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

Enhancing and reducing prosodic cues: A preliminary investigation of turn-end prediction in listeners with hearing loss

Lorenza Zaira Curetti[1], *Hae-Sung Jeon*[2], *Lauren V. Hadley*[1]

[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@uclan.ac.uk, mszlvh1@nottingham.ac.uk

Abstract

In comparison to normal hearing interlocutors, individuals with hearing loss often show delayed and more variable timing of spoken turns in conversation. Limited access to prosodic turn-end cues could contribute to such behaviour. This exploratory study investigates the impact of manipulated utterance-final *f0* contours (with an increased or reduced excursion size) on turn-end prediction for individuals with self-reported hearing loss. Using an online button-press task, we assessed speed and accuracy of predicting a spoken turn end, alongside subjective measures of listening effort and confidence. Results indicate that a decreased *f0* excursion, creating a gradual fall, improves both the speed and accuracy of turn-end prediction. On the other hand, acoustic enhancement of the peak associated with the utterance-final stressed syllable impairs turn-end prediction. These findings highlight that early prosodic cues aligned with linguistic structure facilitate turn-end prediction for individuals with hearing loss.

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

1. Introduction

Effective communication relies on smooth turn taking, with speakers timing their responses precisely to avoid interruptions [1], [2]. This precise timing is facilitated by various linguistic and prosodic cues, indicating the end of one speaker's turn and the opportunity for another to respond [3], [4].

For individuals with hearing loss, however, accessing these prosodic cues may be challenging due to degraded perceptual auditory input [5]. For example, although prosody can support older listeners' word recognition, age-related hearing loss limits the extent to which these cues are exploited [6]. Difficulties recognising emotional prosody serves as another example of such reduced prosodic sensitivity [7]. These difficulties may be further compounded by age-related central auditory processing deficits, which have been linked to impaired underlying functions such as fundamental frequency (*f0*) discrimination [8]. Together, these peripheral and central deficits may also impair recognition of prosodic cues to turn ends, potentially leading to difficulties in conversational engagement.

Indeed, individuals with hearing loss often struggle to participate in turn taking [9], [10]. Fernandez and colleagues recently proposed that such difficulties may partly originate from reduced use of prediction [11]. They suggest three ways that hearing loss could limit predictive processing: increased cognitive demands, the linguistic context being perceived poorly and thus offering a poor basis for prediction, or alternative speech processing strategies being prioritised (e.g., lip reading). Based on this perspective, supporting listeners' abilities to predict and/or detect turn ends could benefit individuals with hearing loss and ultimately facilitate conversational engagement.

This study tests two potential *f0* cues to explore their potential for intervention. First, previous research has shown that exaggerated *f0* contours can improve speech recognition for individuals with hearing loss, particularly in noise [12]. Therefore, a high *f0* peak in the turn-final word may attract listeners' attention and facilitate their word recognition, effectively enhancing turn-end prediction [13]. Second, prosodic cues associated with Intonational Phrase (IP) finality can facilitate turn-end prediction. A series of studies by Bögels and Torreira [15], [16], [17] demonstrated that listeners use IP boundary cues, i.e. lowering in *f0* and final lengthening, to predict turn ends. In their investigations, listeners, who were instructed to press a button at the end of a speaker's turn, were faster to respond to stimuli with intact prosody than to those truncated mid-IP, which were syntactically complete but lacked the IP finality cues. Listeners were also more likely to misinterpret a turn continuation as a turn end in the presence of the IP boundary (as in cross-spliced stimuli that applied IP-final prosody to the first part of a multi-phrase question).

In our experiment (Section 2), we used question utterances with a final *f0* fall. We used a button-press paradigm where participants with self-reported hearing loss were asked to press a button at the moment they expected the question to end, under natural or manipulated (*enhanced* and *reduced*) prosodic conditions. In the *enhanced* condition, the final *f0* peak was heightened, increasing its perceptual prominence, whereas in the *reduced* condition it was reduced, effectively lengthening the *f0* fall. Additionally, we used subjective measures of listening effort [17] and listening confidence [18] to estimate the cognitive impact of our manipulations.

