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

# Predictability and Task-Irrelevant Auditory Distraction: A Perception-Based Functional Approach

*Hae-Sung Jeon, John E. Marsh*

School of Psychology and Humanities, University of Lancashire, UK
HJeon1@lancashire.ac.uk, JEMarsh@lancashire.ac.uk

## Abstract

The Irrelevant Sound Effect (ISE) commonly refers to the disruption of short-term memory tasks by task-irrelevant background sound. Although the ISE is known to be particularly strong with unpredictable and acoustically complex speech, the role of prosodic properties remains underexplored. This paper reports online experiments investigating how two pitch dimensions—acoustic regularity related to pitch change (regular vs irregular) and directional movement (rises vs falls)—influence the ISE. The regular stimuli were deemed more acoustically complex than the irregular stimuli, although this characterisation is debatable. Participants completed short-term memory tasks while they were exposed to manipulated speech strings. The results showed that irregular pitch variation was more disruptive than regular patterns, but no difference was found between rises and falls. Traditionally, recall performance was assumed to be directly determined by the acoustic parameters of background sound. However, our results suggest that an unpredictable distribution of *perceptually* salient acoustic events captures attention and disrupts short-term memory performance. We propose a functional approach grounded in predictability and perceptual grouping rather than spectral change for advancing our understanding of the ISE.

**Index Terms**: pitch, irrelevant sound, perception, attention

## 1. Introduction

Task-irrelevant background sound can interfere with non-auditory cognitive processes, disrupting the processing of visually presented information. This shows cross-modal interactions in cognitive systems. Visual or auditory perception and processing does not operate in isolation, and it is modulated by how the brain links sounds to visual modality. A well-documented example is the Irrelevant Sound Effect (ISE) – the impairment of visual short-term memory tasks, such as serial recall, by background sound [1]. Speech typically produces more disruption than non-speech sounds such as tones or music [2], yet the contribution of prosodic features to this interference remains underexplored.

The present study examines how two pitch dimensions—regularity in pitch change (regular vs irregular) and directional movement (rises vs falls)—influence the ISE. Because previous research on speech-induced ISE has primarily used discrete tones or word-like stimuli (sect. 1), investigating the role of continuously changing prosody can provide new insights. We hypothesised that pitch regularity and direction would modulate short-term memory disruption, although strong directional predictions about disruptive effects related to regularity are not straightforward (sect. 1.2). In this paper, we distinguish between 'fundamental frequency ($f_0$)' as the acoustic parameter and 'pitch' as its corresponding perceptual correlate.

### 1.1. Mechanisms of the Irrelevant Sound Effect

Previous research has shown that different types of task-irrelevant sound disrupt cognitive processing to varying degrees. In general, sounds with high acoustic complexity and unpredictability produce the greatest disruption [3]. However, the magnitude of disruption depends on the interaction between task demands (requiring vs not requiring serial order representation) and the properties of auditory stimuli (chancing- vs steady-states). The Duplex Mechanism account provides a theoretical framework for understanding this interaction, proposing that the ISE arises from two distinct mechanisms: *interference-by-process* and *attentional capture* [4].

First, *interference-by-process* occurs when the processing of task-irrelevant sound conflicts with the cognitive *processes* required by the focal task. For instance, 'changing-state' sounds – sequences involving spectro-temporal variation (e.g., MBCTD... or tone sequences consisting of tones that vary in pitch) – promote the automatic formation of an ordered auditory representation, which disrupts tasks that rely on serial order processing [5, 6]. Consequently, in serial recall, changing-state sounds are more disruptive than steady-state sounds (e.g., MMMMM... or tone sequences with the same pitch). The degree of disruption in serial recall is closely linked to the amount of acoustic change between successive irrelevant items in the background sound [5]. That is, according to the *interference-by-process* account, the ISE in serial recall performance is considered to depend primarily on the *acoustic* profile of the background sound (e.g., the proportion of changes in temporal properties, frequency, or amplitude). Psychoacoustic metrics such as the extent of amplitude or frequency modulation and spectral variability have been suggested as predictors of this 'changing-state' effect [7]. By contrast, short-term memory tasks that do not require serial representation – such as the missing item task, where participants identify a missing item from a sequence (e.g., identifying the missing '5' from numbers 1-7 after being presented with 4, 7, 3, 1, 6, 2) [8] – show little or no vulnerability to 'changing-state' sounds [9, 10, 11].

