Working memory capacity predicts sensitivity to prosodic structure

Burght, C.L. van der; Meyer, A.S.

Working memory capacity predicts sensitivity to prosodic structure

Constantijn L. van der Burght[1,2] & Antje S. Meyer[1,3]

1. Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
2. Leiden University Centre for Linguistics, Leiden University, Leiden, The Netherlands
3. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

Word count: 6245

Max Planck Institute for Psycholinguistics
Wundtlaan 1 | 6525 XD Nijmegen | The Netherlands
stan.vanderburght@mpi.nl

Abstract

Listeners vary in the perception and interpretation of speech prosody (the variations in intonation, loudness, and rhythm of spoken language). The source of this variability is unknown. We investigated whether the ability to recognise and classify prosodic structure is related to working memory (WM) capacity. This hypothesis stems from the tight connection between prosodic and syntactic (grammatical) structure, while processing syntax is known to relate to WN capacity. Healthy adult speakers of Dutch judged prosodic structures in a gating paradigm. The phrases contained early and late intonational cues that signalled whether the phrases contained an internal grouping or not. Listeners also took part in WM (digit span) and processing speed (letter comparison) tasks. There was an interaction between performance in the prosody judgement and WM tasks: high-WM listeners were better at classifying prosodic structure and required less prosodic information to detect the correct structure. There was no interaction between prosody processing and processing speed, suggesting that the interaction between prosodic judgement and WM capacity was not due to motivational or attentional differences. The results demonstrate a close relationship between prosody processing and WM abilities, implying that WM is an important component of prosody processing.

Keywords

Interindividual variability; prosody; intonation; working memory; processing speed

Public significance statement

The melody of a spoken sentence conveys important information about its grammatical structure and intended meaning. The melody can, for instance, determine the interpretation of sentences such as “She will visit Berlin – or Rome and Vienna” or “She will visit Berlin or Rome – and Vienna”. Adult listeners differ in their ability to hear and understand differences in speech melody, but it is unclear what causes these differences. We demonstrate that listeners who have better working memory are also better at perceiving speech melody. This means that the linguistic skill of perceiving speech melody and the broader skill of holding information in working memory are tightly linked. Our results are important for theories about the organisation of the human mind and for understanding impairments of memory and speech.

1. Introduction

Small changes in intonation, rhythm, and intensity of speech (together: prosody) convey important information during sentence processing (Cole, 2015; Cutler et al., 1997; Wagner & Watson, 2010). The interface between prosody and syntactic processing is particularly well described: prosodic and syntactic structure largely correspond to one another (Nespor & Vogel, 1983), meaning that syntactic and prosodic breaks often, but not always, align (Calhoun, 2007; Degano et al., 2024) . As such, prosodic cues can disambiguate structures that are syntactically ambiguous, such as *I read about the repayment with interest,* where *with interest* can refer to *read* or *repayment*. One prosodic phenomenon signalling the correct parsing of such sentences is the intonational phrase boundary, which consists of three acoustic cues: a pause as well as lengthening and a pitch rise on the syllable preceding the pause (Selkirk, 1984). Importantly, prosody processing requires the integration of low-level acoustic information into a sentence representation. For this integration to be successful, it is necessary to hold both the segmental and suprasegmental (i.e., prosodic) information in working memory until a sentence interpretation is formed.

As in many linguistic domains, there is considerable between-listener variability in prosody perception and interpretation (Baumann & Winter, 2018; Roy et al., 2017; van der Burght et al., 2021; van der Burght & Meyer, 2024). Recently, Hansen et al. (2022) showed that listeners differ in how early during sentence processing they can exploit prosodic cues to perceive the sentence structure. The source of such variability has so far remained unexplored. Here, we aimed to investigate whether part of this variability may be attributed to variability in working memory capacity. There are multiple lines of research that suggest this might be the case. First, prosodic phenomena are known to have an effect on memorisation. For example, words that are prosodically marked as prominent are remembered better (Fraundorf et al., 2010, 2012; Zhou et al., 2024) and rhythmic structuring of sequences is known to benefit later recall (Reeves et al., 2000; Sturges & Martin, 1974). Second, a relationship between prosody processing and working memory capacity has previously been found in production studies. For instance, speakers with higher working memory capacity tend to produce longer prosodic phrases (Bishop, 2020) and speakers with low working memory capacity have been shown to have more difficulties with the imitation of prosodic structures (Petrone et al., 2021).

But there is another reason why a link between prosodic and memory skills is plausible. As discussed above, prosody and syntactic processing often align, and syntactic processing, in turn, is known to be modulated by working memory capacity. Specifically, there is evidence that working memory is related to attachment preferences. In a study by Swets et al. (2007), listeners with low working memory capacity tended to create shorter syntactic phrases, preferring high attachment of an ambiguous relative clause. For instance, in *The maid of the princess who scratched herself in public was terribly embarrassed*, attachment to the first noun was preferred by the low-working memory group. In other studies, the opposite pattern was found, meaning that listeners in the low-working memory group preferred attachment to the second noun instead (Felser et al., 2003; Traxler, 2007). Note that in yet other work, the interaction between parsing syntactic structure and working memory capacity was absent (Traxler, 2009). Together, these studies point towards some form of constraint of working memory capacity on syntactic processing. Considering that prosodic cues can be used to modulate attachment preferences (e.g., Carlson et al., 2001, Schafer et al., 2000, Speer et al., 1996), this begs the question whether the memory-syntax connection also manifests itself in a relationship between memory and prosodic boundary perception. So far, no study has investigated whether there is a relationship between the sensitivity to prosodic cues and working memory capacity.

The main goal of the current study was therefore to determine whether individual differences in working memory span predict early sensitivity to prosodic cues. Participants (N=64) took part in a boundary perception task (developed by Hansen et al., 2022), followed by a task tapping working memory capacity (digit span task, forward and backward) and a task tapping processing speed (letter comparison task). The boundary perception task consisted of a gating paradigm in which listeners indicated whether a sequence of three names was either organised as a group of two and a single person (*Joni en Annie # en Frida*) or as three people grouped together (*Joni en Annie en Frida*). The German name sequences of the original study were converted into Dutch and recorded by four native speakers (two female, two male). The speakers were instructed to indicate the grouping structure (with or without bracket) intuitively with a normal speaking voice. The resulting three-name sequences were cut into seven gates. Each gate corresponded to a syllable, so that the first gate contained the least and the final gate the most information. The gates were presented one after the other in increasing length, and participants indicated by button press whether they perceived the structure with or without bracket (presented as line drawings on a computer screen. The study ran online.

Considering the close links between parsing strategies and working memory capacity on the one hand, and between syntactic and prosodic structure on the other, we hypothesised that individual differences in classifying prosodic structure are related to working memory capacity. We were agnostic on the directionality of a possible effect. Listeners with *low* working memory capacity tend to break up sentences into shorter clauses (Swets et al., 2007) and might therefore be particularly sensitive to prosodic cues because they enable them to close a syntactic phrase. Alternatively, listeners with *high* working memory capacity, who are in no hurry to close the phrase, might be more sensitive (or attentive) to prosodic information in the signal: these listeners might be able to afford an attitude of 'wait and see' during sentence processing, holding several interpretations in working memory (MacDonald et al., 1992), and relying on information in the signal for a prosody-driven completion of the syntactic phrase.

