

Research Article

Production and Comprehension of Prosodic Boundary Marking in Persons With Unilateral Brain Lesions

Carola de Beer,^{a,b}  Andrea Hofmann,^a  Frank Regenbrecht,^c Clara Huttenlauch,^a 
Isabell Wartenburger,^a  Hellmuth Obrig,^c  and Sandra Hanne^a 

^aCognitive Sciences, University of Potsdam, Germany ^bFaculty of Linguistics and Literary Studies and Medical School OWL, University of Bielefeld, Germany ^cUniversity Hospital Leipzig and Max Planck Institute for Human Cognitive and Brain Sciences, Germany

ARTICLE INFO**Article History:**

Received May 9, 2022

Revision received July 29, 2022

Accepted August 25, 2022

Editor-in-Chief: Stephen M. Camarata

Editor: Stephen M. Wilson

https://doi.org/10.1044/2022_JSLHR-22-00258

ABSTRACT

Purpose: Persons with unilateral brain damage in the right hemisphere (RH) or left hemisphere (LH) show limitations in processing linguistic prosody, with yet inconclusive results on their ability to process prosodically marked structural boundaries for syntactic ambiguity resolution. We aimed at systematically investigating production and comprehension of three prosodic cues (f_0 range, final lengthening, and pause) at structural boundaries in coordinate sequences in participants with right hemisphere brain damage (RHDP) and participants with left hemisphere brain damage (LHDP).

Method: Twenty RHDP and 15 LHDP participated in our study. Comprehension experiment: Participants and a control group listened to coordinate name sequences with internal grouping by a prosodically marked structural boundary (grouped condition, e.g., “(Gabi und Leni) # und Nina”) or without internal grouping (ungrouped condition, e.g., “Gabi und Leni und Nina”) and had to identify the target condition. The strength and combinations of prosodic cues in the stimuli were manipulated. Production experiment: Participants were asked to produce coordinate sequences in the two conditions (grouped, ungrouped) in two different tasks: a Reading Aloud and a Repetition experiment. Accuracy of participants’ productions was subsequently assessed in a rating study and productions were analyzed with respect to use of prosodic cues.

Results: In the Comprehension experiment, RHDP and LHDP had overall lower identification accuracies than unimpaired control participants and LHDP were found to have particular problems with boundary identification when the pause cue was reduced. In production, LHDP and RHDP employed all three prosodic cues for boundary marking, but struggled to clearly mark prosodic boundaries in 28% of all productions. Both groups showed better performance in reading aloud than in repetition. LHDP relied more on using f_0 range and pause duration to prosodically mark structural boundaries, whereas RHDP employed final lengthening more vigorously than LHDP in reading aloud.

Conclusions: We conclude that processing of linguistic prosody is affected in RHDP and LHDP, but not completely impaired. Therefore, prosody can serve as a relevant communicative resource. However, it should also be considered as a target area for assessment and treatment in both groups.

Supplemental Material: <https://doi.org/10.23641/asha.21637505>

Prosody in Language Processing

Prosody serves a range of relevant functions in language processing and refers to the suprasegmental features of a spoken utterance, for instance, its intonation. The acoustic parameters of the prosodic signal are voice

Correspondence to Carola de Beer: carola.de_beer@uni-bielefeld.de.

Disclosure: The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.

quality, fundamental frequency (f_0), intensity, and durational measures, like the duration of pauses or parts of the spoken utterance (for review, e.g., Cole, 2015). Traditionally, emotional prosody is distinguished from linguistic prosody with the former referring to the use of prosody to express emotions or affects and the latter referring to the use of prosody to fulfil specific linguistic functions. These linguistic functions encompass important semantic as well as syntactic functions in both production and comprehension (Bennett & Elfnér, 2019; Sammler et al., 2015; Sidtis & Van Lancker Sidtis, 2003; Yang & Van Lancker Sidtis, 2016).

The current article focuses on linguistic prosody and, more specifically, on the relation and the processing interface of prosody and syntax as indicated by prosodic phrasing or grouping of elements. We are particularly interested in cases in which linguistic prosody provides the only disambiguating information in otherwise syntactically ambiguous structures. For example, in attachment ambiguities, such as *I met the daughter (#1) of the colonel (#2) who was on the balcony*, the position and the strength of the prosodic boundary (either #1 or #2) determine the subject of the relative clause (*the colonel* or *the daughter*) and can therefore be used to resolve the syntactic ambiguity (Frazier et al., 2006). Other, more simple examples are lists of coordinated nouns, such as a list of colors (e.g., “(pink) # (and black and green)” vs. “(pink and black) # (and green)” vs. “(pink and black and green),” Aasland & Baum, 2003), meal orders (e.g., “(tuna) # (salad) # (and wine)” vs. “(tuna-salad) # (and wine),” Zhang, 2012), or coordinated name sequences (e.g., “(Lola or Mona) # (and Lena)” vs. “(Lola) # (or Mona and Lena),” Petrone et al., 2017). Here, the function of the prosodic boundaries (indicated by #) is to group two nouns of the sequence together while the third one is separated (indicated by the parentheses). Such structures are ideal for the study of prosodic processing as they are relatively simple and easy to control for length and phonetic/phonological features.

Prosodic Processing in Coordinated Sequences

There is a fair number of studies that used coordinate sequences to elicit prosody in unimpaired speakers (e.g., Huttenlauch et al., 2021; Kentner & Féry, 2013; Petrone et al., 2017; Wagner, 2010). These studies have revealed that mainly three prosodic cues are produced by unimpaired individuals to mark structural boundaries prosodically: pause duration at the prosodic boundary, final lengthening of the element preceding the prosodic boundary, and an increased f_0 on the preboundary element(s) (e.g., for German: Huttenlauch et al., 2021, Kentner & Féry, 2013, Petrone et al., 2017).

Using coordinate name sequences, Huttenlauch et al. (2021) focused on individual differences in prosodic cue production in healthy young individuals and found

that not all prosodic cues are realized alike by all speakers. While some individuals employed pause duration, final lengthening, and an increase in f_0 range to mark the prosodic boundary, others mainly used a combination of pause and f_0 range, but no final lengthening. However, despite these differences in individual cue patterns, naïve listeners perceived the intended meaning well. On the comprehension side, studies looked for specific brain potentials in the electroencephalogram, namely, the occurrence of the “closure positive shift” (CPS), a positivity elicited by prosodic phrase boundaries. It has been shown, that the CPS is not elicited by stimuli in which the pause cue to prosodic phrase boundaries had been removed (Steinhauer et al., 1999). In such stimuli without a pause, Holzgrefe-Lang et al. (2016) found that final lengthening and f_0 rise together must be present for prosodic boundary identification and for the elicitation of the CPS, whereas the isolated occurrence of only f_0 or final lengthening is not sufficient. Thus, if we are interested in better understanding the underlying mechanisms of prosody processing, we should consider the impact of the single prosodic cues in both comprehension and production.

Prosody Processing in Individuals With Unilateral Brain Lesions

In the past, a range of different theories have been developed concerning the role of the right (RH) and the left hemisphere (LH) for prosody processing. Early theories assumed prosody to be primarily processed in the RH (e.g., Dykstra et al., 1995; Shipley-Brown et al., 1988), whereas the functional load hypothesis (Colsher et al., 1987; Gandour et al., 1995; e.g., van Lancker, 1980) suggested the RH to be dominant in processing emotional prosody while the LH is responsible for processing linguistic prosody. The “dynamic dual pathway model” assumes that semantic and syntactic information is processed by the LH, whereas the RH mainly processes sentence-level prosody (Friederici & Alter, 2004). Similarly, cue-dependent theories hypothesize that lateralization of prosodic processing is determined by prosodic cue characteristics, like the LH mainly processing temporal cues and the RH mainly processing spectral cues (e.g., Van Lancker & Sidtis, 1992; Zatorre & Belin, 2001). The related “asymmetric sampling in time” model assumes that the RH is responsible for processing of slower changing acoustic parameters, like prosody at the sentence level, whereas the LH primarily processes fast-changing acoustic parameters (e.g., Poeppel, 2003).

If prosody processing is lateralized differently in the two hemispheres, we should expect dissociating patterns of breakdown in prosody processing in persons with left versus right hemisphere brain damage (LHDP vs. RHDP). Several studies have explored production and comprehension of

linguistic prosody in these two populations and we will briefly summarize their findings.

Studies investigating the use of prosody for syntactic disambiguation in language production in RHDP and LHDP, for instance, in attachment ambiguities (Baum et al., 2001; Shah et al., 2006; Walker et al., 2004, 2009), suggest that RHDP and LHDP are capable of producing prosody for syntactic disambiguation, but their prosodic realizations differ from unimpaired persons. With respect to marking of prosodic boundaries in coordinated sequences, English speaking RHDP and LHDP have been shown to employ the three prosodic cues, f_0 range, pause, and final lengthening (Baum et al., 1997), but they used final lengthening for boundary marking more often than a group of unimpaired controls. In addition, pause was used more extensively by LHDP compared with control participants (CPs). No clear group effects were found with respect to the use of f_0 range. These data speak against a clear lateralization of prosody production in general but they indicate that brain lesions to either the LH or RH might affect the realization of the distinct prosodic cues in different ways.

When looking at the comprehension side, previous literature suggests that the ability to resolve syntactic ambiguities by means of prosody is impaired in both RHDP and LHDP (i.e., lower response accuracies and slower reaction times or delayed event-related potential (ERP) responses as compared with unimpaired participants; Balasubramanian & Max, 2005; Baum & Dwivedi, 2003; DeDe, 2012; Hoyte et al., 2004; Perkins et al., 1996; Sheppard et al., 2017, 2019; Walker et al., 2001, 2002). In particular, for both groups, accuracy in identifying prosodic boundaries in natural productions of coordinated sequences was found to be lower than in unimpaired controls (Baum et al., 1997). Moreover, RHDP's identification accuracy was lower than that of LHDP. When it comes to the use of the different prosodic cues in comprehension, the role of durational cues for boundary identification in RHDP and LHDP was studied by Aasland and Baum (2003) by systematically manipulating the strength of the pause and the final lengthening cue at boundary positions. Overall, they found lower identification accuracies in RHDP and LHDP than in unimpaired controls and no differences between RHDP and LHDP. Notably, identification accuracy in LHDP increased with longer pause durations in the stimuli, whereas this was not the case in RHDP. Recent studies with LHDP highlight the importance of attentional aspects and of intact left hemispheric frontal and parietal regions in processing sentence prosody for comprehension of complex sentences (LaCroix et al., 2020).

In summary, RHDP and LHDP are able to identify and to produce prosodic boundaries in coordinated sequences, but their performance differs from unimpaired controls. However, there are some shortcomings. First of all, although some studies tried to disentangle the

influence of LH and RH lesions on the role of the different prosodic cues in language processing, not all cues were tested systematically in individuals with brain damage. In particular, studies that systematically investigate the specific relevance of f_0 range are lacking. Secondly, only very few studies have explored prosodic production and comprehension within the same participants. Hence, only a few studies systematically compared the influence of lesions in the RH or LH on prosody production versus prosody comprehension *within* participants. Only such within participant investigations would allow for conclusions about potential parallel limitations or dissociations of prosodic comprehension and production in persons with unilateral brain damage. Thirdly, given the diversity of the applied methods, the role of the specific tasks, and stimulus materials used in order to test prosody processing in individuals with brain damage is not clear yet. For instance, in some production studies the material was presented visually and participants were requested to read out loud (Baum et al., 1997, 2001; Baum & Pell, 1997; Bélanger et al., 2009; Seddoh, 2008; Shah et al., 2006; Walker et al., 2004, 2009) and in one study participants also had to repeat what they heard before (Baum & Pell, 1997). So far, possible dissociations between tasks tapping into rather free prosody production without any prosodic input (e.g., reading aloud) and tasks that also involve receptive processing of prosody prior to production (as is the case with repetition) have not been explored. The only study we are aware of (Baum & Pell, 1997) did not focus on prosody production in syntactic ambiguities but on the production of affective prosody and on the distinction between sentence types in production. The two tasks, however, differ with respect to their dependence on working memory demands: For the repetition task, verbal working memory is needed to encode and store the prosodic structure of the auditory input string during its presentation as well as afterwards until the stimulus has been repeated completely. The reading aloud task is less demanding for working memory since the written stimuli are available throughout the whole production phase until they have been read out fully. In such reading tasks, the targeted prosodic structuring is often indicated by parentheses in the written stimuli. A final shortcoming relates to the languages investigated so far: the majority of data is from English speaking individuals and it is unclear whether their findings generalize to other languages. For instance, in French or Korean, as opposed to German and English, phrasal stress always coincides with the phrasal boundary (e.g., van Ommen et al., 2020). Furthermore, in French, phrase boundaries are characterized by regular rising and lengthening patterns that align with the phrase final syllable, whereas in German, there is more variability in the use of prosodic boundary cues (see, e.g., Huttenlauch et al., 2021).

We, therefore, intend to systematically investigate the influence of LH versus RH lesions on the use of all three prosodic cues in both production and comprehension controlling for potential task effects (reading aloud vs. repetition in production) and stimulus effects (manipulation of the strength and combinations of prosodic cues in comprehension).

Aims, Research Questions, and Hypotheses

Given the research gaps identified above, in light of the high relevance of prosody for successful communication and, even more, for the disambiguation of otherwise syntactically ambiguous utterances, we aim to assess processing of prosodic cues at structural boundaries in German-speaking RHDP and LHDP in a within-participant design, both, in language comprehension and production. We also aim to explore the role of each of the three cues (f_0 range, final lengthening, and pause) on the use of prosody in language comprehension by systematically manipulating the combinations and strength of these cues. Furthermore, we aim to explore task effects in prosody production by comparing productions elicited in both, a reading aloud and a repetition task. Our study, thus, focuses on the following research questions and hypotheses.

