Laboratory Phonology
Journal of the Association for Laboratory Phonology

Huttenlauch, C., et al. 2021 Production of prosodic cues in coordinate name sequences addressing varying interlocutors. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12(1): 1, pp. 1–31. DOI: https://doi.org/10.5334/labphon.221

JOURNAL ARTICLE

Production of prosodic cues in coordinate name sequences addressing varying interlocutors

Clara Huttenlauch, Carola de Beer, Sandra Hanne and Isabell Wartenburger
Cognitive Sciences, Department of Linguistics, University of Potsdam, Potsdam, DE
Corresponding author: Isabell Wartenburger (isabell.wartenburger@uni-potsdam.de)

Prosodic boundaries can be used to disambiguate the syntactic structure of coordinated name sequences (*coordinates*). To answer the question whether disambiguating prosody is produced in a situationally dependent or independent manner and to contribute to our understanding of the nature of the prosody-syntax link, we systematically explored variability in the prosody of boundary productions of coordinates evoked by different contextual settings in a referential communication task. Our analysis focused on prosodic boundaries produced to distinguish sequences with different syntactic structures (i.e., with or without internal grouping of the constituents). In German, these prosodic boundaries are indicated by three major prosodic cues: f0-range, final lengthening, and pause. In line with the Proximity/Anti-Proximity principle of the syntax-prosody model by Kentner and Féry (2013), speakers clearly use all three cues for constituent grouping and prosodically mark groups within and at their right boundary, indicating that prosodic phrasing is not a local phenomenon. Intra-individually, we found a rather stable prosodic pattern across contexts. However, inter-individually speakers differed from each other with respect to the prosodic cue combinations that they (consistently) used to mark the boundaries. Overall, our data speak in favour of a close link between syntax and prosody and for situational independence of disambiguating prosody.

Keywords: Prosodic boundaries; prosodic cues; coordinates; varying interlocutors; variability; f0; duration; pre-final lengthening; pause

1. Introduction

Syntactic ambiguities, like the internal grouping of sequences, see example (1), are a common phenomenon in many languages. In spoken language, such ambiguities can be resolved by prosodic phrasing, phonetically indicated by modified prosodic cues. If the answer to the question *Who will bring a spare bike for the trip?* were (1), the lexical string alone would not clearly indicate whether there will be one, or two, or three bikes. This is because the phrase has three possible readings depending on the grouping of the coordinated names: One bike could be brought by all three persons together, or two of them could bring one bike together and another person brings a second bike, or each of them could bring their own bike, respectively.

(1) Caro and Lea and Jana.

The syntactic grouping of the names in (1), however, can be disambiguated by prosodic cues which lead to the perception of a boundary that will be referred to as prosodic boundary (Frazier, Carlson, & Clifton, 2006; Holzgrefe-Lang, 2017; Kentner & Féry, 2013; Wagner, 2005), marking the intended syntactic grouping. As such, there is a close

link between syntax and prosody from the perspective of the listener/interlocutor and the speaker as well. At the phonetic level, in German, the language studied here, three main cues in two domains are used for prosodic boundary marking in spoken language production: in the tonal domain, pitch change, mostly realized as a rise in fundamental frequency (f0) and in the durational domain, lengthening of the syllable or segment immediately preceding the boundary (final lengthening) and pause at the boundary (for German: Gollrad, Sommerfeld, & Kügler, 2010; Kentner & Féry, 2013; Peters, Kohler, & Wesener, 2005; Petrone et al., 2017). The pitch change is operationalized as fundamental frequency, abbreviated to f0 and used interchangeably with pitch in this paper, even though pitch refers to the perceptual correlate and f0 to the acoustic measure. Pitch changes have been shown to be relevant in coordinates already in the seminal works of Ladd (1986) for English and van den Berg, Gussenhoven, and Rietveld (1992) for Dutch.

These examples illustrate that syntactic structure and prosody are closely related to each other. However, it is still a matter of debate, how this link is represented in the linguistic system: whether the phonology-syntax mapping follows a fixed, categorical, phonological hierarchy in which certain syntactic categories are mapped to certain phonological units, such as the phonological phrase or the intonational phrase with particular phonetic characteristics (e.g., Nespor & Vogel, 1986), or whether this mapping is more flexible and characterized by rather relative or gradient phonetic correlates (e.g., Wagner, 2005). Moreover, it is being discussed which function disambiguating prosody actually fulfils in the situation in which it is being produced, that is, whether it is produced mainly for the sake of the interlocutors or for the speakers themselves (e.g., Speer, Warren, & Schafer, 2011). The latter case would point towards situational independence of prosodic realizations whereas the first scenario would indicate that prosody production is situationally dependent.

To address the question whether disambiguating prosody is produced in a situationally dependent or independent manner and to contribute to our understanding of the nature of the prosody-syntax link, we will compare prosodic realizations in varying situations between and within individuals. Specifically, we study inter- and intra-individual variability in spoken productions of name sequences in German coordinated with *und* (English *and*), hereafter referred to as *coordinates*. We will focus on two conditions of these coordinates: one without internal grouping referred to as *nobrack* (see 2) and another condition with internal grouping, in which the first two names are grouped together in one sequence and the third name is a separate sequence, referred to as *brack* (see 3). For easier reading, brackets around the grouped names will indicate the structure. Regarding the question of the number of bikes, in (2) there would be one spare bike while in (3) there would be two spare bikes.

(2) without internal grouping (nobrack): [Moni und Lilli und Manu]

(3) with internal grouping (brack): [Moni und Lilli] und Manu

In the following we will briefly introduce previous findings on the functional role of disambiguating prosody (1.1) and on individual variability in prosody production (1.2). Then we summarize theories on the prosodic phrasing in coordinates (1.3).

1.1. Function of disambiguating prosody: For the speaker or for the interlocutor

The function of (disambiguating) prosody concerns, in short, the question whether prosody is produced mainly for the interlocutors or for the speakers themselves. This goes in line with the question in how far the prosodic realization of an utterance is

dependent or independent from the actual situation in which it is being produced. If there is a rather direct link between syntax and prosody, disambiguating prosody should be 'automatically' present in any case–independent of the situation. However, if prosody is less automatically connected to the structural properties of the utterance, but used in a more controlled way by the speaker to support the interlocutor's parsing of an ambiguous utterance, then the use of prosody may vary more depending on the situation and/or properties of the interlocutor. The latter assumption can be subsumed under models of 'situational dependence' and the former under models of 'situational independence.'

Situationally dependent models, on the one hand, assume that prosodic realizations depend on the actual communicative situation. Prosodic cues are only necessary, and therefore expected, if the speaker is aware of the ambiguity and the possible misunderstanding of the interlocutor and if the context does not provide other, non-prosodic disambiguating cues. Those other cues can be linguistic or non-linguistic. Models assuming situational dependence of prosodic realizations predict that speakers use prosody differently when addressing interlocutors with different needs or, more generally, that speakers use prosody differently in different communicative or contextual situations. Situational dependence supports the view that prosody is realized for the interlocutor–to help them derive the intended meaning. In that sense, the speech planning mechanism would be required to foreshadow for any stage of the upcoming speech whether it is in fact ambiguous and lacks disambiguating cues of any kind in order to evaluate the necessity for disambiguation (Speer et al., 2011, p. 87f.). Furthermore, in a strict interpretation of context dependence of prosodic cues, their occurrence would then be more likely in situations which do not provide any disambiguating information and, thus, they should appear rather inconsistent and infrequent (Speer et al., 2011, p. 36f.). This inconsistency, however, would render them unreliable for perception (see e.g., Kraljic & Brennan, 2005, p. 196).

Situationally independent models, on the other hand, assume that prosodic realizations are largely independent from actual interlocutors or the communicative/contextual situation. Under such accounts prosodic cues are produced automatically and their realization is affected by grammatical factors such as phrase structure, information status, or phonological length (Kraljic & Brennan, 2005; Speer et al., 2011, p. 37). In this view, prosody is not primarily realized for the interlocutor, but more automatically 'for' the speaker. Since prosodic cues are interpreted as depending on linguistic factors, their occurrence should be rather common and frequent, which would make them reliable for perception (Kraljic & Brennan, 2005; Speer et al., 2011). In the following we introduce some exemplar studies which support either the situationally dependent or the situationally independent account.

Allbritton, McKoon, and Ratcliff (1996) addressed the issue of situational in/dependence by testing whether untrained, naïve speakers (versus trained speakers) would spontaneously use prosody to resolve syntactic ambiguities in various kinds of sentence types. The speakers were instructed to read aloud "as if you were telling someone a story that you wanted them to understand" (Allbritton et al., 1996, p. 716). It turned out that most naïve and trained speakers did not prosodically disambiguate most of the sentences. Only if the instruction made them aware of the ambiguity and asked them explicitly to produce two different versions, trained speakers used prosody for disambiguation. This can be interpreted as a finding supporting situational dependence. The authors concluded that either the role of prosodic cues for conveying the underlying syntactic structure is limited or laboratory recordings cannot be generalized to real-world settings (Allbritton et al., 1996, p. 732). Applying a more real-world setting, namely a game-like interactive referential communication task, Snedeker and Trueswell (2003) confirmed the hypothesis that the relation between syntax and prosody is mediated by the context. In their study, naïve participants produced clear prosodic groupings of attachment ambiguities ("Tap the

frog with the flower") only in situations in which the context did not provide sufficient information to situationally disambiguate the two possible meanings. The authors concluded that "speakers produce [prosodic cues] primarily when they appear to be necessary for clear communication" (Snedeker & Trueswell, 2003, p. 128). Based on these two example studies, one could conclude that the use of prosody for disambiguation depends on the awareness of the speaker about the ambiguities and/or on whether the actual context made both readings plausible. Prosody is thus mainly used for the interlocutor (cf. audience design, Bell, 1984).

In contrast, others found evidence in favour of situational independence: Using a co-operative interactive game-board task, similar to Snedeker and Trueswell (2003), Schafer, Speer, Warren, and White (2000) and Speer et al. (2011) found no evidence for a dependency of prosodic cues on situational disambiguation or discourse factors using global attachment ambiguities and temporal closure ambiguities. The speakers produced prosodic cues independent of the communicative situation (even in only locally ambiguous sentences), but still with some flexibility or variability in the choice of cues (Speer et al., 2011). This was also confirmed in another interactive game-like study design by Kraljic and Brennan (2005), who also found overall limited effects of the context. In their interactive setting involving attachment ambiguities ("Put the dog in the basket on the star"), speakers produced clear prosodic cues for disambiguation, irrespective of the needs of their interlocutors (i.e., regardless of whether the contextual setting provided disambiguating information or not) and irrespective of whether an interlocutor was present or not. With respect to the function of prosody, they conclude that prosodic marking emerges from the level of planning and articulation, that is, prosody is not produced dependent on the situation but rather automatically and situationally independent. Similarly, for coordinate name sequences, Wagner (2005) found that prosodic boundaries are produced independent of the context and independent of the need of the interlocutors to comprehend, which implicates that prosody is mainly used 'for' the speakers themselves, in an automatic manner.

In sum, there is evidence supporting either of the two accounts on the function of prosody. The differential findings might be related to task differences (e.g., instruction, presence of an interlocutor, degree of interaction between speaker and interlocutor, potential for misunderstandings, awareness of the ambiguities) or the complexity or length of the to-be-produced structures (e.g., longer utterances in Speer et al., 2011 than in Snedeker & Trueswell, 2003). For a detailed discussion on these differences see Kraljic and Brennan (2005), Snedeker and Trueswell (2003), and Speer et al. (2011). For the option of intermediate positions between situational dependence and independence see also Speer et al. (2011, p. 37f.).