*Attentional capture*, in contrast, refers to distraction caused by a sound diverting attention away from the focal task, regardless of the cognitive processes involved [4]. Two classes of attentional capture have been identified. *Specific attentional capture* occurs when the sound has inherent attention-diverting significance (e.g., one's own name or food-related sounds when hungry [12]). *Aspecific attentional capture* occurs when contextual factors make a sound salient (e.g. A in BBAB).

A key finding that supports the theoretical distinction between *interference-by-process* and *attentional capture* mechanisms is that missing-item tasks are immune to changing-state effects but nevertheless substantially disrupted by


10.21437/SpeechProsody.2026-143

unpredictable deviants [13]. Accordingly, the Duplex Mechanism Account predicts a task-specific dissociation: changing-state sounds should impair serial recall more than non-serial tasks, whereas deviant or unpredictable events should disrupt both.

### 1.2. Pitch variation and attention

Previous studies have investigated the effect of emotional prosody—with naturally exaggerated prosodic properties and indexical cues—on the ISE [14, 15]. However, these findings cannot be assumed to generalise to studies using systematically manipulated prosodic features. The present study explores how pitch regularity (regular vs irregular) and pitch movement direction (rises vs falls) influence short-term memory performance in serial and missing-item tasks. It is important to note that the predictions are not straightforward because the relationship between pitch patterns and the acoustically defined steady-state vs changing-state distinction (sect. 1.1) is unclear.

Because prosodic regularity can be interpreted either as continuous acoustic change or as perceptual predictability, strong directional predictions about its disruptive effects are not straightforward a priori. Classification of 'regular' pitch variation is ambiguous (Fig. 1). Their acoustically repeated $f_0$ fluctuation might qualify them as 'changing-state'. Alternatively, repeating rises or falls could be perceived as 'steady-state' because the pattern is predictable across successive syllables. 'Irregular' pitch sequences show less $f_0$ fluctuation than regular sequences and thus approximate 'steady-state' more closely. However, the 'irregular' sequences may disrupt performance by reducing predictability and capturing attention [e.g., 16, 17; cf. 18]. Under the traditional, acoustics-based 'changing-state' definition (sect. 1.1), stimuli with 'regular' pitch rises or falls should impair serial recall more than 'irregular' variation.

We further note that changes of similar physical magnitude can differ in perceptual changes. Pitch height and direction matter: rises and high pitch are generally perceptually more salient than falls and low pitch. Listeners show improved discrimination for pitch peaks (rise-fall) height compared to valleys (fall-rise) [19, 20], and rising pitch tends to capture attention more effectively than falling pitch [21]. Consequently, background sequences rich in rising or high-pitched events should produce greater disruption than sequences dominated by falling or low-pitched events. The 'irregular' sequences contained irregular (pseudorandom) rises (Fig. 1c) or irregular high pitches (Fig. 1d). The predictability and pitch direction effects may combine additively. Under the attentional capture account (sect. 1.1), the 'irregular-rise' sequences are expected to produce the strongest disruption across both serial-recall and missing item tasks.

To summarise, we predict that pitch regularity (regular vs irregular) and direction (rises vs falls) will modulate short-term memory disruption. Whether the acoustically defined 'regular' patterns (Fig. 1a, b) used here behave as 'steady-state' or 'changing-state' stimuli remains an open empirical question. A finding of similar effects across task types would imply that these pitch dimensions do not produce the level of ordered change typically required for a classic changing-state effect. Conversely, if 'irregular' sequences prove to be especially disruptive, this would highlight the importance of both predictability and perceptual salience in influencing the ISE.