We are not aware of theories that would predict that working memory should not be involved at all in our task, as every cognitive task of some complexity requires at least *some* working memory capacity. Our question was therefore whether we would observe, within a sample of healthy adult native speakers, that variations in WM capacity would impact performance in the prosody task. As mentioned before, earlier work on prosody production (Bishop, 2020) and imitation (Petrone et al., 2021) has indicated that WM plays a role in these tasks. Furthermore, WM has been implied to be involved in perceptual learning of intonation contours (Kapatsinski et al., 2017). However, these tasks were invariably complex, probing learning abilities or both comprehension and subsequent imitation (i.e., the retention and transform of phonological representations from comprehension to production). Our task was much simpler and required neither learning nor memorisation as such. Importantly, the syntactic parsing involved was trivial. Thus, our task specifically tapped into the mapping of acoustic-phonetic cues onto prosodic templates. Since this process occurs continuously during everyday speech perception (McQueen & Cutler, 2010), it might be a highly overlearned process, requiring little to no processing capacity (see Hartsuiker & Moors (2017) for a discussion). This would predict that, in a sample of healthy adult speakers, variations in WM would not impact on the classification of prosodic structure. Our work therefore contributes to a better understanding of a fundamental issue of current psycholinguistics, which is the engagement of domain-general executive processes in linguistic tasks. Here, we centre this question on prosody processing, a core component of spoken language comprehension.

2. Methods

The study consisted of three tasks: a prosodic structure classification task, a working memory task, and a processing speed task.

2.1. Participants

64 native speakers of Dutch took part in all three tasks of the study (49 female, 15 male; mean age: 23.4 years; range 18-31 years). All participants entered the final analysis. The sample size was chosen to be in line with previous work and to be a multiple of eight to counterbalance the stimulus lists and response buttons. Participants reported no hearing or language disorders. They were recruited from the Max Planck Institute for Psycholinguistics participant database, were paid €15,00 for participation, and gave informed consent prior to the experiment. The study was approved by the Ethics Committee Faculty of Social Sciences, Radboud university, Netherlands.

The sample size was chosen to be in line with previous work and to be a multiple of eight to counterbalance the stimulus lists and response buttons. To determine the minimal effect size that can be detected with our sample (N=64), we performed an effect size sensitivity analysis (Giner-Sorolla et al., 2024) using the simr package (Green & MacLeod, 2016) in R. Specifically, we ran a power simulation on the Generalised Linear Mixed Model from our main analysis. In an a priori power analysis (or sample size determination analysis), sample sizes are varied while effect sizes are kept constant. In this effect size sensitivity analysis, we instead estimated power for a range of effect sizes while keeping sample size fixed at 64. To this end, we adapted the model estimate of one of our main interaction terms of interest (the interaction between prosodic classification at gates 5-4 and working memory performance). This demonstrated that the smallest possible effect size at 80% power is 1.19 (expressed as log odds). This suggests that small effect sizes were still detectable with the current sample size at a power of 80%. The script and output of this effect size sensitivity analysis can be found on OSF.

2.2. Materials and procedure

The study ran online in summer 2024, and participants accessed the experiment as a web application developed by the technical group at the Max Planck Institute for Psycholinguistics (‘Frinex’, FRramework for INteractive EXperiments). Participants were encouraged to use headphones. After giving informed consent, the boundary perception task started.

2.2.1. Prosodic structure classification task

The prosodic structure classification task was adapted from the study by Hansen and colleagues (2022), following the original study design as closely as possible. All deviations from the original study are explicitly stated. In this task, participants listened to three-name sequences and judged whether the names contained an internal grouping (grouped as two names followed by one) or not (three names grouped together).

2.2.1.1. Materials

The stimulus sequences were adapted from their German counterparts (developed in Holzgrefe-Lang et al., 2016 and Huttenlauch et al., 2021) and consisted of three Dutch proper nouns separated by conjunctions (e.g., *Joni en Annie en Frida*). The three names can be interpreted to either form a group of two people with the final person grouped separately (Figure 1A) or a group of three people (Figure 1B). The distinction between these two grouping structures can be marked prosodically: the sequence can be formed as two intonation phrases, in which case the two phrases are separated by an IPB (Selkirk, 1984). Alternatively, the sequence is formed as one intonation phrase, without separation by an IPB.

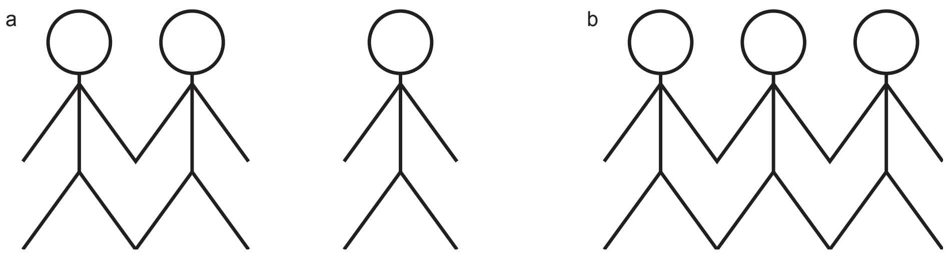

Figure 1: Schematic representations (adapted from Huttenlauch et al., 2021) of the bracket (a) and no bracket (b) conditions. This illustration functioned as response prompt during the experiment.

All proper nouns were Dutch disyllabic names with word-initial stress. A total of 24 name sequences were constructed, consisting of six names that could occur in the first two positions (*Annie, Jannie, Joni, Laurie, Lorie, Tobi*) and three names that could occur in the final position (*Frida, Maru, Mira*). All names consisted of sonorant speech sounds to facilitate tracking of the fundamental frequency (F0), the acoustic correlate of pitch. The stimuli were divided over four native speakers of Dutch (two female, two male), the same number of speakers as in Hansen et al. (2022). They were instructed to realise each stimulus sentence in two alternative groupings (once with and once without bracket). Each speaker therefore recorded a total of twelve utterances. The recordings were made in a sound-proof recording booth and the digitised speech samples (sampling frequency: 44100 Hz) were further manipulated and analysed in Praat (Boersma & Weenink, 2022). To evaluate whether the stimuli conveyed the grouping structures as intended, three native Dutch speakers who did not participate in the study categorised each version of the 24 name sequences as including or not including a bracket (accuracy >90%). For the gating paradigm, the stimuli were cut into seven gates, each gate corresponding to a syllable separated at the segmental boundaries (Turk et al., 2006). As each subsequent gate included an additional syllable, the first gate contained the least and the final gate the most information (note that the final gate contained the final two syllables; Figure 2).

To analyse the pitch cues each speaker used to prosodically realise the two grouping structures, time-normalised pitch contours were created per speaker (Figure 3). First, vocalisations were manually inspected to remove pitch points inadvertently marked by Praat as vocalisations (Cangemi, 2015). F0 contours were then smoothed using ProsodyPro (Xu, 2013) in Praat (Boersma & Weenink, 2022) using a smoothing window of 70 ms. Time-normalised contours were created at 10 samples per gate, yielding 70 F0 points per utterance. Finally, mean F0 contours were obtained per condition per speaker. These analyses indicate that each speaker used intonational cues distinguishing the two grouping structures.

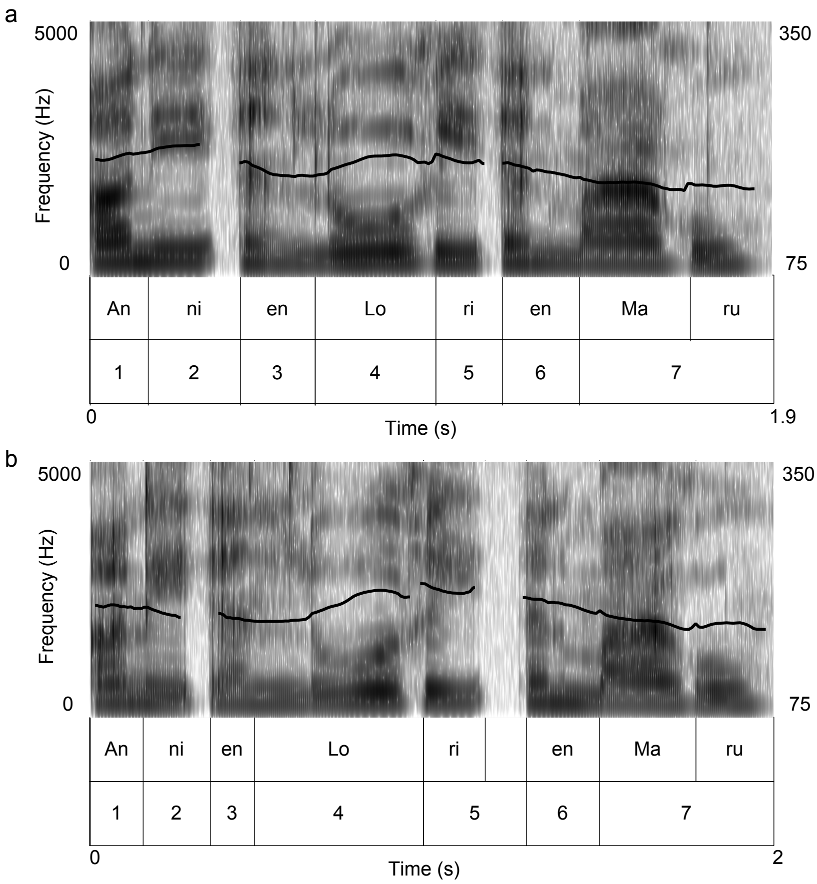

Figure 2: Spectrograms and pitch contours for the sentence conditions without boundary (a) and with boundary (b). The lower tier indicates the gate number. Note the pause in the with-boundary condition at the end of gate 5.

