

Towards an understanding of prosodic cue weighting for turn-end classification in older adults with varying hearing abilities

Lorenza Zaira Curetti¹, Hae-Sung Jeon², Lauren V. Hadley¹

¹ Hearing Sciences - Scottish Section, School of Medicine, University of Nottingham, Glasgow, UK

² School of Psychology and Humanities, University of Lancashire, Preston, UK

zaira.curetti@nottingham.ac.uk, HJeon1@lancashire.ac.uk, lauren.hadley1@nottingham.ac.uk

Abstract

In conversation, interlocutors use prosody to identify when a partner's utterance is complete. Sensorineural hearing loss may limit access to relevant prosodic cues or alter their weighting. This exploratory study examined if older listeners with hearing loss differ from those with typical hearing in their use of temporal, intensity, and pitch-related information to judge utterance completion. Participants classified declarative questions as *finished* or *continuing* based solely on prosody. The acoustic properties of the utterances were quantified in terms of boundary strength (final word duration and intensity) and pitch movement (duration and excursion of final rise). Both listener groups classified utterances above chance. However, to identify *finished* utterances, listeners with typical hearing relied on duration and intensity while those with hearing loss relied on pitch movement. We suggest that hearing loss reshapes the prosodic dimensions listeners rely on for turn-end perception.

Index Terms: turn-end prediction, prosody, hearing loss

1. Introduction

Smooth conversational turn-taking relies on a listener's ability to predict when a speaker's turn will end [1]. This prediction is largely mediated by the perception of phrase boundaries, where prosodic cues signal whether an utterance is complete or projects further talk [2]. More variable turn taking timing for interlocutors with hearing loss could indicate a reduced ability to interpret these cues [3]. In the present study, we therefore examine how older adults with and without hearing loss use prosodic cues to determine whether an utterance has finished.

Prosodically, phrase boundaries are marked by a multitude of acoustic cues, including pitch movements (e.g., rises at the end of declarative questions), final lengthening, and intensity modulation [4], [5], [6]. In English, strong intonational phrase (IP) boundaries typically show greater phrase-final lengthening and weakening (i.e., increased duration and reduced intensity), than weaker intermediate phrase (ip) boundaries [4], [5], [6], [7]. Pitch movements further support such boundary distinction. Contour shape and excursion are primary cues to boundary contrasts in English [8], with utterances approaching an end often exhibiting final pitch movements that differ systematically from those that are continuing (e.g., longer or larger rises for complete declarative questions). Together, such cues contribute to listeners' judgments of completion in perceptual tasks, helping them disambiguate turn ends from continuations, e.g., [1].

Hearing loss may alter the perception and thus use of various prosodic cues, with reduced spectral resolution and altered dynamic range affecting access to both pitch and intensity information [9]. Indeed hearing loss has been linked to reduced prosodic sensitivity for emotion recognition [10], as well as to intonational and stress contrasts [11]. Yet hearing loss does not affect all acoustic dimensions equally: while reduced spectral resolution and impaired temporal fine structure processing can limit sensitivity to pitch, sensitivity to slower temporal-envelope cues is often relatively preserved [9], [12].

Furthermore, perceptual decisions about turn completion can emerge from a variety of acoustic dimensions, and how listeners weight these dimensions depends on their perceptual reliability. For example, when a cue is uninformative or highly variable, listeners may dynamically reweight their relative contribution in favour of cues that are the most consistent [13], [14]. Evidence suggests that listeners with hearing loss and cochlear implant users can shift cue-weighting strategies to facilitate speech perception [15], [16], [17], including relying more on temporal-envelope information and down-weighting fine pitch variation. Thus, hearing loss may not simply reduce the ability to use prosodic information; it may alter the acoustic dimensions listeners use to make perceptual judgments.

Turn-taking offers a particularly informative context for examining potential shifts in cue weighting, as predicting turn completion requires rapid integration of multiple prosodic dimensions in real time. Although conversational timing differences have been documented in individuals with hearing loss [3], the perceptual basis of these differences remains unclear. If turn-ends can be signalled through temporal, amplitude, and pitch-related information, then limited access to any such cues may either lead to poorer identification of turn ends, or to changes in how turn ends are identified.

