The role of prosody and hand gestures in the perception of boundaries in speech

Manon Lelandais, Gabriel Thiberge

The role of prosody and hand gestures in the perception of boundaries in speech

Manon Lelandais [a], [b] & Gabriel Thiberge [a], [c]

[a]Université Paris Cité, 85 boulevard Saint-Germain 75006 Paris, France.
[b]UR 3967 CLILLAC-ARP, Université Paris Cité, 8 place Paul Ricoeur, 75013, Paris, France.
[c]UMR 7110 Laboratoire de Linguistique Formelle, Case Postale 7031, 5, rue Thomas Mann, F-75205 Paris Cedex 13.
manon.lelandais@u-paris.fr

gthg@tuta.io

Corresponding author: Manon Lelandais

Author's version, Accepted Manuscript

Lelandais, M. & Thiberge, G. 2023. "The role of prosody and gesture in the perception of boundaries in speech". *Speech Communication* 150: 41-65. **https://doi.org/10.1016/j.specom.2023.05.001**

Abstract

This paper investigates the use of prosodic, gestural, and syntactic information in the perception of boundaries in extracts of spontaneous speech in British English. Experiment 1 aimed at investigating the effect of prosody on naive participants' perception of boundary strength. 13 naive listeners had to rate boundary strength for 64 extracts on a 5-point scale. The stimuli all contained three tone-units, the second being a syntactic subordinate construction, which was established as a variable. The prosodic cues at the boundary between the tone-units were also established as variables, and were subject to manipulation (addition of a single cue associated with the perception of a prosodic boundary). Experiment 2 aimed at assessing the effect of gesture on naive participants' perception of boundary strength. In Experiment 2, 24 naive listeners had to measure boundary strength for 24 extracts on a 5-point scale. The stimuli all contained three tone-units, the second being a syntactic subordinate construction, which was established as a variable. The hand gestures produced in co-occurrence with the tone-units were established as variables, and were subject to manipulation. Results show that prosody modulates perceived boundary strength, but not gesture, based on the variables we included. Silent pauses have the strongest effect on perceived boundary strength, but final syllabic lengthening and pitch reset also have separate effects as single predictors. Our data also shows a trend concerning the production of two identical hand gestures in terms of configuration and trajectory.

Keywords: perception, speech boundaries, prosody, gesture, subordination.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Transcription conventions

one line of transcription corresponds to one tone-unit

#	pause
[…]	illustrated gestural activity
(1)	left co-text (tone-unit)
(2)	subordinate construction
(3)	right co-text (tone-unit)

Data repository

https://osf.io/j8zux/?view_only=35b3cea05f1548daa5171a7957097b68

1. Introduction

This article discusses the use of verbal, vocal, and gestural information in the perception of boundaries in spontaneous speech (British English). We investigated the strength of perceived syntactic boundaries by manipulating prosodic (Experiment 1) and gestural cues (Experiment 2), in stimuli extracted from a multimodal corpus of spontaneous speech. The term "boundary" refers to a perceptual break expressed with verbal, prosodic and / or gestural cues between two parts of speech.

During speech perception in spontaneous conversation, co-speakers rely on a wide range of simultaneous cues to support comprehension, from prosodic and gestural information to semantic context. Some of these perceptual cues facilitate the segmentation process of continuous streams of information into sub-units. As part of this process, prosodic phrasing is known to play an important part in the segmentation of acoustic information (Kuang, Pik Yu Chan, and Rhee, 2022). The perception of prosodic boundaries is influenced by a number of acoustic cues at the final and initial points of domains, involving both rhythm and melody (Astésano *et al.*, 2012; Barth-Weingarten, 2016). The established acoustic correlates to the perception of a prosodic boundary, in English and in some other languages, are the presence of a silent pause (*e.g.* Yoon, Cole, and Hasegawa-Johnson, 2007), final syllabic lengthening (*e.g.*

Scott, 1982; Turk & Shattuck-Hufnagel, 2007), initial pitch upstep[1] (’t Hart *et al.*, 1990; Wagner & Watson, 2010, pp. 907–910), as well as the presence of falling or rising tones (Pierrehumbert, 1980; Portes, 2002)[2].

Prosodic phrasing is also directly related to linguistic processing taking place in other domains, such as syntactic parsing (Kuang, Pik Yu Chan, and Rhee, 2022). While listeners have been shown to use prosodic boundaries to locate syntactic boundaries or resolve syntactic ambiguity (Beach 1991; Watson & Gibson 2004), much less is known about the reciprocal, *i.e.* whether syntax modulates listeners' perception of prosodic boundaries (Kuang, Pik Yu Chan, and Rhee, 2022). Cole, Mo, and Baek (2010) as well as Buxò-Lugo & Watson (2016) found that syntactic cues do influence the perception of prosodic boundaries. However, the link between prosodic and syntactic boundaries is usually documented and recognized for independent or complex clauses (Cole, Mo, and Baek, 2010; Simon & Christodoulides, 2016; Fromont, Soto-Faraco, and Biau, 2017) rather than smaller constituents, given that the theories having led to the categorization of hierarchies for the prosodic and syntactic domains show different granularities (see Kuang, Pik Yu Chan, and Rhee, 2022 for details).

Co-speakers also rely on gestural information to support comprehension in face-to-face conversation (*e.g.* Debreslioska *et al.*, 2013; Perniss & Özyürek, 2015; Masson-Carro, Goudbeek, and Krahmer, 2016). Some of the visual perceptual cues used by co-speakers have been shown to facilitate the segmentation process of speech (*e.g.* Barkhuysen, Krahmer, and Swerts, 2008; De Kok & Heylen, 2009). However, in linguistics, the segmentation function of gestures in speech has mostly been documented for those gestural cues that carry strong prosodic aspects (*e.g.* Dimitrova *et al.*, 2016), and for major boundaries such as end-of-(conversational)-turn signals (*e.g.* Granström, House, and Lundeberg, 1999; De Kok & Heylen, 2009). Much less is known about the segmentation of conversational speech containing smaller-scale boundaries such as clause or tone-unit[3] boundaries.

[1] Throughout a vocal paragraph (i.e. a group of tone-units forming a global intonation contour; Cruttenden, 1986), F0 height naturally decreases progressively. A pitch upstep disrupts this declination line and often indicates a change in paragraphs.

[2] Other cues frequently reported by the literature include domain-final pitch lowering (Beckman & Pierrehumbert, 1986) as well as domain-final and -initial cues related to duration and voice quality (Crowhurst, 2018). These cues are not addressed in the present article.