In terms of hypotheses, we expected participants to respond faster (i.e., earlier) and more accurately (i.e., closer to the turn end) in the manipulated conditions, relative to the *natural* condition. If impaired predictive processing and cognitive demands are indeed interlinked for individuals with hearing loss [11], we also hypothesised that providing support for prediction through the prosodic manipulations would reduce the cognitive resources allocated to the listening task, in turn enhancing confidence in listening.

2. Methods

2.1. Participants

We used the Prolific platform (prolific.co) for recruitment, and 36 participants (21 female; mean age 61.3 yrs) were included in the study. Pre-screening required participants to be English nationals aged 55–75 yrs residing in England, with English as

 10.21437/SpeechProsody.2026-181

their first language, and with self-reported hearing difficulties. Five additional participants were recruited but removed for failing the screening checks (two reporting Welsh nationality, three reporting their hearing to be "very good" [19]), and two for outlying responses (see Section 3).

2.2. Stimuli

The stimuli consisted of 189 general knowledge *wh*-questions recorded by a young Southern English male talker, who was instructed to deliver them in a natural and conversational manner (see [20] for elicitation method). They included 144 critical one-part "target" questions (e.g., *What is the name of a <u>planet</u>?*), which ended in either a two- or three-syllable word stressed on the first syllable (N = 72 each), and featured utterance-final falling intonations, as is common for English *wh*-questions [21].

There were three experimental conditions (*natural*, *enhanced*, and *reduced*), differing in the shape of utterance-final *f0* contour (see Section 2.3). To ensure that conditions were counterbalanced across participants, target questions were divided into three lists, which were assigned in systematic rotation to the prosodic conditions. Participants therefore encountered a third of the targets in each condition, with each target occurring once. An equal number of targets ending in a two- or three-syllable word were included in each list.

A further 45 two-part "filler" questions (e.g., *Which insects typically live in a <u>colony</u>, and have a queen?*) were added to the lists (N = 15 per list). Fillers included a turn continuation, to introduce intonational variability. Only naturally produced intonation was used for the filler questions.

To ensure all questions were of comparable perceived difficulty, we pre-tested 20 additional participants from the same population (7 female; mean age 64.3 yrs), who verbally answered and rated the *wh*-questions on a 7-point Likert-type scale. Only questions with response times (RTs) and ratings within the median range (interquartile range method, k = 1.5) were included in the final pool.

2.3. Prosodic conditions

Prosodic manipulations were performed exclusively on the *f0* contour of the target questions using an extraction-resynthesis procedure [22]. Prior to resynthesis, *f0* was extracted for all target questions using Praat v6.4.26 [23], through a two-round adaptive procedure involving an initial estimation of the *f0* contour over a broad frequency range of 75–600 Hz, followed by a second extraction over a narrower range, adjusted based on the *f0* properties of the initial estimate. This reduces octave errors, producing a smoother and more reliable *f0* contour.

Following *f0* extraction, we employed a user-guided procedure to detect the onset of the target's final word, based on an auditory and visual inspection of the waveform, spectrogram, and *f0* contour. An iterative peak detection algorithm was then used to localise the *f0* peak associated with the stressed syllable of the final word. This was defined as the maximum *f0* within the final word, with adjacent *f0* values (one point before and one point after) within a tolerance of ±8 Hz. The *f0* contours were then manipulated in MATLAB as discussed below, and resynthesised using Praat.

To ensure no unintended biases, we conducted a second pre-test and asked 19 additional participants from the same population (6 female; mean age 64.3 yrs) to rate the naturalness of all stimuli, including those which were not resynthesised, on a 1-9 Likert-type scale, with 1 being very natural and 9 very unnatural. A 95% equivalence testing of mean ratings (with a 10% equivalence zone) confirmed all conditions were perceived as equally natural (mean ratings: *natural* = 2.1, *enhanced* = 2.1, *reduced* = 2.4).