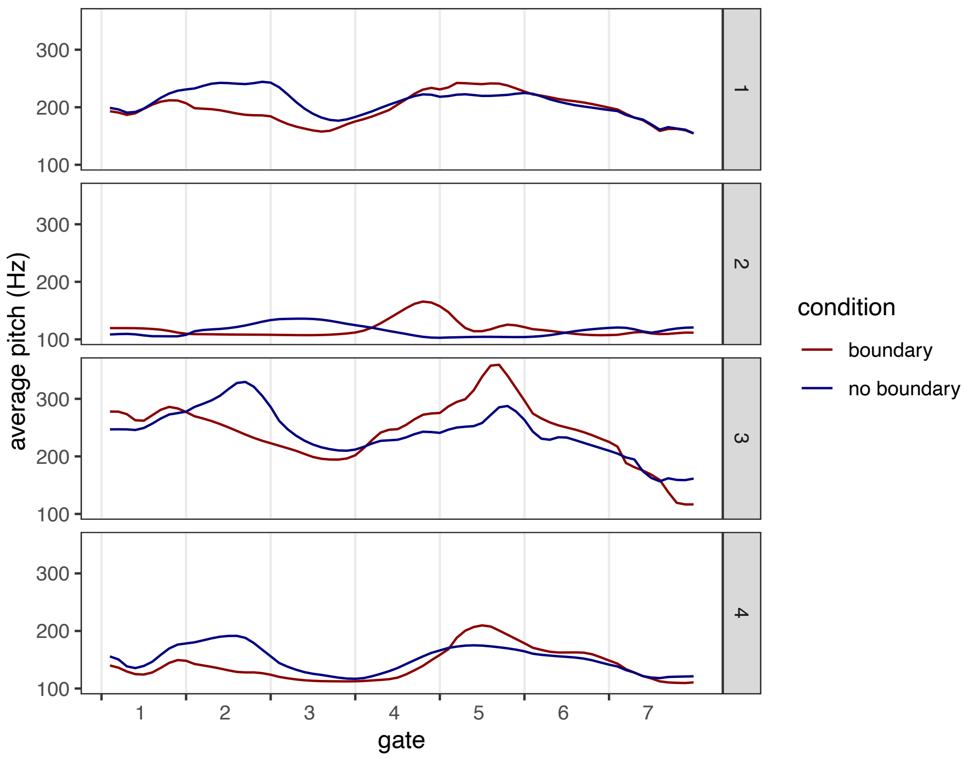

Figure 3: Time-normalised pitch contours per speaker and boundary condition.

2.2.1.2. Procedure

The task started with two demonstration trials that aimed to convey how the two grouping structures sounded and were represented visually. In these trials, two sequences (one of each grouping condition) were played in full, accompanied by the corresponding visual grouping structure. The demonstration trials consisted of two unique name sequences using the same names as in the actual experiment but recorded by a new speaker. Subsequently, a practice phase started. This followed the gating paradigm as was used in the main experiment. Here, two novel name sequences were played as individual gates. The gates were presented one after another in succession, and after each gate participants pressed the left or right arrow key to convey whether they had perceived the structure with or without bracket (presented as line drawings on the left or right side of the computer screen, Figure 1). Following the two practice sentences, the main experiment started. The task included 48 experimental sentences, each presented in seven gates, amounting to 338 trials, and ran for approximately 25 minutes. The allocation of the grouping conditions to the response buttons was counterbalanced across participants[1].

2.2.1.3. Trial structure

Each trial started with a fixation cross presented on the centre of the screen. After 300 ms, the auditory stimulus was played. As soon as the audio ended, the fixation cross was replaced by the two visual grouping illustrations (Figure 1). Participants were instructed that they could already give their response before the end of the stimulus. Upon giving their response, participants were given feedback that their response had been collected by displaying a frame around the visual grouping that matched their response. After a 700 ms interval, the next trial started. At the end of the boundary perception task, participants continued with the working memory task.

2.2.2. *Working memory task*

The working memory task was taken from a publicly available test battery for language skills (Hintz et al., 2024), a computerised version of the digit span test. In this task, participants heard a sequence of digits and were asked to type in the digits either in the order of presentation (forward version) or in reversed order (backward version). The first two trials of each version (considered practice trials) contained two-digit sequences. Then sequences of increasing length were played. When at least one of two consecutive trials was correct, the sequence was extended by one digit. The task ended either when two consecutive trials for a given sequence length were recalled incorrectly or when the maximum sequence length was reached (nine for the forward, eight for the backward version). The trial structure of the tasks consisted of a 2000 ms fixation cross presented in the centre of the screen. Subsequently, the sequence of spoken digits was played while the fixation cross remained on screen. The interval between the digits was approximately 500 ms. After the complete sequence was played, a response field at the bottom of the screen appeared, prompting participants to enter their response. The dependent measure used in the current study was a score that combined both task versions (the sum of correct responses per version). At the end of the working memory task, participants continued with the letter comparison task.

2.2.3. Letter comparison task

This task was taken from the same test battery as the working memory task (Hintz et al., 2024). It had previously been adapted by Huettig and Janse (2015) from earlier work and has been shown to measure processing speed (Salthouse, 1996). In this task, participants were presented with pairs of letter strings and decided whether or not the two strings were identical. In the first block, pairs of three-letter strings (e.g., BRT) were shown, and in the second block six-letter strings (e.g., BTYWUT). The letter strings were presented in a monospaced font (Courier New) with font size 70 underneath each other, 300 pixels apart. Participants were instructed to indicate whether the two letter strings were the same or different by button press (using the left-hand button for *different* and right-hand button for *same*), as quickly and accurately as possible. The task started with six three-letter practice trials, followed by two blocks containing 24 trials each. The proportion of *same* and *different* trials was 50:50. Trials started with a fixation cross shown for 600 ms, followed by the two letter strings until a response was made. After a pause of 1000 ms the next trial started. The dependent measure used in the current study was response speed, averaged over all correct responses.