[3] We refer to tone-units in accordance with the British description of intonation (*e.g.* Crystal 1969; Wells 2006), based on the form of global intonational contours. A tone-unit is a complete coherent intonation contour. It comprises at least one syllable, necessarily the nuclear syllable. The melodic movement starts on the nucleus and can spread to post-nuclear syllables, if any.

Although gaze direction, head movement, and eyebrow movement play a large role in speech segmentation (Argyle & Cook, 1976; Granström & House, 2005; De Kok & Heylen, 2009; Rienks, Poppe, and Heylen, 2010), only hand gestures are discussed in the present paper as they were the only cue tested in this study. We do not focus on the prosodic dimension of beat gestures (see for instance Holle *et al.*, 2012; Dimitrova *et al.*, 2016; Biau, Fromont, and Soto-Faraco, 2018), but on representational hand gestures (*i.e.* gestures describing or representing objects, actions, or ideas).

Gesture production has been shown to be coupled with syntactic packaging choices (*e.g.* Özyürek *et al.*, 2005; Kita *et al.*, 2007; Fritz *et al.*, 2019), including those of clause packaging. These studies nonetheless targeted verb or event encoding rather than different types of clauses as such. McNeill (1992) proposed that speakers generally produce one representational hand gesture per clause. However, it is not clear how complex clauses (for instance a main clause and a subordinate clause) were counted in the study.

The potential of two successive hand gestures with different configurations and trajectories to create a boundary in discourse has been mentioned several times in the literature (Streeck, 2009; Enfield, 2009; Calbris, 2011), but has never been experimentally tested. Hilton *et al.* (2019) carried out experimental work on the perception of (speech) prosody and (nonspeech) gesture. They showed that the segmentation of speech and (nonspeech) action sequences performed with hand gestures both elicit Closure Positive Shifts, indicating similar electrophysiological correlates of boundary processing for speech and visually presented non-speech action sequences.

The two perception experiments in the present paper were designed as a follow-up of production studies on the vocal and gestural characteristics of syntactic subordination in spontaneous speech (see for instance Lelandais & Ferré, 2016). These production studies showed that subordinate structures displayed different degrees of prosodic and gestural boundary depending on their syntactic type, namely appositive clauses, restrictive relative clauses, and adverbial clauses (see also Auran & Loock, 2006 for prosody).

Among these three syntactic types, appositive clauses were found to be produced with the biggest combination of prosodic and gesture boundaries (Lelandais, 2020). In this context as in spontaneous speech in general, variations in duration and silent pauses are the preferred prosodic cues to mark a boundary in discourse (*e.g.* Choi, Hasegawa-Johnson, and Cole, 2005; Mo, Cole, and Lee, 2008; Mo & Cole, 2010), as well as initial pitch upsteps (Collier, de Pijper,

and Sanderman, 1993). As far as gestures are concerned, it was found in a production study (Lelandais, 2020) that speakers use hand gesture parameters to express boundaries in discourse, such as changes in form, direction, and coordinates (*e.g.* Enfield, 2009; Frederiksen, 2016; Streeck, 2009).

Yet, the data collected on the production of subordinate constructions does not document the influence of different prosodic and gestural cues on the perception of syntactic boundaries, especially in the case of different types of subordinate clauses. Moreover, they do not determine whether the preferred cues for discourse segmentation in production are also the preferred cues in perception.

Subordination is still a challenge for Natural Language Processing and discourse modelling (Chen, Alexopoulou, and Tsimpli, 2021). Yet compared to the vast amount of research on subordination either from the point of view of syntax alone or from that of pragmatics, the vocal and gestural contributions to subordination are often left out, just as data coming from spontaneous conversation. The study of subordination in spontaneous speech from a multimodal point of view gives new perspectives on the flexibility of discourse planning and modelling. More information on real-time discourse production with a particular focus on boundaries benefits such areas of study as Natural Language Processing (Biron *et al.*, 2021).

One of the main methodological innovations of the present paper is the use of audio stimuli extracted from spontaneous speech. The speech extracts have all been produced in dialogue settings, in front of a co-speaker.

This paper investigates the perceived strength of syntactic boundaries by manipulating prosodic and gestural cues in extracts of spontaneous speech. We specifically test:

1) Whether prosody modulates the strength of perceived syntactic boundaries. Based on previous research, we predict that prosody will modulate the strength of perceived syntactic boundaries.

2) Whether gesture modulates the strength of perceived syntactic boundaries. To our knowledge, no previous research in perception has investigated the link between (non-beat) gestures and boundary perception. However, in a production study (Lelandais, 2020), we found that speakers used hand gesture parameters to express a boundary. We therefore predict that changes in the configuration of representational hand gestures can modulate perceived syntactic boundaries.

A third question is dependent on a positive outcome for question 1 or 2. 3) Which (prosodic and/or gestural) cues are involved in the modulation of perceived boundary strength?

Experiment 1 aims at testing naive participants' ability to assess boundary strength with syntactic and prosodic cues. In Experiment 1, 13 naive listeners had to measure boundary strength for 64 extracts on a 5-point scale. The stimuli all contained three tone-units, the second being a syntactic subordinate construction, which was established as a variable. The prosodic cues at the boundary between the tone-units were also established as variables, and were subject to manipulation (addition of a single cue associated with the perception of a prosodic boundary). The stimuli also contained filtered speech obliterating lexical and syntactic content while keeping syllabic structure and intonation, in order to distinguish prosodic from lexico-syntactic input.

Experiment 2 aims at testing naive participants' ability to assess boundary strength with syntactic and gestural cues only. In Experiment 2, 24 naive listeners had to measure boundary strength for 24 extracts on a 5-point scale. The stimuli all contained three tone-units, the second being a syntactic subordinate construction, which was established as a variable. The gestural cues produced in co-occurrence with the first two tone-units were also established as variables, and were subject to manipulation. Condition 1 features the production of one single hand gesture in overlap with tone-units (1) and (2). Condition 2 features two identical hand gestures, one produced in co-occurrence with tone-unit (1), the other produced in co-occurrence with tone-unit (2). Condition 3 features two different hand gestures in terms of trajectory and configuration, one produced in co-occurrence with tone-unit (1), the other produced in co-occurrence with tone-unit (2).

After describing the use of boundaries and perceived boundary strength in prosodic and gesture perception studies, we detail the designs of Experiment 1 and Experiment 2. We then give the respective results of our two experiments. We finally comment them altogether in a general discussion and conclusion.