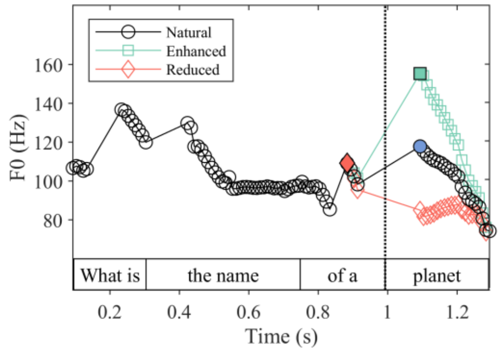

Figure 1: *f0 contours for the natural (black circles), enhanced (green squares), and reduced (red diamonds) conditions. The vertical dotted line indicates final word onset; filled markers mark the final f0 peaks of each condition. High f0 values for 'the' and 'of a' were caused by segmental perturbation.*

2.3.1. *Natural*

The *natural* condition featured unaltered *f0* contours, and served as the baseline for the manipulated conditions. The *natural* questions had a high pitch accent (H*) associated with the utterance-final word ('planet', Fig. 1) and a gradual *f0* fall.

2.3.2. *Enhanced*

The *enhanced* condition was designed to increase the prominence of the *f0* peak of the utterance-final stressed syllable by 5 semitones relative to its original value. The *f0* contour around the peak was gradually ramped up over 200 ms (or from the first sampled point > 200 ms), and down over 300 ms (or until the end of the question). This manipulation created a high pitch accent sounding more emphatic compared to the natural counterpart (Fig. 1, green squares).

2.3.3. *Reduced*

For the *reduced* condition, the final *f0* peak of the high pitch accent was reduced to a value only 10% above the final *f0* value (i.e., that at the end of the question). Values between the reduced *f0* peak and the utterance-final *f0* were proportionally reduced to preserve their relative fluctuations, but were kept no lower than 1% below the final *f0* value to avoid excessive drops in pitch. The *f0* contour preceding the reduced peak was smoothly ramped down over 200 ms (or from the first sampled point > 200 ms). The resynthesised stimuli preserved a slight *f0* rise before the utterance-final word ('of a' in Fig. 1, red diamonds) and a high accent (a downstepped H*) associated with the utterance-final word, but perceived pitch was close to the low level of the speaker's pitch range. The slight rise did not create a perceived pitch accent before the utterance-final word. Instead, the manipulation created an earlier start point for the utterance-final *f0* fall, which now began before the final word as opposed to from the stressed syllable of the final word (as in the *natural* and *enhanced* conditions).

2.4. Procedure

The NOTE platform (surveys.hearing.nottingham.ac.uk) was used to create and host the experiment online. Prior to starting the experiment, participants were instructed to set their loudspeakers to a comfortable level, and to ensure they were in a quiet environment. The use of headphones was not allowed.

The experiment comprised two repeatable practice trials, and three experimental blocks of 63 trials each. Each block included 16 targets per prosodic condition, and 15 fillers, all randomly selected from the pre-defined lists of stimuli (see Section 2.2). The order of presentation in each block was randomised, except for the first trial, which always featured a filler question to account for potential initial timing inaccuracies, common in online testing.

Each trial began at the press of a Start button. A 1 kHz tone lasting 100 ms, followed by a 400 ms silence, signalled question onset. Participants were instructed to listen carefully to the question, and to press the Stop button via a mouse click "as soon as you expect the speaker to be finished talking" [24]. The audio continued to play until the end of the question, to prevent participants from hesitating to press the button early for fear of interrupting the question. Participants were then asked to rate their listening effort (*How hard did you work to understand what was said?*) [25], and listening confidence (*How confident are you that you heard everything correctly?*) [26]. Both ratings were made on Likert-type scales ranging from 1 (*Very hard/Not at all confident*) to 10 (*Not at all hard/Very confident*). To ensure speech comprehension and attentiveness, participants were also asked to type their answers to the questions in a text box. Their typed answers, however, were not included in any formal analysis.