## 2. Experiments

### 2.1. Experiment 1: Serial Recall

#### 2.1.1. Methodology

*Participants* Thirty-two monolingual native speakers of British English (aged 18-30) living in the UK were recruited via Prolific (www.prolific.com). All reported normal hearing and normal or corrected-to-normal vision.

*Materials* Background sounds were generated from a single 'ma' syllable spoken with flat pitch by a female Chinese speaker and used as the basis for resynthesis [22]. Each sound string consisted of 17 'ma' tokens concatenated to yield a total duration of 8-9 seconds. $F_0$ was manipulated with Praat [23] to create four string types, crossing Regularity (regular vs pseudorandom) and Direction (rises vs falls). Across all strings, the $f_0$ range was held constant at 50 Hz.

For regular-rise strings (Fig. 1a), an $f_0$ rise was repeated across all vowels. Pseudorandom-rise strings (Fig. 1c) were created such that the position of $f_0$ rises was unpredictable. Regular-fall and pseudorandom-fall strings were also created following the same principles (Figs 1b, d). Each 'pseudorandom' string contained five deviants (i.e., abrupt rises or falls). Multiple strings were created for each experimental condition, with $f_0$ onset values varying between 180 and 270 Hz. Consequently, no two strings were acoustically identical within an experimental condition; each differed in baseline $f_0$ and/or the placement of rises or falls.

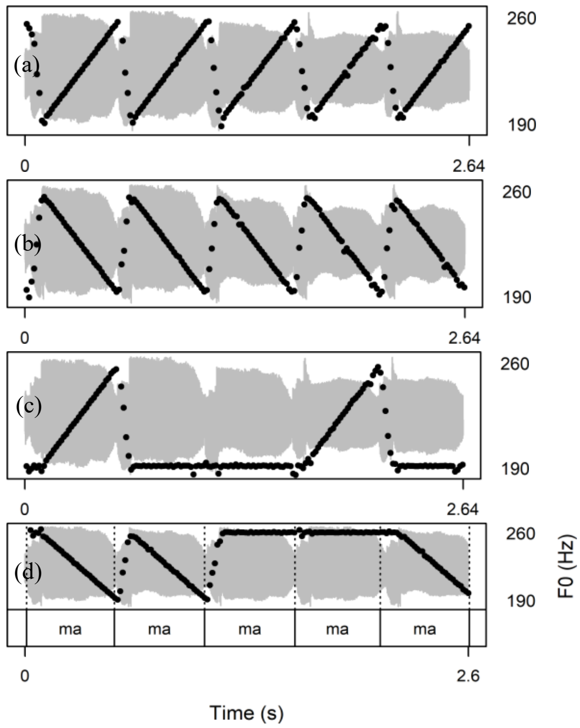


Figure 1: *Example strings used as background sound. (a) regular-rise, (b) regular-fall, (c) pseudorandom-rise, and (d) pseudorandom-fall.*

*Procedure* Serial recall experiments were conducted online using lab.js programs hosted on Open Lab [24]. Participants were required to wear headphones and to pass a headphone screening task [25]. Before the main task, they heard a calibration sound (broadband white noise) and were instructed to adjust their volume to a loud but comfortable level.

The main experiment comprised five auditory conditions: quiet, regular-rise, regular-fall, pseudorandom-rise and pseudorandom-fall (16 trials per condition; 80 trials total). Each trial began with a fixation cross (+), followed by a sequence of eight to-be-remembered digits (sampled without replacement from the range 1-8). Digits were presented in a pseudorandom order, avoiding ascending or descending runs longer than two items. Each digit appeared centrally for 350ms with a 450ms interstimulus interval. In sound trials, the background sound sequence started with the first digit.