2. Theoretical background

2.1 Prosody

2.1.1 The notion of boundary in prosodic studies

Numerous oral corpora include a segmentation of the data in various prosodic units, partly relying on an annotation of prosodic "boundaries" (*e.g.* Svartvik & Quirk, 1980; Auran

et al., 2005; Simon & Christodoulides, 2016). These units are either manually annotated by experts, or automatically detected based on a set of acoustic features. Although their size is variable depending on the phonetic and phonological theories used to segment the data (*e.g.* British School of intonation, see Crystal (1969); Wells (2006); Autosegmental Metrical framework, see Pierrehumbert (1980); Silverman *et al.* (1992); Beckman, Hirschberg, and Shattuck-Hufnagel (2005); IPO –Institute for Perception Research, see 't Hart, Collier, and Cohen (1990); Collier, de Pijper, and Sanderman (1993); de Pijper & Sanderman (1994)), these units are essential to the study of the relations between prosodic, syntactic, and discourse phenomena (Simon & Christodoulides, 2016).

Recent studies on the prosodic perception of subordinate constructions are scarce, but Fromont, Soto-Faraco, and Biau (2017) show that prosodic cues alone are not enough to modify the preferential interpretation of a clause. However, the latter article only targets the fact that prosodic boundaries can be optional in certain cases, since a given syntactic structure can have a variety of equally acceptable intonational phrasings (Watson & Gibson, 2005). In the same direction, Frazier, Clifton Jr, and Carlson (2004) as well as Pynte (2006) found that any syntactic edge is a potential location for a prosodic phrase boundary. Cole, Mo, and Baek (2010) and more recently Kuang, Pik Yu Chan, and Rhee (2022) found that syntactic phrasing had an independent effect on boundary perception for English listeners.

2.1.2 The use of prosodic boundaries in perception studies

In this study, boundary strength is apprehended as a perceptual notion which is directly measurable in a test with naive listeners. These naive listeners cannot explicitly refer to syntactic, phonological, or prosodic phenomena and structures. This was inspired by several studies on the perception of prosodic boundaries by naive listeners (de Pijper & Sanderman, 1994; Mo, Cole, and Lee, 2008; Cole, Mo, and Baek, 2010) evaluating the correlation between a number of prosodic cues and judgements about the presence of prosodic boundaries.

The term of "prosodic boundary" is mainly used to define a perceptual break (Cho & Hirst, 2006) produced by vocal means between two units. The perception of boundaries in spontaneous speech has been studied in Dutch (Streefkerk, Pols, and ten Bosch, 1997; Buhmann *et al.*, 2002), Swedish (Swerts, 1997), American English (Yoon *et al.*, 2004; Mo, Cole, and Lee, 2008; Mo & Cole, 2010; Cole, Mahrt, and Hualde, 2014), French (Smith, 2009; Astésano *et al.*, 2012; Roux *et al.*, 2016; Simon & Christodoulides, 2016), Kabylian and Hebrew (Mettouchi *et*

al., 2007), as well as in Korean (Cho & Hirst, 2006) and Mandarin (Yang & Wang, 2002; Li & Yang, 2009).

Although Ladd (2008, p. 288) defines boundaries as hard to identify and describe with consistency, these studies show altogether that listeners can efficiently and consistently perceive boundaries in spontaneous speech. Some articles have compared the ratings of naive listeners with those of expert annotators (Buhmann *et al.*, 2002; Amir, Silber-Varod, and Izre'el, 2004) working with the ToBI annotation system (Silverman *et al.*, 1992). Testing the robustness and consistency of annotations, these studies show a strong agreement rate between naive and expert annotators (Auran *et al.*, 2005; Pagel *et al.*, 1995; Simon & Christodoulides, 2016). The study led by Roy, Cole, and Mahrt (2017) also shows that the prosodic factors influencing boundary perception by naive listeners (*i.e.* silent pause duration and word phone rate duration) are the same factors as those identified in production for studies realised in experimental conditions.

Boundary perception tests realised by naive listeners are extremely informative on discourse segmentation and its interpretation. Although annotating boundaries as a task inevitably elicits metalinguistic judgements, using naive listeners avoids strong theoretical biases since subjects are not trained for the task. They also show spontaneous decision-making in a controlled context.

2.1.3 Prosodic cues linked to the perception of a break

The presence of a silent pause (*e.g.* Yoon, Cole, and Hasegawa-Johnson, 2007), final syllabic lengthening (*e.g.* Mo, Cole, and Lee, 2008), the presence of tones (*i.e.* falling or rising contours with a large amplitude; Portes, 2002; Smith, 2009), and initial pitch upstep (Wagner & Watson, 2010, pp. 907–910) have all been linked to the perception of a prosodic break.

One of the most useful features across languages in the identification of prosodic boundaries is the presence of a silent pause (Wightman *et al.*, 1992; Carlson & Swerts, 2003; Yoon, Cole, and Hasegawa-Johnson, 2007; Wagner & Watson, 2010; Roy, Cole, and Mahrt, 2017). However, many previous studies are based on read speech (*e.g.* de Pijper & Sanderman, 1994). The correlation between the presence of a silent pause and the perception of a prosodic boundary is less direct in spontaneous speech, since a silent pause can also be a mark of hesitation (Mertens & Simon, 2013). In map-task and television debates extracts, Smith (2009) shows that both silent and filled pauses lasting more than 150 milliseconds influence the perception of a prosodic boundary, but do not represent a viable indicator. In works on French

(Duez, 1985) and Korean (Cho & Hirst, 2006), silent pauses become a decisive cue when longer than 200 milliseconds. Swerts (1997) observes in spontaneous monologues in Swedish that longer pauses (more than 250 milliseconds) tend to be associated with the perception of stronger boundaries. Likewise, the influence of other prosodic cues increases when pause duration decreases (Lehiste, 1979).

Another rhythmic cue playing an important role in discourse segmentation is final syllabic lengthening. In conversational English, the stressed vowels positioned before boundaries are significantly longer than those positioned elsewhere (Kreiman, 1982; Mo, 2008). Final syllabic lengthening is strongly correlated with the perception of a boundary. It is even the most prevailing cue in a study on conversational Hebrew (Amir, Silber-Varod, and Izre'el, 2004). However, the authors signal that weighing the importance of this particular cue among others remains difficult since final syllabic lengthening usually appears in co-occurrence with silent pauses and / or initial pitch upsteps.