Processing of Prosodic Cues in Language Comprehension

- 1) Do German-speaking RHDP and LHDP differ in their ability to identify prosodically marked structural boundaries in coordinate name sequences and how is this ability influenced by varying strength and combinations of f_0 range, final lengthening, and pause at the prosodic boundary? More specifically, are RHDP and LHDP able to identify structural boundaries marked by (a) f_0 range or final lengthening in isolation or by (b) the combination of f_0 range and final lengthening with either a minimal or a maximal pause cue?
- 2) Do RHDP and LHDP perform differently from a group of unimpaired CPs?

Building upon research on prosodic boundary identification in RHDP, LHDP, and unimpaired speakers (Aasland & Baum, 2003; Baum et al., 1997; Holzgrefe-Lang et al., 2016), we formulate the following hypotheses: first hypothesis (H1): We hypothesize that identification accuracy in prosodic boundary comprehension will be lower in RHDP and LHDP as compared with a group of unimpaired participants (Aasland & Baum, 2003; Baum et al., 1997). Second hypothesis (H2): For all participant groups, accuracy is expected to be lower for coordinates in which prosodic boundaries are marked by final lengthening and f_0 range in isolation as compared with stimuli in which boundaries are realized by a combination of both cues (Holzgrefe-Lang et al., 2016). Third hypothesis (H3):

We further hypothesize that RHDP and LHDP will show deficits in using durational cues compared with unimpaired controls (Aasland & Baum, 2003). Fourth hypothesis (H4): When comparing LHDP to RHDP, we expect lower overall accuracies and also a stronger decline in accuracy with decreasing durational cues in LHDP than in RHDP (Aasland & Baum, 2003).

Use of Prosodic Cues in Language Production

- 3) Do German speaking RHDP and LHDP differ from each other in their use of the prosodic cues f_0 range, final lengthening, and pause for marking prosodic boundaries in production of coordinate name sequences?
- 4) Does the use of the three prosodic cues for boundary marking by RHDP and LHDP differ in a reading aloud task (free production without prior prosodic input) in contrast to a repetition task (prosody production with prior receptive processing of prosody)?

Fifth hypothesis (H5): We assume RHDP and LHDP to use all three prosodic cues under investigation to mark prosodic boundaries (Baum et al., 1997). We state no hypotheses on group differences in the use of the three prosodic cues for syntactic disambiguation since previous studies have led to inconclusive results (Baum et al., 1997, 2001; Shah et al., 2006), especially the role of the f_0 cue for prosodic boundary marking by RHDP and LHDP is still unclear. With respect to hypotheses about differences between the tasks used in the two production experiments (reading aloud vs. repetition), Baum and Pell (1997) found some task-specific effects when comparing the use of prosody for distinction of sentence types in reading aloud versus repetition. Concerning prosodic boundary marking in coordinates, we are not aware of studies that have compared these tasks. Building upon what is known, however, about general language and cognitive limitations in LHDP versus RHDP (and despite the observations by Baum & Pell, 1997), sixth hypothesis (H6): We assume RHDP to perform better in reading aloud than in repetition, due to the lower working memory load of the reading aloud task and the often reported working memory difficulties in RHDP (Blake, 2017). In case of LHDP, no clear assumptions are made, since both reading aloud as well as repetition can be affected by LH lesions: Limited working memory capacities in LHDP as well as impairments in receptive and/or productive prosody processing can affect performance in the repetition task, whereas reading capacities can be impaired as well in LHDP. Recent results suggest that limited working memory capacities in LHDP are associated with problems in processing utterances with list prosody, but not with regular sentence prosody (LaCroix et al., 2020). Overall, our study intends to shed more light on the lateralization of prosody, as the above introduced

theories predict contrary lateralization patterns. For instance, the functional load hypothesis (Chobor & Brown, 1987; Colsher et al., 1987; Cooper et al., 1984; Gandour et al., 1995; Geigenberger & Ziegler, 2001; Hird & Kirsner, 1993; Hughes et al., 1983; van Lancker, 1980; Wunderlich et al., 2003) would assume that linguistic prosody is mainly processed by the LH; hence, LHDP should be outperformed by RHDP and controls. In contrast, the dynamic dual pathway model (Friederici & Alter, 2004) as well as the asymmetric sampling in time model (Poeppel, 2003) would predict that RHDP will be outperformed by LHDP and controls.

Method

We investigated comprehension and production of f_0 range, final lengthening, and pause duration as prosodic cues for boundary marking in RHDP and LHDP in a comprehension experiment (boundary identification) and in two production experiments (reading aloud and repetition). In the comprehension experiment, a group of neurologically healthy CPs was investigated in addition to the two groups of individuals with brain lesions.

Participants

The study comprised 35 participants with brain lesions (ischemic or hemorrhagic stroke, see Figure 1 for a lesion overlay of 19 RHDP and 10 LHDP participants) that split up into 20 RHDP (11 men, nine women, $M_{\text{age}} = 56$ years, $SD = 11$, range: 24–77) and 15 LHDP (12 men, three women, $M_{\text{age}} = 54$ years, $SD = 14$, range: 40–87) and were either recruited at the University Hospital in Leipzig or via a participant database at the University of Potsdam. In addition, a group of 20 neurologically unimpaired CP participated in the Comprehension experiment (10 men, 10 women, $M_{\text{age}} = 58$ years, $SD = 9.26$, range: 42–72). The three participant groups were matched for age (RHDP vs. LHDP: $t = .39$, $p = .702$; RHDP vs. CP: $t = .7$, $p = .49$; LHDP vs. CP: $t = .978$, $p = .34$).

All participants were right-handed (assessed using a German translation of the Edinburgh Handedness Inventory, Oldfield, 1971), native speakers of German, and had passed an audiometric screening with < 35 dB hearing level at frequencies between 350 and 1000 Hz in their better ear.¹ Two additional participants took part in the study (2 RHDP), but had to be excluded due to left-handedness ($n = 1$) and a stroke during language development ($n = 1$).

¹We tested this range of frequencies, since these were the frequencies present in our audio stimuli. By this, we made sure that our participants' performance in Experiments 2 and 3 was not influenced by hearing loss for the relevant frequencies.

To check for possible exclusion criteria in the participants with acquired brain lesion, we further assessed the presence of dysarthria (Bogenhausener Dysarthrieskalen, BoDyS, Ziegler et al., 2018), apraxia of speech (HWL-kompakt, Ziegler et al., 2020), as well as their abilities in reading aloud of single words and auditory discrimination of nonwords (LEMO 2.0, Test 8: Reading aloud of regular and irregular words and Test 1: Auditory discrimination of nonword pairs, Stadie et al., 2013). Participants with LHD also completed an aphasia assessment battery (Aachen Aphasia Test, Huber et al., 1983; or Aphasie-Check-Liste, Kalbe et al., 2005). Participants who according to these tests showed presence of dysarthria or apraxia of speech and/or accuracy below 85% in the single word reading test were excluded from the production experiments. For the Comprehension experiment, we excluded participants who performed $< 85\%$ correct on auditory nonword discrimination. Table 1 presents demographic data for LHDP and RHDP and whether they participated in each of the three experiments.

Additionally, we administered a battery of tests to assess participants' cognitive capacities and control for their potential influence on participants' performance in our experiments. Therefore, we ran the digit and block span forwards and backwards from the German version of the revised Wechsler-Memory Scale (Härting et al., 2000) to assess working memory capacities. The digit-symbol-substitution test of the German adaptation of Wechsler Adult Intelligence Scale–Fourth Edition (Petermann, 2012) was employed to assess processing speed, and the Trail Making Test from the German version of the Consortium to Establish a Registry for Alzheimer's Disease-Plus Battery (Aebi, 2002) to test executive functions. In our current study, we will focus on potential correlations of working memory capacities (digit span tasks,² please see Supplemental Material S1 for individual digit span scores) on the performance in our experiments involving receptive auditory processing (i.e., the Repetition and Comprehension experiments).

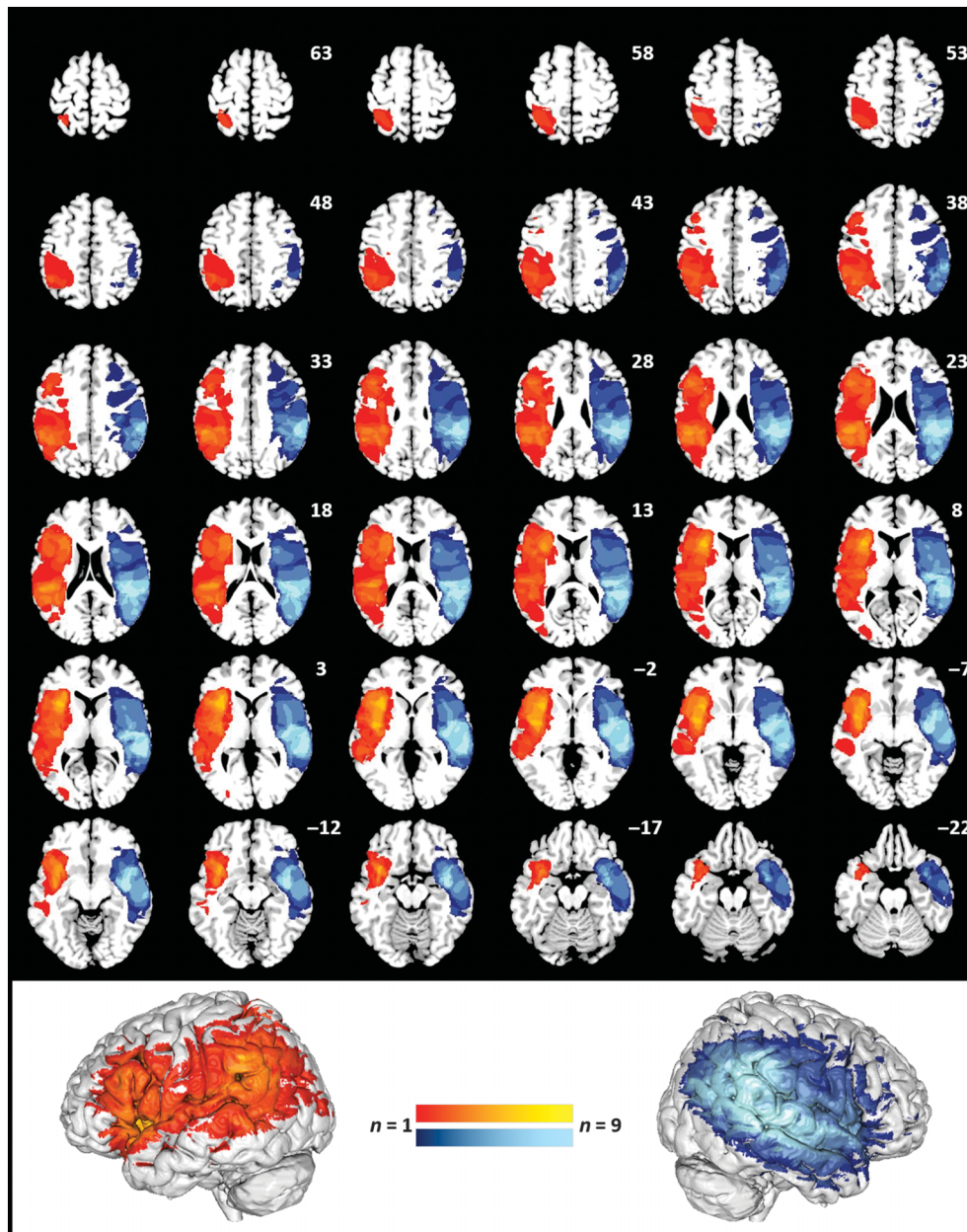
All participants were informed about the study procedures as well as about data protection and they gave written informed consent prior to the start of the experiments. The study was approved by the ethics committee of the University of Potsdam (72/2016) and by the Medical Faculty of the University of Leipzig (reference number: 144-18ek with amendment).

Material

The items in this study were coordinated name sequences, each consisting of three disyllabic trochaic German names coordinated by *und* (and), presented in two different conditions:

²The block span tasks are not included in this article, since those could not be administered with the CP of the Comprehension experiment due to the COVID-19-pandemic.

Figure 1. Overlay of the lesion patterns. Participants with a left hemisphere lesion are depicted in red yellow ($n = 10$); participants with a right hemisphere lesion in shades of blue ($n = 19$). The lesions cover parts of the language network including the inferior frontal gyrus, the posterior superior temporal gyrus, and also the temporo-parietal junction. Lesions were manually delineated on each slice of the T1 images using MRlcron (Rorden & Brett, 2000). Fluid-attenuated inversion recovery images served as a reference. For normalization and transformation of the lesion masks into standard stereotactic space, the “clinical toolbox” (<http://www.nitrc.org/projects/clinicaltbx/>) in SPM8 (fil.ion.ucl.ac.uk/spm) was used. It applies the unified segmentation approach (Ashburner & Friston, 2005), restricting estimation of normalization parameters to healthy tissue (Brett et al., 2001).