#### 2.1.2. *Analysis and results*

Correct response ratios were computed by collapsing across serial positions and trials. Each serial position was scored as 1 for a correct response and 0 points for an incorrect response. Bayesian Linear Mixed Models was constructed using R Version 4. 5. 1 [26], R Studio Build 513 [27] and the package “brms” Version 2.22.0 [28]. Default flat priors were used [family = student; intercept, student_t (3, 0.7, 2.5); parameters, student_t (3, 0, 2.5)]. The number of iterations was set to 5000 with the random seed set to 1500. Experimental data and analysis codes are available upon request. Below model parameters are reported with 95% Credible Intervals (CIs).

For the first-stage model, Condition (quiet, regular-rise, regular-fall, pseudorandom-rise, pseudorandom-fall, see Fig. 2) was used as a fixed predictor with participant as a random intercept. This model showed that for all experimental conditions, correct response rates were significantly lower compared to the ‘quiet’ condition (intercept, b = 0.76, SE = 0.03, CI [0.71, 0.82]); regular-rise, b = -0.05, SE = 0.02, CI [-0.08, -0.02]; regular-fall, b = -0.07, SE = 0.02, CI [-0.10, -0.04]); pseudorandom-rise, b = -0.09, SE = 0.02, CI [-0.12, -0.06]; pseudorandom-fall, b = -0.10, SE = 0.02, CI [-0.13, -0.07]). This confirms that any form of irrelevant background sound impaired serial recall compared to quiet.

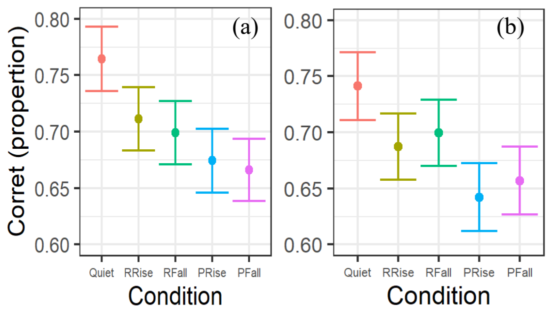


Figure 2: *Estimate proportion of correct responses and their 95% CIs for (a) Experiment 1 (serial recall) and (b) Experiment 2 (missing item) for quiet, regular-rise (RRise), regular-fall (RFall), pseudorandom-rise (PRise) and pseudorandom-fall (PFall).*

The second-stage model was constructed with Regularity (regular, pseudorandom) and Direction (fall, rise), and their interaction as fixed predictors. They were sum-coded (rise = -1, fall = 1; regular = -1, pseudorandom = 1). With sum-coding, the intercept is in the middle of the two predictor categories. Participant was incorporated as a random intercept. The model (intercept, b = 0.69, SE = 0.03, CI [0.63, 0.74]) showed a significant effect of Regularity (b = -0.02, SE = 0, CI [-0.03, -0.01]). Fig. 2a shows that predicted means of correct responses are higher for ‘regular’ compared to ‘pseudorandom’ conditions. However, the effects of Direction (b = -0.01, SE = 0, CI [-0.01, 0]) and the Regularity × Direction interaction (b = 0, SE = 0, CI [-0.01, 0.01]) were not significant.

### 2.2. Experiment 2: Non-serial missing item task

#### 2.2.1. *Methodology*

Sixty monolingual native speakers of British English (aged 18-30) were recruited as participants via the same procedures and eligibility criteria as in Experiment 1. Individuals who took part in Experiment 1 were not eligible to participate. Background sounds were generated using the same resynthesis and pitch-manipulation procedures as in Experiment 1. The experimental setup matched Experiment 1 except that a missing-item task replaced serial recall. On each trial, participants viewed a sequence of digits (1-9) presented one at a time. After the sequence, they were asked to select the digit that had not appeared (e.g. identifying 6 as missing from the sequence 3, 1, 7, 8, 4, 9, 5, 2) [cf. 29, 30]. The experiment comprised 90 trials (18 per auditory condition).