3. Results

Confidence and effort ratings were at ceiling across all conditions, likely due to the relatively favourable listening conditions (speech-in-quiet). Therefore, our analysis focused solely on RTs, estimated as the difference between question offset and button press (i) as a signed value indicating early vs late responses (speed), and (ii) as an absolute value indicating pure distance from offset (accuracy). Note that speed and accuracy are complementary aspects of prediction, reflecting (i) how early one commits to a prediction and (ii) how closely that prediction aligns to the actual turn end.

Analyses were conducted on RT medians to reduce the influence of skewed distributions. Extreme outliers were identified using the interquartile range method (k = 3 where k is the rank), and any participant or trial detected were removed prior to subsequent analyses. We evaluated the effect of the manipulations on RTs with a Friedman Test, a non-parametric approach robust to outliers and non-normality, using the *friedman.test* function of the *stats* package in R v4.2.2 [27]. Wilcoxon signed-rank tests were used for follow-up analyses and run using the *wilcox.test* function. Effects are reported with Bonferroni-adjusted p-values ($\alpha = 0.017$).

3.1. Response speed and accuracy

Consistent with our hypotheses, RTs statistically differed between prosodic conditions for both speed ($\chi^2(2) = 6.17$, p = 0.046) and accuracy ($\chi^2(2) = 17.06$, p < 0.001). For speed, post-hoc pairwise comparisons revealed a significant effect (see Fig. 2, left), with median RTs of 215 ms for the *natural* condition, 224 ms for the *enhanced* condition, and 193 ms for the *reduced* condition. RTs were significantly slower in the *enhanced* condition compared to the *reduced* condition (Z = -2.94, p = 0.008). However, no significant differences were observed between the *natural* condition and either the *enhanced* (Z = -2.04, p = 0.12) or *reduced* (Z = -1.63, p = 0.31) conditions.

A similar pattern emerged for accuracy (median RTs: *natural* = 264 ms, *enhanced* = 293 ms, *reduced* = 253 ms) (see Fig. 2, right). Participants were significantly more accurate at detecting turn ends in the *reduced* condition than in the *enhanced* condition (Z = -3.68, p < 0.001), and when responding to natural stimuli than to enhanced ones (Z = -2.53, p = 0.03). The difference between the *natural* and *reduced* conditions remained non-significant (Z = -2.15, p = 0.09).

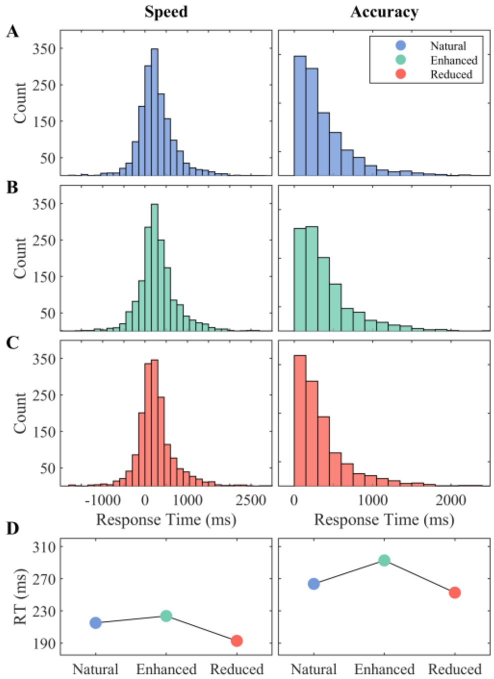

Figure 2: *RT distributions (left: speed; right: accuracy) for the (A) natural, (B) enhanced, (C) reduced conditions. The bottom panel (D) displays the median RTs for each condition.*