Example 1:

Condition 1: No internal grouping/bracketing (ungrouped): *Gabi und Leni und Nina*

Condition 2: With internal grouping/bracketing (grouped): *(Gabi und Leni) und Nina*

The parentheses in the grouped condition indicate that the first two names were grouped together, whereas there was no such internal grouping in stimuli of the ungrouped condition. For visualizing this difference between the two conditions in the comprehension and Reading Aloud experiment, we also used pictograms (see Figure 2). The production experiments both encompassed

Table 1. Demographic data for participants with right hemisphere brain damage (RHDP) and left hemisphere brain damage (LHDP).

| Participant ID/lesion site | Age | Gender | Months post onset | Main results of aphasia assessment | Reading aloud | Repetition | Comprehension |
|----------------------------|-----|--------|-------------------|------------------------------------|----------------|----------------|----------------|
| | | | | | Inclusion Y/N | | |
| RHDP01 | 38 | F | 13 | n/a | Y | Y | Y |
| RHDP02 ^a | 52 | M | 13 | n/a | Y | Y | Y |
| RHDP03 | 59 | F | 15 | n/a | Y | Y | Y |
| RHDP04 | 58 | M | 6 | n/a | Y | Y | Y |
| RHDP05 | 24 | M | 6 | n/a | Y | Y | Y |
| RHDP06 | 50 | M | 15 | n/a | N ¹ | N ¹ | Y |
| RHDP07 | 61 | F | 7 | n/a | N ¹ | N ¹ | Y |
| RHDP08 | 49 | F | 12 | n/a | Y | Y | Y |
| RHDP09 | 57 | M | 7 | n/a | Y | Y | Y |
| RHDP11 | 64 | M | 116 | n/a | N ⁶ | Y | Y |
| RHDP12 | 62 | M | 95 | n/a | Y | Y | Y |
| RHDP13 | 77 | F | 257 | n/a | Y | Y | Y |
| RHDP14 | 66 | M | 181 | n/a | Y | Y | Y |
| RHDP15 | 55 | F | 106 | n/a | Y | Y | Y |
| RHDP16 | 54 | F | 52 | n/a | Y | Y | Y |
| RHDP17 | 67 | M | 24 | n/a | Y | Y | Y |
| RHDP18 | 54 | F | 82 | n/a | Y | Y | Y |
| RHDP19 | 64 | M | 70 | n/a | N ³ | N ³ | Y |
| RHDP20 | 50 | M | 6 | n/a | Y | N ⁶ | Y |
| RHDP22 | 58 | F | 14 | n/a | Y | Y | Y |
| LHDP01 | 47 | M | 69 | AAT syndrome: Wernicke | Y | N ⁵ | Y |
| LHDP02 | 56 | M | 13 | AAT syndrome: Anomic | Y | N ⁴ | N ⁴ |
| LHDP03 | 51 | M | 11 | AAT syndrome: Wernicke | Y | Y | Y |
| LHDP04 ^a | 71 | M | 11 | ACL score: 134 | Y | Y | Y |
| LHDP06 | 65 | M | 36 | AAT syndrome: Wernicke | N ² | N ² | Y |
| LHDP07 ^a | 48 | F | 242 | ACL score: 125 | Y | Y | Y |
| LHDP08 ^a | 78 | M | 62 | AAT syndrome: Wernicke | Y | N ⁵ | Y |
| LHDP09 | 44 | M | 74 | AAT syndrome: Anomic | Y | Y | Y |
| LHDP10 | 55 | M | 11 | AAT syndrome: Anomic | Y | Y | Y |
| LHDP11 | 49 | M | 3 | AAT syndrome: Anomic | Y | Y | Y |
| LHDP12 | 41 | M | 10 | AAT syndrome: Anomic | Y | N ⁵ | Y |
| LHDP13 | 41 | M | 115 | AAT syndrome: Anomic | Y | Y | Y |
| LHDP14 | 40 | F | 15 | AAT syndrome: Broca | Y | Y | Y |
| LHDP15 ^a | 87 | M | 219 | ACL score: 110 | Y | Y | Y |
| LHDP16 ^a | 43 | F | 132 | ACL score: 126 | Y | Y | Y |

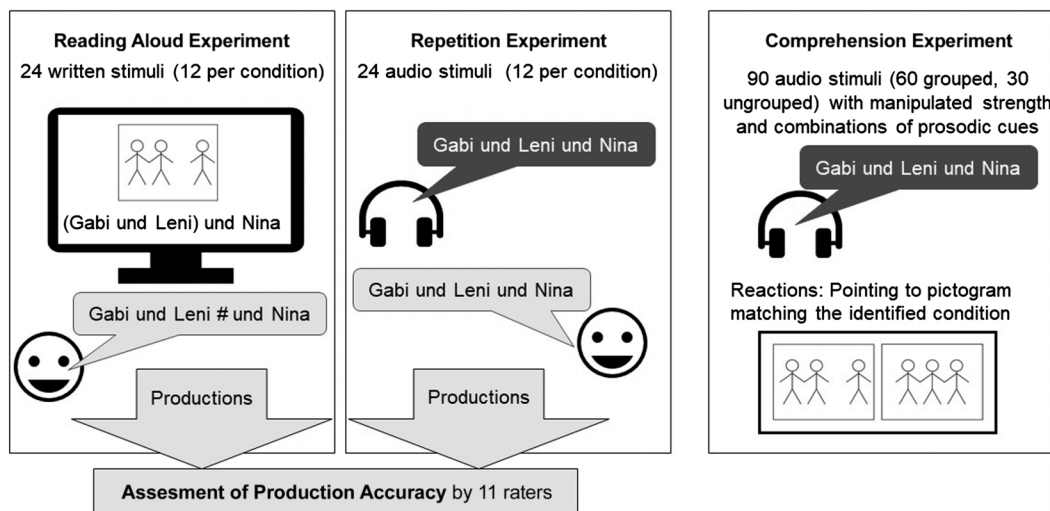
Note. ^aNo lesion data available/not included in lesion overlay (see Figure 1). Reasons for exclusion: ¹Presence of dysarthria. ²Presence of apraxia of speech. ³Presence of a cold affecting voice quality. ⁴< 85% correct in LEMO Test 1. ⁵Abortion of experiment. ⁶Technical problems during recording. RH = right hemisphere; LH = left hemisphere; MCA = mid cerebral artery; ACA = anterior cerebral artery; ICA = internal cerebral artery; AAT = Aachen Aphasia Test; ACL = Aphasia-Check-Liste; F = female; M = male; Y = yes; N = no; n/a = not applicable.

24 items, 12 in each condition with six different name sequences each of which was presented twice per condition.

The audio stimuli presented in the Comprehension experiment were manipulated using Praat software (Boersma & Weenink, 1992) with respect to their combinations and the strength of the different prosodic cues (f_0 range, final lengthening, and pause). A phonetically trained female speaker recorded five different coordinate name sequences in both conditions 5 times (5 name sequences \times 2 conditions \times 5 versions each = 50 recordings). She was told to produce each name sequence naturally without any exaggeration. The total of 50 recordings enabled us to extract a range of values for each of the prosodic cues and thus accounting for variability in natural productions. Final lengthening refers to the duration of the final (second) vowel of name2 in relation to the overall

duration of name2 (see note in Table 2). For the calculation of the f_0 range, the f_0 movement from the stressed to the unstressed vowel of name2 was annotated according to Braun (2006) marking low and high points. In the case of a rising movement, the low point precedes the high point, in the case of a falling movement, the order is reversed. For a rise, the low point was set at the so-called elbow whereas the high point was set at the first high point of the peak at or around the vowel. Note that a rise may well start before the vowel and the high point may be outside the vowel. We confirmed that the trained speaker produced the relevant prosodic cues comparable to a healthy population sample of $n = 15$ young adults in the baseline condition of the study by Huttenlauch et al. (2021). In this study, the mean ratio for relative lengthening on name2 in the ungrouped condition was 35.2% (minimal = 23.0%, maximal =

Figure 2. Overview of experimental tasks and outcome measures in the three experiments: experiments exploring prosodic cue production (Reading Aloud experiment, Repetition experiment, left panels) and the experiment investigating the use of prosodic cues in language comprehension (right panel).



51.9%) and in the grouped condition it was 45% (minimal = 29.9%, maximal = 58.4%). The respective values of the trained female speaker's productions lie within this range (see Table 2). Based on these analyses, we constructed the audio items for the Comprehension experiment by systematically combining minimal or maximal magnitudes/strengths of the three prosodic cues (see Table 2). The recordings of items in the grouped condition were used to construct a total of 60 manipulated audio items with five different levels of manipulation, whereby there were 12 items within each manipulation level. For a given level, either all three cues appeared

maximally (max3), or two cues were maximal while one cue was minimal (maxLR), or two cues were minimal and the third one was maximal (maxL, maxR), or finally, all three cues were minimal (min3). From the spoken coordinates of the ungrouped condition, a total of 30 items was developed with two levels of manipulations, that is 15 audio files in which two cues were maximal while one cue was minimal (maxLR) and 15 audio files in which all three cues were minimal (min3). Table 2 provides an overview of the seven levels of manipulations and on the respective combinations and strengths of prosodic cues.

Table 2. Combinations of prosodic cue values in the different manipulation levels.

| Condition | Manipulation level | f_0 range ^a (R) | Lengthening ^b (L) | Pause ^c | Number of items |
|---------------------|---------------------------|------------------------------|------------------------------|--------------------|-----------------|
| Ungrouped condition | | | | | |
| | Range of extracted values | 3–7 | 27–45 | min = max = 0 | |
| | maxLR | 7 | 45 | 0 | 15 |
| | min3 | 3 | 27 | 0 | 15 |
| Grouped condition | | | | | |
| | Range of extracted values | 9–12 | 36–52 | 3–23 | |
| | max3 | 12 | 52 | 23 | 12 |
| | maxLR | 12 | 52 | 3 | 12 |
| | maxL | 9 | 52 | 3 | 12 |
| | maxR | 12 | 36 | 3 | 12 |
| | min3 | 9 | 36 | 3 | 12 |

Note. R = fundamental frequency (f_0) range; L = lengthening. The combinations of abbreviations identify the respective level of manipulation. Minimal (min) and maximal (max) cue values were extracted from the trained speaker's productions of coordinates in the grouped and ungrouped condition. Respective manipulated items were used in the comprehension experiment. max3 = all three cues appeared maximally; maxLR = lengthening and f_0 range were maximal while pause was minimal; maxL = lengthening was maximal while the two other cues were minimal; maxR = f_0 range was maximal while the two other cues were minimal. min3 = all three cues were minimal.

^a f_0 range on name2 (in semitones). ^bLength/duration of the final vowel of name2 in relation to the overall duration of name2 (in %). ^cPause duration at the prosodic boundary after name2 in relation to overall utterance duration (in %).

To control for order effects of the presented stimuli, two differently ordered lists of stimuli were used in each of our three experiments. In the Comprehension experiment, we also randomized the position of the two pictograms (i.e., the pictogram of the ungrouped condition was either presented on the left side or on the right side of the sheet).

General Procedure

Data collection was performed in two sessions (each approximately 60 min) with a break of at least 1 week in between. The Reading Aloud experiment was always conducted in Session 1 as the first one of three experiments since the two other experiments included audio stimuli of the coordinated name sequences that might have an influence on participants' productions. In the second session, the Comprehension experiment was always performed as the second and the Repetition experiment as the third experiment (see Figure 2).³

During each of the three experiments, participants wore a headset (AKG HSC 271) for presentation of the audio stimuli (for the Comprehension experiment and the Repetition experiment) and for recording of participants' productions. We used an audio interface (Steinberg UR22 MK2) for recording of participants' productions using the software audacity (<https://www.audacity.de/>, Version 2.2.1). Praat software (Boersma & Weenink, 1992) was used for the presentation of the audio stimuli.

Procedure of the Production Experiments

Reading Aloud experiment. Participants were presented with one written name sequence at a time using a Microsoft PowerPoint presentation (<https://www.microsoft.com>) on a Dell laptop computer screen (screen size: 14 in., resolution: 1920 × 1080, font size: 36, font: Calibri) either in the ungrouped condition (names without bracketing/without internal grouping) or in the grouped condition (bracketing/internal grouping of first two names). The written name sequences were always presented along with the matching pictogram (see Figure 2, left panel). Participants were instructed that they would see three written names on the screen and that they should read out the names in such a way, that the interlocutor (i.e., the investigator) could understand precisely, if the three persons would all be coming together or if two persons are coming together and one person is coming alone. There was no further instruction on how to produce the prosodic boundary

and the different prosodic cues were not mentioned during the instructions, but the participants were asked to clearly pronounce the difference between the two conditions. After providing participants with the task instructions, they were asked to read out loud a list of all disyllabic names used in the experiment to familiarize them with all the names. Before the start of the experiment, each participant produced six practice items to become familiar with the task and to ensure adequate task comprehension.

Repetition experiment. Participants heard natural recordings of 24 coordinated name sequences spoken by a phonetically trained speaker. They were instructed to carefully listen to the utterances and to repeat them exactly as they had heard them. If requested by the participant, they could redo an item and the reaction to the second trial was recorded. In case that participants seriously struggled with remembering the names in the presented audio items, they were allowed to replace them with any disyllabic name or to always produce the same name within a coordinate, but to make sure that the perceived pattern was repeated as accurately as possible. Prior to the start of the experiment, participants were presented with four practice items and they were given time to resolve uncertainties.

Procedure of the Comprehension Experiment

Participants were instructed to carefully listen to the manipulated audio stimuli and had to identify whether the stimulus belonged to the condition with or without grouping (grouped or ungrouped). Participants were instructed to decide if the three persons named in an item were all coming together or if the first two persons were grouped together. Participants indicated their decision by pointing to the matching pictogram on a sheet of paper in front of them (see Figure 2, right panel) and these reactions were recorded by the investigator in written form on a response sheet. Before the start of the experiment, participants were presented with four practice items (two per condition) for which nonmanipulated audio files were played to ensure full understanding of the task. If the practice items were not identified correctly, the examiner reinstructed the task and the practice items were repeated. Subsequently, the experimental items were presented with a short break after half of the items. One repetition per audio item was allowed, if requested by the participant.