1.2. Individual variability in prosody production

We now turn from the group level to the individual level and to the question of whether individuals vary in their prosodic realizations. Variability between or within speakers is interesting for two reasons. First, if all speakers reliably use disambiguating prosody to distinguish between coordinates without and with grouping (example [2] versus [3]), this would indicate a close link between syntax and prosody (e.g., Nespor & Vogel, 1986) and situational independence (e.g., Speer et al., 2011). If, on top of the disambiguation, we would also find variability across speakers in how they realize prosodic boundaries, this would add evidence that the link between syntax and phonology is relational rather than categorical (e.g., Clifton, Carlson, & Frazier, 2002; Wagner, 2005). Second, if speakers do not reliably disambiguate the different syntactic structures (example [2] versus [3]) or show within-speaker variability in different contextual settings, this would speak in

favour of situationally dependent models–and against a close prosody-syntax link. So far, the issue of inter-individual variability concerning prosodic boundary production in coordinates has been explored only scarcely (one exception, for German, being the work by Petrone et al., 2017, or, for English, the findings by Allbritton et al., 1996, and Lehiste, 1973), and variability induced by different situational contexts has, to the best of our knowledge, not been studied yet.

Regarding *variability between speakers (*i.e., *inter-individual variability)*, Petrone et al. (2017) found that their speakers differed in how the prosodic boundary was realized in coordinates with internal grouping (i.e., they found multiple types of prosodic boundaries): Only two out of 12 speakers consistently used the same f0-contour, namely a rise. Although production of a rise was also the predominant contour in six further participants, these additionally employed a high plateau. Another three speakers varied between rise, high plateau, and final fall to different degrees and one speaker produced either rises or falls. Using similar three-name sequences, Lehiste (1973) reported that two (English) speakers differed in how they used durational cues for disambiguation (insertion of a pause versus lengthening of the coordinating element).

With respect to *variability within speakers (*i.e., *intra-individual variability)* induced by contextual settings, specifically concerning the type of interlocutor, previous research has focused on differences in prosodic realizations when children, elderly adults, or non-native speakers are being addressed in comparison to young adult native speakers. Most studies take the speech addressed to an adult native speaker of the language under investigation as a baseline for comparisons. For easier reading, we will refrain from mentioning this adult baseline in the following. For example, for attachment disambiguation, Kempe et al. (2010) reported lengthened vowels when English-speaking adults addressed two–four-year-old real or imaginary children and, in addition, found longer pause durations. Other studies investigated intra-individual variability in prosodic information per se (i.e., not focusing on disambiguating prosody): Biersack, Kempe, and Knapton (2005) reported an increased pitch range and higher f0-maxima as well as longer durations due to the lengthening of vowels in semi-spontaneous speech addressed to a two-year-old imaginary child in English. DePaulo and Coleman (1986) also reported longer pauses in spontaneous English speech addressing a six-year-old child. When it comes to prosodic cues in speech addressing a non-native interlocutor, results are inconclusive: While one study involving English speakers found no differences (DePaulo & Coleman, 1986), another one found a lowered speech rate due to lengthened pauses (Biersack et al., 2005), and Smith (2007) reported an increased f0-range and segmental modifications leading to a more emphatic style in French. Regarding prosodic cues when addressing elderly interlocutors in English, Kemper, Vandeputte, Rice, Cheung, and Gubarchuk (1995) reported a slower speech rate due to prolonged vowels and more frequent pauses in spontaneous speech of a map task with a physically present interlocutor. Although expected, they did not find exaggerated pitch ranges. For German, Thimm, Rademacher, and Kruse (1998) also reported more pauses as well as more variation in intonation in spoken explanations of an alarm clock when a positively stereotyped elderly person was addressed as opposed to a young adult.

As an alternative to the experimental manipulation of type of the (imaginary or real) interlocutor, some studies varied the contextual setting via the presence or absence of noise. Speech production in noisy environments leads to increased f0-values and f0-range, increased signal amplitude, increased word or segment durations, and spectral changes such as smaller spectral slope (Davis, Kim, Grauwinkel, & Mixdorff, 2006; Folk & Schiel, 2011; Garnier, Bailly, Dohen, Welby, & Lœvenbruck, 2006; Jessen, Köster, & Gfroerer, 2003; Junqua, 1993, 1996; Landgraf, Schmidt, Köhler-Kaeß, Niebuhr, & John, 2017; Lu & Cooke, 2008; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988; Varadarajan & Hansen,

2006; Zollinger & Brumm, 2011). These noise-dependent changes are summarized under the term Lombard speech, tracing back to Étienne Lombard who first described the noise-dependent increase in speech amplitude for French (Lombard, 1911; as cited in Zollinger & Brumm, 2011). Lombard speech is also described as a source of inter- and intra-speaker variability (Jessen et al., 2003; Junqua, 1993; Stanton, Jamieson, & Allen, 1988). For a recent review on the neural mechanisms of the Lombard effect in humans and animals see (Luo, Hage, & Moss, 2018).

1.3. Prosody of coordinates (in German)

As our study specifically investigates the prosody of coordinates, we will briefly review some relevant models on prosodic phrasing in coordinates which have been proposed in the past. We will focus on the Proximity/Similarity model (Kentner & Féry, 2013) since it has been tested with German speakers in similar structures as we use in the current study.

With respect to the question how coordinates are prosodically phrased in English, Taglicht (1998) formulated the 'Coordination Constraint.' It specifies that the same hierarchical level of intonational boundaries must be applied to all elements at the same syntactic level. Watson and Gibson (2004) argue in their 'Left hand side/Right hand side Boundary hypothesis' that the likelihood of the presence of a prosodic boundary depends on the size of the preceding and following constituents because of processing demands: For larger constituents, the speaker needs more refractory time to recover from the preceding constituent and more time to plan the upcoming constituent, respectively. Wagner (2005, 2010) demonstrates that, in coordinate structures, the relative strength of the boundary reflects the level of embedding at the syntactic level and thereby confirms the close match between prosody and syntax in coordinate structures. According to Wagner (2010, p. 186) more deeply embedded constituents "are separated from each other by weaker boundaries than constituents that are less deeply embedded" and "constituents separated by relatively weaker boundaries are perceived as grouping together." In a similar vein, Kentner and Féry (2013) developed a model that aims to account for both, processing demands, depending on the constituent size or complexity, and demands of the syntactic structure, depending on the depth of syntactic embedding. Their so-called Proximity/Similarity model assumes two principles that "interact to shape the prosody of syntactic structures" (Kentner & Féry, 2013, p. 283) such as coordinated name sequences. Proximity is related to the syntactic constituent structure and states that "adjacent elements which are syntactically grouped together into one constituent should be realized in close proximity" (Kentner & Féry, 2013, p. 282). Similarity is related to the depth of syntactic embedding and refers to the idea that "constituents at the same level of embedding should be realized in a similar way, that is, they should be similar in pitch and duration, irrespective of their inherent complexity"; this principle is comparable to the models of Taglicht (1998) and Wagner (2005, 2010) which assume that elements at the same syntactic level are prosodically matched.

For the structures used in the present study ([2] and [3]), the Proximity/Similarity model makes the following predictions: In coordinates with internal grouping, the principle of Proximity predicts a weakening of the prosodic cues–compared to a coordinate without internal grouping–at an element x if the neighbouring element to the right is part of the same group as x (cf. the first name, i.e., Moni, in [3]). Reversely, Anti-Proximity predicts a strengthening of a prosodic boundary if the right-adjacent element of x does not form a group with x (cf. the second name, i.e., Lilli, in [3]). The Similarity principle predicts that a simplex element at the same level of embedding as a complex constituent is, for instance, lengthened to adjust its duration to the length of a complex constituent. This would be

relevant for groupings in which the first name is followed by a complex sequence on the right (cf. Moni und [Lilli und Manu]). As the current study will focus on coordinates with an internal grouping of the first two names (as in [3]), the Similarity principle will not be discussed any further. In coordinates without internal grouping (considered the baseline form; see [2]), all names are expected to be separated by boundaries of the same strength.

With boundary cue weakening, Kentner and Féry (2013) refer to the use of lower pitch and shorter durations on the first grouped element compared to a non-grouped baseline, while a strengthened boundary at the right edge of the grouped element is expressed by a higher boundary tone and longer durations. On the final element of the coordinates (cf. Manu in [2] and [3]), Kentner and Féry (2013) observed neutralization in duration and f0-movement. The findings of increased duration of the word preceding the boundary and a possible pause along with higher pitch at the prosodic boundary have been confirmed in a further study on elicited coordinate productions in German (Petrone et al., 2017) and are in line with results on prosodic marking of syntactic boundaries in spontaneous German speech (Peters et al., 2005). What is still unclear with respect to the Proximity/Similarity account is whether its assumptions also hold situationally independent. Until now, variations in prosodic phrasing of coordinates within speakers across different situations/interlocutors have not been explored sufficiently. In addition, it is unclear to which extent there is variability across speakers and, specifically, whether the speakers differ in how they use and combine the different prosodic cues to mark the prosodic boundaries.

Therefore, in our study, we use coordinate name sequences ([2] and [3]) to replicate the findings of Kentner and Féry (2013) under different contextual settings, that is, in different situations. At the same time, due to the focus on (2) and (3), our data will not be sufficient to adjudicate among the models briefly introduced above in this section. Thus, the current study will not directly contribute to the question as to whether–or to which extent–processing demands or the level of syntactic embedding drive prosodic realizations. Instead, the main focus of our study is on inter- and intra-individual variability and its limits in prosodic boundary production, the relation of the different prosodic boundary cues to one another, and on the situational in/dependence of prosodic phrasing.

In summary, the functional role of disambiguating prosody or its situational in/dependence has been studied by means of the presence or absence of contextual effects on prosodic realizations–but remains largely inconclusive. The fact that participants are aware of an ambiguity, the task setting (reading-out loud versus interactive setting with real versus imagined interlocutors), the type of ambiguity (e.g., attachment ambiguities versus pragmatic ambiguities), the length of the to-be-produced utterance, the type of interlocutor (e.g., child versus adult), and other contextual factors, such as absence/presence of noise, seem to influence if and how individuals use prosody to disambiguate syntactic structures. We are going to address the question of situational in/dependence by comparing prosodic realizations in varying situations within individuals. In addition, we will explore differences between speakers as they will give us further insights into the nature of the prosody-syntax link.

1.4. Aims and hypotheses

In this study we systematically explore inter- and intra-individual variability in the production of prosodic boundaries to get insights into the prosody-syntax relation and the function of prosody. According to situationally dependent models of prosodic phrasing (Allbritton et al., 1996; Snedeker & Trueswell, 2003; cf. audience design hypothesis, Bell, 1984) we would predict that, if speakers use prosody to disambiguate different syntactic structures at all (e.g., because they are aware of the ambiguity and/or because an interlocutor is present in

the communicative situation), they vary considerably in their prosodic productions between interlocutors with different needs. Contrary, according to situationally independent models of prosodic phrasing (Kraljic & Brennan, 2005; Schafer et al., 2000; Speer et al., 2011), we would predict that speakers use prosody to disambiguate different syntactic structures in any event, because they are doing it 'automatically' during speech planning stages. The prosodic realizations should hence be rather clear between conditions without and with internal grouping (example [2] versus [3]), and consistent across different interlocutors– although some variability between speakers is also expected (Speer et al., 2011, p. 88ff.).

We argue that the issues of within-speaker situational in/dependence and of between-speaker in/variability are related to the underlying nature of the prosody-syntax link: If there is a fixed relationship (and dependency) between prosody and syntax, we would predict that speakers 'automatically' produce prosodic boundaries in a rather fixed or stable manner to disambiguate the syntactic structure, irrespective of the situation they are confronted with (i.e., situationally independent). If, at the same time, the relationship between syntax and phonology is relational or gradient (e.g., Wagner, 2005), we would additionally predict some variability between speakers with respect to the phonetic correlates they employ to disambiguate the syntactic structure.

Our study thus explores the effect of the type of the interlocutor and presence/absence of noise on variability between and within speakers' prosodic boundary realizations in a controlled, semi-interactive setting. Specifically, the speakers are asked to utter coordinates with versus without internal grouping (such as [3] and [2]). The five different contextual settings will henceforth be referred to as *contexts*. The contexts involve four different female interlocutors: a young adult (YOUNG), a child (CHILD), an elderly adult (ELDERLY), and a young non-native adult speaker of German (NON-NATIVE) and a noisy environment (the young adult with white background noise, NOISE).