In this exploratory study, we address this gap by directly testing how hearing ability affects the use of different prosodic cues for classifying turn ends. Listeners with differing hearing abilities judged whether short utterances were complete or were going to continue. The acoustic properties of the stimuli were parameterised along dimensions associated with utterance-final boundary strength (final word duration and intensity), and pitch movement (duration and excursion of the final pitch rise). By linking perceptual classifications to these acoustic dimensions, we aimed to assess not only whether classification behaviour differs across groups, but also whether any differences reflect shifts in the relative weighting of prosodic cues.

2. Methods

2.1. Participants

Participants were recruited online via Prolific (prolific.co). A total of 112 met the inclusion criteria (aged 55–75, English nationality, native English speaker, resident in England), and passed data-quality screening (see Section 2.3). Of these, 55 had typical hearing (TH; 27 male; mean age 61.3 yrs) and 57 had hearing loss (HL; 32 male; mean age 60.5 yrs). The HL group included 14 bilateral hearing aid users (average 9.1 yrs of use; range 1–35 yrs). Participants completed the study unaided.

2.2. Stimuli

The materials comprised 40 items, each realised in two prosodic versions (*finished* and *continuing*), yielding 80 recorded stimuli in total. All utterances were declarative questions recorded by a male native speaker of Standard Southern British English, who had expertise in intonational analysis and read the questions at a comfortable speaking rate.

Each item contained a one- to three-syllable target word stressed on the first syllable. In the *finished* stimuli the target word occurred in utterance-final position (e.g., *She plays football?*), whereas in the *continuing* stimuli it occurred at a continuation point (e.g., *She plays football for England?*). The *continuing* stimuli were truncated after the target word to make them semantically identical to the *finished* ones while preserving their original prosodic properties. Stimulus versions were designed to contrast two prosodic properties: (1) the strength of the prosodic boundary at the right edge of the target word, and (2) the trajectory of pitch movement across the target word. For *finished* items, the target word was aligned with the right edge of an IP, whereas in *continuing* items it was aligned with the right edge of an ip and showed no substantial final lengthening or pause.

In both conditions, the speaker kept the intonation of the words preceding the target word as consistent as possible, maintaining a gently rising intonation ('She plays' in Fig. 1), or all stimuli, the stressed syllable of the target word was pitch accented, carrying a fall-rise f_0 contour; the contour shape was held constant within each *finished-continuing* pair. In the *finished* versions, the accent was followed by a steeper and more immediate rise (end of 'football' in Fig. 1A). In the *continuing* versions, by contrast, this accent was followed by a rise forming a plateau or gentle upward trajectory, leading into the utterance-final rise ('for England' in Fig. 1B). The speaker aimed to produce consistent prosodic differences between the *finished-continuing* stimuli across all pairs. All stimuli were loudness-normalised to -23 Loudness Units Relative to Full Scale (per [18]), and gated with 5 ms onsets and offsets to remove edge transients.

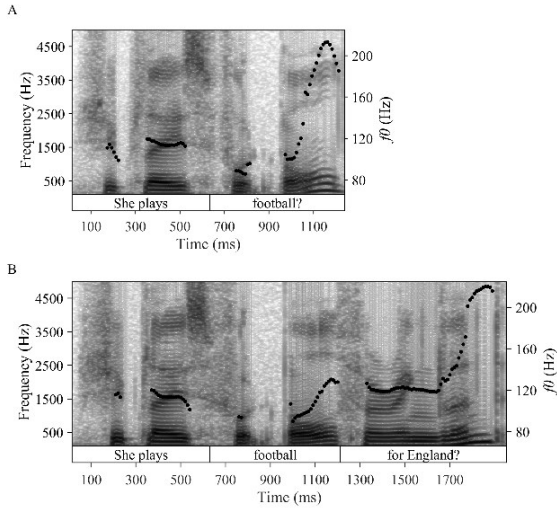


Figure 1: Spectrogram and f_0 contours for a (A) *finished* and (B) *continuing* stimulus.