3.2. Effect of the *f0* fall duration

To further investigate the benefit of the *reduced* condition, we examined whether RTs were related to the timing of the utterance-final *f0* fall in the *natural* and *reduced* conditions (i.e., conditions with natural peak heights). For *natural*, the time interval was marked from the *f0* peak of the final word to the end of the question. For *reduced*, the onset was marked from the preceding *f0* maximum (i.e., the most recent local maximum with a minimum prominence of 6 Hz, 'of a' in Fig. 1). This maximum occurred on average 210 ms before the onset of the utterance-final word.

To assess this relationship, we performed two Spearman's rank correlations using the *cor.test* function in R. Both speed (ρ = -0.13, p = 0.03) and accuracy (ρ = -0.05, p = 0.41) showed weak negative correlations, with only speed reaching significance (see Fig. 3). This suggests that the earlier onset of the *f0* fall may have contributed to faster RTs, without affecting their alignment with the predicted turn-end. Note that while including *enhanced* data strengthened both correlations (speed:

ρ = -0.13, p = 0.005; accuracy: ρ = -0.09, p = 0.06), we used only *natural* and *reduced* data to ensure conclusions were drawn based on naturally occurring peak heights.

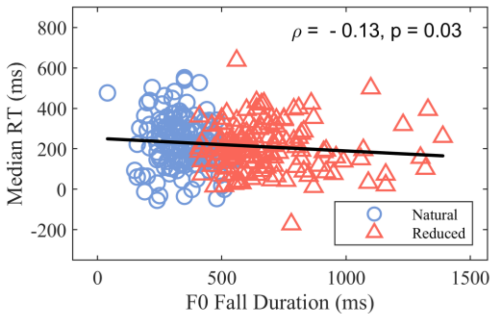

Figure 3: *Scatter plots of median RT speed and f0 fall duration for each target question, with a linear regression line and Spearman's statistics.*

4. Discussion

This study examined whether utterance-final prosodic manipulations influence turn-end prediction for individuals with self-reported hearing loss. Using a paradigm where participants had to detect question ends via button presses, we examined response speed and accuracy under *natural*, *enhanced* (heightened final *f0* peak), and *reduced* (reduced final *f0* peak and earlier onset of the *f0* fall) conditions. We also assessed self-reported listening effort and listening confidence to evaluate the cognitive impact of our manipulations.

We hypothesised that having access to an enhanced *f0* peak height of the utterance-final word, or to IP-finality cues including a reduced *f0* peak and a lengthened *f0* fall would improve turn-end prediction. We observed the facilitating effect for the *reduced* condition only; the *f0* IP-finality cues led to significantly faster and more accurate RTs compared to the *enhanced* condition, and numerical (but non-significantly) faster and more accurate RTs compared to the naturally spoken, unmanipulated utterances. In contrast, enhancing the peak associated with the utterance-final stressed syllable introduced uncertainty (shown through lack of accuracy), and appeared to impair RT speed. Together, these results suggest that the early onset of the *f0* fall as an IP-finality cue is more beneficial for turn-end prediction than increased *f0* cue prominence.

Our findings align with previous research suggesting that early prosodic cues can be used to generate earlier or more reliable turn-end predictions [16]. We found significantly faster turn-end predictions when the final pitch accent was realised with a reduced height (*reduced*), providing listeners with an earlier boundary-related cue, compared to when the *f0* peak height was increased (*enhanced*). This points to the early availability of IP-finality cues as the critical factor supporting turn-end prediction, and suggests that early *f0* cues may contribute to a more predictive, rather than purely reactive, processing of turn ends (see [29], [30]).