Melodic cues (*i.e.* F0 cues) also influence prosodic boundary perception. However, some studies examining the role of intonational contours indicate that the correlation between a falling contour vs. a rising contour and the perception of a boundary is unstable (Simon & Christodoulides, 2016). De Pijper & Sanderman (1994) notice that melodic discontinuity is the only vocal cue occurring in isolation to create boundaries. In French, Portes (2002) shows that prosodic boundaries perceived as weak are associated with rising contours while prosodic boundaries perceived as strong are associated with falling contours. Smith (2009) notes however that the amplitude of the F0 movement is a better correlate of boundary strength than the tone direction.

Finally, pitch upstep has also been analysed as a prosodic boundary cue (’t Hart, Collier, and Cohen, 1990), whose weight is subject to debate (Wagner & Watson, 2010). Similar to final syllabic lengthening, pitch upstep is usually produced as part of a cluster of other boundary cues.

Assessing the impact of a single prosodic boundary cue on perception is often reported as a difficult task, since prosodic boundary cues are usually used in combination. Stronger boundaries will be perceived in combinations of several cues (*e.g.* silent pauses combined with melodic discontinuity; de Pijper & Sanderman, 1994). A positive correlation has additionally been found between the number of prosodic boundary cues on a specific syntactic / prosodic constituent and its syntactic / phonological weight (Blaauw, 1994).

Although intensity, glottalization, and the articulatory configuration of the vocal tract also play a great role in the perception of boundaries (Mo, 2008; Barth-Weingarten, 2016), these cues are not assessed in the present paper. Disfluencies are also excluded from the stimuli and the analysis, since our stimuli are extremely short and we include many variables.

2.1.4 Prosodic boundary strength

One of the questions motivating our perception test concerns the number of boundary degrees that can consistently be perceived by naive listeners. The number of categories is generally low in experimental scales (de Pijper & Sanderman, 1994), as Grover *et al.* (1998) and Auran *et al.* (2005) show that a 4-degree scale is enough to transcribe boundary strength consistently.

The number of degrees generally depends on the theoretical principles held by researchers concerning prosodic boundaries. The influence of the Autosegmental Metrical model of phonology led to the elaboration of a system for transcribing and annotating prosody, ToBI (Tone and Break Indices; Pierrehumbert, 1980), which integrates a tonal annotation system along with an annotation system of boundaries on a 5-point scale (Silverman *et al.*, 1992).

Other studies choose to identify consensus boundaries at positions where a certain proportion of participants have identified a boundary (Auran *et al.*, 2005; Smith, 2009). Boundary strength can also be calculated as the proportion of subjects having indicated a boundary at any given position, and expressed as a value between 0 and 1 (Cole, Mo, and Baek, 2010; Simon & Christodoulides, 2016).

2.2 Gesture

2.2.1 Boundaries in gesture studies

In gesture studies and linguistics, the notion of boundary was first used as part of the investigation of the link between gesture and speech units, which was conducted with data coming from production, essentially in psycholinguistics and cognitive sciences (*e.g.* McNeill & Duncan, 2000). It is known that co-speech gestures are semantically (*e.g.* McNeill & Duncan, 2000; McNeill, 2005), temporally (*e.g.* Chui, 2005), and structurally (*e.g.* Kita & Özyürek, 2003; Lewandoswki & Özçalışkan, 2018) linked to speech. This coordination takes place on many levels (Bernardis & Gentilucci, 2006), especially at the prosodic (Mendoza-Denton & Jannedy, 2011) and syntactic levels (Fritz *et al.*, 2019).

It has for instance been shown that gesture apexes are temporally aligned with pitch accents (Kendon, 1983; Loehr, 2004; Mendoza-Denton & Jannedy, 2011), and that final articulatory lengthening at the end of prosodic units does not only affect oral gestures, but manual gestures as well (in that both the final syllable of an intonation unit and the co-occurring gesture are lengthened; Esteve-Gibert & Prieto, 2013; Wagner, Malisz, and Kopp, 2014; Krivokapić, 2014).

Depending on the variables under study and on the task imposed to participants, the performance of participants in boundary perception studies on gesture can be assessed with answers to a questionnaire (Granström, House, and Lundeberg, 1999; House, Beskow, and Granström, 2001), reaction time to stimuli (Barkhuysen, Krahmer, and Swerts, 2008; Kelly, Özyürek, and Maris, 2010), oculometry (Oben & Brône, 2015), evoked potentials (Dimitrova *et al.*, 2016), EEG (*i.e.* electroencephalography; (Meyer *et al.*, 2012), or with fMRI (Biau *et al.*, 2016).

Other experimental work on the synchronization between prosodic cues and gestural cues in the perception of speech use modelling to comparative ends with extracts from spontaneous interaction (*e.g.* Schlangen, 2006; Atterer, Baumann, and Schlangen, 2008; Barkhuysen, Krahmer, and Swerts, 2008). These studies work at implementing models on virtual conversational agents (*e.g.* Cassell *et al.*, 2001). In these studies, the acceptability of prosodic and gestural cues can be judged by naive participants, either with a questionnaire (Barkhuysen, Krahmer, and Swerts, 2008) or by their ability to interact with virtual agents (Granström & House, 2005; De Kok & Heylen, 2009). This stream of work is more particularly interested in multimodal cues for backchannels.

The use of gestural cues in the expression of speech boundaries is also supported by studies on co-speakers' attention to gesture in interaction. Many studies show that gestural cues are essential in the construction and management of common ground in interaction (Parrill & Kimbara, 2006; Holler & Bavelas, 2017), and that speakers and co-speakers alike rely on these gestural cues (Hoetjes, *et al.*, 2015; Oben & Brône, 2016). It has for instance been shown that speakers adapt their gestures to co-speakers, relying on signals communicated by co-speakers (Debreslioska *et al.*, 2013; Perniss & Özyürek, 2015; Masson-Carro, Goudbeek, and Krahmer, 2016).

2.2.2 Gestural cues linked to the perception of a break

This study focuses on the capacity of hand gestures to mark a boundary in speech, based on the fact that little research has been carried out on representational hand gestures (*i.e.* gestures describing or representing objects, actions, or ideas) and on their forms and trajectories in the context of speech segmentation, as opposed to gaze, head, and eyebrow movement (Argyle & Cook, 1976; Granström & House, 2005; De Kok & Heylen, 2009; Rienks, Poppe, & Heylen, 2010). In perception studies, these variables are easy to manipulate in the case of modelling and implementation on virtual conversational agents. However, our auditory stimuli are issued from conversational speech and were temporally synchronised with gestures from other participants. Experiment 2 is the follow-up of a production study highlighting the role of hand gestures in the production of boundaries in speech (Lelandais, 2020).