2.3. Procedure

The study was part of a larger online study comprising two sessions (~20 min each). This data was collected in the first session. Participants were asked to wear headphones and set volume to a comfortable level at the start of the experiment.

2.3.1. Digit Triple Test (DTT)

Hearing was assessed using an online DTT [19], in which listeners listen to and then recall quasi-random digit triplets presented in speech-shaped noise. Digits ranged from zero (spoken as 'oh') to nine, excluding the disyllabic 'seven'. There were 120 possible triplets, selected at random. The signal-to-noise ratio (SNR) was adapted towards an individual's threshold following a one-up/one-down procedure across 25 trials, starting at -16 dB and ranging from -25 to -6 dB.

Participants entered their responses using an on-screen keypad. The DTT score was defined as the mean SNR over trials 7-25.

Participants were classified as TH or HL based on a DTT score cutoff of -17.1, corresponding approximately to a 20 dB high-frequency pure-tone loss [19]. The median DTT score for the whole sample was -16.9 dB (range -24.3 to -6 dB), confirming balanced groups despite variation in listening setup.

2.3.2. Classification task

The classification task comprised three practice trials followed by 44 experimental trials. Of these, 40 were main trials, and four were catch trials that served as audibility/attention checks.

All trials required participants to press a Start button to begin stimulus presentation. The main trials included 20 *finished* and 20 *continuing* stimuli. Stimulus version (*finished/continuing*) was counterbalanced such that each participant heard only one version of each target word. Trial order was randomised, with the constraint that no more than three *finished/continuing* stimuli could occur consecutively. After stimulus playback, participants had to judge whether the talker sounded like "they are going to continue talking (you're only hearing the first part of a longer sentence that has been cut), or whether they are finished". They indicated their choice by selecting: 'Talker will continue' or 'Talker is finished'. The stimuli could be replayed without limit.

Catch trials occurred every seven main trials, and differed only by including a 200 ms 1 kHz tone, for which participants were instructed to press a 'beep' button. The same catch stimuli were used for all participants, with version (2 *finished*, 2 *continuing*) and order randomised. Only participants that identified three out of the four catch trials were retained.

2.3.3. Prosodic measures

Four acoustic measures were extracted from each stimulus to capture the key prosodic differences between *finished* and *continuing* stimuli:

- T_{tw} : the duration of the target word (ms).
- dB_{tw} : the mean intensity of the target word (dB).
- T_{rise} : the time from the onset of the final upward pitch movement to its maximum (ms).
- ST_{rise} : the excursion in T_{rise} (semitones ST).

T_{tw} and dB_{tw} were selected as acoustic correlates of boundary strength at the right edge of the target word [4], [5]. Since intonational movement can extend beyond lexical word boundaries, the target word's edges were adjusted where necessary to ensure the relevant pitch contour was fully included. T_{tw} and dB_{tw} were then extracted from these adjusted intervals using ProsodyPro in Praat [20], [21]. To characterise pitch movement, we used T_{rise} (the duration of the utterance's final rise) and ST_{rise} (its pitch excursion).

Table 1: Medians (Mdn), Interquartile Ranges (IQR), and Spearman's correlations (ρ) for the prosodic measures.

	Mdn (IQR)	dB_{tw}	T_{tw}	T_{rise}	ST_{rise}
<i>finished</i>					
dB_{tw} (dB)	70.9 (1.7)	-	-0.45**	-0.15	0.06
T_{tw} (ms)	644 (136)		-	0.35*	0.13
T_{rise} (ms)	261 (127)			-	0.14
ST_{rise} (ST)	14.7 (2.2)				-
<i>continuing</i>					
dB_{tw} (dB)	71.1 (1.7)	-	-0.15	-0.06	-0.51**
T_{tw} (ms)	562 (122)		-	0.56**	0.42**
T_{rise} (ms)	212 (111)			-	0.37**
ST_{rise} (ST)	5.1 (1.8)				-