#### 2.2.2. *Analysis and results*

The statistical approach was identical to Experiment 1. The dependent variable was the proportion of correct responses in each auditory condition, averaged across trials per participant. The first-stage model showed that for all conditions with background sound (Fig. 2b), correct response rates were significantly lower compared to the ‘quiet’ condition (intercept, b = 0.74, SE = 0.03, CI [0.68, 0.80]; regular-rise [b = -0.05, SE = 0.02, CI [-0.09, -0.02]); regular-fall [b = -0.04, SE = 0.02, CI [-0.08, -0.01]), pseudorandom-rise [b = -0.10, SE = 0.02, CI [-0.14, -0.06]), pseudorandom-fall [b = -0.08, SE = 0.02, CI [-0.12, -0.05]). Thus, the presence of background sound impaired missing-item task performance compared with quiet.

For the second-stage model, the number of iterations was increased to 8000 to address a minor convergence issue. The model output (intercept, b = 0.67, SE = 0.03, CI [0.61, 0.73]) showed a significant effect of Regularity (b = -0.02, SE = 0.01, CI [-0.03, -0.01]). Although Fig. 2b suggests that sounds containing falls may have been less disruptive than those containing rises, the effect of Direction (b = 0.01, SE = 0.01, CI [-0.01, 0.02]) and the Regularity × Direction interaction (b = 0, SE = 0.01, CI [-0.01, 0.01]) were not significant.

## 3. Discussion

The present study confirmed that background speech varying in pitch impairs short-term memory relative to quiet, replicating the well-established disruptive effect of human speech [31]. We expected that if ‘regular’ pitch variation functions as ‘changing-state’ (see Fig. 1 and sect. 1.1), its disruption would be stronger for serial recall than for a non-serial missing-item task; if pitch ‘irregularity’ chiefly captures attention, it should affect both tasks similarly. In fact, the results showed that ‘irregular’ sequences were more disruptive than ‘regular’ ones for both tasks, and no reliable interaction with the task type emerged. (A supplementary Bayesian model including Task [Experiments 1 vs 2] confirmed that the Regularity × Task interaction was not credible [95% CI spanning zero]. This analysis is not reported here due to space limitations.) This pattern accords with the idea that unpredictable, salient acoustic events capture attention irrespective of a task's serial demands and *aspecific attentional capture* (sect. 1.1) was in operation.

However, our findings raise an important conceptual issue concerning the applicability of the traditional changing- vs steady-state distinction to prosodic materials. In classic ISE work, these categories are defined acoustically by the spectro-temporal properties determined by discrete changes (e.g., MBCTD... or tone sequences consisting of tones differing in pitch levels); however, perceived prosodic organisation may not correlate with acoustic measures commonly used in ISE research. For our 'regular' sequences, although the $f_0$ changed every syllable, the repetition of the same rising or falling $f_0$ contour created strong predictability. By contrast, the 'pseudorandom' sequences introduced local violations and unpredictability that functionally resemble deviant events, although its prosodic variation was less complex than the regular sequences. We further note that some previous studies involve circular reasoning, classifying stimuli as 'changing-state' based on whether they produce the 'changing-state' effect rather than on independent stimulus-specific properties.

To avoid the issues concerning acoustics-perception mismatch and circularity, we propose a functional definition grounded in predictability and perceptual salience rather than raw spectral change. Under this view, prosodic irregularity disrupts performance by challenging the formation of a coherent auditory stream and eliciting attentional diversion, whereas predictable prosody behaves functionally like 'steady-state'. This reframing clarifies how prosody should be treated within the duplex mechanism framework in future ISE research. Specifically, further studies should examine the role of pitch variation under different levels of perceptual coherence (e.g. varying grouping structure or distribution of perceptually salient events) and determine whether prosodic predictability can override the effects of discrete changes observed for traditional 'changing-state' stimuli (e.g. MBCTD...).