Very few perception studies have addressed the segmentation value of hand gestures. Some perception experiments were conducted on beat gestures (which typically feature a downward movement on a vertical axis) and have found that they help syntactic segmentation and focalization (Holle *et al.*, 2012; Dimitrova *et al.*, 2016; Biau, Fromont, and Soto-Faraco, 2018). These studies clearly highlight the segmenting ability of hand beats. However, these experiments feature read speech or extracts from political speech. Data is virtually inexistent on conversational speech, as well as on other types of hand gestures, for instance those that do not show such a strong prosodic value. To our knowledge, no study focuses on discourse segmentation with representational gestures.

The rest of the knowledge concerning hand gestures and their link with speech segmentation comes from production studies, which have mentioned that two successive hand gestures that are different in terms of configuration and trajectory can create a boundary in discourse (Streeck, 2009; Enfield, 2009; Calbris, 2011). This is explained by the fact that some gesture features participate in the creation of cohesion in discourse (Hoetjes *et al.*, 2015; Perniss & Özyürek, 2015). In this view, the repetition of a same hand gesture can convey continuity, in the same manner that two different speech segments can be linked if they are produced in co-occurrence with one same hand gesture. McNeill (*e.g.* 2005) describes this phenomenon in terms of "catchment", *i.e.* a group of hand gestures presenting a same formal recurrence in their configuration.

The results of these studies show that speakers also signal boundaries in speech by gestural means. However, given the weak number of gesture perception studies, only a small number of these gestural boundary cues have been shown to manifestly be interpreted as such by co-speakers. Questions remain unanswered concerning the weight of gestural cues compared

to that of prosodic cues. While participants have been shown to be able to perceive boundaries with auditory cues, their capacity to do so with gestural cues has to be further documented. Barkhuysen, Krahmer, and Swerts (2008) have for instance shown that boundaries are more easily identified with audiovisual cues or audio cues only than with visual cues only. Studies have focused either on the influence of gestural cues with that of prosodic cues in the expression of prominence (*e.g.* Hadar *et al.*, 1983; Krahmer & Swerts, 2007; Swerts & Krahmer, 2008), or in the interpretation of statements or questions (*e.g.* Borràs-Comes & Prieto, 2011; Cruz, Swerts, and Frota, 2017). These papers have shown that participants perceive linguistic functions through the combination of auditory and gestural cues and that they benefit from this combination, but that participants rely more on auditory cues than on gestural cues.

2.2.3 Gestural boundary strength

While prosodic perception studies widely use gradual scales measuring the strength of perceived boundaries (*e.g.* Grover *et al.*, 1998; Auran *et al.*, 2005), gesture perception studies do not use these scales. The notion of boundary strength only appears in studies mixing prosodic and gestural boundary cues, such as that by Granström, House, and Lundeberg (1999), and is used indirectly in House, Beskow, and Granström (2001) for the multimodal expression of prominence, measured as the proportion of subjects indicating that they have perceived a boundary at a specific location.

However, in prosodic perception studies, boundary strength is directly questioned with specific point scales that participants can tick at the location they perceived a boundary. These point scales are adapted from the ToBI system of prosodic annotation (Silverman *et al.*, 1992; Brugos, Shattuck-Hufnagel, and Veilleux, 2006), and are relatively precise. The degrees of strength are adapted to the naive participants as they do not refer to any verbal, prosodic, or gesture unit or structure. The levels usually include no boundary, weak boundary, uncertain, boundary, and strong boundary.

3. Corpus and methods

3.1 Design of Experiment 1

3.1.1 Stimuli preparation and manipulation

The audio stimuli were extracted from the ENVID corpus (Lelandais & Ferré, 2016), a 2-hour video collection of spontaneous conversational British English (8 women, 2 men; mean age = 22 years old). This collaborative corpus gathers video recordings realized in soundproof

studios between 2000 and 2012, making up a total of 5 dialogues (2 hours and 10 minutes). Each participant had a lavalier microphone, which provided two separate audio tracks. Two audio files corresponding to each microphone were created in a .wav format, so as to facilitate the analysis of overlapping speech.

The ENVID corpus is annotated with syntactic, prosodic, and gestural information. It was transcribed in Praat (Boersma & Weenink, 2021) using a standard orthographic transcription of tone-units, in which subordinate constructions were localised and coded on a separate track. A total of 303 constructions were annotated in the corpus, which represents 10.09% of the total speaking time (*i.e.* 2.68 form/min): 83 restrictive relative clauses (1.88% of speaking time), 161 adverbial clauses (3.46% of speaking time), and 59 appositive relative clauses (1.23% of speaking time).

They represent the three most widespread types of finite clauses functioning as syntactic modifiers[4] (*e.g.* Huddleston & Pullum, 2006) in the ENVID corpus. In example (1), the adverbial clause "when he was first starting teaching" restricts the temporal frame in which the referential elements must be understood. This paper focuses on adverbial clauses introduced by "when". The results and conclusions given in this study for adverbial clauses only concern such clauses.

(1) **Adverbial clause (**Transcription conventions are provided before the Introduction section of the paper)

Zoe	(1)	my dad used to teach in Hebburn
	(2)	**when he was first starting teaching**
	(3)	and he was getting harassed by all the pupils (laughs)

While an adverbial clause modifies another clause, a restrictive relative clause modifies a nominal expression. In (2), the restrictive relative clause "that were there" increases the relevance of "the Spanish girls", creating a subcategory for this referent.

[4] In syntactic studies, "modifiers" refer to elements specifying or elaborating upon some primary features (Halliday, 1985), often described as additions to propositional content in the host or embedding structure (Huddleston & Pullum, 2006). Modifiers are not inherently presupposed by their head. They supplement the head with additional information.

(2) **Restrictive relative clause**

Joey	(1)	the Spanish girls
	(2)	**that were there #**
	(3)	on our second one

This paper focuses on restrictive relative clauses introduced by "Ø"[5] and "that" as relative pronouns. This construction allows speakers to provide the co-speaker with more complex information about the antecedent than in non-relative structures, without the co-speaker having trouble processing it.

Lastly, appositive relative clauses are not invoked to single out a nominal referent, but to make an additional comment about it. In (3), the appositive relative clause "which was horrible" comments upon "a place called Tropicana".