* $p < 0.01$, ** $p < 0.001$

Descriptive statistics for all prosodic measures, along with their inter-correlations, are summarised in Table 1. As expected, T_{tw} and dB_{tw} were negatively correlated for *finished* items, consistent with the known co-occurrence of phrase-final lengthening and intensity shifts [6]. *Finished* items also exhibited clear final-word lengthening (median difference ~ 80 ms), and substantially larger pitch excursions (median difference ~ 9 ST) than *continuing* items, aligning with descriptions of IP (rather than ip) boundaries in English [4], [8]. The correlation between T_{rise} and ST_{rise} for *continuing* items likely reflects the tendency for longer rises to involve larger pitch excursions [8].

2.3.4. Statistical analysis

Performance was evaluated separately at the participant and item levels. First, sensitivity and response bias were estimated separately for TH and HL using the *dprime* function of the *psycho* package in R v4.5.2 [22], [23]. Because the task required listeners to detect prosodic boundary cues, *finished* stimuli (with stronger boundaries) were treated as signal and *continuing* stimuli as noise. Group differences were assessed using Mann-Whitney U tests, and performance relative to chance using Wilcoxon signed-rank tests against zero, implemented with the *wilcox.test* function.

Second, to assess which prosodic measures predicted listeners' performance, we fitted rank-based regression models using the *Rfit* package [24], separately for each listener group (TH, HL) and stimulus version (*finished*, *continuing*). Cue use within each listener group was assessed separately due to limited power for higher-order interaction terms. *Finished* and *continuing* stimuli were analysed separately as they differ systematically in their prosodic structures, and modelling them together would risk masking condition-specific effects.

The full model included predictors from both prosodic dimensions: $accuracy \sim dB_{tw} \times T_{tw} + T_{rise} \times ST_{rise}$. Accuracy was defined as the proportion of correct responses per item, and all prosodic measures were z-transformed prior to modelling. Interactions were included to account for non-additive effects within each prosodic dimension. Although some predictors were moderately correlated (see Table 1), none exceeded conventional thresholds for problematic multicollinearity ($p > 0.8$), and variance inflation factors from equivalent linear models were all below 1.8. All predictors were therefore retained in the initial models.

Model selection proceeded via backward elimination using drop-in-dispersion tests using the *drop.test* function. The final models therefore retained only predictors whose removal significantly reduced model fit.

Where significant interactions were found, simple slopes analyses were conducted following [25]. This involved re-centring one predictor at ± 1 standard deviation from the mean. The model was then re-estimated to evaluate the effect of the primary predictor at lower and higher levels of the interacting variable. Statistical significance was evaluated at $\alpha = 0.05$.

3. Results

3.1.1. Sensitivity and response bias

Sensitivity did not significantly differ between groups ($p = 0.12$), and both groups performed above chance (TH: $d' = 0.28$, $p < 0.001$; HL: $d' = 0.10$, $p = 0.03$), with mean accuracy levels of 61% and 54%, respectively.

Response bias also did not differ between groups ($p = 0.74$). However, both groups showed a small but significant tendency to identify utterances as *finished* (TH: $C = -0.08$, $p = 0.016$; HL: $C = -0.09$, $p = 0.015$).

3.1.2. Item-level predictors

For *continuing* stimuli, no predictors reached significance in either group (all $p \geq 0.32$), and overall model fit was weak (TH:

$R^2 = 0.07$; HL: $R^2 = 0.09$), indicating that the prosodic measures did not reliably predict responses in this condition.

By contrast, both group's *finished* models resulted in moderate fit (TH: $R^2 = 0.28$, $p = 0.007$; HL: $R^2 = 0.24$, $p = 0.005$), but incorporated different prosodic cues (see Table 2).

Table 2: Rank-based regression results for *finished* stimuli for TH and HL. Predictors were z-standardised.