2.3. Data analysis

Statistical analysis was performed with logistic generalised linear mixed models (GLMM; Baayen et al., 2008) using the lme4 package (Bates et al., 2015) in R, version 4.4.1 (R Core Team, 2024). Accuracy in the prosody task, determined as the proportion of correct bracket/no bracket responses, was modelled as a function of *gate* to test whether increasing amounts of available prosodic information improved classification performance. To test for an effect of working memory capacity on performance in the prosody task, we ran an additional model which included *gate* and *digit span high score* as well as their interaction as fixed effects. To test for an effect of processing speed on performance in the prosodic task, this model also contained *mean response time* in the letter comparison task as fixed effect as well as the interaction between *mean response time* and *gate*. Finally, to explore prosodic structure classification per speaker, separate models with *gate* as main effect were run on parts of the data that included stimuli from each of the four speakers only. In all models, the predictor *gate* was contrast coded using a sliding difference contrast, so that each level of the factor was compared to its preceding level (gate 2 vs. gate 1, gate 3 vs. gate 2, etc.) Predictors for the working memory task (high score across the backward and forward versions) and letter comparison task (mean response time over correct trials) were centred and scaled before they were included in the GLMM. We aimed for maximal random effects structures (Barr et al., 2013) which, due to convergence issues, were simplified by removing first the interaction terms followed by the main effects from the random effects structure. This resulted in the inclusion of by-item and by-participant random intercepts in all models. The intercept-only models, too, led to convergence warnings. To assess whether these convergence issues impacted the validity of our inferences, we ran each model with all optimisers available in the lme4 package. Because the different optimisers yielded highly similar estimates, the convergence warnings were considered false positives (Bates et al., 2015). Finally, to assess a possible correlation between working memory and processing speed tasks, we ran a Pearson's correlation between the two performance indices per participant. For each fitted model, generalised variance inflation factors (VIFs) were checked using the *car* package (Fox & Weisberg, 2019). All generalised VIFs were lower than 3, confirming that there were no multicollinearity problems.

To distinguish between situations where there is evidence for a null effect vs. a lack of evidence, we ran a Bayesian mixed effects model with a logit link function using the *brms* package (Bürkner, 2017). The model structure of this model was identical to the frequentist

GLMM. Priors were centred at 0 because we were agnostic on the direction of the interaction effects. To verify that the results were not influenced by the priors for our model, we fit models with a narrower prior (SD = .1) and a wider prior (SD .5). We set the same priors for the interactions between tasks as well as for the by-participant and by-item random intercepts.

In all analyses, *p*-values smaller than .05 were considered significant. Data visualisation was done using *ggplot* (Wickham, 2016) and *ggeffects* (Lüdecke, 2018). To visualize possible effects of each task on the prosodic task, predicted values were plotted at each gate. To illustrate between-participant differences, the *ggeffects* package plots predicted values for representative values of low, mid, and high performers. These representative values corresponded to the predictor's lower quartile, median, and upper quartile, respectively.

2.4. Transparency and openness

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study, and we follow the journal article reporting standards by Appelbaum et al. (2018). Raw data and analysis scripts for the study are available on OSF (https://osf.io/ygbvm/?view_only=c6849b75e1a542d8a13008c80f8fbfda). All software and packages employed are reported throughout the Methods section. The design and analysis of this study were not pre-registered.

3. Results

In the prosodic structure classification task, there was a significant effect of gate on accuracy. This effect became significant from the 3rd gate onward (gate 3-4: $ß$ = 0.241, SE = 0.072, t = 3.354, p = 0.001; gate 4-5: $ß$ = 0.424, SE = 0.073, t = 5.772, $p < .001$; gate 5-6: $ß$ = 0.634, SE = 0.078, t = 8.114, $p < .001$; gate 6-7: $ß$ = 0.487, SE = 0.086, t = 5.645, $p < .001$; see Table 1). The effect sizes at each gate, expressed as odds ratios, were 1.14 (gate 1-2), 1.10 (gate 2-3), 1.27 (gate 3-4), 1.53 (gate 4-5), 1.89 (gate5-6), and 1.63 (gate 6-7). This suggests that from gate 3 onward, classification accuracy improved with each subsequent gate. If we compare performance in the prosody judgement task to the pitch cues that were available in the intonational contours of each speaker (Figure 3), the 2nd and 3rd gates already showed a different intonational pattern according to the prosodic grouping condition. This apparently helped listeners to improve on their prosodic judgements from the 3rd gate onward.

In the analysis of a possible relationship between prosodic structure classification and working memory, there was a significant interaction between working memory capacity and gate from the 4th gate onward (gate 4-5: $ß = 0.183$, $SE = 0.06$, $t = 3.04$, $p < .001$; gate 5-6: $ß = 0.237$, $SE = 0.07$, $t = 3.363$, $p = 0.001$; gate 6-7: $ß = 0.238$, $SE = 0.088$, $t = 2.71$, $p = 0.007$; see Table 2). The effect sizes for the interaction terms at each gate, expressed as odds ratios, were 1.01 (gate 1-2), 1.03 (gate 2-3), 1.04 (gate 3-4), 1.20 (gate 4-5), 1.27 (gate 5-6), and 1.27 (gate 6-7). As the first interaction to be significant is between digit span performance and gates 4-5, this indicates that differences in working memory had an effect from gate 5 onward. Specifically, participants with a higher working memory span showed a steeper increase in prosodic structure classification with each subsequent gate (Figure 4). This result indicates that listeners with higher working memory capacity were overall better at classifying the prosodic structure. In addition, this shows that listeners with high working memory capacity required less information from the prosodic signal to reach above-chance accuracy in detecting the correct structure. Note that significant above-chance level performance in this experiment corresponded to accuracy levels above 65%. Put differently, participants with high working memory capacity identified the correct prosodic structure earlier (at gate 5) than the low working memory participants (at gate 6). Up until gate 4, the interactions between gate and working memory capacity were not significant (all p-values > .05).

In the analysis of a possible relationship between prosodic boundary perception and processing speed, there were no significant interactions between prosodic boundary perception and letter comparison performance (all *p*-values > .05; see Table 2). The effect sizes for the interaction terms at each gate, expressed as odds ratios, were 1.01 (gate 1-2), 0.97 (gate 2-3), 0.98 (gate3-4), 1.05 (gate 4-5), 1.10 (gate 5-6), and 1.05 (gate 6-7). This suggests that, unlike working memory capacity, processing speed was not associated with the classification of prosodic structure (Figure 5). A Bayesian analysis provided evidence for this null effect, since the credible intervals for the interactions between prosodic classification performance and processing speed crossed 0 (gate 5-4: $ß = 0.05$ [-0.06, 0.16]; gate 6-5: $ß = 0.09$ [-0.03, 0.22]; gate 7-6: $ß = 0.04$ [-0.10, 0.19]). In contrast, the credible intervals for the interactions between prosodic classification and working memory performance excluded 0 (gate 5-4: β = 0.18 [0.07, 0.30]; gate 6-5: β = 0.24 [0.10, 0.37]; gate 7-6: β = 0.23 [0.05, 0.40]). We report results from the model with wider priors, but the results from the model with narrower priors were qualitatively the same. The model output from the Bayesian mixed models can be found on OSF. Finally, there was a significant correlation between performance in the working memory and processing speed tasks ($r(62) = -0.365$; $p = 0.003$), indicating that participants with faster response times in the letter comparison task tended to have higher working memory scores.

To investigate possible across-speaker differences in performance in the prosodic structure task, we ran separate models per speaker (Figures A1-4). Following visual inspection, there seemed to be no obvious effects of different speakers on prosodic structure classification.