(3) **Appositive relative clause**

Beth	(1)	and then we went into # a place called Tropicana #
	(2)	**which was horrible (laughs) #**
	(3)	it's on Saint Mary street

This study focuses on appositive relative clauses introduced by "which" as a conjunction.

The selection targeted occurrences without an interruption, surrounded with immediate left and right co-texts other than a single silent pause yielding the speaking turn. The selected occurrences were classified according to their syntactic type in Praat (restrictive relative clause, adverbial clause, appositive relative clause).

In order to establish reliability of the clause type classification (restrictive relative clause, adverbial clause, appositive relative clause), a second coder judged 20% of the data that had been classified by the original coder. The second coder is also a specialist of the field. The agreement between coders was 100%.

The corpus is segmented into tone-units, according to the British school of intonation (Crystal, 1969; Wells, 2006) based on dynamic pitch contours. The Momel-Intsint algorithm (Hirst, 2007; Bigi, 2012) was used for the automatic annotation of the F0 target points in the signal (Figure 1). Annotations are made in two respects: the algorithm notes pitch height (in Hz) on target syllables, which allows us to calculate mean F0 values for specific segments. The

[5] "Ø" is the zero relative pronoun, marking its ellipsis from a clause.

algorithm also codes symbolic (relative) values of intonation, in which each measured F0 value is compared to preceding ones, *i.e.* significant changes in the F0 curve either regarding the speaker's pitch range (Top, Bottom) or regarding the neighbouring tones or sequences of tones (Upstep, Downstep, Same, Low, High).

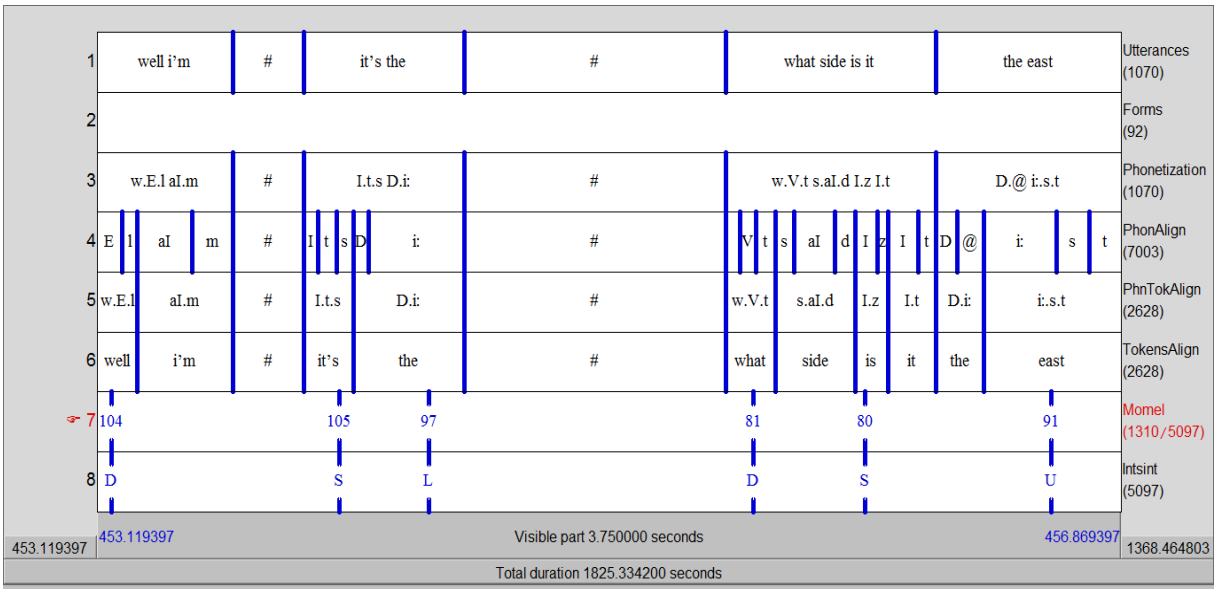

Figure 1. Prosodic transcription of the ENVID corpus in Praat. Tiers 7 and 8 show the automatic annotations performed with Momel-Intsint.

Within each segment of the sequences under study, the nature of each nuclear contour (fall; fall-rise; rise; rise-fall; flat) was also coded manually. Pitch key was annotated in regards to each speaker's specific range (high; mid; low) on both the whole segments (L, SC, R) and the boundary (initial and final) syllables in these segments. In order to establish reliability of the nuclear contour classification, a second coder judged 20% of the data that had been classified by the original coder. The second coder is also a specialist of the field. The agreement between coders was 81.9%.

64 audio stimuli were used (mean duration = 4 seconds), all of them containing three tone-units, the second being a subordinate construction, which was established as a variable (appositive relative clauses, adverbial clauses, restrictive relative clauses). Each extract can then be described as a (1)-(2)-(3) sequence. Example (4) shows a stimulus containing an adverbial clause. It can be heard as Audio file 1 in the Extra material section.

(4) Stimulus containing an adverbial clause

(1) especially in class

(2) **when someone says something funny**

(3) and i'm always dying

These extracts did not feature any prosodic cue associated with a boundary (*e.g.* silent or filled pause, final syllabic lengthening, falling tone) between tone-units (1) and (2), and this was validated by a pre-test on 6 native speakers of British English, who were also naive to the experiment and to prosodic studies[6]. Each extract has been played twice by the experimenter on a computer. Listeners had to draw a vertical line on the orthographic transcription of each extract (without any punctuation mark or jump in line) if they perceived any break of some kind during the extract. A boundary was broadly defined as "anything that acts as a separator between parts of speech". No boundary has been perceived by any participant (agreement rate = 100%).

The prosodic cues at the boundary between the first and second tone-units were established as variables and were subject to manipulation in Praat (Boersma & Weenink, 2021) by one of the authors. Manipulations included adding 1) a silent pause between (1) and (2); 2) final syllable lengthening at the end of (1); 3) a falling tone at the end of (1); 4) a rising tone at the end of (1); a pitch upstep at the beginning of (2). A table detailing the total number of items for each manipulation is available in the Extra material section (Table A). The same items appear in their filtered and non-filtered versions. The acoustic manipulations on prosodic target sites were done on the same items. The acoustic manipulations were implemented on each syntactic category (3 types of clauses). Given the short duration of an average tone-unit, tone-unit (3) provides meaningful context for listeners to avoid any effect resulting from a gating paradigm. A detailed list of each item is available in the Extra material section, along with the spectrographic representations of both authentic and manipulated items.