Predictor	β	SE	t-stat	p-value
TH ($R^2 = 0.28$)				
dB_{tw}	-0.033	0.013	-2.36	0.023
T_{tw}	0.016	0.014	1.17	0.248
$dB_{tw} \times T_{tw}$	0.036	0.012	2.88	0.007
HL ($R^2 = 0.24$)				
T_{rise}	-0.032	0.012	-2.54	0.015
ST_{rise}	-0.028	0.012	-2.27	0.029

For TH listeners (final model: $accuracy \sim dB_{tw} \times T_{tw}$), we observed a significant interaction between dB_{tw} and T_{tw} ($\beta = 0.036$, $p = 0.007$), alongside a negative main effect of dB_{tw} ($\beta = -0.033$, $p = 0.023$). Simple slopes analysis revealed that at lower T_{tw} (-1 SD), dB_{tw} was a significant negative predictor of accuracy ($\beta = -0.069$, $p = 0.002$), such that accuracy decreased (approximately from 0.8 to 0.5 in Fig. 2) as dB_{tw} increased. By contrast, at higher T_{tw} ($+1$ SD), the effect of dB_{tw} was not significant ($\beta = 0.003$, $p = 0.878$), and accuracy remained relatively high (~ 0.7 in Fig. 2). These results indicate that the effect of dB_{tw} varied as a function of word duration, with high dB_{tw} hindering performance when target words were shorter, but having no impact when words were sufficiently lengthened.

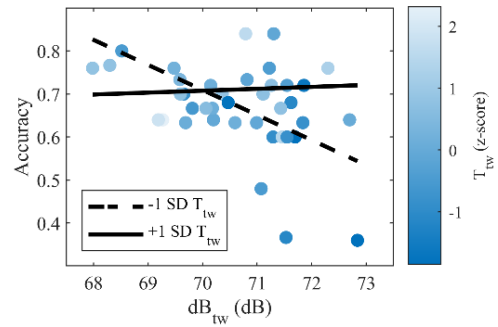


Figure 2: Relationship between dB_{tw} and accuracy for TH, shaded by T_{tw} (z-scored), with regression lines for low (-1 SD) and high ($+1$ SD) T_{tw} .

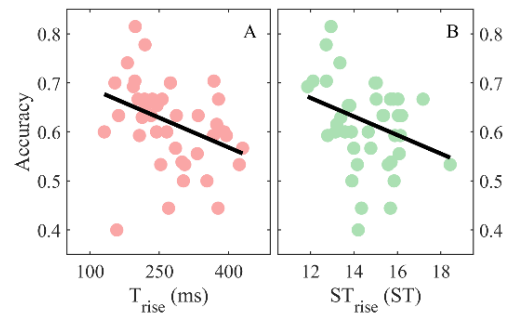


Figure 3: Relationship between (A) T_{rise} or (B) ST_{rise} and accuracy for HL. Lines represent regression estimates.

Conversely, for HL listeners accuracy was predicted by rise measures only (final model: $accuracy \sim T_{rise} + ST_{rise}$). Both T_{rise} ($\beta = -0.032$, $p = 0.015$) and ST_{rise} ($\beta = -0.028$, $p = 0.029$) showed negative effects, suggesting higher accuracy for stimuli with shorter final rises and smaller pitch excursions (Fig. 3).

4. Discussion

This study examined how older adult listeners with differing hearing abilities classify utterances as finished or continuing based solely on prosodic information. Listeners heard naturally produced declarative questions ending at an IP boundary (*finished*), or semantically identical utterances truncated at an ip boundary (*continuing*). We assessed overall sensitivity (d') to acoustic markers of (1) boundary strength at the right edge of the utterance-final target word (the target word's duration T_{tw} and intensity dB_{tw}), and (2) the trajectory of pitch movement across the target word (the final rise's duration T_{rise} and pitch excursion ST_{rise}). We then examined how these cues were associated with classification accuracy within each group.