Speakers are completely aware of the intended syntactic grouping of the coordinates and are asked to utter the name sequences in such a way that the different virtual interlocutors can resolve them. We will focus on the prosodic cues f0-range, final lengthening, and pause at/after the first and the second name, as these are known to be modulated to indicate prosodic boundaries. The results will be discussed referring to the Proximity/Similarity model of syntax-prosody mapping (Kentner & Féry, 2013). We will describe the interplay and combined use of the prosodic cues of prosodic boundaries and how these are affected by inter- and intra-individual variability as these will contribute to our understanding of the prosody-syntax relation and the functional role of disambiguating prosody.

Our research questions are as follows:

(Q1) Prosodic disambiguation of coordinates: Can the findings of previous studies concerning differences in the use of f0-range, final lengthening, and pause on the first and on the second name in coordinates without internal grouping, such as (2), and with internal grouping of the first two names, such as (3), be replicated?
(Q2) General context-dependent prosodic variability: To what extent do these prosodic boundary cues vary in the five different contexts?
(Q3) Inter-speaker variability: Do different speakers show different patterns in their combined use of the three prosodic cues within contexts?
(Q4) Intra-speaker variability: Do speakers show different patterns in their combined use of the three prosodic cues between contexts?

Regarding Q1, based on the literature outlined in 1.3, we expect speakers to mark the difference between coordinates with (3) and without (2) internal grouping in line with the

Proximity/Similarity model by Kentner and Féry (2013). More specifically, we expect a prosodic boundary realized by an increase of final lengthening and an increased f0-range at the right edge of the group (i.e., on the second name), as well as the insertion of a pause after the grouping in (3) compared to (2). On the first name, we expect a decrease in final lengthening and a smaller f0-range in (3) compared to (2).

With respect to Q2, we confront speakers with five different contexts (YOUNG [baseline], CHILD, ELDERLY, NON-NATIVE, and NOISE) to disentangle the question of situational in/dependence of prosodic variability. If speakers vary their productions between contexts, we expect those variations to be in line with the literature mentioned in 1.2: We expect speakers to mark the difference between conditions with and without internal grouping in a more pronounced way in the non-baseline contexts. If the findings of prosodic cue modifications for contextual situations in different sentence types are transferable to coordinates and if the modifications found for English and French speakers also hold for German, we expect an increase in segmental and pause durations as well as increased f0-ranges for CHILD and ELDERLY. Due to inconsistent findings in previous studies regarding a non-native interlocutor, we explore this context at a rather exploratory level and refrain from a specific hypothesis. Regarding the presence of noise (NOISE), the literature predicts an increase in f0 and segment durations. Note that the NOISE condition is the only condition, which directly affects the speaker, because the virtual interlocutor *and* the speaker are confronted with the noise.

With Q3 and Q4, we expect to further disentangle the nature of the prosody-syntax link. Regarding Q3, between-speaker in/variability in the combined use, that is, in the interplay of the prosodic cues, will inform us about the type of link between syntax and phonology (i.e., fixed and categorical or relative and allowing for some flexibility). Regarding Q4, within-speaker in/variability will give further insights into situational in/dependence of prosodic cues on the individual level. These two research questions will be addressed in an exploratory manner.

Overall, marked differences between contexts would speak in favour of situationally dependent models and their absence for models of situational independence (given speakers would prosodically disambiguate the conditions with and without grouping). If speakers show inter-individual variability in how they employ and combine prosodic cues at the surface to mark prosodic boundaries, this would speak in favour of models of relative boundary strength (e.g., Wagner, 2005).

2. Methods

2.1. Participants

Sixteen monolingual German native speakers (sex: 13 female, 2 male, 1 other; age range: 19–34 years, mean: 25.75 years, *SD*: 4.6) took part in the study. They were recruited at the University of Potsdam and were reimbursed or received course credits. Written informed consent was obtained from all participants prior to the study. They were naïve to the purpose of the study. The procedure for this study was approved by the Ethics Committee of the University of Potsdam (approval number 72/2016). Participants (henceforth *speakers*) reported normal or corrected to normal vision. Normal hearing was also confirmed by a hearing screening using an audiometer (Hortmann DA 324 series).

2.2. Stimuli

2.2.1. Items

Stimuli were taken from Holzgrefe-Lang et al. (2016) and consisted of six sequences of three German names coordinated by *und* (English *and*). Each name sequence appeared in two conditions: without internal grouping (4) or grouping the first two names together

(5), resulting in 12 items overall. Grouping was visually indicated by parentheses around the grouped names (see [5]).

(4) Name1 and Name2 and Name3

(5) (Name1 and Name2) and Name3

A total of nine different names was used. Six of these occurred as Name1 and as Name2 and ended in the high frontal vowel /i/ (Moni, Lilli, Leni, Nelli, Mimmi, or Manni) in order to decrease glottalization. The remaining three names (Manu, Nina, or Lola) ended either in /u/ or in /a/ and occurred only in the position of Name3. The names were controlled for number of syllables (disyllabic), syllable structure (trochaic), and sonority of the segments (only sonorant material was used to allow for better pitch tracking). Two corpora (*Google Ngram Viewer*, https://books.google.com/ngrams retrieved on 06.08.2020 and dlexDB Heister et al., 2011) confirmed that the name combinations we used (e.g., 'Moni und Lilli') were all non-frequent (no hits).

2.2.2. Contexts

The five contexts (YOUNG, CHILD, ELDERLY, NON-NATIVE, NOISE; see **Figure 1** and **Table 1**) were evoked by videos, giving the speakers a visual-auditory impression of their interlocutors. The noise for the NOISE context was created in Praat (Boersma & Weenink, 2017) using the formula *randomGauss(0,0.7)*. For each context, the corresponding interlocutor appeared in two video clips (introduction and instruction) and produced a trigger question (see below).

In the introduction video, each interlocutor presented her character in a few sentences, talking about her fictional demographic background, including information on her name, age, origin, occupation, place of living, and some interests (**Figure 1**; for the exact wording of their presentation see Appendix A). (True) demographic data of the interlocutors are given in **Table 1**. Note, however, that these data were unknown to the speakers of the production study.

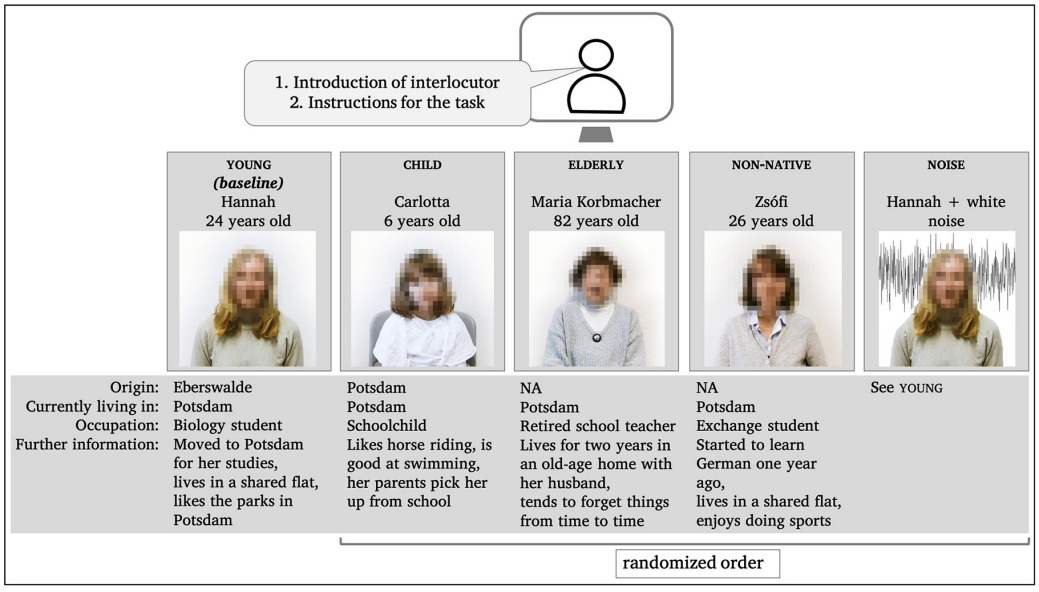

Figure 1: Pictures and fictional names, ages, origins, and further information of the interlocutors present in the five contexts. *Note*: Faces were not pixelated in the experiment; noise was presented auditorily.

In the instruction video, the interlocutor instructed the speakers to utter the name sequences in a way that would allow the interlocutor "to understand as rapidly and accurately as possible who is coming together." The wording of the instruction for the task was nearly the same for all contexts but the adult interlocutors addressed the speakers in the formal way using German *Sie* (you), while the child used the informal *Du* (you), which reflects prescriptive German pronoun use.

For the noise context (NOISE), the young interlocutor was exposed to the same white noise that was later played to the speakers in the recording session. She heard the noise via in-ear headphones which were invisible in the video clip. Instead of presenting herself again, she reminded the speakers of who she was and that they should do the task with her again. Furthermore, she commented on the noise in the background and repeated the instruction for the task, adding to the usual wording that she was going to be the interlocutor *again*.

In order to reduce the influence of non-person specific factors to a minimum, the interlocutors in the videos wore similar unicoloured clothes and were all seated in front of a light neutral background (**Figure 1**). They were asked to look into the camera and to talk with few gestures and little moving. The introduction and instruction videos had comparable durations (cf. **Table 1**). In order to trigger the production of the name sequences and to remind the speaker of their interlocutor, the speakers were played the question *Wer kommt?* ('who is coming?') produced by the respective interlocutor of each context. The trigger questions had a mean duration of 0.94 seconds (*SD*: 0.028 sec; see **Table 1**) and preceded each trial (see *2.3. Procedure*).

2.3. Procedure

Before the start of the recording session, the white noise was played to the speakers for one second to familiarize them with the sound to be played in the NOISE condition in order to prevent surprisal or scare effects during the experiment. The experiment then

Table 1: Information on the five contexts.

	YOUNG	CHILD	ELDERLY	NON-NATIVE	NOISE
(True) demographic data of the person behind the character of the fictional interlocutor (unbeknownst to the speakers)					
age (in years)	21	7	89	32	See YOUNG
mother tongue	German	German	German	Hungarian	
origin	Berlin-Brandenburg area	moved to Berlin-Brandenburg area at the age of 4	Berlin-Brandenburg area	Hungary	
currently living in	Berlin-Brandenburg area	Berlin-Brandenburg area	Berlin-Brandenburg area	Berlin-Brandenburg area (in Germany for < 3 years)	
Technical details of videos: Durations in seconds					
introduction video	28	18	41	30	17
instruction video	21	19	35	24	22
trigger question *wer kommt?* (*'who is coming'*)	0.892	0.956	0.961	0.932	0.953

started with a practice phase (four items which were not used in the actual experiment) followed by the test phase. The test phase consisted of five blocks, corresponding to the five experimental contexts (YOUNG, CHILD, ELDERLY, NON-NATIVE, NOISE; see above) in which speakers were asked to produce the coordinated name sequences in the two conditions, that is, with or without internal grouping. Each of the 12 items was presented in each context, hence, speakers produced each item five times. The YOUNG context, as the baseline context, was always presented first; the other four contexts were presented in randomized order (cf. **Figure 1**). In each context, items were pseudo-randomized using different lists. No more than two items of the same condition followed one another. In addition, Name1 and Name2 were never repeated in two subsequent trials.

Each block started with the two video clips and during the test phase, for each trial, speakers saw a fixation cross on the screen while they heard the trigger question *Wer kommt?* ('who is coming?') via headphones. After 1000 ms, the fixation cross was replaced by the visual presentation of a name sequence (i.e., the item) in one of the two conditions which stayed on the screen for 5000 ms. The sound recording started together with the presentation of the name sequence and continued for 1000 ms after the names disappeared; see **Figure 2**.

Recordings took place in a sound-attenuated booth in the acoustics laboratory of the University of Potsdam via an Alesis io12 interface. Speakers wore a headset HSC 271 (AKG Acoustics by Harman, www.akg.com) with over-ear headphones and a condenser microphone and were seated in front of a wide screen monitor with 1920 x 1200 resolution and saw the stimuli in Arial font of size 50. The experiment was run from a Dell laptop standing outside of the recording booth using the software Presentation (*Neurobehavioural Systems*, https://www.neurobs.com/; Version 20.1).

After the recording session, speakers completed some questionnaires which will not be analyzed.