Data Analysis

Assessment of Condition Accuracy in the Production Experiments

To objectively assess accuracy of the productions in the reading aloud and the Repetition experiment, we ran a rating study involving all name sequences produced by our participants. A group of eleven neurotypical listeners who were naïve to the aims of the study rated the

³We consider a spill-over effect from the Comprehension experiment to the Repetition experiment unlikely, as keeping the audio stimuli in mind for a short time to react to them was already very demanding for our participants. However, we cannot exclude potential order effects.

productions of the LHDP and RHDP with respect to the target condition (grouped vs. ungrouped). The order of the audio recording was pseudorandomized using four differently ordered lists. All raters evaluated all items of one participant one after the other, that is, the productions of each participant were presented block wise. Raters were instructed to wear headphones and to mark their responses in a table. Raters' responses were scored in terms of correct identification of the target condition. If the rater's response was congruent with the target condition the rating was coded as one (otherwise as zero). Therefore, for each production the rating accuracy could be a score between 0 and 11, representing the number of ratings adhering with the intended condition among the 11 raters. Finally, all productions with rating scores above chance level (i.e., > 9 according to a binomial test) were categorized as correct productions (i.e., accuracy = 1) whereas productions with rating scores at nine or below (i.e., not above chance) were categorized as incorrect productions (i.e., accuracy = 0).

Prosodic Analysis of Production Data

Each produced name sequence from the reading aloud and the Repetition experiment (1.368 productions overall) was analyzed using Praat (Boersma & Weenink, 1992) as follows: Firstly, productions were segmented into single phonemes and all produced segments as well as filled and unfilled pauses were annotated by a trained annotator. The highest and lowest points of the f_0 curve on name2 were automatically determined and annotated and checked by the annotator. The annotated values were automatically extracted by Praat and transformed for further analysis. Following Huttenlauch et al. (2021), f_0 range was our measure for the difference between the minimum and maximum f_0 value on name2 in semitones. The durational cues were transformed into relative values, that is, *pause duration* was measured as the duration of the pause after name2 in relation to the duration of the complete utterance, and *final lengthening* was defined as the duration of the final segment (vowel) in name2 relative to the duration of name2. The "und" (and) was annotated separately and only entered the analysis as part of the complete utterance duration and therefore as part of our measure of pause duration. Following Kentner and Féry (2013, p. 288), the f_0 movement on "und" is not considered in the analysis.

Due to their cognitive and language processing difficulties, some participants struggled to produce the name sequences in the intended manner. We excluded items from subsequent analyses, if the phrase structure of the name sequence was disrupted by:

- abortion of the utterance or of single elements within the utterance

- missing elements
- hesitations
- repetition of elements (name1, name2, or *und*)
- voice or speech disturbances (e.g., whispering or stuttering)
- insufficient sound quality of the recordings

Applying these criteria, a total of 80 productions were excluded from all analyses (5.85% of all data; RHDP: 16; LHDP: 64). This was done before the rating study assessing accuracy of the productions. Our subsequent analyses, thus, included 711 productions from the Reading Aloud experiment (RHDP: 404; LHDP: 307) and 577 productions from the Repetition experiment (RHDP: 372, LHDP: 205).

Not all production errors, however, lead to exclusion of the whole item since productions with a continuous speech flow and intonation pattern were included in our analyses despite the following types of production errors: elisions (elision of up to 1/3 of the segments/phonemes), substitutions (substitution of up to 1/3 segments/phonemes), additions (addition of up to 1/3 segments/phonemes), metatheses, distortions, filled pauses/interjections, wrong name (wrong disyllabic trochaic name, e.g., *Leni* instead of *Gabi*), or doubling of the initial or final element (segment, syllable or name; in this case, the erroneously produced segment was cut out before further analysis; e.g., in "*Moni...Moni und Lilli und Lisa*" the first *Moni* was not analyzed). Such errors were annotated on an extra tier in the Praat scripts.

Statistical Analysis

For comprehension data, we ran two logistic regression models on participant accuracy (correct identification of condition: grouped, ungrouped) using the *glmer* function in the *lme4* package in R (Bates et al., 2014). For extracting confidence intervals and *p* values for each model parameter we used the *lmerTest* function (Kuznetsova et al., 2017). We applied a systematic process of model reduction, which we have established in our lab based on procedures and work flows specified in collaboration with colleagues and statisticians within our department. Here, firstly we iteratively reduced model complexity in terms of the random effects structure following the concept of parsimony in mixed models (Bates et al., 2015, see also Matuschek et al., 2017). This was supported by the use of the *RePsychLing* package in R for a random-effects principal components analysis (*rePCA* function, Baayen et al., 2015). Secondly, we validated that a reduced model did not inherit a significant loss in goodness-of-fit. To this, we compared the different reduced models with each other and each of them also to the maximally complex model in terms of both fixed and random effects using the *ANOVA* function for model comparisons in R

(Chambers & Hastie, 1992). If a reduced model showed a better model fit in terms of a smaller Akaike information criterion (Akaike, 1998) it was chosen as the final model. This was also done in cases in which the reduced model did not account for significantly more variance than the previous model in the log-likelihood ratio test because the reduced model still was less complex in its random effects structure (see scripts on OSF, <https://bit.ly/3jzAZL6>, for details). For the comprehension task, we were interested in how accuracy in identifying the correct condition (correct identification coded as 1) was affected by different combinations of cue manipulations (i.e., by the five manipulation levels in the grouped condition: max3, maxLR, maxR, maxL, and min3; and the two manipulation levels in the ungrouped condition: maxLR, min3) and if any effect depended on the participant group (three-level predictor group: CP, LHDP, and RHDP). The models included main effects and the interaction for the predictors group and manipulation level as fixed effects. Furthermore, the random effects structure contained random intercepts both for subjects and items, random slopes for the effect of manipulation within subjects and random slopes for the effect of group within items. For the grouped model with five manipulation levels, the predictors group and manipulation were each coded with sliding-difference (repeated) contrasts that allowed us to directly test the difference between condition means of neighboring predictor levels (Schad et al., 2020). For the predictor group, the difference between condition means of CP and each of the other groups, respectively, was compared. For the model on data from the ungrouped condition with two manipulation levels, contrast coding for group was the same as in the grouped model but the predictor manipulation was coded with scaled sum-to-zero contrasts (min3 coded as -0.5 , maxLR coded as $+0.5$) testing the difference of condition means.

For the analysis of production data, we first ran a logistic regression in R (R Development Core Team, 2018) on participants' production accuracy (categorical, 0 = *incorrect*, 1 = *correct*) using the `glmer` function. The model included main effects and the interaction for the categorical predictors condition (grouped, ungrouped), production accuracy (correct, incorrect), group (LHDP, RHDP), and experiment (reading aloud, repetition) as fixed effects, whereas the random effects structure contained random intercepts for subjects as well as random slopes for the main effect and interaction of condition and experiment within subjects.

For the prosodic measures, we ran three separate linear mixed-effects regression models analyzing the use of each of the three prosodic cues (f_0 range, final lengthening, pause on/after name2) with the `lmer` function in the `lme4` package (Bates et al., 2014). We were interested in how the predictors condition, production accuracy, group,

and experiment influenced the production of each prosodic cue. The model specification based on our research hypotheses was selected prior to any data analysis testing only for outcomes that were justified by theory as described in our hypotheses (even if only exploratory in nature). Besides the main effects per predictor, we specified the interactions as follows: Condition \times Accuracy, Condition \times Group \times Experiment, and Group \times Experiment \times Accuracy. We also included random intercepts for items and subjects and all random slopes in relation to within-item and within-subject variance.

All predictors were coded with scaled sum-to-zero contrasts for all models analyzing production data (first level in alphabetical order coded as $+0.5$, second level coded as -0.5) testing the difference of level means (see Schad et al., 2020). The application of the contrast coding procedures described above allowed us to estimate the respective coefficients as overall effects across the levels of all other predictors, defining the intercept term as the grand mean across all predictor levels; in addition to direct comparisons of two condition means within a predictor. To further explore and extract information on interaction terms, we used the `emmeans` package (Lenth et al., 2018) that allows for automatized comparison of all possible predictor level pairs and interaction contrasts. The required alpha level adjustment for (post hoc) multiple comparisons was applied by means of the Bonferroni method.

Finally, to test for potential correlations of working memory capacities with performance in (a) the Repetition and (b) the Comprehension experiment regarding lesion side (group) and condition, we used Spearman's rho to correlate the digit span composite scores (mean of digit span forwards and backwards percentile ranks) of each participant with (a) their mean production accuracy in the Repetition experiment and (b) their mean production accuracy in the Comprehension experiment.

Results

Results of the Comprehension Experiment

Figure 3 provides an overview of the descriptive data for response accuracies per participant group (CP, LHDP, and RHDP), condition (grouped, ungrouped) and level of manipulation (see Table 2). Statistical analyses are reported per condition below as well as results of the correlational analyses.

Comprehension Accuracy in the Ungrouped Condition

The output of the model (see Table 3) in the ungrouped condition revealed a statistically significant interaction of group and manipulation. Note that the

Figure 3. Response accuracies (correct identification of target condition) in the Comprehension experiment of control participants (CP; red square), participants with left hemisphere brain damage (LHDP; green triangle), and participants with right hemisphere brain damage (RHDP; blue circle) in the grouped condition (left facet), and the ungrouped condition (right facet) per level of manipulation. Bullets = mean. Whiskers = standard error. The horizontal black lines indicate the upper and lower bounds of chance range. Note that the upper and lower bound of chance range are different across the two conditions since chance range is calculated based on the different numbers of items in the two conditions. max3 = all three cues appeared maximally; maxLR = lengthening and f_0 range were maximal while pause was minimal; maxL = lengthening was maximal while the two other cues were minimal; maxR = f_0 range was maximal while the two other cues were minimal; min3 = all three cues were minimal.

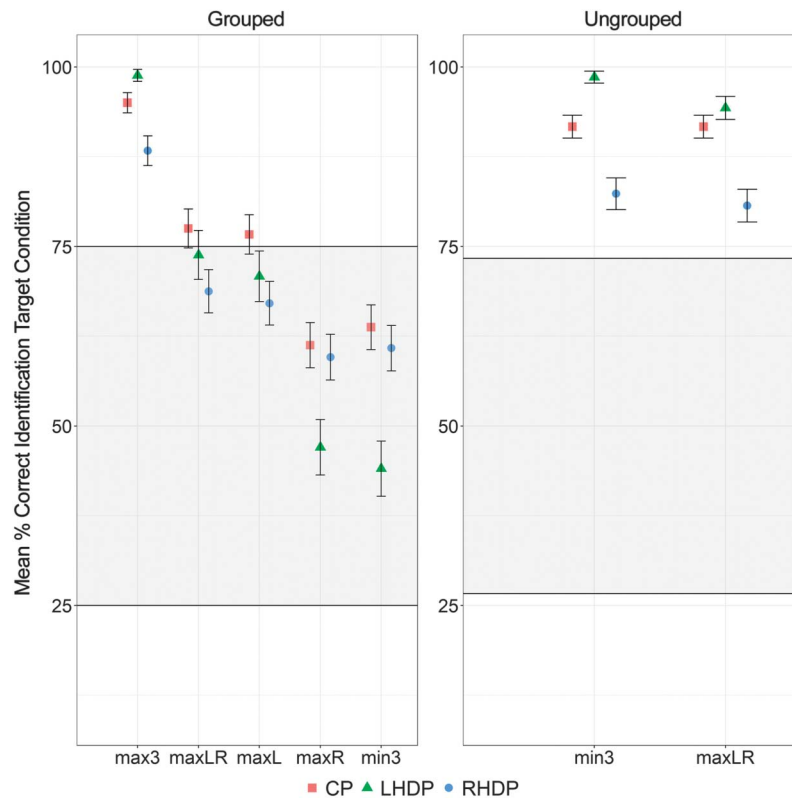


Table 3. Model estimates for accuracy in the ungrouped condition of the Comprehension experiment.

| Predictors | Log odds (CI) | SE | z | p |
|-----------------------------------|---------------|------|-------|------------------|
| (Intercept) | 3.48 | 0.35 | 9.94 | < .001 |
| [mean of all condition levels] | (2.79–4.16) | | | |
| Manip | –0.57 | 0.32 | –1.80 | .071 |
| [maxLR–min3] | (–1.19–0.05) | | | |
| Group | –1.36 | 0.80 | –1.70 | .088 |
| [CP–LHDP] | (–2.93–0.20) | | | |
| Group | –1.08 | 0.64 | –1.68 | .093 |
| [RHDP–CP] | (–2.35–0.18) | | | |
| manip:group | 1.57 | 0.75 | 2.08 | .038 |
| [maxLR–min3:CP–LHDP] | (0.09–3.05) | | | |
| manip:group | –0.18 | 0.44 | –0.41 | .684 |
| [maxLR–min3:RHDP–CP] | (–1.04–0.68) | | | |
| Observations | 1620 | | | |
| Marginal R^2 /conditional R^2 | 0.141/0.569 | | | |

Note. Statistically significant effects are marked in bold ($p < .05$). Manip = manipulation; maxLR = lengthening and f_0 range were maximal while pause was minimal; min3 = all three cues were minimal; CP = control participants; LHDP = participants with left hemisphere brain damage; RHDP = participants with right hemisphere brain damage.

estimates of the model only cover comparisons between predefined contrasts (e.g., LHDP vs. controls, RHDP vs. controls) as overall effects across the levels of the other predictors. The interaction terms for all other levels of predictors (e.g., LHDP vs. RHDP) are extracted in post hoc tests adjusting for multiple comparisons by means of the Bonferroni method. These post hoc tests indicated that for LHDP the difference between the manipulation levels min3 and maxLR was larger than for CP ($z = -2.077$, $p = .038$). Post hoc tests further revealed that LHDP were more accurate than RHDP at the min3 level ($z = 3.283$, $p = .015$). None of the other post hoc comparisons yielded statistically significant results.