Silent pauses were added between tone-units (1) and (2) with the Praat sound editor. Since each stimulus was extracted from a 30-minute dialogue, a 500-millisecond pause extracted from each corresponding interaction (without any speech or noise) was inserted between the two tone-units in each extract. The 500-millisecond threshold was chosen based on the observation of extra-constituent silent pauses in our corpus (mean duration = 0.56 seconds; median = 0.55 seconds). We wanted the inserted silent pauses to be both representative of our corpus and unambiguous. We also followed previous studies showing that silent pauses become a decisive cue when they reach a longer duration than 200 milliseconds. Given the short

[6] As a reviewer noted, example (4) exhibits perceptible final lengthening on the target word 'class' to expert ears. The literature on prosodic boundary perception has attested the difficulty for naive listeners to perceive lower levels of prosodic boundaries, but they still might process the information. Hence, the authentic, non-manipulated stimuli used in Experiment 1 possibly all exhibit prosodic boundary cues, albeit weak (given the speakers' fast articulatory rate). This has an impact on the manipulations dedicated to artificially create prosodic boundary cues. We still chose to proceed with these authentic stimuli, since they passed the pre-test with naive listeners.

duration of our stimuli, we also made sure that silent pauses were still shorter than the duration of the whole speech extract.

Final syllable lengthening was realized with the manipulation function in Praat. We relied on Duez (1985) showing that a 50% lengthening of the syllable is perceptible as a boundary by listeners. The duration of each final stressed syllable of our original stimuli was extracted and lengthened by 50%, with addition of duration points in Praat. An example of a stimulus containing a lengthened final syllable can be heard as Audio file 2 in the Extra material section.

The addition of falling and rising tones was made with the pitch stylization function in Praat. F0 trajectory is modified on a target syllable. The realized excursion starts at the onset of the syllable in question, and finishes after its coda. The excursion range was defined in semi-tones based on the studies made by ’t Hart (1981), Rietveld & Gussenhoven (1985), and Kakouros & Räsänen (2016b). Rietveld & Gussenhoven (1985) show that a variation becomes perceptible from 1.5 semitone. We chose a variable excursion range between 1.5 and 4 semitones depending on each stimulus, in order to obtain a perceptible difference with the pitch movement remaining natural. Adjustments in pitch height were made on the whole length of extracts when the change in tone affected other prosodic parameters (*e.g.* reset of the melodic curve, octave jump). The mean F0 height in each extract was measured before and after manipulation, in order to limit any global perceptive change on the extract curve. An example of a stimulus containing an added falling tone can be heard as Audio file 3 in the Extra material section, while Audio extract 4 is a stimulus containing an added rising tone.

Initial pitch upstep was made on (2)'s initial syllable. The difference in height between (1)'s final syllable and (2)'s initial syllable was also defined in semitones. It varies between 4 and 6 semitones depending on each extract, since the minimum threshold of 4 semitones gave the clearest and most natural difference (Kakouros & Räsänen, 2016a). Similar to the modifications on pitch movement, adjustments in F0 height were made when the pitch upstep affected other prosodic parameters (*e.g.* octave jump). An example of a stimulus containing an initial pitch upstep can be heard as Audio file 5 in the Extra material section.

Every stimulus was also resynthesized in another set to obliterate lexical and syntactic content while keeping syllabic structure and intonation (32 filtered stimuli + 32 unfiltered stimuli). This allowed us to make lexico-syntactic information inaccessible to participants, in

order to study the influence of each modality on boundary perception. We used the Pass Hann band in Praat (Figure 2), with a frequency from 0 to 450/500 hertz (for speech produced by men and women, respectively). The obtained sound signal being very low in amplitude, each extract was then amplified (+ 5 dB) in Audacity.

We considered filtering to be the most appropriate method for speech delexicalization, since Experiment 1 focuses on the contribution of a whole set of prosodic cues, and that this method does not dissociate rhythm from intonation. When listening to filtered speech, participants have access to melodic (*e.g.* nuclear tones) and temporal (*e.g.* pauses, speech rate, duration) cues. However, the influence of some suprasegmental cues is neutralized and it is impossible to identify segments. Since we do not manipulate or control intensity in this study, some phonotactic cues can nonetheless show through the segments' relative intensity. We thus chose to vary the filtered frequency in function of each extract, making sure that phonotactic information remains unavailable. The filtered stimuli were checked for available semantic content with an external listener, expert to the field of prosody. An example of a filtered stimulus can be heard as Audio file 6 in the Extra material section.

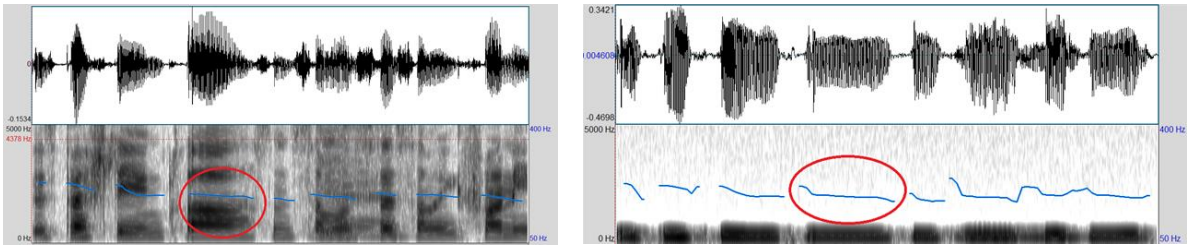

Figure 2. Manipulated final lengthening at the end of (1) on unfiltered (left) and filtered (right) stimuli.

10 distractors have been added to the set. The selection criteria for distractors were the same as those for the authentic stimuli (speech coming from one speaker, clear signal without any noise), except that the prosodic boundary cues are not controlled.

3.1.2 Procedure

Our 64 stimuli (54 stimuli + 10 distractors) were randomized and presented sequentially to 24 British participants (aged 19 to 45, mean = 24 years old) via a specifically designed web interface, eSurv (2017), permitting to run perception experiments through the internet using a standard web browser. The test lasted approximately 40 minutes.

Most of the participants were recruited online, via mailing lists and social media, while one third of them were recruited by one of the authors in person in Bristol, UK. For the latter third, the test was realised in presence of the experimenter. Preliminary questions on the web interface secured the fact that all participants were native speakers of English, had no hearing or visual deficiency, had no experience in prosodic / gesture annotation, and had headphones plugged in. 11 participants (all recruited online) were excluded from the analysis, for not fulfilling one or more of these criteria[7] (total number of participants after exclusion: 13).