The perception-based functional approach explains why pitch movement direction had no effect. Although rises and high pitch are known to preferentially capture attention [19, 20, 21], no reliable direction effect (rises vs falls) emerged. This null effect likely stems from the equivalent predictability of the regular-rise and regular-fall conditions (Fig. 1a, b): regular rises and falls were equally predictable and thus equally tolerable. Similarly, the pseudorandom-rise stimuli (Fig. 1c) and pseudorandom-fall stimuli (Fig. 1d) both contained unpredictably distributed perceptually salient events – rises for pseudorandom-rise and high pitches for pseudorandom-fall, making their levels of unpredictability comparable. The predictability of perceptually salient events outweighs their pitch direction in determining distraction.

That said, several stimulus- and task-related factors in the current design may have prevented a clearer interaction between stimulus and task types from emerging. First, the modest disruption produced by our regular stimuli could reflect a steady-state effect arising from predictable prosody. Steady-state sequences are usually minimally disruptive, yet they can impair performance on non-serial tasks [e.g. 32]. Then, the pseudorandom sequences may have superimposed a deviation effect capturing attention—local violations of the established contour—on top of the baseline steady-state interference discussed above. Our pseudorandom sequences differed from classic changing-state sequences in that the segmental contents were controlled (repeating 'ma' syllables) as for classic steady-state sequences, their $f_0$ range (maximum–minimum ≈ 50 Hz) was tightly controlled, while multiple small violations were distributed across sequences. Such controlled, limited variation may be insufficient to elicit the strong serial-order interference typical of the changing-state effect [2], while they could lead to performance decrement across both tasks.

It is also possible that the absence of an interaction between stimulus and task types partly stem from strategy differences linked to a hybrid mechanism of *interference-by-process* and *attentional capture*. Some participants in the missing-item experiment may have engaged in serial rehearsal, which would render that task susceptible to process-based interference [9, 33]. If so, both interference-by-process and attentional capture could have contributed in parallel, producing a blended disruption pattern. Recent work on emotional prosody supports such a hybrid view. For instance, sudden prosodic changes (e.g. from neutral and exaggerated prosody) can yield disruption patterns intermediate between classical changing-state and attention capture effects, implying that prosody may modulates the duplex mechanisms (sect. 1.1) simultaneously [14, 15].

Collectively, the findings suggest that pitch variation in speech may engage a mixture of mechanisms, operating at the interface of predictability-driven interference and salience-driven attentional diversion. Predictable sequences behave as steady-state stimuli that modestly interferes with encoding of visually presented information, whereas unpredictable sequences introduce deviation-like violations that provoke attentional diversion. If some participants also relied on serial rehearsal, a weak process-interference component could further blur the expected duplex dissociation. These factors could explain why disruption appeared across both tasks.

# 4. Conclusions

The present study examined how pitch 'regularity' and movement direction shape the Irrelevant Sound Effect (ISE). Across two experiments, background speech varying in pitch disrupted short-term memory relative to quiet, confirming the robustness of the ISE. Irregular (pseudorandom) pitch sequences impaired both serial recall and missing-item performance more than regular sequences, whereas pitch direction (rises vs falls) exerted no reliable influence.

The results diverge from the Duplex Account's prediction of a task-specific dissociation—greater disruption for serial than for non-serial tasks when a changing-state stream is present—and instead is consistent with a predominantly attentional-diversion account. Nevertheless, several features of the stimuli and tasks suggest a hybrid explanation. The 'regular' sequences behaved as predictable, steady-state speech that modestly interfered with encoding, while the 'irregular' sequences introduced local contour violations resembling deviation effects that provoked attentional capture. If some participants in the non-serial task engaged in serial rehearsal, a small process-interference component may also have contributed, further blurring the duplex dissociation.