Comprehension Accuracy in the Grouped Condition

Overall, the analysis revealed statistically significant main effects for all levels of the factor manipulation (see Table 4 for details). Furthermore, we found a statistically significant interaction between group and manipulation. This was specified in post hoc tests indicating that in

LHDP the difference between the level max3 and maxLR was larger than in CP ($z = 2.241$, $p = .025$). This indicates, that the decrease of response accuracy from max3 to maxLR was stronger in LHDP than in CP. Post hoc tests further revealed the following results: In all three participant groups, the differences between the level that was most clearly indicating a grouped condition (max3, marked by all three cues in their maximal strength) and the following three manipulation levels was statistically significant: max3 > maxL, max3 > maxR, and max3 > min3. In addition, only the LHDP group showed statistically significant differences between the following other manipulation levels: max3 > maxLR and maxLR > min3 (see Table 5, for an overview). That is, the performance of the LHDP group was more affected by the cue manipulations compared with the other two groups.

Effects of Working Memory

When checking for correlations between response accuracies and the digit-span composite scores per group, we found a statistically significant positive correlation

Table 4. Model estimates for accuracy in the grouped condition of the comprehension experiment.

| Predictors | Log odds (CI) | SE | z | p |
|-----------------------------------|-----------------|------|-------|------------------|
| (Intercept) | 1.81 | 0.30 | 6.09 | < .001 |
| [mean of all condition levels] | (1.23–2.39) | | | |
| Manip | –3.02 | 0.72 | –4.19 | < .001 |
| [maxLR–max3] | (–4.43 – –1.61) | | | |
| Manip | –1.47 | 0.41 | –3.60 | < .001 |
| [maxR–maxLR] | (–2.26 – –0.67) | | | |
| Manip | 1.11 | 0.39 | 2.87 | .004 |
| [maxL–maxR] | (0.35–1.87) | | | |
| Manip | –1.03 | 0.39 | –2.67 | .008 |
| [min3–maxL] | (–1.79 – –0.27) | | | |
| Group | 0.25 | 0.67 | 0.37 | .711 |
| [CP–LHDP] | (–1.06–1.56) | | | |
| Group | –0.64 | 0.58 | –1.10 | .274 |
| [RHDP–CP] | (–1.78–0.50) | | | |
| manip:group | 2.65 | 1.18 | 2.24 | .025 |
| [maxLR–max3:CP–LHDP] | (0.33–4.97) | | | |
| manip:group | 0.66 | 0.60 | 1.10 | .272 |
| [maxR–maxLR:CP–LHDP] | (–0.51–1.82) | | | |
| manip:group | –0.56 | 0.54 | –1.03 | .303 |
| [maxL–maxR:CP–LHDP] | (–1.62–0.50) | | | |
| manip:group | 1.02 | 0.55 | 1.87 | .061 |
| [min3–maxL:CP–LHDP] | (–0.05–2.10) | | | |
| manip:group | 0.35 | 0.72 | 0.48 | .632 |
| [maxLR–max3:RHDP–CP] | (–1.07–1.76) | | | |
| manip:group | 0.69 | 0.47 | 1.46 | .143 |
| [maxR–maxLR:RHDP–CP] | (–0.23–1.61) | | | |
| manip:group | –0.62 | 0.42 | –1.47 | .141 |
| [maxL–maxR:RHDP–CP] | (–1.45–0.21) | | | |
| manip:group | 0.51 | 0.42 | 1.22 | .223 |
| [min3–maxL:RHDP–CP] | (–0.31–1.33) | | | |
| Observations | 3240 | | | |
| Marginal R^2 /conditional R^2 | 0.464/NA | | | |

Note. Statistically significant effects are marked in bold ($p < .05$). Manip = manipulation; maxLR = lengthening and f_0 range were maximal while pause was minimal; max3 = all three cues appeared maximally; maxR = f_0 range was maximal while the two other cues were minimal; maxL = lengthening was maximal while the two other cues were minimal; CP = control participants; LHDP = participants with left hemisphere brain damage; RHDP = participants with right hemisphere brain damage; NA = not applicable.

Table 5. Post hoc comparisons between levels of manipulation per participant group for accuracy in the grouped condition of the Comprehension experiment (only comparisons of manipulations levels with statistically significant effects in at least one participant group).

| Manipulation level comparisons | CP | LHDP | RHDP |
|--------------------------------|----------------------------|----------------------------|----------------------------|
| maxLR vs. max3 | n.s. | $z = -3.793$ $p = .016$ | n.s. |
| maxL vs. max3 | $z = -3.567$ $p = .038$ | $z = -4.188$ $p = .003$ | $z = -3.555$ $p = .04$ |
| maxR vs. max3 | $z = -4.610$ $p < .001$ | $z = -5.216$ $p < .001$ | $z = -3.807$ $p = .015$ |
| min3 vs. max3 | $z = -4.268$ $p = .002$ | $z = -5.361$ $p < .001$ | $z = -3.583$ $p = .036$ |
| min3 vs. maxLR | n.s. | $z = -3.691$ $p = .024$ | n.s. |

Note. p value adjustment = Bonferroni method; CP = control participants; LHDP = participants with left hemisphere brain damage; RHDP = participants with right hemisphere brain damage; maxLR = lengthening and f_0 range were maximal while pause was minimal; max3 = all three cues appeared maximally; maxL = lengthening was maximal while the two other cues were minimal; maxR = f_0 range was maximal while the two other cues were minimal. min3 = all three cues were minimal; n.s. = not significant.

between the digit-span composite score and response accuracy for the CP ($R = .64$, $p = .003$) and a tendency for a positive correlation in the RHDP ($r_s = .42$, $p = .065$), indicating higher accuracies in the Comprehension experiment in participants with better working memory in the CP and RHDP. No correlation was found for the LHDP ($r_s = .31$, $p = .28$).

Results of the Production Experiments (Reading Aloud and Repetition, Performed Only in RHDP and LHDP)

Production Accuracy (Reading Aloud and Repetition)

The model (see Table 6, for model output) of the production accuracy revealed a main effect of experiment that

was included in a statistically significant interaction of the factors condition and experiment. These effects were specified in post hoc testing indicating that, in both groups, production accuracy in the Reading Aloud experiment was significantly higher than in the Repetition experiment, but only in the grouped condition. In other words, both groups had the most problems in correctly repeating the prosodically marked structural boundaries. Figure 4 illustrates these results.

Results of Prosodic Analyses (Reading Aloud and Repetition Experiment)

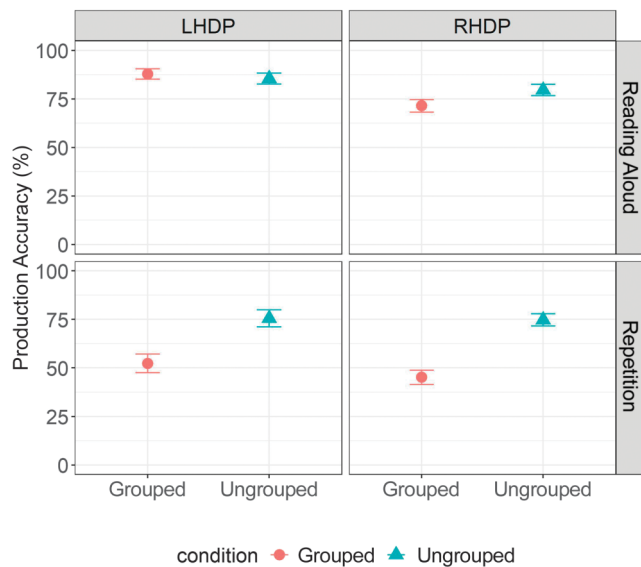
Statistical analyses on the use of the different prosodic cues were separated for correct vs incorrect productions and are described below. Figure 5 illustrates the descriptive data of the three prosodic cues in participants'

Table 6. Model estimates (log-odds scale) for the production accuracy.

| Predictors | Estimates (CI) | SE | z | p |
|---|----------------|------|------|------------------|
| (Intercept) | 1.57 | 0.36 | 4.37 | < .001 |
| [mean of all condition levels] | (0.87–2.27) | | | |
| Condition | 0.19 | 0.66 | 0.28 | .777 |
| [grouped-ungrouped] | (–1.10–1.47) | | | |
| Group | 0.79 | 0.66 | 1.20 | .230 |
| [LHDP-RHDP] | (–0.50–2.09) | | | |
| Experiment | 2.31 | 0.56 | 4.11 | < .001 |
| [read-repeat] | (1.21–3.42) | | | |
| Condition × Group | 1.08 | 1.19 | 0.91 | .362 |
| [grouped-ungrouped:LHDP-RHDP] | (–1.24–3.40) | | | |
| Condition × Experiment | 3.49 | 0.92 | 3.79 | < .001 |
| [grouped-ungrouped:read-repeat] | (1.69–5.30) | | | |
| Group × Experiment | 1.50 | 0.99 | 1.52 | .129 |
| [LHDP-RHDP:read-repeat] | (–0.44–3.43) | | | |
| Condition × Group × Experiment | 2.28 | 1.49 | 1.53 | .125 |
| [grouped-ungrouped: LHDP-RHDP: read-repeat] | (–0.64–5.20) | | | |
| Observations | 1289 | | | |
| Marginal R^2 /conditional R^2 | 0.427/NA | | | |

Note. Statistically significant effects are marked in bold ($p < .05$). LHDP = participants with left hemisphere brain damage; RHDP = participants with right hemisphere brain damage; NA = not applicable.

Figure 4. Accuracy in the two production experiments by condition and participant group. Grouped condition (red circle), ungrouped condition (blue triangle) in participants with left hemisphere brain damage (LHDP; left facet) and participants with right hemisphere brain damage (RHDP; right facet) per experimental task (first row = Reading Aloud experiment, second row = Repetition experiment). Bullets = mean. Whiskers = standard error.



productions per group, divided by experiment, condition, and production accuracy.

f_0 range. For the f_0 cue, we found main effects of condition (grouped > ungrouped) and experiment (reading aloud > repetition) as well as statistically significant two-way interactions of Condition \times Accuracy, Condition \times Experiment, and Group \times Experiment (see Table 7).

The Condition \times Accuracy interaction indicates that in correctly rated items, we found a statistically significant difference in the f_0 range on name2 between the two conditions. On average participants produced a higher f_0 range on name2 in the grouped than in the ungrouped condition. Post hoc tests confirmed the condition effect for correct productions ($t = 9.790$, $p < .001$) and effects of accuracy for both conditions (grouped: $t = 5.225$, $p < .001$; ungrouped: $t = -3.562$, $p = .004$).

The Condition \times Experiment interaction suggests that the increase in f_0 range in the grouped versus ungrouped condition was stronger in the Reading Aloud than in the Repetition experiment. Post hoc tests revealed effects of condition in both experiments (reading aloud: $t = 7.227$, $p < .001$; repetition: $t = 4.696$, $p = .0002$) as well as effects of experiment in both conditions (grouped: $t = 7.016$, $p < .001$; ungrouped: $t = 5.082$, $p < .001$).

The Group \times Experiment interaction suggests that the increase in f_0 range in the Reading Aloud versus

Repetition experiment was stronger in LHDP than in RHDP. The difference in the increase was confirmed post hoc in terms of a group effect in the Reading Aloud experiment ($t = 2.944$, $p = .033$) and an effect of experiment in the group of LHDP ($t = 7.139$, $p < .001$).

Final lengthening. The model for the final lengthening cue (see Table 8) revealed statistically significant main effects of condition (grouped > ungrouped) and group (LHDP > RHDP) as well as statistically significant interactions of Condition \times Accuracy and Experiment \times Group.

The Condition \times Accuracy interaction suggests stronger final lengthening in the grouped than in the ungrouped condition for correct productions. Post hoc tests confirmed these effects ($t = 5.608$, $p < .001$ for the condition-effect in correct productions; for the effect of accuracy in the two conditions: grouped: $t = 4.547$, $p < .001$; ungrouped: $t = -3.949$, $p = .002$).

The Experiment \times Group interaction indicates stronger final lengthening in the Reading Aloud experiment as opposed to the Repetition experiment for RHDP, but not for LHDP. This is supported by post hoc tests revealing a group effect in the Repetition experiment.