2.4. Perception check

2.4.1. Procedure

In the production study described above, we recorded a total of 960 individual productions: 6 name sequences * 2 conditions * 5 contexts * 16 speakers. In order to verify whether the *intended* internal structure (i.e., the *grouping of constituents*) is congruent with the structure perceived by other naïve listeners, we ran a perception check of all 960 productions. Note that *intended* refers to the conditions with or without parentheses presented to the speakers on the screen in the production study. We lack information about the intention of the speakers at the time of production.

The perception check encompassed 32 listeners, who had not taken part in the production experiment (sex: 22 female, 10 male; age range: 18–41 years, mean: 24.25 years, *SD*: 5.8).

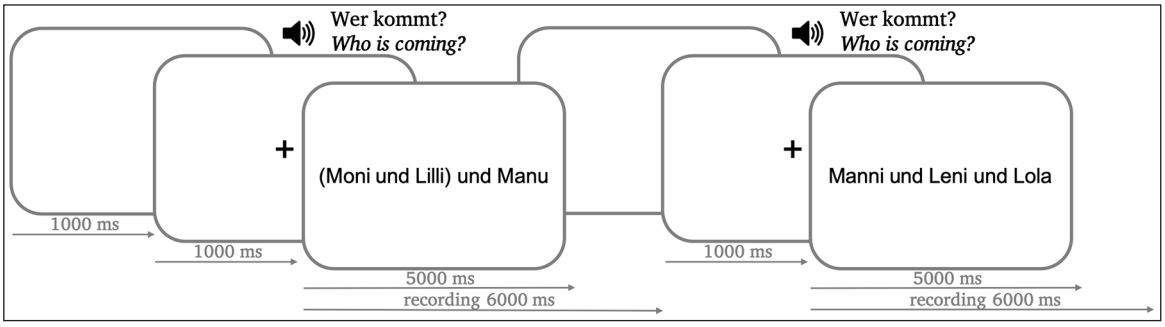

Figure 2: Experimental setting and timing of two trials.

They were recruited at the University of Potsdam and were reimbursed or received course credits. Another 10 listeners took part, but had to be excluded from the analysis due to technical problems (n = 9) or German as a non-native language (n = 1).

Each listener judged a set of 267 out of the 960 productions, which consisted of the total 60 productions of 4 speakers (4 speakers * 60 recordings = 240) plus a subset of 27 productions from various speakers. The subset of 27 productions was judged by all listeners and constituted a semi-random sample of all productions, containing at least one production of each speaker and of each context. Furthermore, the subset included three productions which to the first author seemed to mismatch between intended and perceived grouping. The perception check started with the presentation of the subset, followed by the remaining 240 productions of four speakers, each presented in a block. The 960 productions of the 16 speakers were judged in four testing lists (4 lists * 240 recordings = 960). Each list, and therefore the productions of each speaker, was judged by eight listeners (8 listeners * 4 lists = 32 listeners in total). Each of the four lists contained some productions twice, those which were part of the subset and the following 240 productions. In the case of repetitions, only the first judgement was considered.

The perception check was run in sessions with several listeners at the same time. Two pictograms with three persons each were used to depict the two conditions (**Figure 3**, picture A without and picture B with internal grouping). The task was twofold: to listen to each production and (1) to choose the matching pictogram (i.e., to identify the condition) and (2) to indicate the most probable addressee the name sequence was uttered to (young adult, child, elderly, non-native, in noise; i.e., to identify the context).

2.4.2. Analysis and results

First, for each listener, we counted the number of congruent rates (i.e., correct identifications of the intended grouping/condition and context, referred to as *hit-rate*). Following standard assumptions on the exclusion of data points (e.g., Howell, 1998), if, for a given listener, the hit-rate was 2 *SD* below the mean hit-rate of all listeners, all ratings of this listener were excluded. This was the case for one out of the 32 listeners, thus ratings from altogether 31 listeners, rendering 8067 ratings overall entered the perception check analyses (the 8067 ratings result from 27 productions in the subset * (rated by) 32 listeners + 933 remaining productions * (rated by) 8 listeners (= 8328) – 261 ratings of the excluded listener).

Second, for each individual production (n = 960), we calculated the ratio of the hit-rate to the number of total rates. We used the ratio instead of the absolute hit-rate since individual productions were rated by a varying number of listeners (in the subset: 31 listeners, for

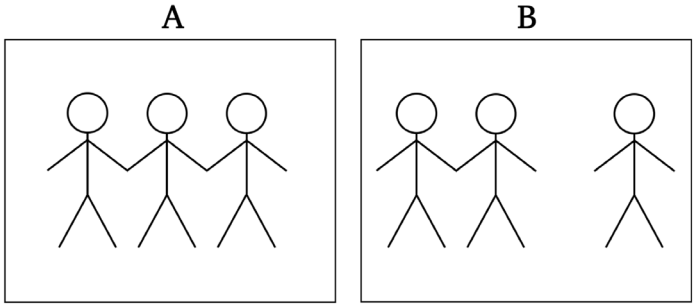

Figure 3: Pictograms used in the perception check. Picture A depicts the condition without internal grouping, picture B the condition with internal grouping.

the rest of the productions: 8 listeners, or 7 in the case of the excluded listener). We then calculated the mean ratio of all productions as well as the standard deviation.

In what follows, we will report the ratings of condition and context separately. Only the ratings of condition influenced the exclusion of individual productions: Productions for which the ratio of the hit-rate was more than 2 *SD* below the overall mean ratio of hit-rates were excluded for further analyses.

With respect to the rating of condition, the mean ratio was 0.936 (*SD*: 0.1545), and we thus used an accuracy cut-off level of 0.627. Applying this criterion, 38 productions were excluded since their ratios fell more than 2 *SD* below the mean (3 productions of the subset and 35 of the remaining productions). Nevertheless, the majority of all productions was perceived with the intended grouping (689 out of 960).

A closer look at the excluded items reveals that productions with internal grouping were twice as often not perceived as intended (25 with internal grouping, 13 without internal grouping). Looking at the context of the excluded items, we observed that most incongruent rates involved productions produced in the NOISE context (n = 15, with n = 11 in condition brack), followed by productions produced in the YOUNG context (n = 10, with n = 7 in condition brack), the CHILD context (n = 7, with n = 2 in condition brack), the NON-NATIVE context (n = 5, with n = 4 in condition brack), and the ELDERLY context (n = 1 in condition brack).

Regarding the rating of the probable interlocutor (i.e., the listeners had to select the context in which the coordinates were most probably produced), the hit-rates are overall much lower than for condition. For only 21 out of the 960 productions (2%), all listeners correctly identified the context. A closer look at these 21 productions reveals that 17 of them were productions in the NOISE context and the other four in the YOUNG context. An extended analysis revealed that, overall, in 12% of the productions (119 out of 960 productions), at least 75% of the listeners perceived the context in the intended way. These 119 productions are distributed across the five contexts as follows: 65 stem from the NOISE context, 44 from the YOUNG context, 8 from the CHILD context, 1 from the ELDERLY, and 1 from the NON-NATIVE context.

2.5. Segmentation and measurements

In addition to the 38 productions excluded based on the perception check, production data of one speaker was excluded completely, because this speaker did not comply with the task specified in the instructions for the experiment. Visual inspection revealed that the speaker–consciously or not–misinterpreted the whole experimental setting: The productions include quirky, inconsistent prosodic behaviour in the use of the prosodic cues we are interested in. Thus, a total of 96 productions (10% of the overall data) were excluded, consisting of the 38 items following the perception check and 58 productions of the excluded speaker (note that two of their productions are included in the 38 excluded items of the perception check). The remaining data comprise 864 productions from 15 speakers (sex: 13 female, 1 male, 1 other; age range: 19–34 years, mean: 25.47 years, *SD*: 4.6). **Table 2** provides the distribution of the number of remaining productions across contexts and conditions.

Table 2: Number of productions entering statistical analyses across contexts and conditions in the final data set.

	YOUNG	CHILD	ELDERLY	NON-NATIVE	NOISE
nobrack	87	85	90	90	87
brack	83	88	89	86	79

For further analyses, segment boundaries and pauses were manually labelled following the criteria in Turk, Nakai, and Sugahara (2006) using Praat (Boersma & Weenink, 2017). In unclear cases, the boundary between the last vowel of Name1 and Name2, and the following *und,* respectively, was set to the mid of the F2 transition. The end of the utterance was set to the point where the intensity profile fell below 50 dB. The f0-minima on the first syllable and the f0-maxima on the second syllable of both, Name1 and Name2, were annotated. For phonetic analyses, we extracted three acoustic measures regarding duration and f0 each on Name1 and Name2: *rise, final lengthening*, and *pause*. The variable rise captures the range between the f0-minimum and the f0-maximum on NameX calculated in semitones (st; formula used for calculation: $12*\log_2(f0_{max}/f0_{min})$). Rise and f0-range will be used interchangeably in this paper when referring to the f0-measurements taken in the study. The second variable captures the final lengthening on each name (in %) and was calculated by dividing the duration of the final vowel in NameX by the duration of NameX. The variable pause captures the relative duration (in %) of the possible pause following NameX and was calculated by dividing the duration of the pause after NameX by the duration of the whole utterance. Relative values for durational measurements were chosen in order to normalize for differences in speech rate.

Another method to transcribe prosodic boundaries based on f0 would be GToBI (Grice, Baumann, & Benzmüller, 2005; Grice & Baumann, 2002), the German adaption of the ToBI system based on autosegmental-metrical theory of intonation (Ladd, 2008 and references therein) and originally established for American English (Silverman et al., 1992). Since our main focus is on the combined realization of several acoustic cues, we opted for an analysis that can be applied to tonal and durational cues equally.

2.6. Statistical analysis of prosodic disambiguation and general context variability

For each dependent variable (rise, final lengthening, pause) on Name1 and Name2 we ran separate linear mixed-effects regression models using the function *lmer* from the R (R Development Core Team, 2018) packages *lme4* (Bates, Mächler, Bolker, & Walker, 2015) and *lmerTest* (Kuznetsova, Brockhoff, & Bojensen Christensen, 2017). Context was entered as an independent variable and four contrasts were coded comparing each of the contexts CHILD, ELDERLY, NON-NATIVE, and NOISE against YOUNG (baseline) using the general inverse (Schad, Hohenstein, Vasishth, & Kliegl, 2018). The model, thus, estimates the difference in the dependent variables between addressing the child compared to the young adult (CHILD versus YOUNG), addressing the elderly compared to the young adult (ELDERLY versus YOUNG), addressing the non-native speaker of German compared to the young (native German-speaking) adult (NON-NATIVE versus YOUNG), and addressing the young adult in the presence of noise compared to a non-noisy environment (NOISE versus YOUNG). For final lengthening and rise, condition was coded with a sum contrast, with the condition brack coded as 1 and the condition nobrack as -1. Pause was modelled for the condition brack only, due to the absence of a pause after Name2 in most nobrack productions (i.e., values of zero in the dataset).

For model fitting, we always started with a maximal model including the interaction of context and condition as fixed-effects terms (except for the pause measure), as well as a random-effects structure with all possible principal components and correlation parameters associated with the four within-subject contrasts (CHILD versus YOUNG, ELDERLY versus YOUNG, NON-NATIVE versus YOUNG, NOISE versus YOUNG). Following the approach outlined in Bates, Kliegl, Vasishth, and Baayen (2015), in order to avoid overfitting of the random effects structure, we fitted the corresponding zero correlation parameter model using the *double-bar* ('||') syntax. The complexity of the random-effects structure was then reduced in a step-wise manner, dropping those components with a

proportion of variance close to zero in a random-effects Principal Component Analysis (using the *rePCA* function in the *RePsychLing* package, Baayen, Bates, Kliegl, & Vasishth, 2015). We assessed improvements in model fit of the maximal model and the zero correlation parameter models using the log-likelihood ratio test and comparisons of the Akaike Information Criterion. For the zero correlation parameter model with the best fit, we returned to a model that included correlations of random effects (i.e., the *single-bar* syntax). In cases in which the reduction of variance components in the zero correlation parameter model did not lead to a better fit than the fit of the maximal model, we kept the maximal model. If, however, the maximal model did not converge (which happened for the pause measure) or if the maximal model had a high degree of correlations in the fixed effects, and the degree of correlations was less pronounced in the zero correlation parameter model, we kept the suppression of the random effects' correlations (i.e., we did not return to the *single-bar* syntax).