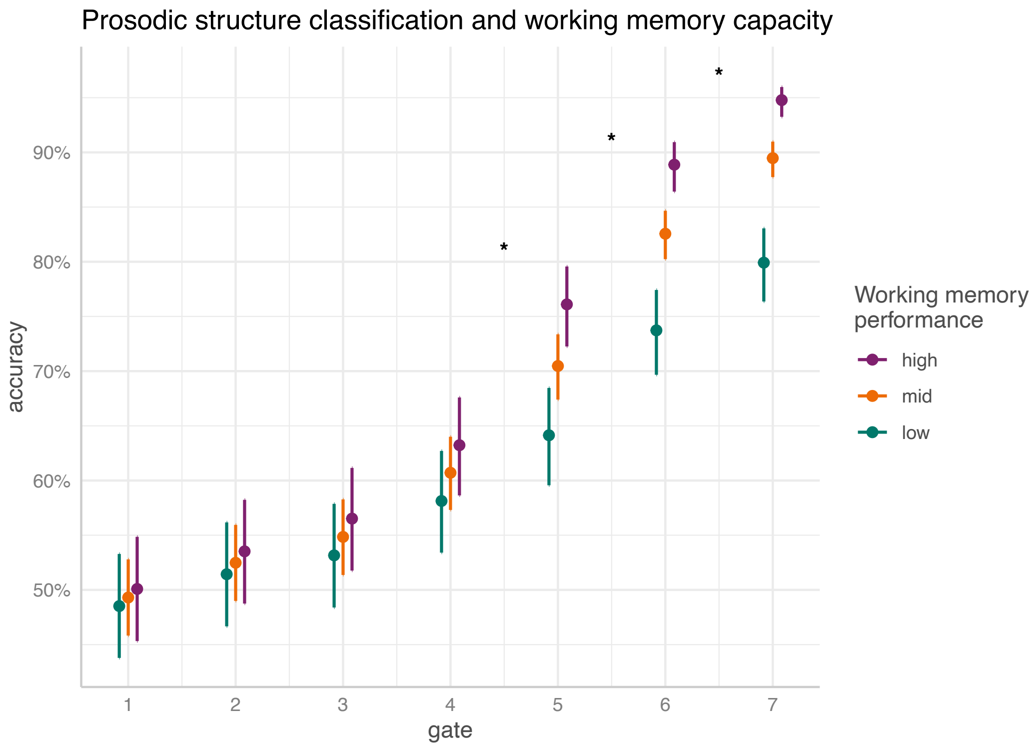

Figure 4: The relationship between performance in the prosodic structure classification and working memory (WM) tasks presented across gates. Dots represent predicted probabilities and error bars standard deviations. Note that the statistical analysis included memory performance as continuous predictor and that the predicted values for high, mid, and low performers is for illustrative purposes only (see Methods for details). Asterisks mark statistically significant interactions.

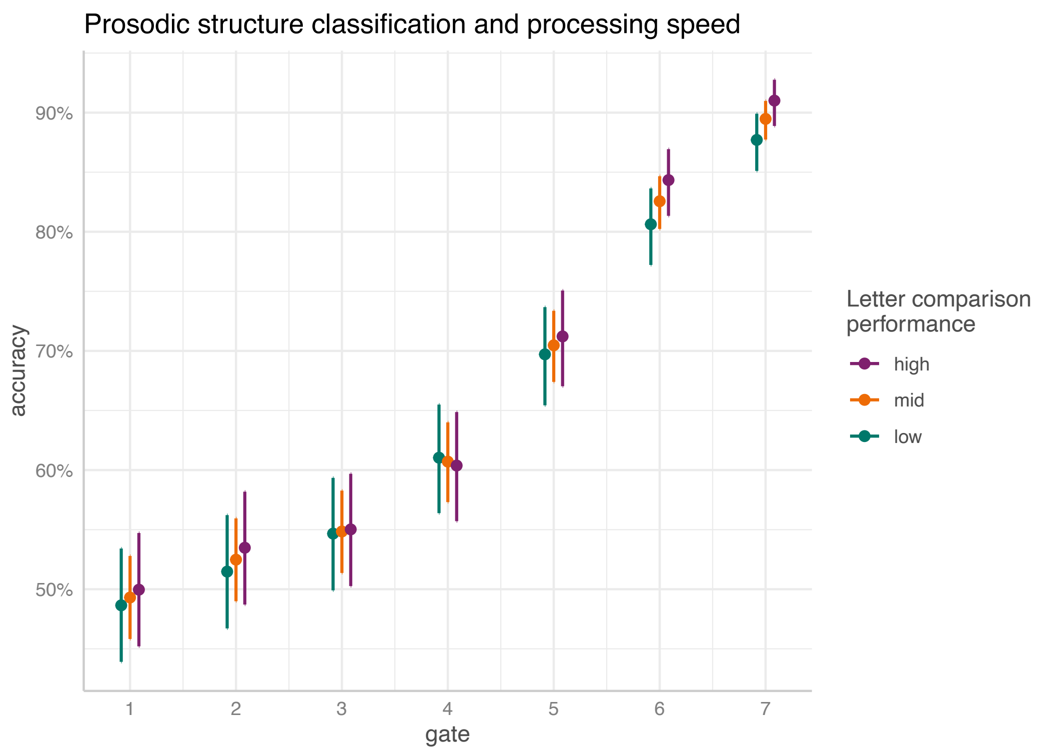

Figure 5: The relationship between performance in the prosodic structure classification and processing speed tasks presented across gates. Dots represent predicted probabilities and error bars standard deviations. Note that the statistical analysis included letter comparison performance as continuous predictor and that the predicted values for high, mid, and low performers is for illustrative purposes only (see Methods for details).

Table 1: Output of GLMM with gate as predictors of accuracy in the prosodic structure classification task. The random effects structure included random intercepts by item and participant.

term	estimate	std. error	*t*-value	*p*-value
intercept	0.720	0.058	12.412	<0.001
gate2-1	0.128	0.071	1.791	0.073
gate3-2	0.095	0.071	1.334	0.182
gate4-3	0.241	0.072	3.354	0.001
gate5-4	0.424	0.073	5.772	<0.001
gate6-5	0.634	0.078	8.114	<0.001
gate7-6	0.487	0.086	5.645	<0.001

Table 2: Output of GLMM with gate, digit span, and letter comparison scores as predictors of accuracy in the prosodic structure classification task. The random effects structure included random intercepts by item and participant.

term	estimate	std. error	*t*-value	*p*-value
intercept	0.752	0.054	13.827	<0.001
gate2-1	0.127	0.072	1.769	0.077
gate3-2	0.095	0.072	1.322	0.186
gate4-3	0.241	0.072	3.327	0.001
gate5-4	0.435	0.074	5.858	<0.001
gate6-5	0.685	0.081	8.508	<0.001
gate7-6	0.585	0.093	6.260	<0.001
digit span	0.262	0.057	4.587	<0.001
letter comparison	0.058	0.057	1.017	0.309
gate2-1 * digit span	0.011	0.057	0.189	0.850
gate3-2 * digit span	0.026	0.057	0.463	0.643
gate4-3 * digit span	0.040	0.057	0.691	0.489
gate5-4 * digit span	0.183	0.060	3.040	0.002
gate6-5 * digit span	0.237	0.070	3.363	0.001
gate7-6 * digit span	0.238	0.088	2.710	0.007
gate2-1 * letter comparison	0.014	0.056	0.249	0.803
gate3-2 * letter comparison	-0.033	0.056	-0.591	0.555
gate4-3 * letter comparison	-0.021	0.057	-0.373	0.709
gate5-4 * letter comparison	0.050	0.059	0.861	0.389
gate6-5 * letter comparison	0.094	0.064	1.452	0.147
gate7-6 * letter comparison	0.046	0.074	0.629	0.530

4. Discussion

This study investigated whether working memory capacity can explain between-listener differences in prosody processing. Participants took part in a gating experiment that tested the ability to perceive and categorise prosodic structure: they judged the internal grouping of three-name sequences, presented in gates, and had to rely on the prosodic cues available to make their judgement. Subsequently, they performed a working memory test (digit span, forward and backward) and a processing speed test (letter comparison) that functioned as control task. The results showed that listeners with higher working memory capacity performed better in the prosodic structure classification task, detecting the correct prosodic structure earlier (requiring less information from the speech signal) as well as showing overall better performance. Importantly, individual differences in processing speed had no effect on performance in the prosodic task, suggesting that the interaction with working memory cannot be explained by mere between-listener differences in processing speed, attention, or motivation. The current study thus shows that variability in the detection of prosodic structure (Hansen et al., 2022) can in part be explained by variability in working memory capacity. The result is in line with recent developmental research reporting that children with higher working memory capacity are better at discriminating prosodic structures in a language unknown to them (Stepanov et al., 2020). This implies that listeners with higher working memory capacity are either more sensitive to prosodic cues or better at matching the sensory input to a syntactic or prosodic structural representation. While previous work has highlighted working memory constraints on sentence processing (Felser et al., 2003; Swets et al., 2007; Traxler, 2007) and on prosody production (Bishop, 2020), this is the first study to demonstrate a link between working memory and prosody perception in adult listeners.