Overall, the findings indicate that predictability and perceptual salience, rather than acoustic change *per se*, may be the primary determinants of auditory distraction. We therefore propose a perception-based functional approach that reframes the changing- vs steady-state distinction in terms of perceived predictability and salience. Future work should parametrically manipulate the perceptual dimensions of prosodic cues to clarify how they interact with discrete variation in irrelevant sounds.

# 5. References


[1] H. A. Colle and A. Welsh, “Acoustic masking in primary memory,” *Journal of Verbal Learning and Verbal Behavior*, vol. 15, no. 1, pp. 17–31, 1976.

[2] W. Ellermeier and K. Zimmer, “The psychoacoustics of the irrelevant sound effect,” *Acoustical Science and Technology*, vol. 35, no. 1, pp. 10–16, 2014.

[3] R.W. Hughes, S. Tremblay, and D. M. Jones, “Disruption by speech of serial short-term memory: The role of changing-state vowels,” *Psychonomic Bulletin & Review*, vol. 12, no. 5, pp. 886–890, 2005.

[4] R. W. Hughes, “Auditory distraction: A duplex-mechanism account,” *PsyCh Journal*, vol. 3, no. 1, pp. 30–41, 2014.

[5] D. Jones, C. Madden, and C. Miles, “Privileged access by irrelevant speech to short-term memory: The role of changing state,” *The Quarterly Journal of Experimental Psychology Section A*, vol. 44, no. 4, pp. 645–669, 1992.

[6] D. M. Jones and S. Tremblay, “Interference in memory by process or content? A reply to Neath (2000),” *Psychonomic Bulletin & Review*, vol. 7, no. 3, 550–558, 2000.

[7] S. J. Schlittmeier, T. Weißgerber, S. Kerber, H. Fastl, and J. Hellbrück, “Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength,” *Attention, Perception & Psychophysics*, vol. 74, no. 1, pp. 194–203, 2011.

[8] J. E. Marsh, J. Yang, P. Qualter, C. Richardson, N. Perham, F. Vachon, and R. W. Hughes, “Postcategorical auditory distraction in short-term memory: Insights from increased task load and task type,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 44, no. 6, pp. 882–897, 2018.

[9] C. P. Beaman and D. M. Jones, “Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 23, no. 2, pp. 459–471, 1997.

[10] R. W. Hughes and J. E. Marsh, “When is forewarned forearmed? Predicting auditory distraction in short-term memory,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 46, no. 3, pp. 427–442, 2020.

[11] F. Kattner, S. Hanl, L. Paul and W. Ellermeier, “Task-specific auditory distraction in serial recall and mental arithmetic,” *Memory & Cognition*, vol. 51, no. 4, pp. 930–951, 2023.

[12] J. P. Röer, R. Bell, and A. Buchner, “Self-relevance increases the irrelevant sound effect: Attentional disruption by one’s own name,” *Journal of Cognitive Psychology*, vol. 25, no. 8, pp. 925–931, 2013.

[13] R. W. Hughes, F. Vachon, F., and D. M. Jones, “Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 33, no. 6, pp. 1050–1061, 2007.

[14] F. Kattner and W. Ellermeier, “Emotional prosody of task-irrelevant speech interferes with the retention of serial order,” *Journal of Experimental Psychology: Human Perception & Performance*, vol. 44, pp. 1303– 1312, 2018.

[15] F. Kattner, M. Fischer, A. L. Caling, S. Cremona, A. Ihle, T. Hodgson, and J. Föcker, “The disruptive effects of changing-state sound and emotional prosody on verbal short-term memory in blind, visually impaired, and sighted listeners,” *Journal of Cognitive Psychology*, vol. 36, no. 1, pp. 28–41, 2024.

[16] K. Bohn, J. Knaus, R. Wiese, and U. Domahs, “The influence of rhythmic (ir)regularities on speech processing: Evidence from an ERP study on German phrases,” *Neuropsychologia*, vol. 51, no. 4, pp. 760–771, 2013.