Furthermore, the results of our correlational analysis verified the advantage of the early onset of the *f0* fall. Similar to other temporal cues, such as the number of syllables in a final word [1], advanced *f0* cues may function as an early warning signal of the approaching turn end, providing additional time for predictions about when it will occur. However, this temporal preparation may not necessarily fine-tune the accuracy of those predictions. Although we found the *reduced* condition to improve accuracy overall, we did not see a correlation between accuracy and *f0* fall duration. Thus, whilst early *f0* cues may support turn-end predictions, their effect may be modulated by a combination of other prosodic or syntactic cues.

Alternatively, the difference in speed and accuracy between the *enhanced* and *reduced* conditions could reflect disruption from *enhanced* (rather than facilitation from *reduced*). Notably, enhancing the *f0* peak in fact led to greater response uncertainty, with listeners being statistically less accurate and, to some extent, slower than in the *natural* condition. This differs from studies where enhanced *f0* aided prominence or intelligibility rather than turn-timing [12], [13]. One possible explanation is that more prominent peaks may attract attention in a way that disrupts a listener's ability to clearly interpret turn-end cues. For instance, English speakers with normal hearing are likely to interpret a high accent peak associated with the utterance-final word as a cue to continuation and a low accent peak as a cue to turn completion [30]. This could cause hesitation or confusion, leading to delayed or inaccurate responses. Nevertheless, the fact that the *enhanced* condition produced effects opposite to those of the *reduced* condition indicates that our reported effects are not merely a result of the artificial nature of our manipulations, but rather stem from other cognitive processes.

Overall, our findings suggest that the predictive abilities of individuals with hearing loss can be strengthened, or partially restored, via acoustic manipulations that may have the potential to improve conversational dynamics [11]. However, this study represents an initial step in exploring the role of prosodic manipulations in turn-end prediction, and several methodological limitations should be considered. First, the study was conducted online, which introduces variability in RT measurements and limits control over experimental conditions [31], [32]. Thus, although we employed a within-subject design and time lags tend to remain consistent within browsers [31], variability between devices may have added noise to our data. Additionally, relying on self-reported hearing abilities without formal audiological assessments means our sample may not fully represent the typical population of adults with age-related hearing loss, and the favourable listening conditions of our stimuli (e.g., isolated targets, no background noise) likely posed minimal processing demands (reflected in ceiling rating scores). Our task may therefore have been easier for the listeners than intended, potentially masking more subtle effects of our manipulations. Future studies should address these limitations, as real-world conversations often occur in noisier environments in which the benefits of advancing prosodic cues might become more apparent, e.g., [33].

5. Conclusions

This study offers new insights into the relationship between prosodic turn-end cues and hearing loss, an area that remains largely unexplored. By experimentally manipulating the timing and prominence of utterance-final *f0* cues, we found indications that the earlier appearance of finality cues allows for faster and more accurate turn-end predictions. Our preliminary findings point to the duration of *f0* falls as a potential factor influencing turn-end prediction, making them a promising target for future investigations. This work opens new avenues for supporting conversational turn taking for individuals with hearing loss.

6. References

[1] L. Magyari, J. P. De Ruiter, and S. C. Levinson, ‘Temporal Preparation for Speaking in Question-Answer Sequences’, *Front. Psychol.*, vol. 8, p. 211, 2017, doi: 10.3389/fpsyg.2017.00211.

[2] S. Garrod and M. J. Pickering, ‘The use of content and timing to predict turn transitions’, *Front. Psychol.*, vol. 6, Jun. 2015, doi: 10.3389/fpsyg.2015.00751.

[3] A. Gravano and J. Hirschberg, ‘Turn-taking cues in task-oriented dialogue’, *Comput. Speech Lang.*, vol. 25, no. 3, pp. 601–634, Jul. 2011, doi: 10.1016/j.csl.2010.10.003.

[4] R. E. Corps, C. Gambi, and M. J. Pickering, ‘Coordinating Utterances During Turn-Taking: The Role of Prediction, Response Preparation, and Articulation’, *Discourse Process.*, vol. 55, no. 2, pp. 230–240, Feb. 2018, doi: 10.1080/0163853X.2017.1330031.