Both groups showed above-chance sensitivity, indicating that prosodic cues alone were sufficient to support classification. Additionally, both groups showed a slight bias towards classifying items as *finished*. Given that IP boundaries are typically marked by stronger and more consistent acoustic cues than ip boundaries [6], [7], this bias is not surprising in a decontextualized listening task. In the absence of lexical, syntactic, or pragmatic context, listeners may be more inclined to treat a clearly marked prosodic contour as complete (since it did indeed end at that point). Under such conditions, listeners may primarily be scanning for completion cues, rather than cues specific to continuation (which are more subtle). This lack of use of a specific 'completion template' may also explain why no single acoustic measure consistently predicted *continuing* responses. Crucially, however, despite comparable sensitivity and bias between groups, they differed in the acoustic dimensions predicting their responses. This suggests that rather than hearing loss leading to a general loss of sensitivity to prosodic information, it alters the prosodic cues listeners use.

For TH listeners, dB_{tw} showed a significant negative main effect, with lower intensity values increasing the likelihood of correctly identifying *finished* items. This is consistent with descriptions of phrase-final weakening, where intensity often trails off towards the end of complete utterances [6]. Interestingly, however, intensity did not operate independently of duration. The significant $dB_{tw} \times T_{tw}$ interaction indicates that the effect of intensity was modulated by the duration of the target word. For longer target-word durations (solid line in Fig. 2), the effect of intensity was absent. However, for shorter target-word durations (dashed line in Fig. 2), increased intensity was associated with a substantial drop in *finished* accuracy.

One plausible interpretation is that listeners treated duration and intensity as complementary cues to boundary strength. Final lengthening is among the most robust correlates of prosodic boundaries [4], [5]. Thus, substantial word-final lengthening may have served as a sufficiently strong cue to completion, reducing the need to rely on intensity. A similar interdependence has been reported in prominence perception, where duration and intensity jointly contribute to stress judgments, but intensity rarely functions as a reliable cue on its own [26]. In contrast, it was when the target words were shorter that listeners relied more strongly on intensity, with lower intensity values (consistent with a phrase-final 'trailing off') supporting *finished* judgments. This pattern aligns with models of speech perception in which multiple acoustic dimensions are integrated probabilistically, and listeners increase reliance on secondary cues when primary cues are weak or ambiguous [13], [14]. This interpretation may also explain why pitch-rise measures did not significantly predict TH listeners' responses, since in other prosodic contexts (e.g., statements) a larger or sharper rise can reflect a continuation cue [7], [8]. The ambiguity of this cue may thus have led to down-weighting.

Unlike TH listeners, HL listeners did not show sensitivity to intensity or duration differences. While the former may have been expected, the latter was more surprising. Reduced dynamic range and altered loudness growth associated with sensorineural hearing loss may diminish the perceptual salience of amplitude-based cues [9], leading to down-weighting of intensity cues in favour of alternative dimensions. However, the absence of a durational effect was unexpected given that sensitivity to slower temporal-envelope cues (such as overall word duration) is often relatively preserved in sensorineural hearing loss [9]. A speculative explanation for such lack of reliance on these cues could be that both word lengthening and phrase-final weakening require an internalised expectation against which the heard input can be compared (i.e., a prediction of how long or loud the word *should have been*). This is a relatively complex cognitive process, as a word can only be judged to be lengthened by accounting for both the speech rate and syllable count, and weakened based on the prior speech level (rather than against absolute thresholds). If access to, or confidence in, this internal model of speech is weakened, such cues may be perceived as less informative, leading to reliance on dimensions that can be evaluated more directly [13], [14].

For HL listeners, it was the pitch slope metrics that predicted accuracy of classifying *finished* items. Pitch movement can be seen as a cue that can be evaluated more directly than word-lengthening/phrase-final weakening, since it involves tracking a dynamic trajectory rather than comparing input against an internalised expectation. Despite well-documented reductions in spectral resolution in sensorineural hearing loss [9], dynamic fundamental frequency contours can remain perceptually accessible even when fine spectral detail is degraded, provided that low-rate envelope cues are preserved [27]. HL listeners may therefore have strategically shifted towards pitch movement not because of a generalised inability to process T_{tw} , but because T_{rise} or ST_{rise} offered a more comparatively robust cue.