The current results have implications for models of spoken sentence processing as well as those of prosody processing. Relevant for sentence processing models, the results suggest that the cues listeners rely on during sentence processing are in part shaped by working memory capacity. Listeners with lower working memory capacity, who were less accurate at classifying prosodic structure, might rely more on cues other than prosodic ones when disambiguating sentence structure (e.g., semantic or morphosyntactic cues). Such cue-weighting differences between listeners have been found previously in the preferences for prosodic as compared to syntactic or semantic information (Baumann & Winter, 2018). The implication for theories of prosodic processing is that prosodic skills are likely not only determined by factors such as low-level acoustic abilities but depend on working memory capacity as well. Recent theories of prosody processing indeed posit that memory plays a prominent role, for example when listeners establish talker-dependent exemplars of prosodic categories based on episodic memory traces (Kurumada & Roettger, 2022). The current study demonstrates that any processing model of prosody perception should accommodate working memory. Working memory refers to the ability to hold information online, but it also involves the capacity to manipulate and combine representations (Baddeley, 2003b). To categorise intonational patterns, pitch contours must be held in working memory before being mapped onto an intonation template, e.g., of the prosodic structure with or without internal grouping. This might explain why in the current study limits on working memory capacity also limited the capacity to perceive and identify prosodic structure. However, as responses in the prosodic task could be given immediately upon perceiving each gate, the memory load from holding the contours in working memory was probably limited. More likely, the necessity to map contours onto intonation templates explains the prosody-working memory relationship.

Considering the tight link between syntax and working memory discussed in the Introduction, one might argue that the relationship between prosody processing and working memory found here was in fact a syntax-working memory relationship in disguise. However, in the current task syntactic processing was trivial, with participants judging repeatedly whether or not a three-name sequence had an internal grouping structure. In this sense, the current task was a prosodic template-matching task rather than one involving much prosody-syntax mapping. In any case, the syntactic demands involved in this task were altogether limited. We therefore propose that the relationship between the prosodic and working memory tasks represents the fact that during prosody processing, intonational, rhythmic, and intensity cues must be perceived and held in working memory before they can be mapped onto abstract prosodic templates with or without intonational boundary.

Listeners with strong prosodic abilities may detect intonational cues earlier and may be better at using these cues to guide the parser. These listeners may also be inclined to break up sequences of auditory stimuli into chunks and to retain these chunks, having good working memory abilities as a consequence. Indeed, the application of an (implicit) prosodic structure may have been useful in the digit span task. In the digit span task, an auditory sequence of digits needs to be remembered and entered on a keyboard. In the forward version of the task the sequence must be reproduced as is, while in the backward version the sequence must be reproduced in reverse order. In both versions, the digits are held in and retrieved from auditory working memory, with an additional manipulation required in the backward version. While the test is assumed to engage core processes of verbal working memory (Baddeley, 2003a), implicit prosodic structuring might help structuring the sequence in a verbal memory buffer. Indeed, there is some evidence that rhythmic grouping of digits aids the recall of digits, in particular those immediately preceding or following a pause (Reeves et al., 2000). Similarly, recognition and immediate recall of (non-linguistic) tone sequences were better for rhythmic patterns than non-rhythmic patterns, as shown in a same-different judgement task and in written recall (Sturges & Martin, 1974). In addition to rhythmical chunking, serial recall of digit strings has been argued to be aided by intonation as well (Savino et al., 2020). It may therefore be that applying an implicit prosodic structure was beneficial for performance in the digit span task. Provided that listeners' implicit and explicit prosodic abilities are related (Bishop, 2020), an implicit prosodic prosody strategy during the working memory task could thus have contributed to the relationship between prosodic structure identification and digit span tasks performance.

Above we discussed the role working memory might play in prosody perception as well as a possible advantage of (implicit) prosodic abilities for the working memory task. In addition, it may have been the case that higher working memory capacity was advantageous to performing the prosodic task as such: apart from the importance of working memory for prosody perception as a process, there might be task-specific aspects driving the effect. However, as discussed above, we consider it unlikely that much working memory capacity was required for the prosodic structure classification task. This is because participants could give their response immediately following the presentation of each gate, therefore limiting the necessity to hold prosodic cues in working memory. We therefore argue that the relationship between prosodic and working memory abilities reflects the working memory aspects of processing prosodic contours (as described earlier in the Discussion), rather than task-specific working memory computations involved in this particular paradigm.

Since the prosodic structure classification task closely mirrored the study by Hansen and colleagues (2022), the results invite a crosslinguistic comparison in prosodic boundary processing. In the original study, German listeners were able to detect and classify the appropriate structure at gate 3, whereas in the current study Dutch listeners only showed above-chance performance from gate 5 onwards. This may indicate that the prosodic cues were less salient in the Dutch stimuli than in the German stimuli, that the samples tested in the two studies happened to differ in their prosodic skills, or that, more generally, Dutch listeners rely less on early prosodic cues than the German listeners. Based on the present data, we cannot decide between these options. Concerning possible differences between the stimulus materials, we showed that at least three out of the four speakers providing stimuli for our study distinguished between the two prosodic structures using intonational differences during the early gates. As a result, early categorisation of the structures should have been possible. Concerning skill differences in the samples, we note that overall performance in the Dutch listeners was lower than in their German counterparts, with performance at gate 7, when the full sentence was available, reaching 88% and 97%, respectively. This may mean that the Dutch listeners tested in our study were generally less good at performing the task than the sample tested by Hansen and colleagues.

More interestingly, it is possible that German and Dutch listeners generally differ in their reliance on early prosodic cues. Indeed, recent event-related potential (ERP) evidence suggests that Dutch listeners are particularly sensitive to the pause when processing structures such as the ones used in the current study (Ganga et al., 2024). In contrast, studies using the same ERP measure have shown that in German listeners the pause is a non-essential cue to perceive intonational boundaries (Holzgrefe-Lang et al., 2016). These results suggest that German listeners make more use of intonational and lengthening cues than Dutch listeners. This would explain why listeners in the study in German (Hansen et al., 2022) were able to identify the correct prosodic structure at earlier gates, since the pause cue only occurs towards the end of the three-name sequence.

Conclusion

This study showed that between-listener variability in the perception and classification of prosody can in part be explained by between-listener variability in working memory capacity: Listeners with larger working memory spans were better at detecting whether three-name sequences contained an intonational boundary or not. Between-listener differences in performance in the prosody task could not be explained by differences in processing speed, ruling out general effects of motivation and attention. We tested adult native speakers of Dutch, in line with the target language of this study. Future work should therefore address whether our results generalise to other languages as well as to other prosodic functions, such as prominence marking or question/statement intonation. Regardless, the current results suggest that working memory plays an important role in processing prosody, probably because prosody perception requires the mapping of intonational contours to intonation templates. On a more general note, our results suggest that a domain-general capacity such as WM constrains prosody perception, even though it is a highly over-learned component of sentence comprehension.

Acknowledgements

The authors thank Joanne Ho Yan Chan stimulus preparation, Janay Monen for support in setting up the study, and Veerle Wilms for help with data preprocessing. They thank Kyla McConnell for valuable discussion. The study was funded by the Max Planck Society.

CRediT authorship statement

CvdB: Conceptualisation, Formal analysis, Investigation, Visualisation, Writing – original draft, Writing – review & editing. AM: Conceptualisation, Funding acquisition, Writing – review & editing

References

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal Article Reporting Standards for Quantitative Research in Psychology: The APA Publications and Communications Board Task Force Report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. (2003a). Working memory and language: an overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/s0021-9924(03)00019-4

Baddeley, A. (2003b). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics*, *70*, 20–38. https://doi.org/10.1016/j.wocn.2018.05.004

Bishop, J. (2020). Exploring the Similarity Between Implicit and Explicit Prosody: Prosodic Phrasing and Individual Differences. *Language and Speech*, 23830920972732. https://doi.org/10.1177/0023830920972732

Boersma, P., & Weenink, D. (2022). *Praat* (Version 6.2.14). http://www.praat.org/

Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Calhoun, S. (2007). *Information structure and the prosodic structure of English : a probabilistic relationship*. University of Edinburgh.