[5] M. Karimi-Boroujeni, H. R. Dajani, and C. Giguère, ‘Perception of Prosody in Hearing-Impaired Individuals and Users of Hearing Assistive Devices: An Overview of Recent Advances’, *J. Speech Lang. Hear. Res. JSLHR*, vol. 66, no. 2, pp. 775–789, Feb. 2023, doi: 10.1044/2022_JSLHR-22-00125.

[6] A. Wingfield, K. C. Lindfield, and H. Goodglass, ‘Effects of Age and Hearing Sensitivity on the Use of Prosodic Information in Spoken Word Recognition’, *J. Speech Lang. Hear. Res.*, vol. 43, no. 4, pp. 915–925, Aug. 2000, doi: 10.1044/jslhr.4304.915.

[7] 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 Hear.*, vol. 40, no. 5, p. 1069, Oct. 2019, doi: 10.1097/AUD.0000000000000694.

[8] N. Giroud, M. Keller, S. Hirsiger, V. Dellwo, and M. Meyer, ‘Bridging the brain structure-brain function gap in prosodic speech processing in older adults’, *Neurobiol. Aging*, vol. 80, pp. 116–126, Aug. 2019, doi: 10.1016/j.neurobiolaging.2019.04.017.

[9] E. B. Petersen, E. N. MacDonald, and A. Josefine Munch Sørensen, ‘The Effects of Hearing-Aid Amplification and Noise on Conversational Dynamics Between Normal-Hearing and Hearing-Impaired Talkers’, *Trends Hear.*, vol. 26, p. 23312165221103340, 2022, doi: 10.1177/23312165221103340.

[10] A. J. M. Sørensen, T. Lunner, and E. N. MacDonald, ‘Conversational Dynamics in Task Dialogue Between Interlocutors With and Without Hearing Impairment’, *Trends Hear.*, vol. 28, p. 23312165241296073, Sep. 2024, doi: 10.1177/23312165241296073.

[11] L. B. Fernandez, M. J. Pickering, G. Naylor, and L. V. Hadley, ‘Uses of Linguistic Context in Speech Listening: Does Acquired Hearing Loss Lead to Reduced Engagement of Prediction?’, *Ear Hear.*, vol. 45, no. 5, pp. 1107–1114, Oct. 2024, doi: 10.1097/AUD.0000000000001515.

[12] J. Shen and P. E. Souza, ‘Do Older Listeners With Hearing Loss Benefit From Dynamic Pitch for Speech Recognition in Noise?’, *Am. J. Audiol.*, vol. 26, no. 3S, pp. 462–466, Oct. 2017, doi: 10.1044/2017_AJA-16-0137.

[13] H.-S. Jeon and A. Heinrich, ‘Perception of Pitch Height and Prominence by Old and Young listeners’, in *Speech Prosody 2022*, Lisbon, 2022, pp. 654–658. doi: 10.21437/SpeechProsody.2022-133.

[14] S. Bögels and F. Torreira, ‘Listeners use intonational phrase boundaries to project turn ends in spoken interaction’, *J. Phon.*, vol. 52, pp. 46–57, Sep. 2015, doi: 10.1016/j.wocn.2015.04.004.

[15] S. Bögels and F. Torreira, ‘Turn-end Estimation in Conversational Turn-taking: The Roles of Context and Prosody’, *Discourse Process.*, vol. 58, no. 10, pp. 903–924, Nov. 2021, doi: 10.1080/0163853X.2021.1986664.

[16] F. Torreira and S. Bögels, ‘Vocal reaction times to speech offsets: Implications for processing models of conversational turn-taking’, *J. Phon.*, vol. 94, p. 101175, Sep. 2022, doi: 10.1016/j.wocn.2022.101175.

[17] M. K. Pichora-Fuller *et al.*, ‘Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL)’, *Ear Hear.*, vol. 37, p. 5S, Aug. 2016, doi: 10.1097/AUD.0000000000000312.