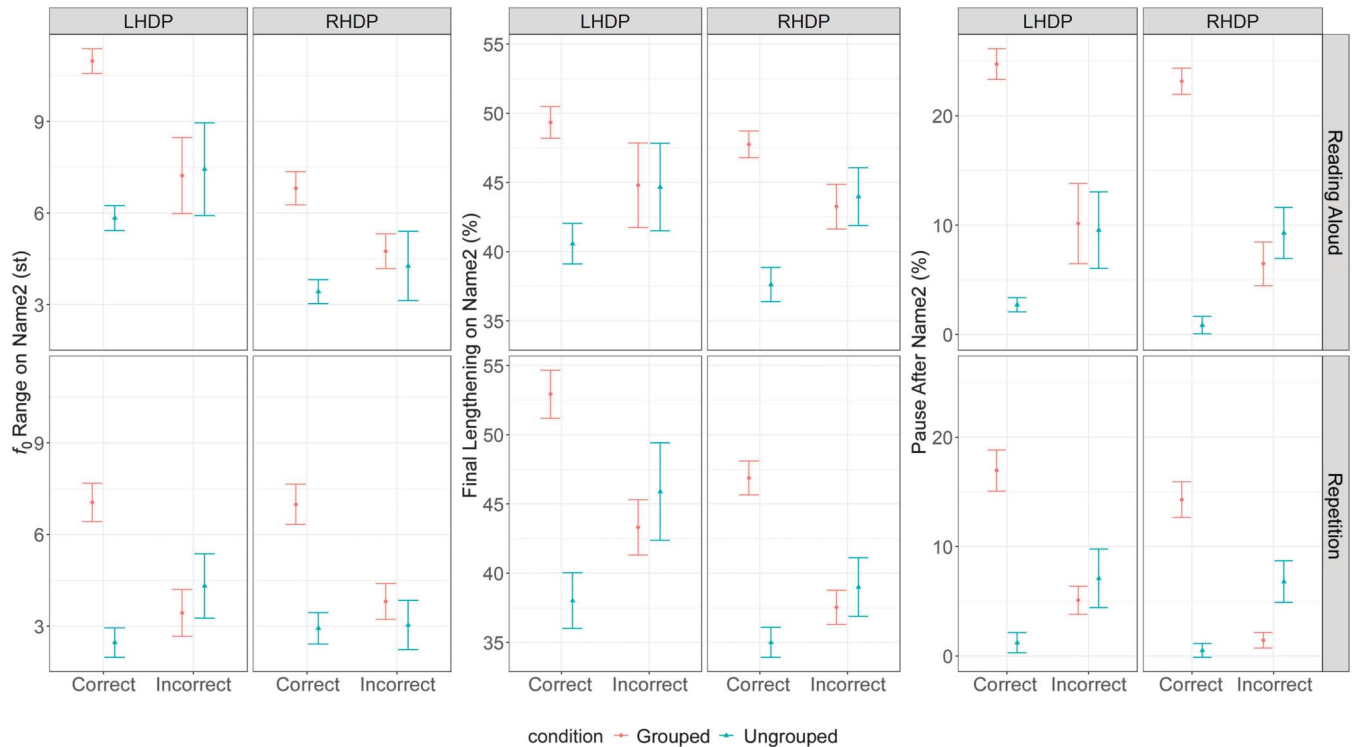
Pause. Table 9 shows the fixed effects extracted from the model exploring the pause cue. There were statistically significant main effects of condition (grouped > ungrouped), accuracy (correct > incorrect), group (LHDP > RHDP), and experiment (reading aloud > repetition), as well as a statistically significant Condition \times Accuracy interaction, indicating an increase of the pause duration in the grouped as opposed to the ungrouped condition for correct productions, but no difference between the two conditions was evident in the incorrect productions. Post hoc tests confirmed that the condition effect was only statistically significant for correct productions ($t = 16.278$, $p < .001$). Effects of accuracy were found for both conditions (grouped: $t = 10.217$, $p < .001$; ungrouped: $t = -6.726$, $p < .001$).

The Condition \times Experiment interaction suggests that there was a stronger difference between the conditions (grouped > ungrouped) in the Reading Aloud experiment as compared with the Repetition experiment. The condition effects were confirmed for both experiments in post hoc tests (reading aloud: $t = 10.547$, $p < .001$; repetition: $t = 6.124$, $p < .001$). An effect of experiment was only found for the grouped condition ($t = 6.601$, $p < .001$), but not for the ungrouped condition, indicating that the differences in the grouped condition were driving the differential effects in the reading aloud versus the Repetition experiment.

Effects of Working Memory

Production accuracy in the repetition experiment did not correlate with the digit-span composite scores

Figure 5. Descriptive data for prosodic cues per condition (grouped, ungrouped), participant group (participants with left hemisphere brain damage [LHDP], participants with right hemisphere brain damage [RHDP]), experiment (reading aloud, repetition) and production accuracy (correct, incorrect). Left facet: f_0 range (difference between highest and lowest f_0 value on name2 in semitones). Mid facet: Final lengthening (duration of the final vowel of name2 relative to duration of name2). Right facet: Pause (pause duration after name2 relative to utterance duration). Bullets = means. Whiskers = standard errors. This figure shows that both groups can clearly produce the three cues to mark the grouping (correct productions). However, incorrect productions reveal that there is inadequate use of each cue in both groups.



for neither group (LHDP: $r_s = .51$, $p = .14$; RHDP: $r_s = .13$, $p = .66$).

Discussion

Summary and Discussion of Main Results in Comprehension (Research Questions 1 and 2)

In the Comprehension experiment, we aimed to investigate if RHDP, LHDP, and CP differ with respect to their accuracy in prosodic boundary identification in coordinate name sequences and whether and how their performance is influenced by the strengths and combinations of the three prosodic cues f_0 range, final lengthening, and pause.

Our H1, assuming that LHDP and RHDP would show overall lower accuracies than CP in using prosodic cues for identification of prosodically marked structural boundaries, is not supported by our data. RHDP were found to have lower accuracies than LHDP and CP only for the items of the ungrouped condition. All other group

differences varied with respect to the strengths and combinations of prosodic cues (i.e., levels of manipulation). Aasland and Baum (2003) as well as Baum et al. (1997) reported lower overall accuracies in RHDP and LHDP as compared with CP for prosodic boundary identification in coordinate color sequences. They found this in natural productions (Baum et al., 1997) as well as with stimuli with manipulated durational cues (Aasland & Baum, 2003). One possible reason for the overall higher accuracies in RHDP and LHDP in our study might lie in the linguistic materials used in the two studies. While Baum et al. and Aasland and Baum used monosyllabic color nouns, we used disyllabic names. It might have been the case that the marking of a prosodic boundary was more easily identified, when the cues unfold on two syllables than on one. This explanation would be in line with the “informative prosodic boundary” account, which assumes that the listener’s interpretation of a prosodic boundary depends on the preceding prosodic information in the utterance they have perceived (Clifton et al., 2002, see also Holzgrefe et al., 2013). To prove or disprove this explanation, thorough investigations of the exact influence of preceding prosodic information on boundary identification in

Table 7. Model estimates for the use of f_0 range on name2 in the production experiments.

| Predictors | Estimates (CI) | SE | <i>t</i> | <i>p</i> |
|---|----------------|------|----------|------------------|
| (Intercept) | 5.26 | 0.36 | 14.61 | < .001 |
| [mean of all condition levels] | (4.56–5.97) | | | |
| Condition | 2.06 | 0.29 | 7.16 | < .001 |
| [grouped-ungrouped] | (1.50–2.63) | | | |
| Accuracy | 0.38 | 0.21 | 1.84 | .066 |
| [correct-incorrect] | (–0.03–0.79) | | | |
| Group | 0.93 | 0.67 | 1.39 | .164 |
| [LHDP-RHDP] | (–0.38–2.25) | | | |
| Experiment | 2.38 | 0.34 | 7.03 | < .001 |
| [read-repeat] | (1.71–3.04) | | | |
| Condition × Accuracy | 3.20 | 0.57 | 5.60 | < .001 |
| [grouped-ungrouped:correct-incorrect] | (2.08–4.33) | | | |
| Condition × Group | 0.05 | 0.60 | 0.09 | .931 |
| [grouped-ungrouped:LHDP-RHDP] | (–1.13–1.23) | | | |
| Condition × Experiment | 0.81 | 0.37 | 2.19 | .029 |
| [grouped-ungrouped:read-repeat] | (0.08–1.53) | | | |
| Group × Experiment | 2.84 | 0.70 | 4.06 | < .001 |
| [LHDP-RHDP:read-repeat] | (1.47–4.22) | | | |
| Accuracy × Group | –0.26 | 0.41 | –0.63 | .530 |
| [correct-incorrect:LHDP-RHDP] | (–1.07–0.55) | | | |
| Accuracy × Experiment | 0.29 | 0.39 | 0.73 | .466 |
| [correct-incorrect:read-repeat] | (–0.48–1.06) | | | |
| Condition × Group × Experiment | 1.42 | 0.82 | 1.72 | .085 |
| [grouped-ungrouped: LHDP-RHDP: read-repeat] | (–0.20–3.04) | | | |
| Accuracy × Group × Experiment | 1.23 | 0.78 | 1.57 | .116 |
| [correct-incorrect: LHDP-RHDP: read-repeat] | (–0.31–2.77) | | | |
| Observations | 1241 | | | |
| Marginal R^2 /conditional R^2 | 0.351/0.607 | | | |

Note. Statistically significant effects are marked in bold ($p < .05$). LHDP = participants with left hemisphere brain damage; RHDP = participants with right hemisphere brain damage.

RHDP and LHDP are needed. Our results do not suggest a general impairment in using prosody for syntactic ambiguity resolution in language comprehension in RHDP or LHDP. This speaks against a strict division of labor between the LH and RH in processing linguistic versus emotional prosody and against the assumption that linguistic prosody is exclusively processed by the LH (as was suggested by the functional load hypothesis, e.g., Chobor & Brown, 1987; Gandour et al., 1995; van Lancker, 1980).

Our H2 was based on the findings by Holzgrefe-Lang et al. (2016) and assumed that, in all participant groups, identification accuracy is better in coordinate sequences in which boundaries are indicated by a combination of final lengthening and f_0 range (manipulation level maxLR) than in trials, in which only lengthening or only f_0 range is maximally present (levels maxL, maxR). However, this was not completely confirmed by our data. In all three groups, accuracy on the level with only lengthening (maxL) was comparable to the level with a combination of lengthening and f_0 range (maxLR). Overall, performance in trials in which the prosodic boundary is indicated mostly by f_0 range (maxR) was poorer than for the level with combined f_0 range and final lengthening (maxLR). Therefore, our data are different from the results of the ERP experiment by Holzgrefe-Lang et al.

(2016), in that the combination of lengthening and f_0 range is necessary for boundary identification, but highlight the informativeness of lengthening information for all groups. This might be due to differences between the cue manipulations applied by Holzgrefe-Lang et al. (2016) and in this study. Holzgrefe-Lang and colleagues inserted the cues final lengthening and f_0 rise into natural productions of coordinate sequences without internal boundaries. In this study, we used natural productions with an internal boundary for the items of the grouped condition (and vice versa) and manipulated the strengths and combinations of the cues at the boundary. Hence, our stimuli were probably more natural with respect to the target conditions (see Hansen et al., 2022) for evidence on the influence of prosodic cues at name1 in coordinate name sequences for boundary identification by neurotypical listeners.

Our third and fourth hypotheses were based on the study of Aasland and Baum (2003) and assumed that both LHDP and RHDP should have particular deficits in boundary identification in the absence of durational cues (H3) and that this is more evident in LHDP than in RHDP (H4). Contradictory to H3, we found generally lower accuracies on the maxR—as opposed to the maxL—level for all three participant groups including CP. We can only speculate about the reasons: Since the CP

Table 8. Model estimates for the use of final lengthening on name2 in the production experiments.

| Predictors | Estimates (CI) | SE | t | p |
|---|-----------------|------|-------|--------|
| (Intercept) | 43.47 | 0.99 | 43.70 | < .001 |
| [mean of all condition levels] | (41.52–45.42) | | | |
| Condition | 4.70 | 1.26 | 3.73 | < .001 |
| [grouped-ungrouped] | (2.23–7.17) | | | |
| Accuracy | –0.11 | 0.68 | –0.16 | .873 |
| [correct-incorrect] | (–1.43–1.22) | | | |
| Group | 3.98 | 1.69 | 2.35 | .019 |
| [LHDP-RHDP] | (0.65–7.30) | | | |
| Experiment | 1.72 | 1.39 | 1.24 | .215 |
| [read-repeat] | (–1.00–4.44) | | | |
| Condition × Accuracy | 9.51 | 1.64 | 5.79 | < .001 |
| [grouped-ungrouped: correct-incorrect] | (6.29–12.73) | | | |
| Condition × Group | –0.44 | 1.54 | –0.29 | .773 |
| [grouped-ungrouped: LHDP-RHDP] | (–3.47–2.58) | | | |
| Condition × Experiment | 0.20 | 2.29 | 0.09 | .931 |
| [grouped-ungrouped: read-repeat] | (–4.30–4.70) | | | |
| Group × Experiment | –4.25 | 1.79 | –2.38 | .017 |
| [LHDP-RHDP: read-repeat] | (–7.76 – –0.75) | | | |
| Accuracy × Group | –1.73 | 1.25 | –1.38 | .168 |
| [correct-incorrect: LHDP-RHDP] | (–4.19–0.73) | | | |
| Accuracy × Experiment | –0.75 | 1.36 | –0.55 | .582 |
| [correct-incorr:read-repeat] | (–3.41–1.92) | | | |
| Condition × Group × Experiment | –2.22 | 2.17 | –1.02 | .308 |
| [grouped-ungrouped: LHDP-RHDP: read-repeat] | (–6.48–2.04) | | | |
| Accuracy × Group × Experiment | 4.34 | 2.61 | 1.66 | .097 |
| [correct-incorrect: LHDP-RHDP: read-repeat] | (–0.78–9.47) | | | |
| Observations | 1289 | | | |
| Marginal R^2 /conditional R^2 | 0.434/NA | | | |

Note. Statistically significant effects are marked in bold ($p < .05$). LHDP = participants with left hemisphere brain damage; RHDP = participants with right hemisphere brain damage; NA = not applicable.

had problems with the maxR stimuli, too, one possible reason could be that the isolated presentation of the f_0 cue without the lengthening cue was just too difficult to process. In the study by Aasland and Baum (2003), no stimuli were included in which pitch was the most prominent prosodic cue to be processed. This might have led to the differential results. However, our latter hypothesis (H4) could be partly confirmed: only LHDP show, indeed, that they profit from the presence of the durational pause cue more than the control group: The difference in performance between trials in which the pause cue is maximal versus minimal (max3 vs. maxLR) is larger in LHDP than in CP.

Summary and Discussion of Main Results in Production (Research Questions 3 and 4)

We aimed to figure out how RHDP and LHDP make use of the prosodic cues f_0 range, final lengthening, and pause for prosodic boundary marking in the production of coordinate name sequences. We were further looking into task-specific effects by comparing performance in a reading aloud and a Repetition experiment.

