Analysis," *Journal of Speech, Language, and Hearing Research*, vol. 67, no. 10, pp. 3862–3879, Oct. 2024.

[24] Y. Lin, X. Ye, H. Zhang, F. Xu, J. Zhang, H. Ding, and Y. Zhang, "Category-Sensitive Age-Related Shifts Between Prosodic and Semantic Dominance in Emotion Perception Linked to Cognitive Capacities," *Journal of Speech, Language, and Hearing Research*, vol. 67, no. 12, pp. 4829–4849, Dec. 2024.

[25] 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.

[26] J. Amlerova *et al.*, "Emotional prosody recognition is impaired in Alzheimer's disease," *Alzheimer's Research & Therapy*, vol. 14, no. 1, p. 50, Apr. 2022.

[27] A. Wong *et al.*, "Montreal Cognitive Assessment: One Cutoff Never Fits All," *Stroke*, vol. 46, no. 12, pp. 3547–3550, Dec. 2015.

[28] 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," *American Journal of Alzheimer's Disease & Other Dementias*, vol. 36, 2021.

[29] M. M. Bradley and P. J. Lang, "Measuring emotion: The Self-Assessment Manikin and the Semantic Differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994.

[30] 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.

[31] A. Gramfort *et al.*, "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroscience*, vol. 7, p. 267, Dec. 2013.

[32] 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.

[33] C. Brodbeck *et al.*, "Eelbrain, a Python toolkit for time-continuous analysis with temporal response functions," *Elife*, vol. 12, Nov. 2023.

[34] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models using lme4," Jun. 2014.

[35] 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.

[36] R. V. Lenth, *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*, 2025.