2.7. Exploratory analysis of inter- and intra-speaker variability

Following the statistical analyses for rise, final lengthening, and pause, we further explored the interplay of the three cues in combination. Specifically, we were interested in observable patterns in this interplay that differ between speakers within a given context (inter-speaker variability in cue combinations, here we will focus on the context YOUNG) or within speakers between all contexts (intra-speaker variability in cue combinations). Since pause after Name1 was not used by all speakers (see below) and since Name2 is the critical element before the syntactic boundary we decided to do the exploratory analysis of cue combinations on Name2.

We developed a classification system which was applied to each individual cue within each speaker and context, resulting in two parameters as indicators for how effectively each cue distinguishes between the brack and nobrack condition. In order to determine the degree of distinction between conditions, we estimated for each cue within speaker and context the statistical probability of the respective cue distinguishing between the two conditions using a Mann-Whitney U-Test (Mann & Whitney, 1947). The two parameters were (1) the *p*-value of the Mann-Whitney U-Test computed in Matlab (*MATLAB*, 2019) and (2) the common language effect size (CLES, McGraw & Wong, 1992). The CLES returns a value between 0 and 1, and indicates the probability that a random pair of data points belongs to two independent groups. Thus, a value of 1 for the CLES of our comparisons refers to a case in which this cue clearly separates the two conditions (brack and nobrack) from each other. For our analysis, we differentiated between three types of distinction (**Table 3**): (i) *clear distinction* (abbreviated to C) for cases in which the Mann-Whitney U-Test returns a p-value $< .05$ and the CLES $= 1$, (ii) *partial distinction* (abbreviated to P) for comparisons with a Mann-Whitney U-Test resulting in a p-value $< .05$ and a CLES < 1), and (iii) *no distinction* (abbreviated to N) for cases in which the Mann-Whitney U-Test returns a p-value $> .05$, meaning that this cue does not separate the two conditions.

Table 3: Criteria for the three possible types of distinction of a cue between the two conditions used for the exploratory analysis.

	Estimated probability of the Mann-Whitney U-Test	Common language effect size (CLES)
Clear distinction (C)	$p < .05$	CLES $= 1$
Partial distinction (P)	$p < .05$	CLES < 1
No distinction (N)	$p > .05$	–

In order to explore possible patterns in the use of cues, the individual types of distinction of each of the three cues were combined for each speaker and context. The three cues and their three distinction types combine to 27 possible patterns shown in **Table 4**. The cues are always given in the following order: rise, final lengthening (abbreviated to 'lengthening' to ease the reading), and pause, altogether shortened to RLP and given as subscript at the end of the pattern label. For example, if all three cues are clearly used to distinguish between the two conditions, this pattern is characterized as CCC_{RLP}, as given in the upper leftmost cell of **Table 4**. If rise and pause are clearly used and final lengthening is partially used to distinguish between the two conditions, the pattern is called CPC_{RLP}, as given in the third cell from the left in the upper row of **Table 4**. The raw data of the three cues on Name2 is given in two-dimensional space in the appendices B–D, plotting the magnitude of two of the three cues, each, separated for speaker and context. In order to cover all possible combinations, we visualized three comparisons: (a) pause on the x-axis to rise on the y-axis (Appendix B), (b) final lengthening on the x-axis to pause on the y-axis (Appendix C), and (c) final lengthening on the x-axis to rise on the y-axis (Appendix D).

Table 4: Matrix of possible cue combinations (patterns) of the cues rise (R), final lengthening (L), and pause (P) (in this order). Differentiation between three types of distinction: clear distinction (C), partial distinction (P), no distinction (N).

Rise Lengthening Pause (RLP):	CCC	CCP	CPC	CPP	CCN	CNC	CNN	CPN	CNP
Rise Lengthening Pause (RLP):	PPP	PPC	PCP	PCC	PPN	PNP	PNN	PCN	PNC
Rise Lengthening Pause (RLP):	NNN	NNP	NPN	NPP	NNC	NCN	NCC	NCP	NPC

3. Results

3.1. Statistical analyses of prosodic disambiguation and general context-dependent variability of individual prosodic cues on Name1

3.1.1. Rise on Name1

For rise, the estimates for the fixed-effects were extracted from the maximal model (**Table 5**). Regarding prosodic disambiguation of coordinates, we found a main effect of condition. On average, speakers produced an f0-range on Name1 which was 3 st smaller in the brack compared to the nobrack condition. Regarding general context variability, there was a marginally significant interaction between condition and the CHILD and the ELDERLY contexts, indicating that speakers showed a tendency to decrease their f0-range in the brack condition even more when speaking to a child and to an elderly adult.

Table 5: Estimates of the model for rise on Name1 (i.e., f0-range on Name1). Statistically significant effects are marked in bold ($p < .05$).

Predictor	Estimate	*SE*	*t*-value	*p*-value
condition_brack	−1.53	0.27	−5.76	**< .001**
CHILD VS. YOUNG	0.30	0.21	1.43	.168
ELDERLY VS. YOUNG	0.41	0.33	1.24	.233
NON-NATIVE VS. YOUNG	0.12	0.32	0.39	.704
NOISE VS. YOUNG	0.68	0.38	1.80	.091
condition_brack: CHILD VS. YOUNG	−0.45	0.23	−1.98	.064
condition_brack: ELDERLY VS. YOUNG	−0.34	0.18	−1.94	.064
condition_brack: NON-NATIVE VS. YOUNG	−0.33	0.21	−1.53	.142
condition_brack: NOISE VS. YOUNG	0.09	0.22	0.43	.675

3.1.2. Final lengthening on Name1

For final lengthening, the estimates for the fixed-effects were extracted from the maximal model (**Table 6**). Regarding prosodic disambiguation of coordinates, there was a main effect of condition, indicating that the duration of the final segment of Name1 was shorter in the brack compared to the nobrack condition. With respect to general context variability, we found a marginally significant interaction between condition and the NON-NATIVE context, indicating that the difference between nobrack and brack was larger when speakers addressed the non-native adult, since the duration of the final segment tended to be even shorter in the brack condition.

3.1.3. Pause on Name1

Since six out of 15 speakers did not produce a pause after Name1 in the brack nor in the nobrack condition and there were only three speakers who produced a pause in each of the contexts, we did not run any statistical analyses of pause duration after Name1.

3.2. Statistical Analyses of prosodic disambiguation and general context-dependent variability of individual prosodic cues on Name2

3.2.1. Rise on Name2

For rise, the estimates for the fixed-effects were extracted from the maximal model (**Table 7**). We found a main effect of condition and a main effect of the contexts CHILD

Table 6: Estimates of the model for final lengthening on Name1. Statistically significant effects are marked in bold ($p < .05$).

Predictor	Estimate	*SE*	*t*-value	*p*-value
condition_brack	−2.87	0.49	−5.85	**<.001**
CHILD VS. YOUNG	−0.33	0.70	−0.47	.644
ELDERLY VS. YOUNG	1.41	0.93	1.51	.152
NON-NATIVE VS. YOUNG	0.64	0.62	1.04	.305
NOISE VS. YOUNG	−0.06	0.92	−0.07	.949
condition_brack: CHILD VS. YOUNG	−1.02	0.58	−1.75	.085
condition_brack: ELDERLY VS. YOUNG	−0.99	0.57	−1.75	.084
condition_brack: NON-NATIVE VS. YOUNG	−1.23	0.62	−1.96	.058
condition_brack: NOISE VS. YOUNG	−0.72	0.57	−1.25	.213

Table 7: Estimates of the model for rise on Name2 (i.e., f0-range on Name2). Statistically significant effects are marked in bold ($p < .05$).

Predictor	Estimate	*SE*	*t*-value	*p*-value
condition_brack	2.86	0.27	10.77	**<.001**
CHILD VS. YOUNG	1.10	0.31	3.56	**.003**
ELDERLY VS. YOUNG	0.94	0.36	2.59	**.021**
NON-NATIVE VS. YOUNG	0.59	0.31	1.89	.078
NOISE VS. YOUNG	0.53	0.30	1.76	.097
condition_brack: CHILD VS. YOUNG	0.03	0.20	0.16	.874
condition_brack: ELDERLY VS. YOUNG	0.09	0.22	0.40	.694
condition_brack: NON-NATIVE VS. YOUNG	0.04	0.24	0.18	.856
condition_brack: NOISE VS. YOUNG	−0.09	0.26	−0.33	.746

and ELDERLY. Regarding prosodic disambiguation, speakers, overall, produced a larger f0-range in the brack than in the nobrack condition. Regarding contexts, when addressing the child as well as the elderly person, speakers increased the f0-range of the rise compared to addressing the young adult (cf. **Figure 4** and left panel of **Figure 5**). For a subset of 13 female speakers this increased f0-range can be seen in **Figure 4**, where the green and blue (CHILD and ELDERLY context, respectively) dashed and solid lines start below the black lines (YOUNG context) at the beginning of Name2 and rise to a level above the black lines towards the f0-peak of Name2. Note, **Figure 4** cannot be compared directly to the results of the statistical model, since values in Hertz are plotted in the figure, while the model is calculated on semitones and the figure contains only data of a subset of the speakers. A similar, though statistically non-significant tendency is observable for the other two contexts, addressing the non-native speaker and in noise. The model revealed no statistically significant interactions between contexts and condition.

3.2.2. Final lengthening on Name2

For final lengthening, the estimates for the fixed-effects were extracted from a model that included principal components but not the correlation parameters in the random-effects

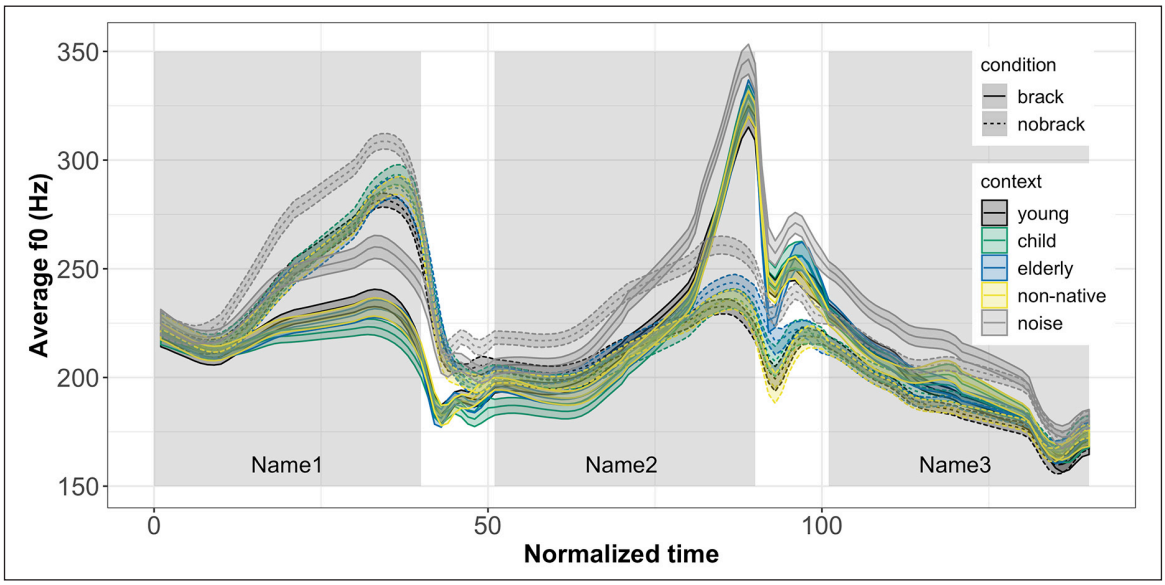

Figure 4: Time-normalized f0-contours (in Hz) of coordinates in brack (solid lines) and nobrack (dashed lines) conditions produced in five contexts (cf. colours) by a subset of 13 female speakers.