Notably however, we were surprised to find a negative relationship between T_{rise} or ST_{rise} and *finished* responses, suggesting that the faster or smaller rises were related to greater accuracy. A possible explanation for this finding is that exaggerated pitch movements could be less reliable with hearing loss due to known distortions in frequency and loudness, making moderate rises comparatively easier to use. Clarifying this pattern will require follow-up work that systematically manipulates rise characteristics.

While our findings indicate substantial differences in turn-end classification approaches between listeners depending on hearing ability, it is important to recognise several limitations. First, the study was conducted online, limiting control over listening environments. Second, groups were classified based on threshold criteria rather than a continuous metric, potentially obscuring graded relationships between hearing and cue weighting. Third, accuracy in this task was only modestly above chance, which may have reduced sensitivity to smaller effects and limited the stability of some estimates. Nevertheless, the systematic differences in cue weighting between groups provide a meaningful starting point for further investigations, which could both further our understanding of speech processing and suggest future targets for hearing technology.

5. Conclusions

This study provides evidence that hearing abilities may affect the weighting of prosodic cues for turn-end judgments. Using a classification task, we found indications that TH listeners relied primarily on duration and intensity cues associated with boundary strength, while HL listeners showed greater reliance on pitch movement. These findings suggest that hearing loss may alter the relative contribution of acoustic dimensions even when overall performance remains above chance, highlighting the need to consider cue-weighting differences when assessing communication difficulties and developing interventions.

6. Generative AI Disclosure

The authors used generative AI solely to support language refinement during manuscript preparation, but content, analysis, and conclusions are entirely the authors' own.