[17] D. Fogerty, V. A. Sevich, and E. W. Healy, “Spectro-temporal glimpsing of speech in noise: Regularity and coherence of masking patterns reduces uncertainty and increases intelligibility,” *The Journal of the Acoustical Society of America*, 148(3), 1552–1566, 2020.

[18] S. Solli, A. Danielsen, S. Leske, A. O. Blenkmann, K. B. Doelling, A.-K. Solbakk, and T. Endestad, “Rhythm-based temporal expectations: Unique contributions of predictability and periodicity,” *Journal of Cognitive Neuroscience*, vol. 37, no. 3, pp. 555–581, 2025.

[19] H.-S. Jeon and A. Heinrich, “Perceptual asymmetry between pitch peaks and valleys,” *Speech Communication*, vol. 140, pp. 109–127, 2022.

[20] H.-S. Jeon, “The perception of intonational peaks and valleys: The effects of plateaux, declination and experimental task,” *Speech Communication*, 173, 103267, 2025.

[21] M. Lialiou, M. Grice, C. T. Röhr, and P. B. Schumacher, “Auditory processing of intonational rises and falls in German: Rises are special in attention orienting,” *Journal of Cognitive Neuroscience*, vol. 36, no. 6, pp. 1099–1122, 2024.

[22] C. Ryu, Mandarin Tone Perception & Production Team, and Michigan State University Libraries, *Tone Perfect: Multimodal Database for Mandarin Chinese*. Accessed 1 January 2022. https://tone.lib.msu.edu/tone/5020

[23] P. Boersma and D. Weenink, Praat: doing phonetics by computer (ver. 6.3.02), 2022, https://www.fon.hum.uva.nl/praat/

[24] Y. Shevchenko, “Open Lab: A web application for running and sharing online experiments,” *Behavior Research Methods*, vol. 54, no. 6, pp. 3118–3125, 2022.

[25] K. J. P. Woods, M. H. Siegel, J. Traer, and J. H. McDermott, “Headphone screening to facilitate web-based auditory experiments,” *Attention, Perception, & Psychophysics*, vol. 79, no. 7, pp. 2064–2072, 2017.

[26] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2025, https://www.R-project.org/

[27] Posit Team, RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA, 2025, http://www.posit.co/

[28] P.-C. Bürkner, “brms: Bayesian Regression Models using “Stan”,” In CRAN: Contributed Packages. The R Foundation, 2015.

[29] H. Buschke, “Relative retention in immediate memory determined by the Missing Scan Method,” *Nature*, vol. 200, no. 4911, pp. 1129–1130, 1963.

[30] S. T. Klapp, E. A. Marshburn, and P. T. Lester, “Short-term memory does not involve the “working memory” of information processing: The demise of a common assumption,” *Journal of Experimental Psychology: General*, vol. 112, no. 2, pp. 240–264, 1983.

[31] P. Salamé and A. Baddeley, “Disruption of short-term memory by unattended speech: Implications for the structure of working memory,” *Journal of Verbal Learning and Verbal Behavior*, vol. 21, no. 2, 150–164, 1982.

[32] E. M. Elliott, R. W. Hughes, A. Briganti, T. N. Joseph, J. E. Marsh, and B. Macken, “Distraction in verbal short-term memory: Insights from developmental differences,” *Journal of Memory and Language*, vol. 88, pp. 39–50, 2016.

[33] R. W. Hughes and J. E. Marsh, “Dissociating two forms of auditory distraction in a novel Stroop serial recall experiment,” *Auditory Perception & Cognition*, vol. 2, no. 3, pp. 129–142, 2019.

[34] F. B. R. Parmentier and C. P. Beaman, “Contrasting effects of changing rhythm and content on auditory distraction in immediate memory,” *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, vol. 69, no. 1, pp. 28–38, 2015.

[35] S. Gorin, “The influence of rhythm on short-term memory for serial order,” *Quarterly Journal of Experimental Psychology*, vol. 73, no. 12, pp. 2071–2092, 2020.