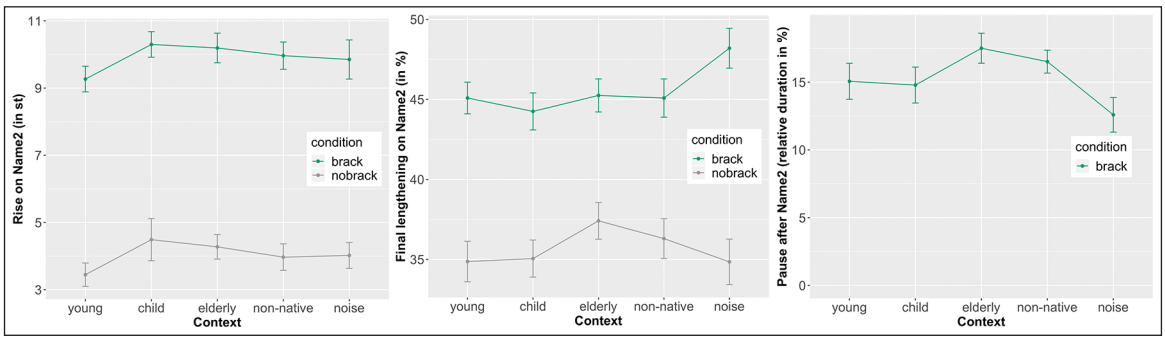

Figure 5: Mean values and 95% confidence intervals for rise (left panel), final lengthening (mid panel), and pause (right panel) on Name2 for each context and condition (green = condition brack, grey = condition nobrack).

structure (**Table 8**). Regarding prosodic disambiguation, the data show a main effect of condition, indicating that speakers marked the brack condition with increased final lengthening compared to the nobrack condition. Regarding general context variability, there was an interaction between condition and the ELDERLY context with a negative estimate and an interaction between condition and the NOISE context with a positive estimate (cf. mid panel **Figure 5**). This indicates that speakers increased the final lengthening when addressing the elderly as opposed to the young interlocutor in the nobrack condition. In the noise context, however, they increased final lengthening in the brack condition, but not in the nobrack condition.

3.2.3. Pause on Name2

For pause, the model was run on a subset containing the brack condition only; the nobrack condition was excluded, due to the large number of zero values. The estimates of the fixed-effects were extracted from the zero correlation parameter model including all variance components (**Table 9**). Regarding general context variability, there was a main effect of the ELDERLY context indicating that speakers produced a longer pause addressing the elderly compared to the young adult interlocutor (cf. right panel **Figure 5**). Additionally, speakers showed a tendency to reduce pause duration in the noisy environment (NOISE), though this was not statistically significant.

3.3. Exploratory analyses of inter- and intra-speaker variability of cue combinations on Name2

The cue combinations for each speaker (cf., y-axis) and context (cf., x-axis) are plotted in **Figure 6**. For each speaker and context, the cell is divided into three rows, with the distinction type of rise given in the uppermost row of the cell, final lengthening in the middle row, and pause in the bottom row. The three types of distinction are represented

Table 8: Estimates of the model for final lengthening on Name2. Statistically significant effects are marked in bold ($p < .05$).

Predictor	Estimate	*SE*	*t*-value	*p*-value
condition_brack	4.96	0.58	8.59	**<.001**
CHILD VS. YOUNG	−0.39	0.58	−0.67	.511
ELDERLY VS. YOUNG	1.34	0.62	2.17	**.039**
NON-NATIVE VS. YOUNG	0.61	0.57	1.07	.293
NOISE VS. YOUNG	1.45	0.71	2.04	.056
condition_brack: CHILD VS. YOUNG	−0.56	0.51	−1.09	.274
condition_brack: ELDERLY VS. YOUNG	−1.24	0.51	−2.43	**.015**
condition_brack: NON-NATIVE VS. YOUNG	−0.65	0.51	−1.27	.205
condition_brack: NOISE VS. YOUNG	1.43	0.52	2.75	**.006**

Table 9: Estimates of the model for pause after Name2. Statistically significant effects are marked in bold ($p < .05$).

Predictor	Estimate	*SE*	*t*-value	*p*-value
CHILD VS. YOUNG	0.02	0.96	0.02	.984
ELDERLY VS. YOUNG	2.67	1.08	2.47	**.025**
NON-NATIVE VS. YOUNG	1.45	0.97	1.5	.153
NOISE VS. YOUNG	−2.56	1.36	−1.87	.08

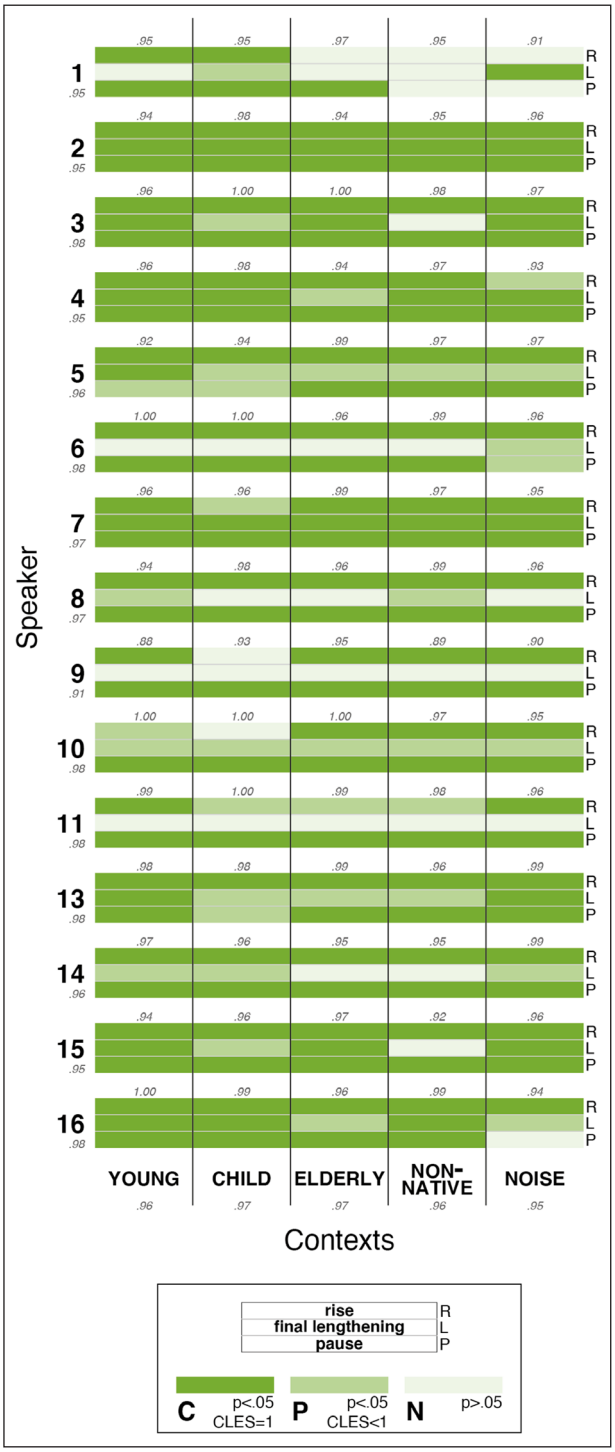

Figure 6: Speaker variability of cue combinations across contexts on Name2, showing the patterns of cue combinations (shades of green) used by individual speakers (y-axis) in the contexts YOUNG, CHILD, ELDERLY, NON-NATIVE, and NOISE (x-axis). The shades of green indicate the type of distinction: full = clear distinction (C), light = partial distinction (P), lightest = no distinction (N). For each speaker, the three rows indicate the different cues (R: rise, L: final lengthening, P: pause). The small numbers in italics indicate the mean ratios of the hit-rates for condition in the perception check (i.e., ratio of correct identifications of the intended grouping to all rates; numbers to the left: average per speaker; lowest line: average per context; above cells: average per speaker per context). For example, speaker 16 clearly distinguishes between the brack and nobrack condition, using all three cues in the YOUNG, CHILD, and NON-NATIVE context, but in the context NOISE the speaker uses final lengthening only partially and pause not at all to distinguish between the two conditions. In the YOUNG context, 100% of the rates in the perception check were congruent rates, while in the NOISE context 94% were congruent rates.

by shading: full colour for clear distinction, light shade for partial distinction, and the lightest shade for no distinction.

Regarding *inter-speaker variability*, we focused on whether there are different patterns of cue combinations between speakers within the young context only, represented in the left-most column of the plot in **Figure 6**. In general, five different patterns are observable: CCC_{RLP} (i.e., all three cues in full green), CNC_{RLP}, CPC_{RLP}, CCP_{RLP}, and PPC_{RLP}, however they differ in number of occurrences. Seven speakers out of 15 (2, 3, 4, 7, 13, 15, and 16) produced the pattern CCC_{RLP}, indicating that all three cues were clearly used to distinguish between the two conditions. A further four speakers (1, 6, 9, and 11) produced the pattern CNC_{RLP}, which means that they clearly distinguished between brack and nobrack using rise and pause, but not using final lengthening. The other three patterns were produced by either two or one speakers. Overall, in four of the five patterns brack and nobrack were clearly distinguished by at least two of the cues. While both rise and pause were used clearly distinctively by 14 out of 15 speakers, only eight speakers used final lengthening in a clearly distinctive way. Notably, the pattern with no distinction in all three cues was never observed in the young context.

Regarding *intra-speaker variability*, we focused on whether speakers vary the patterns of cue combinations when addressing different interlocutors or speaking in noise. For that purpose, we examined the patterns of cue combinations within speaker across contexts (i.e., the three rows in each cell for each speaker across columns in **Figure 6**). For speaker 2, the pattern is identical across all five contexts, thus showing stability in the use of the prosodic cues for distinguishing between the brack and nobrack condition across different contexts. Most other speakers show two or three different patterns across contexts (cf., speakers 6, 7, 8, 9, 11, 14 and 3, 4, 5, 10, 13, 15, 16, respectively). Overall, there is more variability across contexts in the use of final lengthening than in rise and pause, visualized by more varying shading in the middle row of the speaker-specific cells. There is one speaker who in one context shows no distinction between brack and nobrack in any of the three cues (cf., NNN_{RLP} speaker 1, context non-native); for all other speakers and contexts at least one cue is clearly distinctive. In addition, we plotted the mean ratio of the hit-rates for condition in the perception check in **Figure 6** for the respective speakers and contexts. This allows us to get an impression of the relation of the produced types of distinction in the three prosodic cues to how well the prosodic boundaries (i.e., conditions) have been perceived by naïve listeners.

4. Discussion

In the current study, we aimed to gain insights into the situational dependence or independence of disambiguating prosody and to learn more about the nature of the prosody-syntax relation. To this end, we explored the production of prosodic boundaries used to disambiguate coordinated sequences of three names (coordinates) between two conditions: without (nobrack) and with (brack) internal grouping of the first two names. We focussed on the variability induced by speakers and contextual settings, such as interlocutors differing in age and mother tongue, as well as the absence/presence of noise (contexts). Besides the distinction between the two conditions (prosodic disambiguation of coordinates, research question Q1), we were interested in the type and size of cues produced at the prosodic boundaries and whether and how speakers varied in producing them depending on the context. Coordinate productions were elicited by means of a referential communication task with five contexts: addressing a young adult (YOUNG), a child (CHILD), an elderly adult (ELDERLY), a young non-native adult (NON-NATIVE), and the young adult in a noisy environment (NOISE). Variability was addressed on three levels: across speakers between contexts (general context-dependent prosodic variability,

research question Q2), between speakers within contexts (inter-speaker variability of cue combinations, research question Q3), as well as within speakers between contexts (intra-speaker variability of cue combinations, research question Q4).

4.1. Prosodic disambiguation of coordinates (research question Q1)

Our findings replicate previous studies, showing that the internal grouping of coordinates in German is marked by a prosodic boundary consisting of three prosodic cues from the tonal and durational domain: f0-range, final lengthening, and pause. As expected, speakers used prosodic cues on Name1 as well as on Name2 to clearly distinguish between the two conditions. A perception check with naïve listeners showed that the distinction between the conditions was perceptually recoverable: 96% of the productions were correctly recognized as the intended grouping.