7. References

- [1] S. C. Levinson and F. Torreira, 'Timing in turn-taking and its implications for processing models of language', *Front. Psychol.*, vol. 6, Jun. 2015, doi: 10.3389/fpsyg.2015.00731.
- [2] S. Bögels and F. Torreira, 'Turn-end Estimation in Conversational Turn-taking: The Roles of Context and Prosody', *Discourse Processes*, vol. 58, no. 10, pp. 903–924, Nov. 2021, doi: 10.1080/0163853X.2021.1986664.
- [3] A. J. M. Sørensen, T. Lunner, and E. N. MacDonald, 'Conversational Dynamics in Task Dialogue Between Interlocutors With and Without Hearing Impairment', *Trends in Hearing*, vol. 28, p. 23312165241296073, Sep. 2024, doi: 10.1177/23312165241296073.
- [4] M. Wagner and D. G. Watson, 'Experimental and theoretical advances in prosody: A review', *Language and Cognitive Processes*, vol. 25, no. 7–9, pp. 905–945, Sep. 2010, doi: 10.1080/01690961003589492.
- [5] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price, 'Segmental durations in the vicinity of prosodic phrase boundaries', *J. Acoust. Soc. Am.*, vol. 91, no. 3, pp. 1707–1717, Mar. 1992, doi: 10.1121/1.402450.
- [6] A. Gravano and J. Hirschberg, 'Turn-taking cues in task-oriented dialogue', *Computer Speech & Language*, vol. 25, no. 3, pp. 601–634, Jul. 2011, doi: 10.1016/j.csl.2010.10.003.
- [7] J. B. Pierrehumbert and J. Hirschberg, 'The Meaning of Intonational Contours in the Interpretation of Discourse', in *Pronunciation*, Routledge, 2017, pp. 256–295.
- [8] D. R. Ladd, *Intonational Phonology*, 2nd edn. in Cambridge Studies in Linguistics. Cambridge: Cambridge University Press, 2008. doi: 10.1017/CBO9780511808814.
- [9] B. C. J. Moore, *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*. John Wiley & Sons, 2007.
- [10] J. A. Christensen, J. Sis, A. M. Kulkarni, and M. Chatterjee, 'Effects of Age and Hearing Loss on the Recognition of Emotions in Speech', *Ear and Hearing*, vol. 40, no. 5, p. 1069, Oct. 2019, doi: 10.1097/AUD.0000000000000694.
- [11] T. Most, 'Perception of the Prosodic Characteristics of Spoken Language by Individuals With Hearing Loss', in *Oxford Handb. Deaf Stud. Lang.*, M. Marschark and P. E. Spencer, Eds, Oxford University Press, 2015, pp. 79–93. doi: 10.1093/oxfordhb/9780190241414.013.6.
- [12] B. C. J. Moore, S. Mariathasan, and A. P. Şek, 'Effects of Age and Hearing Loss on the Discrimination of Amplitude and Frequency Modulation for 2- and 10-Hz Rates', *Trends in Hearing*, vol. 23, p. 2331216519853963, Jan. 2019, doi: 10.1177/2331216519853963.
- [13] L. L. Holt and A. J. Lotto, 'Cue weighting in auditory categorization: implications for first and second language acquisition', *J Acoust Soc Am*, vol. 119, no. 5 Pt 1, pp. 3059–3071, May 2006, doi: 10.1121/1.2188377.
- [14] J. C. Toscano and B. McMurray, 'Cue Integration With Categories: Weighting Acoustic Cues in Speech Using Unsupervised Learning and Distributional Statistics', *Cognitive Science*, vol. 34, no. 3, pp. 434–464, 2010, doi: 10.1111/j.1551-6709.2009.01077.x.
- [15] P. Souza, F. Gallun, and R. Wright, 'Contributions to Speech-Cue Weighting in Older Adults With Impaired Hearing', *J Speech Lang Hear Res*, vol. 63, no. 1, pp. 334–344, Jan. 2020, doi: 10.1044/2019_JSLHR-19-00176.
- [16] S.-C. Peng, M. Chatterjee, and N. Lu, 'Acoustic Cue Integration in Speech Intonation Recognition With Cochlear Implants', *Trends in Amplification*, vol. 16, no. 2, pp. 67–82, Jun. 2012, doi: 10.1177/1084713812451159.
- [17] M. Chatterjee *et al.*, 'Age-Related Changes in Acoustic Cue Weighting for Emotional Prosody Identification by Adult Listeners', *Ear & Hearing*, vol. 47, no. 2, pp. 558–575, Mar. 2026, doi: 10.1097/AUD.0000000000001755.
- [18] European Broadcasting Union, *Loudness Normalisation and Permitted Maximum Level of Audio Signals*, EBU R 128, 2014.
- [19] M. S. M. G. Vlaming, R. C. MacKinnon, M. Jansen, and D. R. Moore, 'Automated screening for high-frequency hearing loss', *Ear Hear*, vol. 35, no. 6, pp. 667–679, 2014, doi: 10.1097/AUD.0000000000000073.
- [20] Y. Xu, 'ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis', in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 2013, pp. 7–10.
- [21] Boersma, Paul, 'PRAAT, a system for doing phonetics by computer', *Glott International*, vol. 5, no. 9/10, pp. 341–347, 2001.
- [22] R Core Team, *R: A language and environment for statistical computing*. (2025). R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: <https://www.R-project.org/>
- [23] D. Makowski, 'The Psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science', *Journal of Open Source Software*, vol. 3, no. 22, p. 470, 2018, doi: 10.21105/joss.00470.
- [24] J. D. Kloeke and J. W. McKean, 'Rfit: Rank-based Estimation for Linear Models', *The R Journal*, vol. 4, no. 2, pp. 57–64, 2012, doi: 10.32614/RJ-2012-014.
- [25] L. S. Aiken and S. G. West, *Multiple regression: Testing and interpreting interactions*. Sage Publications, Inc, 1991, pp. xi, 212.
- [26] A. E. Turk and J. R. Sawusch, 'The processing of duration and intensity cues to prominence', *J Acoust Soc Am*, vol. 99, no. 6, pp. 3782–3790, Jun. 1996, doi: 10.1121/1.414995.
- [27] M. A. Stone, C. Füllgrabe, and B. C. J. Moore, 'Relative contribution to speech intelligibility of different envelope modulation rates within the speech dynamic range', *J Acoust Soc Am*, vol. 128, no. 4, pp. 2127–2137, Oct. 2010, doi: 10.1121/1.3479546.