Cangemi, F. (2015). *Mausmooth*. http://phonetik.phil-fak.uni-koeln.de/fcangemi.html

Carlson, K., Jr, C. C., & Frazier, L. (2001). Prosodic Boundaries in Adjunct Attachment. *Journal of Memory and Language*, *45*(1), 58–81. https://doi.org/10.1006/jmla.2000.2762

Cole, J. (2015). Prosody in context: a review. *Language, Cognition and Neuroscience*, *30*(1–2), 1–31. https://doi.org/10.1080/23273798.2014.963130

Cutler, A., Dahan, D., & Donselaar, W. van. (1997). Prosody in the comprehension of spoken language: a literature review. *Language and Speech*, *40 (Pt 2)*(2), 141–201. https://doi.org/10.1177/002383099704000203

Degano, G., Donhauser, P. W., Gwilliams, L., Merlo, P., & Golestani, N. (2024). Speech prosody enhances the neural processing of syntax. *Communications Biology*, *7*(1), 748. https://doi.org/10.1038/s42003-024-06444-7

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's Processing of Ambiguous Sentences: A Study of Relative Clause Attachment. *Language Acquisition*, *11*(3), 127–163. https://doi.org/10.1207/s15327817la1103_1

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression*. Sage, Thousand Oaks CA. https://www.john-fox.ca/Companion/

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*(3), 367–386. https://doi.org/10.1016/j.jml.2010.06.004

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2012). The effects of age on the strategic use of pitch accents in memory for discourse: A processing-resource account. *Psychology and Aging*, *27*(1), 88–98. https://doi.org/10.1037/a0024138

Ganga, R., Geutjes, J., Niekerk, E. van, Reshetnikova, V., & Chen, A. (2024). Processing prosodic boundaries in Dutch coordinated constructions. *Speech Prosody 2024*, 985–989. https://doi.org/10.21437/speechprosody.2024-199

Giner-Sorolla, R., Montoya, A. K., Reifman, A., Carpenter, T., Lewis, N. A., Aberson, C. L., Bostyn, D. H., Conrique, B. G., Ng, B. W., Schoemann, A. M., & Soderberg, C. (2024). Power to Detect What? Considerations for Planning and Evaluating Sample Size. *Personality and Social Psychology Review*, *28*(3), 276–301. https://doi.org/10.1177/10888683241228328

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210x.12504

Hansen, M., Huttenlauch, C., de Beer, C., Wartenburger, I., & Hanne, S. (2022). Individual Differences in Early Disambiguation of Prosodic Grouping. *Language and Speech*, 002383092211273. https://doi.org/10.1177/00238309221127374

Hartsuiker, R. J., & Moors, A. (2017). On the automaticity of language processing. In H.-J. Schmid, *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge.* (pp. 201–225). https://doi.org/10.1037/15969-010

Hintz, F., Shkaravska, O., Dijkhuis, M., Hoff, V. van 't, Huijsmans, M., Dongen, R. C. A. van, Voeteé, L. A. B., Trilsbeek, P., McQueen, J. M., & Meyer, A. S. (2024). IDLaS-NL – A platform for running customized studies on individual differences in Dutch language skills via the Internet. *Behavior Research Methods*, *56*(3), 2422–2436. https://doi.org/10.3758/s13428-023-02156-8

Holzgrefe-Lang, J., Wellmann, C., Petrone, C., Räling, R., Truckenbrodt, H., Höhle, B., & Wartenburger, I. (2016). How pitch change and final lengthening cue boundary perception in

German: converging evidence from ERPs and prosodic judgements. *Language, Cognition and Neuroscience*, *31*(7), 904–920. https://doi.org/10.1080/23273798.2016.1157195

Huettig, F., & Janse, E. (2015). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2015.1047459

Huttenlauch, C., de Beer, C., Hanne, S., & Wartenburger, I. (2021). Production of prosodic cues in coordinate name sequences addressing varying interlocutors. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *12*(1), 1. https://doi.org/10.5334/labphon.221

Kapatsinski, V., Olejarczuk, P., & Redford, M. A. (2017). Perceptual Learning of Intonation Contour Categories in Adults and 9- to 11-Year-Old Children: Adults Are More Narrow-Minded. *Cognitive Science*, *41*(2), 383–415. https://doi.org/10.1111/cogs.12345

Kurumada, C., & Roettger, T. B. (2022). Thinking probabilistically in the study of intonational speech prosody. *Wiley Interdisciplinary Reviews: Cognitive Science*, *13*(1), e1579. https://doi.org/10.1002/wcs.1579

Lüdecke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*(1), 56–98. https://doi.org/10.1016/0010-0285(92)90003-k

McQueen, J. M., & Cutler, A. (2010). Cognitive Processes in Speech Perception. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (pp. 489–520).

Nespor, M., & Vogel, I. (1983). Prosodic Structure Above the Word. In *Prosody: Models and Measurements* (Berlin, Heidelberg; Vol. 14, pp. 123–140). Prosody: Models and Measurements. https://doi.org/10.1007/978-3-642-69103-4_10

Petrone, C., D'Alessandro, D., & Falk, S. (2021). Working memory differences in prosodic imitation. *Journal of Phonetics*, *89*, 101100. https://doi.org/10.1016/j.wocn.2021.101100

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reeves, C., Schmauder, A. R., & Morris, R. K. (2000). Stress grouping improves performance on an immediate serial list recall task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1638. https://doi.org/10.1037/0278-7393.26.6.1638

Roy, J., Cole, J., & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology*, *8*(1), 22. https://doi.org/10.5334/labphon.108

Salthouse, T. A. (1996). The Processing-Speed Theory of Adult Age Differences in Cognition. *Psychological Review*, *103*(3), 403–428. https://doi.org/10.1037/0033-295x.103.3.403

Savino, M., Winter, B., Bosco, A., & Grice, M. (2020). Intonation does aid serial recall after all. *Psychonomic Bulletin & Review*, *27*(2), 366–372. https://doi.org/10.3758/s13423-019-01708-4

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational Disambiguation in Sentence Production and Comprehension. *Journal of Psycholinguistic Research*, *29*(2), 169–182. https://doi.org/10.1023/a:1005192911512

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. MIT Press.

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, *25*(2), 249–271. https://doi.org/10.1007/bf01708573

Stepanov, A., Kodrič, K. B., & Stateva, P. (2020). The role of working memory in children's ability for prosodic discrimination. *PLoS ONE*, *15*(3), e0229857. https://doi.org/10.1371/journal.pone.0229857

Sturges, P. T., & Martin, J. G. (1974). Rhythmic structure in auditory temporal pattern perception and immediate memory. *Journal of Experimental Psychology*, *102*(3), 377. https://doi.org/10.1037/h0035866

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: a psychometric approach. *Journal of Experimental Psychology: General*, *136*(1), 64–81. https://doi.org/10.1037/0096-3445.136.1.64

Traxler, M. J. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory & Cognition*, *35*(5), 1107–1121. https://doi.org/10.3758/bf03193482

Traxler, M. J. (2009). A Hierarchical Linear Modeling Analysis of Working Memory and Implicit Prosody in the Resolution of Adjunct Attachment Ambiguity. *Journal of Psycholinguistic Research*, *38*(5), 491–509. https://doi.org/10.1007/s10936-009-9102-x

Turk, A., Nakai, S., & Sugahara, M. (2006). *Methods in Empirical Prosody Research*. 1–28. https://doi.org/10.1515/9783110914641.1

van der Burght, C. L., Friederici, A. D., Goucha, T., & Hartwigsen, G. (2021). Pitch accents create dissociable syntactic and semantic expectations during sentence processing. *Cognition*, *212*, 104702. https://doi.org/10.1016/j.cognition.2021.104702

van der Burght, C. L., & Meyer, A. S. (2024). Interindividual variation in weighting prosodic and semantic cues during sentence comprehension – a partial replication of Van der Burght et al. (2021). *Speech Prosody 2024*, 792–796. https://doi.org/10.21437/speechprosody.2024-160

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7–9), 905–945. https://doi.org/10.1080/01690961003589492

Wickham, H. (2016). ggplot2, Elegant Graphics for Data Analysis. *Use R!* https://doi.org/10.1007/978-3-319-24277-4

Xu, Y. (2013). ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody 2013*.