[18] J. Obleser, ‘Metacognition in the listening brain’, *Trends Neurosci.*, vol. 0, no. 0, Jan. 2025, doi: 10.1016/j.tins.2024.12.007.

[19] G. Naylor, L. A. Burke, and J. A. Holman, ‘Covid-19 lockdown affects hearing disability and handicap in diverse ways: a rapid online survey study’, *Ear Hear.*, vol. 41, no. 6, p. 1442, Dec. 2020, doi: 10.1097/AUD.0000000000000948.

[20] R. E. Corps, C. Gambi, and M. J. Pickering, ‘How do listeners time response articulation when answering questions? The role of speech rate’, *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 46, no. 4, pp. 781–802, Apr. 2020, doi: 10.1037/xlm0000759.

[21] J. C. Wells, *English Intonation: An Introduction*. Cambridge University Press, 2006.

[22] P. A. Mesiano, J. Zaar, H. Relaño-Iborra, L. Bramsløw, and T. Dau, ‘Analysis of Voice Fundamental-Frequency Adaptations in Response to Noise and Hearing Impairment’, Jul. 26, 2024, *Social Science Research Network, Rochester, NY*: 4892096. doi: 10.2139/ssrn.4892096.

[23] Boersma, Paul, ‘PRAAT, a system for doing phonetics by computer’, *Glot Int.*, vol. 5, no. 9/10, pp. 341–347, 2001.

[24] R. E. Corps, M. J. Pickering, and C. Gambi, ‘Predicting turn-ends in discourse context’, *Lang. Cogn. Neurosci.*, vol. 34, no. 5, pp. 615–627, May 2019, doi: 10.1080/23273798.2018.1552008.

[25] E. M. Picou and T. A. Ricketts, ‘Increasing motivation changes subjective reports of listening effort and choice of coping strategy’, *Int. J. Audiol.*, vol. 53, no. 6, pp. 418–426, Jun. 2014, doi: 10.3109/14992027.2014.880814.

[26] E. Giovanelli, C. Valzolgher, E. Gessa, M. Todeschini, and F. Pavani, ‘Unmasking the Difficulty of Listening to Talkers With Masks: lessons from the COVID-19 pandemic’, *-Percept.*, vol. 12, no. 2, p. 2041669521998393, Mar. 2021, doi: 10.1177/2041669521998393.

[27] R Core Team, *R: A language and environment for statistical computing*. (2022). R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: https://www.R-project.org/

[28] 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.

[29] M. J. Sjerps and A. S. Meyer, ‘Variation in dual-task performance reveals late initiation of speech planning in turn-taking’, *Cognition*, vol. 136, pp. 304–324, Mar. 2015, doi: 10.1016/j.cognition.2014.10.008.

[30] M. Zellers, ‘Duration and pitch in perception of turn transition by Swedish and English listeners’, in *Proceedings from FONETIK 2014*, Department of Linguistics, Stockholm University, 2014, p. 41.

[31] D. Bridges, A. Pitiot, M. R. MacAskill, and J. W. Peirce, ‘The timing mega-study: comparing a range of experiment generators, both lab-based and online’, *PeerJ*, vol. 8, p. e9414, 2020, doi: 10.7717/peerj.9414.

[32] D. M. Oppenheimer, T. Meyvis, and N. Davidenko, ‘Instructional manipulation checks: Detecting satisficing to increase statistical power’, *J. Exp. Soc. Psychol.*, vol. 45, no. 4, pp. 867–872, Jul. 2009, doi: 10.1016/j.jesp.2009.03.009.

[33] R. McGarrigle, P. Dawes, A. J. Stewart, S. E. Kuchinsky, and K. J. Munro, ‘Pupillometry reveals changes in physiological arousal during a sustained listening task’, *Psychophysiology*, vol. 54, no. 2, pp. 193–203, 2017, doi: 10.1111/psyp.12772.