In reading aloud, LHDP show results that are comparable to a group of unimpaired young CPs (Huttenlauch et al., 2021), whereas RHDP show a tendency to make

more errors than young controls (data reported in Huttenlauch et al., 2021). In addition, we found an interaction of condition and experiment for both groups: For both, LHDP and RHDP, production accuracy was worse for grouped than ungrouped items in the Repetition experiment. This indicates that both groups have task-dependent difficulties in using prosodic cues for prosodic marking of a boundary in production: When the task involves comprehension as well as production of prosodic cues, as is the case in repetition, marking of the boundary in grouped items is difficult. The observed differences between the reading aloud and the Repetition experiment were initially only assumed for RHDP (H6), given the higher working memory demands in the repetition task (Blake, 2017). For the Reading Aloud experiment, the stimuli were presented in their written form with a bracketing of the names indicating the to-be-produced prosodic structure. For the Repetition experiment, participants had to listen to the stimuli, to analyze their structure (boundary, no boundary), and to memorize the lexical items as well as the structure (boundary, no boundary) before uttering the sequence. Therefore, the Repetition experiment is cognitively more demanding than the Reading Aloud experiment and this is clearly mirrored in the data of RHDP. For LHDP, we did not make specific

Table 9. Model estimates for the use of the pause cue after name2 in the two production experiments.

| Predictors | Estimates (CI) | SE | t | p |
|---|----------------|------|-------|--------|
| (Intercept) | 8.97 | 0.56 | 16.06 | < .001 |
| [mean of all condition levels] | (7.88–10.07) | | | |
| Condition | 8.44 | 0.73 | 11.51 | < .001 |
| [grouped-ungrouped] | (7.00–9.87) | | | |
| Accuracy | 2.02 | 0.64 | 3.16 | .002 |
| [correct-incorrect] | (0.77–3.27) | | | |
| Group | 1.92 | 0.86 | 2.24 | .025 |
| [LHDP-RHDP] | (0.24–3.60) | | | |
| Experiment | 4.81 | 0.78 | 6.18 | < .001 |
| [read-repeat] | (3.28–6.34) | | | |
| Condition × Accuracy | 17.47 | 1.51 | 11.59 | < .001 |
| [grouped-ungrouped: correct-incorrect] | (14.52–20.43) | | | |
| Condition × Group | 1.31 | 1.46 | 0.90 | .369 |
| [grouped-ungrouped: LHDP-RHDP] | (–1.55–4.16) | | | |
| Condition × Experiment | 6.98 | 1.22 | 5.74 | < .001 |
| [grouped-ungrouped: read-repeat] | (4.60–9.37) | | | |
| Group × Experiment | –0.59 | 1.52 | –0.39 | .698 |
| [LHDP-RHDP: read-repeat] | (–3.57–2.39) | | | |
| Accuracy × Group | 0.12 | 1.28 | 0.09 | .926 |
| [correct-incorrect: LHDP-RHDP] | (–2.39–2.62) | | | |
| Accuracy × Experiment | –0.15 | 0.92 | –0.17 | .867 |
| [correct-incorrect: read-repeat] | (–1.96–1.66) | | | |
| Condition × Group × Experiment | –0.48 | 2.40 | –0.20 | .841 |
| [grouped-ungrouped: LHDP-RHDP: read-repeat] | (–5.18–4.22) | | | |
| Accuracy × Group × Experiment | 2.46 | 1.84 | 1.34 | .182 |
| [correct-incorrect: LHDP-RHDP: read-repeat] | (–1.15–6.07) | | | |
| Observations | 1289 | | | |
| Marginal R^2 /conditional R^2 | 0.698/0.777 | | | |

Note. Statistically significant effects are marked in bold ($p < .05$). LHDP = participants with left hemisphere brain damage; RHDP = participants with right hemisphere brain damage.

hypotheses on task effects, because so far, studies on this issue are lacking. In addition, both reading aloud as well as repetition could be affected by LH lesions: Limited reading capacities could impact on the reading-aloud task and limited working memory capacities as well as impairments in receptive phonology could impact on the repetition task, although a recent study found that limited working memory in LHDP is not associated with deficits in processing utterances with regular sentence-level prosody (LaCroix et al., 2020). However, our data do show task-specific differences in LHDP that are similar to those found in RHDP. We, thus, assume that for both groups, the finding of task-specific effects in using prosodic cues for boundary marking can be related to the higher working memory demands of the repetition task. In the audio stimuli of our Repetition experiment, participants were presented with natural (i.e., nonmanipulated) productions of a trained speaker. We acknowledge that the observed task-specific differences between the Reading aloud and the Repetition experiments might have been influenced by this audio input. More precisely, participants potentially adjusted their use of prosodic cues to the strength of the prosodic cues in the audio “model stimuli.” We, therefore, checked the cue strength of our participants’ productions for a relation to the cues of the experimental model

stimuli. First, the cues in the model stimuli are clearly different for the two conditions (grouped vs. ungrouped). Second, for f_0 range and final lengthening there was no evidence for an influence of the model stimuli on the participants’ productions. However, for the pause cue, we found that mean pause durations produced by RHDP and LHDP in the Repetition experiment were similar to the mean pause durations in the audio model stimuli and shorter than in the Reading experiment. But the range of pause durations in the participants’ productions was much larger than in the model stimuli, that is, pause duration is more variable in the participants’ productions. We would, therefore, rule out the possibility, that an influence of the model stimuli on participants’ productions was the main reason for the task-specific differences.

With respect to the different prosodic cues, we hypothesized (H5) that RHDP and LHDP would be able to produce all three prosodic cues under investigation (f_0 range, final lengthening, and pause). This is clearly supported by our results. In our analysis of the production data, we took the production accuracy into account and therefore differentiated between correct and incorrect productions to figure out, if the LH or RH stroke specifically impacts the adequate or inadequate use of the single prosodic cues. For correct productions in both groups, we

found that all three cues were increased in the grouped as opposed to the ungrouped condition. In incorrect trials, however, speakers produced *increased* prosodic cues in *ungrouped* items and *decreased* prosodic cues in *grouped* items (i.e., they deviated from the pattern that would be necessary for correct production and comprehension of the boundary). In other words, the raters, who evaluated the accuracy of participants' productions could successfully recover the intended structure (grouped, ungrouped) in those trials, in which the speakers indicated the difference between grouped and ungrouped items by an increase of f_0 range, final lengthening, and pause on/after name2.

Overall, there were 924 correct productions (72%) and 364 incorrect productions (28%), which indicates that RHDP and LHDP are able to employ all three prosodic cues under investigation, and they do so at least in the majority of trials (in around 72% of the cases). However, the high number of incorrect productions also indicates that production of linguistic prosody poses some difficulties for the participants.

Concerning task-specific effects in the use of the different cues, both groups used more f_0 range, more lengthening, and a longer pause in the Reading Aloud versus the Repetition experiment, but the strength of the cues was different between the groups: LHDP used more f_0 range and longer pauses than RHDP, whereas RHDP used more final lengthening than LHDP in the Reading Aloud experiment. Our findings that RHDP and LHDP are able to produce durational prosodic cues as well as f_0 range for syntactic disambiguation are in line with previous studies (e.g., Baum et al., 2001; Baum & Pell, 1997). When participants had to produce prosodic boundaries in sequences like “pink and black and green,” which are comparable to our stimuli, both groups used more final lengthening than a control group and LHDP produced longer pauses than the control group (Baum et al., 1997). Although we have no direct control group for our production experiment (but can relate our findings to those of healthy young controls in Huttenlauch et al.'s [2021] study who used very similar stimuli), it is important to note that LHDP in this study also produced longer pauses than RHDP. This cannot be explained by a potentially general slower speaking style in the individuals with aphasia, as all cues in our study (and in Baum et al., 1997) were relative measures (i.e., the pause at the prosodic boundary after name2 was measured in relation to overall utterance duration or speaking rate).

Discussion of Results in Relation to Lateralization Theories

With respect to the lateralization of prosody processing in the brain, cue-dependent theories (e.g., Van Lancker & Sidsis, 1992; Zatorre & Belin, 2001) would assume that the LH is mainly concerned with processing of durational cues while the RH supports processing of spectral cues.

Under this assumption, LHDP should show decreased sensitivity to durational cues; however, in our study, LHDP showed clearly worse performance on stimuli in which the pause cue was minimized compared with trials that included maximal pause. In line with Aasland and Baum's (2003) study, this finding can be considered as an indication that LHDP are sensitive to durational cues and need the pause information to correctly disambiguate the stimuli. Furthermore, since we did not find an overall deficit in prosody processing in RHDP, our data do not support the “asymmetric sampling in time” model or the “dynamic dual pathway model,” which assume that the RH is responsible for processing of slower changing acoustic parameters, like prosody at the sentence level (e.g., Friederici & Alter, 2004; Poeppel, 2003). However, conclusions regarding lateralization of prosody processing in the brain should be drawn with caution here, since we do not provide detailed analyses of the location or size of the lesions in our sample as well as their relation to participants' performance. The lesion overlays, which we provide in Figure 1 are of merely illustrative value. We did not perform a formal lesion to symptom correlation. This is due to the small sample size and the fact that the groups of participants in whom a lesion delineation was possible was only 10 for the (smaller) LHDP group. It would be interesting, however, in a larger sample, to analyze lesion patterns more specifically (e.g., by a region-based approach as available in some of the packages for lesion-behavior correlations). This might provide insights into whether lesions to specific areas of the left language network and/or their RH homologues will specifically impair aspects of prosodic comprehension and/or production. This may be also informative for overall theories regarding the functional anatomy underlying the different dimensions of prosodic functions (see Sammler et al., 2015, for a more detailed discussion).

Limitations

With respect to the link between prosody production and comprehension, our current study focused on group level results. We did not perform a detailed analysis of the connection between comprehension and production within single individuals, but we correlated comprehension and production scores of the individual speakers in both groups. For LHDP, we found a positive correlation of the comprehension scores and the production accuracy scores in the repetition task for items of the grouped condition ($R = .86$, $p = .0016$), that is, participants who are good in comprehension are also good in repetition. Given the small number of participants, we refrain from discussing this correlation in greater depth. There were no correlations for the reading aloud task or for the group of RHDP. Future research should explore possible dissociations between perception and production of prosodic cues

in single cases. In a similar vein, we have to interpret the lateralization data with great care, since we did not perform a lesion-symptom mapping that would have allowed us to determine particular brain regions relevant for processing of specific prosodic cues. In addition, the lengthening of other parts of the word, such as the first syllable could be analyzed in future work.

Conclusions and Clinical Implications

Overall, our data neither support an exclusive role of the LH nor of the RH for processing of linguistic prosody. RHDP and LHDP show limitations in the comprehension and production of prosodic cues, but no complete impairment of prosodic processing is evident in either of the two groups. Our data furthermore indicate that individual impairments in the use of the single prosodic cues in the different modalities should be considered in future research and in clinical practice (see Hawthorne & Fischer, 2020, for a recent survey on the role of prosody for speech-language therapy, which mainly focuses on language and communication impairments differing from those investigated in this study). In fact, the use of prosody might serve as a good resource for communication. In LHDP, especially the use of durational cues (mainly pause, but also final lengthening) might constitute a resource to increase the comprehensibility of syntactic constructions by marking syntactic junctures, but also to emphasize semantically relevant elements. In production, both groups were principally able to produce the typical prosodic cues to mark prosodic boundaries, but their performance varies: in about 30% of the trials, production of linguistic prosody was incorrect, that is, the target condition could not be recovered by neurotypical listeners. Therefore, speech and language therapists should be aware of this variability in performance and of possible deficits in using prosodic cues for structural disambiguation in LHDP as well as in RHDP.

Data Availability Statement

Our analyzed data and the scripts of our statistical analyses are available on OSF: <https://bit.ly/3jzAZL6>.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) Project ID 317 633 480 - SFB 1287 (PIs: S.H. and I.W.). The authors thank Kim Lawatsch for her support during data collection. They further thank Marie Hansen and Kathleen Schneider for helping with data collection and

analysis. They also acknowledge Anna Laurinavichyute for her help with the statistical analyses.

References

- Aasland, W. A., & Baum, S. R. (2003). Temporal parameters as cues to phrasal boundaries: A comparison of processing by left- and right-hemisphere brain-damaged individuals. *Brain and Language*, 87(3), 385–399. [https://doi.org/10.1016/S0093-934X\(03\)00138-X](https://doi.org/10.1016/S0093-934X(03)00138-X)
- Aebi, C. (2002). *Validierung der neuropsychologischen Testbatterie CERAD-NP: Eine Multi-Center Studie* [Doctoral dissertation, University of Basel]. https://www.memoryclinic.ch/fileadmin/user_upload/Memory_Clinic/Literatur/2002/Aebi_2002.pdf
- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). Springer. https://doi.org/10.1007/978-1-4612-1694-0_15
- Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, 26(3), 839–851. <https://doi.org/10.1016/j.neuroimage.2005.02.018>
- Baayen, H., Bates, D., Kliegl, R., & Vasishth, S. (2015). *RePsychLing: Data sets from psychology and linguistics experiments*, (Version R package version 0.0.4) [Computer software].
- Balasubramanian, V., & Max, L. (2005). Hemispheric processing of prosody. In J. Syka & M. M. Merzenich (Eds.), *Plasticity and signal representation in the auditory system* (pp. 271–276). Springer. https://doi.org/10.1007/0-387-23181-1_25
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. <http://arxiv.org/pdf/1506.04967v2>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2014). *Fitting linear mixed-effects models using lme4*. <http://arxiv.org/pdf/1406.5823v1>
- Baum, S. R., & Dwivedi, V. D. (2003). Sensitivity to prosodic structure in left- and right-hemisphere-damaged individuals. *Brain and Language*, 87(2), 278–289. [https://doi.org/10.1016/S0093-934X\(03\)00109-3](https://doi.org/10.1016/S0093-934X(03)00109-3)
- Baum, S. R., & Pell, M. D. (1997). Production of affective and linguistic prosody by brain-damaged patients. *Aphasiology*, 11(2), 177–198. <https://doi.org/10.1080/02687039708248463>
- Baum, S. R., Pell, M. D., Leonard, C. L., & Gordon, J. K. (1997). The ability of right- and left-hemisphere-damaged individuals to produce and interpret prosodic cues marking phrasal boundaries. *Language and Speech*, 40(4), 313–330. <https://doi.org/10.1177/002383099704000401>
- Baum, S. R., Pell, M. D., Leonard, C. L., & Gordon, J. K. (2001). Using prosody to resolve temporary syntactic ambiguities in speech production: Acoustic data on brain-damaged speakers. *Clinical Linguistics & Phonetics*, 15(6), 441–456. <https://doi.org/10.1080/02699200110044813>
- Bélanger, N., Baum, S. R., & Titone, D. (2009). Use of prosodic cues in the production of idiomatic and literal sentences by individuals with right- and left-hemisphere damage. *Brain and Language*, 110(1), 38–42. <https://doi.org/10.1016/j.bandl.2009.02.001>
- Bennett, R., & Elfner, E. (2019). The syntax–prosody interface. *Annual Review of Linguistics*, 5(1), 151–171. <https://doi.org/10.1146/annurev-linguistics-011718-012503>
- Blake, M. L. (2017). Right-hemisphere pragmatic disorders. In L. Cummings (Ed.), *Research in Clinical Pragmatics. Perspectives in Pragmatics, Philosophy & Psychology* (Vol. 11, pp. 243–266). Springer. https://doi.org/10.1007/978-3-319-47489-2_10
- Boersma, P., & Weenink, D. (1992). *Praat: Doing phonetics by computer*, (Version 6–1-12) [Computer software].

- Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech*, 49(4), 451–493. <https://doi.org/10.1177/00238309060490040201>
- Brett, M., Leff, A. P., Rorden, C., & Ashburner, J. (2001). Spatial normalization of brain images with focal lesions using cost function masking. *NeuroImage*, 14(2), 486–500. <https://doi.org/10.1006/nimg.2001.0845>
- Chambers, J. M., & Hastie, T. J. (1992). *Statistical Models in S* (Version R package ‘stats’ Version 4.0.5) [Computer software]. Wadsworth & Brooks/Cole Advanced Books & Software.
- Chobor, K. L., & Brown, J. W. (1987). Phoneme and timbre monitoring in left and right cerebrovascular accident patients. *Brain and Language*, 30(2), 278–284. [https://doi.org/10.1016/0093-934X\(87\)90103-9](https://doi.org/10.1016/0093-934X(87)90103-9)
- Clifton, C., Jr., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, 45(2), 87–114. <https://doi.org/10.1177/00238309020450020101>
- 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>
- Colsher, P. L., Cooper, W. E., & Graff-Radford, N. (1987). International variability in the speech of right-hemisphere damaged patients. *Brain and Language*, 32(2), 379–383. [https://doi.org/10.1016/0093-934X\(87\)90134-9](https://doi.org/10.1016/0093-934X(87)90134-9)
- Cooper, W. E., Soares, C., Nicol, J., Michelow, D., & Goloskie, S. (1984). Clausal intonation after unilateral brain damage. *Language and Speech*, 27(1), 17–24. <https://doi.org/10.1177/002383098402700102>
- DeDe, G. (2012). Lexical and prosodic effects on syntactic ambiguity resolution in aphasia. *Journal of Psycholinguistic Research*, 41(5), 387–408. <https://doi.org/10.1007/s10936-011-9191-1>
- Dykstra, K., Gandour, J., & Stark, R. E. (1995). Disruption of prosody after frontal lobe seizures in the non-dominant hemisphere. *Aphasiology*, 9(5), 453–476. <https://doi.org/10.1080/02687039508248709>
- Frazier, L., Carlson, K., & Clifton, C., Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, 10(6), 244–249. <https://doi.org/10.1016/j.tics.2006.04.002>
- Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*, 89(2), 267–276. [https://doi.org/10.1016/S0093-934X\(03\)00351-1](https://doi.org/10.1016/S0093-934X(03)00351-1)
- Gandour, J., Larsen, J., Dechongkit, S., Ponglorpisit, S., & Khunadorn, F. (1995). Speech prosody in affective contexts in Thai patients with right hemisphere lesions. *Brain and Language*, 51(3), 422–443. <https://doi.org/10.1006/brln.1995.1069>
- Geigenberger, A., & Ziegler, W. (2001). Receptive prosodic processing in aphasia. *Aphasiology*, 15(12), 1169–1187. <https://doi.org/10.1080/02687040143000555>
- Hansen, M., Huttenlauch, C., De Beer, C., Wartenburger, I., & Hanne, S. (2022). *Individual differences in early disambiguation of prosodic grouping*. *Language and Speech*. <https://doi.org/10.1177/00238309221127374>
- Härting, C., Markowitsch, H.-J., Neufeld, H., Calabrese, P., Deisinger, K., & Kessler, J. (2000). *WMS-R: Wechsler Gedächtnistest - Revidierte Fassung* [WMS-R: Wechsler Memory Scale-Revised]. Hogrefe.
- Hawthorne, K., & Fischer, S. (2020). Speech-language pathologists and prosody: Clinical practices and barriers. *Journal of Communication Disorders*, 87, 106024. <https://doi.org/10.1016/j.jcomdis.2020.106024>
- Hird, K., & Kirsner, K. (1993). Dysprosody following acquired neurogenic impairment. *Brain and Language*, 45(1), 46–60. <https://doi.org/10.1006/brln.1993.1032>
- Holzgrefe, J., Wellmann, C., Petrone, C., Truckenbrodt, H., Höhle, B., & Wartenburger, I. (2013). Brain response to prosodic boundary cues depends on boundary position. *Frontiers in Psychology*, 4, 421. <https://doi.org/10.3389/fpsyg.2013.00421>
- Holzgrefe-Lang, J., Wellmann, C., Petrone, C., Råling, 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>
- Hoyte, K. J., Kim, A., Brownell, H., & Wingfield, A. (2004). Effects of right hemisphere brain injury on the use of components of prosody for syntactic comprehension. *Brain and Language*, 91(1), 168–169. <https://doi.org/10.1016/j.bandl.2004.06.087>
- Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1983). *Aachen Aphasia Test (AAT)*. Hogrefe.
- Hughes, C. P., Chan, J. L., & Su, M. S. (1983). Aprosodia in Chinese patients with right cerebral hemisphere lesions. *Archives of Neurology*, 40(12), 732–736. <https://doi.org/10.1001/archneur.1983.04050110050007>
- Huttenlauch, C., de Beer, C., Hanne, S., & Wartenburger, I. (2021). Production of prosodic cues in coordinate name sequences addressing varying interlocutors. *Journal of the Association for Laboratory Phonology*, 12(1), 1. <https://doi.org/10.5334/labphon.221>
- Kalbe, E., Reinhold, N., & Kessler, J. (2005). *Aphasie-Check-Liste (ACL)*. ProLog.
- Kentner, G., & Féry, C. (2013). A new approach to prosodic grouping. *The Linguistic Review*, 30(2). <https://doi.org/10.1515/tlr-2013-0009>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTestPackage: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- LaCroix, A. N., Blumenstein, N., Tully, M., Baxter, L. C., & Rogalsky, C. (2020). Effects of prosody on the cognitive and neural resources supporting sentence comprehension: A behavioral and lesion-symptom mapping study. *Brain and Language*, 203, 104756. <https://doi.org/10.1016/j.bandl.2020.104756>
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). *Package “Emmeans”* (R package version 4.0.3) [Computer software].
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Perkins, J. M., Baran, J. A., & Gandour, J. (1996). Hemispheric specialization in processing intonation contours. *Aphasiology*, 10(4), 343–362. <https://doi.org/10.1080/02687039608248416>
- Petermann, F. (2012). *Wechsler Adult Intelligence Scale (4th ed.)*; *Deutschsprachige adaptation nach David Wechsler*. Pearson.
- Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgrefe-Lang, J., Wartenburger, I., & Höhle, B. (2017). Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics*, 61, 71–92. <https://doi.org/10.1016/j.wocn.2017.01.002>
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication*, 41(1), 245–255. [https://doi.org/10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3)

- R Development Core Team.** (2018). *R: A language and environment for statistical computing* (Version R version 4.0.5 (2021-03-31)) [Computer software]. R Foundation for Statistical Computing.
- Rorden, C., & Brett, M.** (2000). Stereotaxic display of brain lesions. *Behavioural Neurology*, 12(4), Article 421719. <https://doi.org/10.1155/2000/421719>
- Sammler, D., Grosbras, M.-H., Anwender, A., Bestelmeyer, P. E. G., & Belin, P.** (2015). Dorsal and ventral pathways for prosody. *Current Biology: CB*, 25(23), 3079–3085. <https://doi.org/10.1016/j.cub.2015.10.009>
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R.** (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110, 104038. <https://doi.org/10.1016/j.jml.2019.104038>
- Seddoh, S. A.** (2008). Conceptualisation of deviations in intonation production in aphasia. *Aphasiology*, 22(12), 1294–1312. <https://doi.org/10.1080/02687030701640289>
- Shah, A. P., Baum, S. R., & Dwivedi, V. D.** (2006). Neural substrates of linguistic prosody: Evidence from syntactic disambiguation in the productions of brain-damaged patients. *Brain and Language*, 96(1), 78–89. <https://doi.org/10.1016/j.bandl.2005.04.005>
- Sheppard, S. M., Love, T., Midgley, K. J., Holcomb, P. J., & Shapiro, L. P.** (2017). Electrophysiology of prosodic and lexical-semantic processing during sentence comprehension in aphasia. *Neuropsychologia*, 107, 9–24. <https://doi.org/10.1016/j.neuropsychologia.2017.10.023>
- Sheppard, S. M., Love, T., Midgley, K. J., Shapiro, L. P., & Holcomb, P. J.** (2019). Using prosody during sentence processing in aphasia: Evidence from temporal neural dynamics. *Neuropsychologia*, 134, 107197. <https://doi.org/10.1016/j.neuropsychologia.2019.107197>
- Shipley-Brown, F., Dingwall, W. O., Berlin, C. I., Yeni-Komshian, G., & Gordon-Salant, S.** (1988). Hemispheric processing of affective and linguistic intonation contours in normal subjects. *Brain and Language*, 33(1), 16–26. [https://doi.org/10.1016/0093-934X\(88\)90051-X](https://doi.org/10.1016/0093-934X(88)90051-X)
- Sidtis, J. J., & Van Lancker Sidtis, D.** (2003). A neurobehavioral approach to dysprosody. *Seminars in Speech and Language*, 24(2), 093–106. <https://doi.org/10.1055/s-2003-38901>
- Stadie, N., Cholewa, J., & de Bleser, R.** (2013). *LEMO 2.0: Lexikon modellorientiert: Diagnostik für Aphasie, Dyslexie und Dysgraphie*. NAT-Verlag.
- Steinhauer, K., Alter, K., & Friederici, A. D.** (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2(2), 191–196. <https://doi.org/10.1038/5757>
- van Lancker, D.** (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Paper in Linguistics*, 13(2), 201–277. <https://doi.org/10.1080/08351818009370498>
- Van Lancker, D., & Sidtis, J. J.** (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research*, 35(5), 963–970. <https://doi.org/10.1044/jshr.3505.963>
- van Ommen, S., Boll-Avetisyan, N., Larraza, S., Wellmann, C., Bijeljac-Babic, R., Höhle, B., & Nazzi, T.** (2020). Language-specific prosodic acquisition: A comparison of phrase boundary perception by French- and German-learning infants. *Journal of Memory and Language*, 112, 104108. <https://doi.org/10.1016/j.jml.2020.104108>
- Wagner, M.** (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory*, 28(1), 183–237. <https://doi.org/10.1007/s11049-009-9086-0>
- Walker, J. P., Daigle, T., & Buzzard, M.** (2002). Hemispheric specialisation in processing prosodic structures: Revisited. *Aphasiology*, 16(12), 1155–1172. <https://doi.org/10.1080/02687030244000392>
- Walker, J. P., Fongemie, K., & Daigle, T.** (2001). Prosodic facilitation in the resolution of syntactic ambiguities in subjects with left and right hemisphere damage. *Brain and Language*, 78(2), 169–196. <https://doi.org/10.1006/brln.2001.2454>
- Walker, J. P., Joseph, L., & Goodman, J.** (2009). The production of linguistic prosody in subjects with aphasia. *Clinical Linguistics & Phonetics*, 23(7), 529–549. <https://doi.org/10.1080/02699200902946944>
- Walker, J. P., Pelletier, R., & Reif, L.** (2004). The production of linguistic prosodic structures in subjects with right hemisphere damage. *Clinical Linguistics & Phonetics*, 18(2), 85–106. <https://doi.org/10.1080/02699200310001596179>
- Wunderlich, A., Ziegler, W., & Geigenberger, A.** (2003). Implicit processing of prosodic information in patients with left and right hemisphere stroke. *Aphasiology*, 17(9), 861–879. <https://doi.org/10.1080/02687030344000283>
- Yang, S.-Y., & Van Lancker-Sidtis, D.** (2016). Production of Korean idiomatic utterances following left- and right-hemisphere damage: Acoustic studies. *Journal of Speech, Language, and Hearing Research*, 59(2), 267–280. https://doi.org/10.1044/2015_JSLHR-L-15-0109
- Zatorre, R. J., & Belin, P.** (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 11(10), 946–953. <https://doi.org/10.1093/cercor/11.10.946>
- Zhang, X.** (2012). *A comparison of cue-weighting in the perception of prosodic phrase boundaries in English and Chinese* [Doctoral dissertation, University of Michigan].
- Ziegler, W., Aichert, I., Staiger, A., & Schimeczek, M.** (2020). *HWL-kompakt*. <https://neurophonetik.de/sprechapraxie-wortlisten>
- Ziegler, W., Schölderle, T., Staiger, A., & Vogel, M.** (2018). *BoDyS: Bogenhausener Dysarthrieskalen Tests für die Neuropsychologie* (1. Auflage ed.). Hogrefe.