The results of the production study are in line with the Proximity and Anti-Proximity principles that form part of the Proximity/Similarity model introduced by Kentner and Féry (2013) and along with this, they are in line with the literature (Taglicht, 1998; Wagner, 2005, 2010; Watson & Gibson, 2004). Thus, our hypothesis (Q1) was confirmed: In the condition with internal grouping compared to the condition without grouping, we found a statistically significant decrease in final lengthening and f0-range on Name1 along with an increase in final lengthening and f0-range on Name2 as well as the insertion of a pause after Name2. In terms of Proximity, durational and tonal cues of Name1 were decreased, indicating that the neighbouring element to the right (i.e., Name2) forms part of the same group. In terms of Anti-Proximity, the prosodic boundary after Name2 was strengthened, indicating that the neighbouring element to the right (i.e., Name3) does not form a group with Name2. This finding also underlines the assumption that prosodic phrasing is not a local phenomenon with changes of prosodic cues occurring only at the prosodic boundary (cf. in our case Name2) but rather depends on globally distributed prosodic changes (cf. in our case Name1 and Name2) (e.g., Clifton et al., 2002; Frazier et al., 2006; Wagner, 2005, 2010).

We further found that speakers use the pause cue in a slightly different way than f0-range and final lengthening in marking the difference between conditions. Following Name2, a pause was mostly absent in the condition without internal grouping, while it was present in the condition with grouping. The pause, thus, appears rather as a categorical than a continuous variable. Since we were interested in differences in pause duration between contexts, however, we kept pause as a continuous variable for our analyses.

Overall, the syntactic structure (with or without internal grouping) was clearly disambiguated by means of prosody. This can be interpreted as evidence in favour of a close link between syntax and prosody.

4.2. General context-dependent prosodic variability (research question Q2)

The current study is, to the best of our knowledge, the first to systematically investigate prosodic variability in production of coordinates across speakers between various contexts to explore the situational in/dependence of disambiguating prosody and to find out whether the principles of Proximity/Anti-Proximity also hold across situations.

At the group level, we found some variability driven by the different contexts. Nevertheless, variability was rather small and not as distinct as expected. In the following, the contexts CHILD, ELDERLY, NON-NATIVE, and NOISE will be discussed individually in comparison to the baseline context (YOUNG).

In the context CHILD, when addressing the child as opposed to the young adult, speakers changed their productions in the tonal domain: They increased the f0-range on Name2 independent of condition. This can be interpreted as an adaptation to the interlocutor, but

without affecting the ease of disambiguation between conditions. The increased f0-range when addressing a child is partly in line with semi-spontaneous speech data from English speakers (Biersack et al., 2005), who additionally showed lengthened vowels. These differences might be due to differences in age of the interlocutor. For a child addressee of the same age as in our study, DePaulo and Coleman (1986) reported longer pauses; a finding that was not evident in our data. A possible explanation for the absence of statistically significant effects in the durational prosodic cues (i.e., final lengthening and pause) in our study might be related to differences in speech style as well as in language-specific factors. Our data were highly restricted with respect to the wording, whereas the data of DePaulo and Coleman (1986) consisted of spontaneous speech and the data of Biersack et al. (2005) of semi-spontaneous speech, both in English.

In the context ELDERLY, when addressing the elderly adult compared to the young adult, speakers modified their speech in the tonal as well as in the durational domain. On Name2, speakers produced an overall larger f0-range in the ELDERLY context along with a longer pause (in the condition with internal grouping). In contrast, final lengthening on Name2 was not used to make the conditions more distinct in the ELDERLY context: Unexpectedly, speakers increased the lengthening in the condition without grouping compared to coordinates addressed to the young adult. Yet, with the increased pause duration, the smaller difference in final lengthening between the conditions was probably levelled out. The findings of increased pause durations and increased f0-ranges, thus, partly confirm our hypotheses and are comparable to observations on other structures in English and German (Kemper et al., 1995; Thimm et al., 1998). Those studies found slower speech due to prolonged vowels and more pauses as well as increased variation in intonation, among other speech adaptations. Regarding the increased number of pauses in the reported studies, again, it needs to be mentioned that the respective data stem from spontaneous speech which probably allows for more pause insertion than the relatively restricted stimuli used in our study. Nevertheless, we suggest that the increased pause durations in our data can be interpreted as comparable speech adaptations. In previous research on speech directed at elderly persons, Kemper, Ferrell, Harden, Finter-Urczyk, & Billington (1998) distinguished two sets of parameters that speakers modify in order to adapt to the needs of their elder interlocutor: semantic and discourse information on the one hand, and fluency, prosody, and grammatical complexity on the other. Kemper et al. (1998, p. 53) discuss that the latter set of parameters does not "appear to benefit" perception, but to the contrary, decreases self-esteem on the side of the interlocutor. This type of speech is referred to as patronizing communication (Kemper et al., 1998; Ryan, Hummert, & Boich, 1995; Thimm et al., 1998; Torrey, Fussell, & Kiesler, 2005) and includes the changes in prosodic cues found in our data.

In the context NON-NATIVE, in response to the non-native interlocutor, the data show no clear effects. This contrasts with reports in the literature, in which non-native speakers were addressed with increased f0-ranges and a more emphatic style compared to native speakers (Smith, 2007).

Finally, in the context NOISE, the interlocutor was the same young adult as in the baseline context. For adaptation to the noise, speakers increased final lengthening on Name2 in the condition with grouping while at the same time, they decreased the relative duration of the following pause. The increase in final lengthening is in line with our hypotheses and findings in the literature, although we would have expected an additional increase in the f0-range. A possible explanation for the unexpected decrease in pause duration is that a silent pause is a less effective cue in a noisy environment than in a quiet one. Instead of a silent pause, speakers lengthened the final segment to mark the boundary. Furthermore, speakers might have tried to fill the noise with their own voice, in order to

distract themselves from the noise. Varadarajan and Hansen (2006) interpreted this result as "a sense of urgency on the part of the speaker […] due to persistent exposure of the environmental noise" (Varadarajan & Hansen, 2006, p. 938).

With respect to our research question Q2, we can conclude that we found only some small differences in the three prosodic boundary cues produced in coordinates elicited in different contexts. In addition, the small differences between the contexts could hardly be discriminated on the perceptual side as shown by the weak performance in the perception check regarding the assignment of productions to the differential contexts: Listeners were not able to reliably identify to whom the utterance was addressed.

With regard to the question of situational in/dependence of prosodic disambiguation, the finding of clear production of a prosodic boundary to disambiguate the conditions with/without grouping (Q1) together with the only small contextual adaptations (Q2) in our data, speaks in favour of situational independence. In the context of our study, the prosodic distinction between coordinates with or without internal grouping might have been considered to be more 'relevant' than a prosodic adaptation to possibly different needs of the interlocutors.

In the following we will discuss two limitations of our study, before turning to research questions Q3 and Q4:

First, another explanation for the fact that the context effects in our production data were smaller than expected might be based on the somewhat artificial design of the study: The interlocutors were auditorily present before the recording of each stimulus, however, there was no feedback of their perceptual performance. A request for repetition or a misunderstanding may have triggered further accommodations in the speech addressed to the interlocutors. As mentioned above, accommodation to possible needs of an interlocutor can also be interpreted as patronizing by the interlocutor, as Kemper et al. (1998) reported for the speech used by young adults when addressing elderly adults. In our study, speakers either may have perceived no need to adapt any further to their interlocutors or they might have been sensible and avoided an over-exaggerated speech style since no feedback was given. This can apply especially for the elderly adult and the young non-native speaker, as they are both adults. Future studies, nevertheless, might want to include feedback of the interlocutors in order to increase the necessity of speakers to adapt to their interlocutors and to make the interaction more natural.

Second, we focused on three particular prosodic boundary cues and, therefore, cannot disregard the possibility that speakers may have produced additional prosodic cues to adapt to their interlocutors. This could, for instance, apply to the NOISE context: The context NOISE was best identified in the perception check (17 out of the 21 productions that were correctly identified by all listeners had been produced in the NOISE context and a total of 65 productions in NOISE was correctly identified by at least 75% of all listeners). This suggests that speakers used additional (prosodic) cues to adapt to the noise. Other studies looking at speech in noise reported, for instance, increased intensity in the presence of noise, as well as spectral changes (e.g., Davis et al., 2006; Junqua, 1996; Landgraf et al., 2017; Lu & Cooke, 2008; Summers et al., 1988). This could be seen as further evidence that disambiguating prosody is not primarily produced for the interlocutor but automatically produced 'for' the speaker during planning and articulation: When speakers are confronted with noise, this might affect the cognitive resources used for the planning and articulation and hence get reflected in their prosodic output. Future studies are needed to test this hypothesis.

In the final two sections, we discuss the results of the exploratory analysis regarding which cue combinations are used by individual speakers to mark the prosodic boundary in the grouped name sequences.

4.3. Inter-speaker variability of cue combinations (research question Q3)

With regard to inter-speaker variability of prosodic cues and cue combinations (i.e., the interplay of prosodic cues) in the YOUNG context, the data show that the majority of the speakers (14 out of 15) employed at least two cues distinctively to mark the prosodic boundary in the condition with grouping on Name2. Furthermore, for 13 speakers these two clearly distinctive cues were rise and pause. To put it simply: The vast majority of speakers clearly used pause and rise on Name2 to distinguish between conditions. In comparison to rise and pause, final lengthening was used more variably in the YOUNG context: Some speakers produced it clearly distinctively, others with partial or no distinction. A post-hoc exploratory visual inspection of the data points that were excluded after the perception check further showed that the 'clear distinction'-pattern in either of the three prosodic cues was beneficial for perception: Often the perception of the non-intended condition went along with one of the three prosodic cues falling within the range of the values of the perceived condition. In other words, if for instance a grouped item was perceived as having no internal grouping, the value of one of the three prosodic cues was more similar to other items without grouping of that speaker than to grouped items.

Overall, most speakers combined at least two cues to clearly disambiguate the conditions, but still, there is some variability between speakers. This speaks in favour of a close relation of syntax and prosody that nonetheless allows for some flexibility in how prosodic boundaries are phonetically realized at the surface (Wagner, 2005, p. 155). Despite this variability between speakers at the phonetic level, the boundaries are easily and reliably detected by the listeners, as shown by the perception check.

4.4. Intra-speaker variability of cue combinations (research question Q4)

This discussion concerns the question whether individual speakers mark the boundaries on Name2 differently in the five contexts. Mirroring the group analysis (see 4.2.), almost half of the speakers (7 out of 15) were stable across contexts also with regard to the relation between cues, as they used one or two patterns only. A closer look at these speakers revealed that the patterns they used mostly contained alternations in one cue only and were, consequently, quite similar to each other. Again, final lengthening emerges as the cue used least distinctively of the three cues investigated, while rise and pause in most cases clearly distinguish between the two conditions–also across contexts. In conclusion, in terms of cue patterns used, the differences between contexts were quite small and individual speakers rather stuck to their individual 'prosodic strategy' of marking the boundaries in the condition with grouping independent of their interlocutor.

Overall, we can summarize that individual speakers showed a limited set of cue patterns with only slight shifts in cue distribution between contexts. Hence, also the analysis of individual speakers in varying contexts is in favour of a relatively limited range of variability or rather stable intra-individual 'prosodic strategies' to disambiguate coordinates with versus without internal grouping. This adds to the notion of situational independence of disambiguating prosody that is produced automatically by the speakers in a rather invariant manner.

5. Conclusion

In conclusion, speakers in our production study used prosodic boundaries to reliably mark constituent grouping in sequences of three coordinated names. At the phonetic level, speakers mainly used f0-range and pause for prosodic disambiguation, while final lengthening was used more flexibly. Across contexts, speakers behaved in accordance to the Proximity/Anti-Proximity principle of the syntax-prosody model by Kentner and

Féry (2013): When the first two names were grouped together, the durational and tonal cues of the first name were weakened, while the boundary on the second name was strengthened. We found only limited contextual effects within speakers, but inter-speaker variability in how the prosodic boundaries were phonetically realized. The data hence indicate a close link between syntax and prosody that is employed independently of the actual communicative situation with some flexibility at the surface.