Zhou, Y., van der Burght, C. L., & Meyer, A. S. (2024). Investigating the role of semantics and perceptual salience in the memory benefit of prosodic prominence. *Speech Prosody 2024*, 1250–1254. https://doi.org/10.21437/speechprosody.2024-252

Appendix

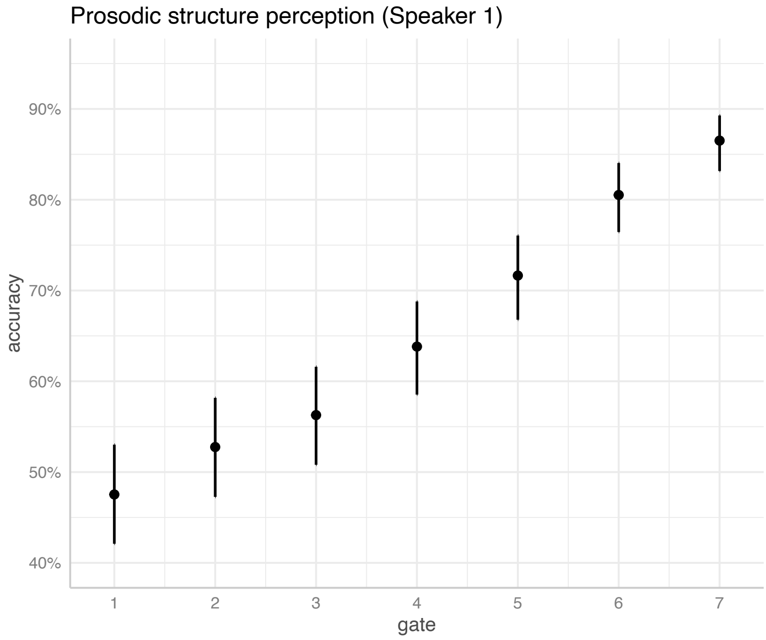

Figure A1: Performance in the prosodic structure classification task for stimuli from speaker 1. Dots represent predicted probabilities and error bars standard deviations.

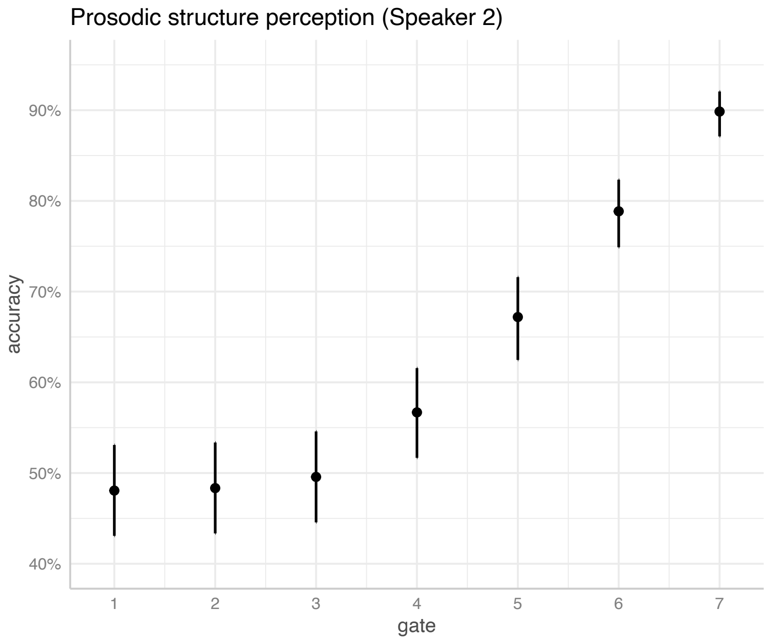

Figure A2: Performance in the prosodic structure classification task for stimuli from speaker 2. Dots represent predicted probabilities and error bars standard deviations.

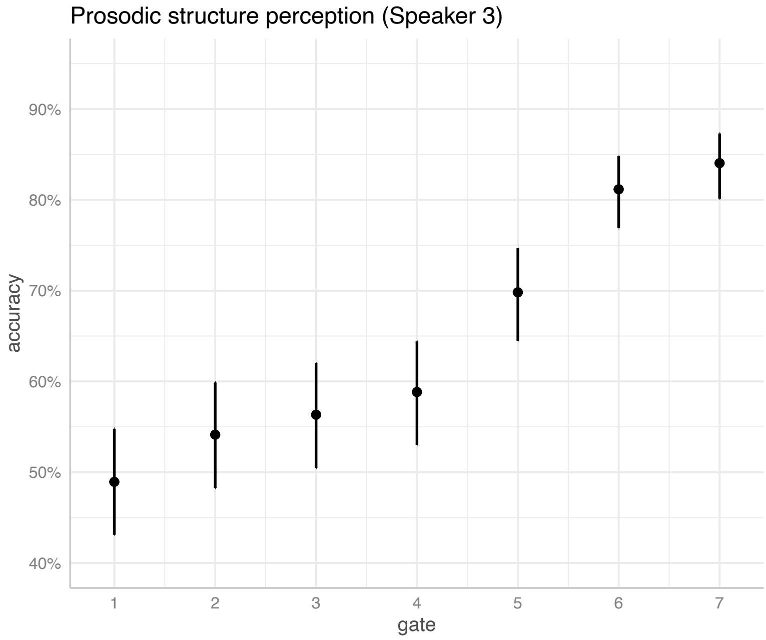

Figure A3: Performance in the prosodic structure classification task for stimuli from speaker 3. Dots represent predicted probabilities and error bars standard deviations.

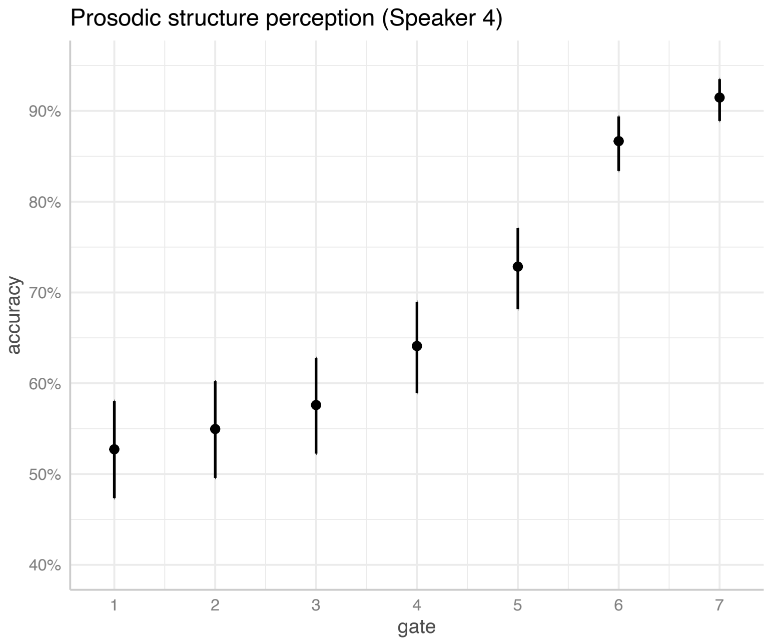

Figure A4: Performance in the prosodic structure classification task for stimuli from speaker 4. Dots represent predicted probabilities and error bars standard deviations.

[1] The results for these two subgroups did not qualitatively differ.