Additional Files

An Open Science Framework project page (https://osf.io/rnxej/) has been created to store the data and code. We additionally provide the following files:

- **Appendix A:** Wording of introduction and instruction in the five contexts. DOI: https://doi.org/10.5334/labphon.221.s1
- **Appendix B:** Comparison pause_rise on Name2 with pause plotted on the x-axis and rise on the y-axis. DOI: https://doi.org/10.5334/labphon.221.s2
- **Appendix C:** Comparison lengthening_pause on Name2 with final lengthening plotted on the x-axis and pause on the y-axis. DOI: https://doi.org/10.5334/labphon.221.s3
- **Appendix D:** Comparison lengthening_rise on Name2 with final lengthening plotted on the x-axis and rise on the y-axis. DOI: https://doi.org/10.5334/labphon.221.s4

Acknowledgements

We would like to thank J. Ries for his help with setting up the experiment and for the support with the exploratory analysis, as well as L. Junack, A. Hofmann, and E. Weiß for their help with data acquisition and pre-processing, and D. Schad and R. Kliegl for their advice on data analysis. Another thank goes to the four interlocutors for lending us their voices and faces. We also thank B. Höhle and two anonymous reviewers as well as the associate editor for providing valuable comments to an earlier version of this manuscript.

Funding Information

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 317 633 480 – SFB 1287, Project B01.

Competing Interests

The authors have no competing interest to declare.

References

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 714–735. DOI: https://doi.org/10.1037/0278-7393.22.3.714

Baayen, H., Bates, D., Kliegl, R., & Vasishth, S. (2015). *RePsychLing: Data sets from Psychology and Linguistics experiments* (0.0.4) [Computer software]. Retrieved from https://github.com/dmbates/RePsychLing

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. Retrieved from ArXiv:1506.04967v2. https://arxiv.org/abs/1506.04967v2

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204. DOI: https://doi.org/10.1017/S004740450001037X

Biersack, S., Kempe, V., & Knapton, L. (2005). Fine-tuning speech registers: A comparison of the prosodic features of child-directed and foreigner-directed speech. *Ninth European Conference on Speech Communication and Technology*, 2401–2404. Retrieved from http://www.interspeech2005.org/

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer* (6.0.32) [Computer software]. Retrieved from www.praat.org

Clifton, C., Jr., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, *45*(2), 87–114. DOI: https://doi.org/10.1177/00238309020450020101

Davis, C., Kim, J., Grauwinkel, K., & Mixdorff, H. (2006). Lombard speech: Auditory (A), Visual (V) and AV effects. *Proceedings of the Third International Conference on Speech Prosody*, 248–252. Retrieved from http://www.isca-speech.org/archive

DePaulo, B. M., & Coleman, L. M. (1986). Talking to children, foreigners, and retarded adults. *Journal of Personality and Social Psychology*, *51*(5), 945–959. DOI: https://doi.org/10.1037/0022-3514.51.5.945

Folk, L., & Schiel, F. (2011). The Lombard effect in spontaneous dialog speech. *Proceedings of the Interspeech*, 2701–2704.

Frazier, L., Carlson, K., & Clifton, C., Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*(6), 244–249. DOI: https://doi.org/10.1016/j.tics.2006.04.002

Garnier, M., Bailly, L., Dohen, M., Welby, P., & Lœvenbruck, H. (2006). An acoustic and articulatory study of Lombard speech: Global effects on the utterance. *Ninth International Conference on Spoken Language Processing*, 2246–2249.

Gollrad, A., Sommerfeld, E., & Kügler, F. (2010). Prosodic cue weighting in disambiguation: Case ambiguity in German. *Speech Prosody 2010-Fifth International Conference*. *Speech Prosody 2010-Fifth International Conference*, Chicago, IL.

Google Ngram Viewer. (n.d.). Retrieved from https://books.google.com/ngrams

Grice, M., & Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte*, *191*, 267–298.

Grice, M., Baumann, S., & Benzmüller, R. (2005). German intonation in autosegmental-metrical phonology. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*, 55–83. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199249633.003.0003

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). DlexDB: Eine lexikalische Datenbank für die psychologische und linguistische Forschung. In *Psychologische Rundschau, 62*, 10–20. Hogrefe Verlag. DOI: https://doi.org/10.1026/0033-3042/a000029

Holzgrefe-Lang, J. (2017). *Prosodic phrase boundary perception in adults and infants* (Doctoral dissertation, University of Potsdam). Retrieved from https://publishup.uni-potsdam.de/frontdoor/index/index/docId/40594

Holzgrefe-Lang, J., Wellmann, C., Petrone, C., Räling, R., Truckenbrodt, H., Höhle, B., & Wartenburger, I. (2016). How pitch change and final lengthening cue boundary perception in German: Converging evidence from ERPs and prosodic judgements. *Language, Cognition and Neuroscience*, *31*(7), 904–920. DOI: https://doi.org/10.1080/23273798.2016.1157195

Howell, D. C. (1998). *Statistical methods in human sciences*. Wadsworth.

Jessen, M., Köster, O., & Gfroerer, S. (2003). Effect of increased vocal effort on average and range of fundamental frequency in a sample of 100 German-speaking male subjects. *Reports of the 15th International Congress of Phonetic Sciences*, 1623–1626.

Junqua, J.-C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America*, *93*(1), 510–524. DOI: https://doi.org/10.1121/1.405631

Junqua, J.-C. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication*, *20*(1–2), 13–22. DOI: https://doi.org/10.1016/S0167-6393(96)00041-6

Kempe, V., Schaeffler, S., & Thoresen, J. C. (2010). Prosodic disambiguation in child-directed speech. *Journal of Memory and Language*, *62*(2), 204–225. DOI: https://doi.org/10.1016/j.jml.2009.11.006

Kemper, S., Ferrell, P., Harden, T., Finter-Urczyk, A., & Billington, C. (1998). Use of elderspeak by young and older adults to impaired and unimpaired listeners. *Aging, Neuropsychology, and Cognition*, *5*(1), 43–55. DOI: https://doi.org/10.1076/anec.5.1.43.22

Kemper, S., Vandeputte, D., Rice, K., Cheung, H., & Gubarchuk, J. (1995). Speech adjustments to aging during a referential communication task. *Journal of Language and Social Psychology*, *14*(1–2), 40–59. DOI: https://doi.org/10.1177/0261927X95141003

Kentner, G., & Féry, C. (2013). A new approach to prosodic grouping. *The Linguistic Review*, *30*(2), 277–311. DOI: https://doi.org/10.1515/tlr-2013-0009

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, *50*(2), 194–231. DOI: https://doi.org/10.1016/j.cogpsych.2004.08.002

Kuznetsova, A., Brockhoff, P. B., & Bojensen Christensen, R. H. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

Ladd, D. R. (1986). Intonational phrasing: The case for recursive prosodic structure. *Phonology Yearbook*, *3*, 311–340. DOI: https://doi.org/10.1017/S0952675700000671

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511808814

Landgraf, R., Schmidt, G., Köhler-Kaeß, J., Niebuhr, O., & John, T. (2017). More noise, less talk: The impact of driving noise and in-car communication systems on acoustic-prosodic parameters in dialogue. *43. Deutsche Jahrestagung für Akustik (DAGA)*, 1485–1488.

Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *The Journal of the Acoustical Society of America*, *54*(5), 1228–1234. DOI: https://doi.org/10.1121/1.1914379

Lombard, É. (1911). Le signe de l'élévation de la voix. *Annales des Maladies de l'Oreille et du Larynx*, *37*, 101–109.

Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America*, *124*(5), 3261–3275. DOI: https://doi.org/10.1121/1.2990705

Luo, J., Hage, S. R., & Moss, C. F. (2018). The Lombard effect: From acoustics to neural mechanisms. *Trends in Neurosciences*, *41*(12), 938–949. DOI: https://doi.org/10.1016/j.tins.2018.07.011

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. DOI: https://doi.org/10.1214/aoms/1177730491

MATLAB (9.6.0.1135713 (R2019a) Update 3). (2019). [Computer software]. The MathWorks Inc.

McGraw, K. O., & Wong, S. (1992). A common language effect size statistic. *Psychological Bulletin*, *111*(2), 361–365. DOI: https://doi.org/10.1037/0033-2909.111.2.361

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Foris.

Neurobehavioural Systems (20.1). (2018). [Computer software]. Retrieved from https://www.neurobs.com/

Peters, B., Kohler, K. J., & Wesener, T. (2005). Phonetische Merkmale prosodischer Phrasierung in Deutscher Spontansprache. In K. J. Kohler, F. Kleber, & B. Peters (Eds.), *Prosodic structures in German spontaneous speech, 35*, 143–184. AIPUK, IPDS Kiel.

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. DOI: https://doi.org/10.1016/j.wocn.2017.01.002

R Development Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Ryan, E. B., Hummert, M. L., & Boich, L. H. (1995). Communication predicaments of aging: Patronizing behavior toward older adults. *Journal of Language and Social Psychology*, *14*(1–2), 144–166. DOI: https://doi.org/10.1177/0261927X95141008

Schad, D. J., Hohenstein, S., Vasishth, S., & Kliegl, R. (2018). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *ArXiv Preprint ArXiv:1807.10451*.

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, *29*(2), 169–182. DOI: https://doi.org/10.1023/A:1005192911512

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). *ToBI: A standard for labeling English prosody*. 867–870.

Smith, C. L. (2007). Prosodic accommodation by French speakers to a non-native interlocutor. *Proceedings of the XVIth International Congress of Phonetic Sciences*, 1081–1084.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103–130. DOI: https://doi.org/10.1016/S0749-596X(02)00519-3

Speer, S. R., Warren, P., & Schafer, A. J. (2011). Situationally independent prosodic phrasing. *Laboratory Phonology*, *2*(1), 35–98. DOI: https://doi.org/10.1515/labphon.2011.002

Stanton, B. J., Jamieson, L. H., & Allen, G. D. (1988). *Acoustic-phonetic analysis of loud and Lombard speech in simulated cockpit conditions*. 331–334. DOI: https://doi.org/10.1109/ICASSP.1988.196583

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, *84*(3), 917–928. DOI: https://doi.org/10.1121/1.396660

Taglicht, J. (1998). Constraints on intonational phrasing in English. *Journal of Linguistics*, *34*(1), 181–211. DOI: https://doi.org/10.1017/S0022226797006877

Thimm, C., Rademacher, U., & Kruse, L. (1998). Age stereotypes and patronizing messages: Features of age-adapted speech in technical instructions to the elderly. *Journal of Applied Communication Research*, *26*(1), 66–82. DOI: https://doi.org/10.1080/00909889809365492

Torrey, C., Fussell, S. R., & Kiesler, S. B. (2005). Appropriate accommodations: Speech technologies and the needs of older adults. *Caring Machines: AI in Eldercare: Papers from the AAAI Fall Symposium*, 99.

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schließer (Eds.), *Methods in empirical prosody research*, *3*, 1–28. Walter de Gruyter. DOI: https://doi.org/10.1515/9783110914641.1

van den Berg, R., Gussenhoven, C., & Rietveld, T. (1992). Downstep in Dutch: Implications for a model. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (Vol. 335, p. 359). Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511519918.015

Varadarajan, V. S., & Hansen, J. H. (2006). Analysis of Lombard effect under different types and levels of noise with application to in-set speaker ID systems. *INTERSPEECH-2006*, 937–940.

Wagner, M. (2005). *Prosody and recursion* (Doctoral dissertation, Massachusetts Institute of Technology, Dept. of Linguistics and Philosophy). Retrieved from http://hdl.handle.net/1721.1/33713

Wagner, M. (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory*, *28*(1), 183–237. DOI: https://doi.org/10.1007/s11049-009-9086-0

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, *19*(6), 713–755. DOI: https://doi.org/10.1080/01690960444000070

Zollinger, S. A., & Brumm, H. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour*, *148*(11–13), 1173–1198. DOI: https://doi.org/10.1163/000579511X605759

How to cite this article: Huttenlauch, C., de Beer, C., Hanne, S., & Wartenburger, I. 2021 Production of prosodic cues in coordinate name sequences addressing varying interlocutors. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12(1): 1, pp. 1–31. DOI: https://doi.org/10.5334/labphon.221

Submitted: 20 August 2019 **Accepted:** 01 December 2020 **Published:** 25 January 2021