Participants were first presented with a short description of the study, then had to answer the preliminary questions. They were then presented with the stimuli. Each sound extract featured its orthographic transcription without any punctation mark (see **Error! Reference source not found.** below), a star indicating the location of the boundary to be identified. The delexicalized stimuli read "X*X". A boundary was broadly defined as "anything that acts as a separator between parts of speech". Participants had to tick one of the five boxes of the point scale to rate boundary strength, as the instructions indicated ("Please rate the presence or absence of a boundary at *"). In line with the boundary detection systems developed in Silverman *et al.* (1992) and in Carlson & Swerts (2003) using detailed point scales, we included an "uncertain" category, since our paradigm did not contain any forced-choice task in limited time and focused on the strength of boundaries. Participants could play the sound file any number of times they wanted. Figure 3 shows the presentation of the first stimulus on the test interface.

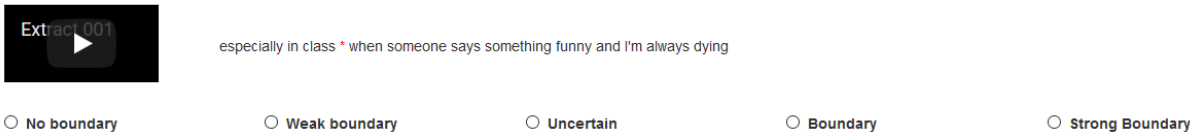

Figure 3. Test interface on eSurv with an (unfiltered) sound extract and its transcription followed by the 5-point scale.

3.1.3 Data analysis

[7] Two cases were observed in these exclusions. First, a number of online participants indicated they were a native speaker of a different language from English, and proceeded with the experiment. Second, some online participants consistently ticked the same box (*i.e.* perceptual category) throughout the entire set of items in less than a minute, in the sole purpose of completing the experiment.

In Experiment 1, our predictions are that (1) perceived prosodic boundary strength varies in function of syntactic type; (2) stronger boundaries are perceived in stimuli containing appositive clauses.

Because of the ordinal nature of the dependent variable (ratings ranging from 0 to 5), we performed a series of Bayesian Cumulative Link Models (CLM) using the R 4.2.1 statistical programming language (R Core Team, 2022) and the *brms* 2.17.0 package (Carpenter *et al.*, 2017; Bürkner, 2021) on the data to account for the degree of influence of each separate variable on perceived boundary strength. All data frames, model scripts, and posterior distributions are available in the Extra material section. Participant and Item were coded as random factors. Type was coded as a three-level, centred predictor (appositive; adverbial; relative), with "relative" as the reference level. All binary predictors were mean-centred for a more accurate interpretation of the coefficient yielded by the models (Brehm & Alday, 2020).

We used Cumulative Link Models within a Bayesian analysis framework. Bayesian modelling offers multiple advantages as opposed to more classical Frequentist methods. They are for instance more appropriate when dealing with small datasets, because models show less convergence failures, even with a maximal random effects structure (Barr *et al.*, 2013). For more information on the benefits of using a Bayesian framework in cognitive and related sciences, see Sorensen, Hohenstein, & Vasishth (2016). Based on the observed data and a set of priors (more or less precise expectations on what the model should look like), a Bayesian model yields posterior distributions of possible values for each of the parameters that are estimated. Because of the exploratory nature of the experiments and because of the relatively low number of observations, we chose not to specify the priors beyond the default, non-informative priors used by brms; this means the yielded posterior distributions stay quite close to the observed data. The posterior distributions correspond to the interval and point estimates of the response (perceived boundary strength) for each variable when the level of the reference variable is set to zero. A mean value of 0 for these estimated parameters (^ß) means there is no effect of the manipulated variable (IV) on the variable of interest (DV). This estimated mean value comes along with credible intervals (Crl), here of 95%, which represent the values between which there is a 0.95 probability of finding the real value of the estimated parameter (Morey *et al.*, 2016; Winter, Duffy, and Littlemore, 2020). We also report the probability of the estimated parameter being superior or inferior to 0 ($P(ß) > 0$ or $P(ß) < 0$), as a further indication of the strength with which the data support the existence of an effect. Drawing on Engelmann *et al.* (2019) and Pozniak & Burnett (2021), we do not interpret the results as either significant

or not, since evidence for an effect of a given parameter varies in strength depending on multiple parameters (*e.g.* settings related to the zero reference level, and credible intervals). However, we chose the following interpretation thresholds:

- if the probability of the effect of the parameter to be different from 0 is $\geq .9$, the effect on the dependent variable will be considered reliable;
- if the probability of the effect of the parameter to be different from 0 is $\geq .75$, the effect on the dependent variable will still be considered meaningful, but not as robust;
- if the probability of the effect of the parameter to be different from 0 is $< .75$, the effect on the dependent variable will not be considered reliable.

Note that these thresholds were chosen in consideration with the small size of our dataset. They help qualify the reliability of posterior distributions to give robust information about general tendencies in our population. In this sense, they do not directly measure the magnitude of the effects.

All statistical analyses were conducted using the *tidyverse* 1.3.1 (Wickham *et al.*, 2019), *dplyr* 1.0.9 (Wickham *et al.*, 2022), *shinystan* 2.6.0 (Stan Development Team, 2017), and *sjPlot* 2.8.10 (Lüdecke, 2021) packages for data processing and visualization. We used the *ggplot2* 3.3.6 (Wickham, 2016), *ggridges* 0.5.3 (Wilke, 2021), and *ggstance* 0.3.5 (Henry, 2020) packages to plot density ridges for posterior distributions and build violin plots for the response variable. The seed of each model was chosen arbitrarily, in function of the time at which the model was first run.

3.2 Design of Experiment 2

Experiment 1 having investigated the impact of several prosodic cues on boundary perception in speech containing subordinate clauses, Experiment 2 gathers audiovisual stimuli which do not include any prosodic boundary cue, and in which only hand gestures are manipulated. It explores the impact of boundary cues only expressed through gestures, in particular those realized with the hand.

3.2.1 Stimuli preparation and manipulation

The speech stimuli were extracted from the ENVID corpus. Six audio extracts were selected (4 from female speakers, 2 from male speakers), with an average duration of 4 seconds. Five of these extracts are the same as those used in Experiment 1 (authentic extracts), while one

of these extracts does not belong to the set previously used (we chose a longer extract in order to implement gestures). Each stimulus contained three tone-units: one corresponding to the subordinate construction (2 appositive relative clauses, 2 adverbial clauses, 2 restrictive relative clauses), one corresponding to its left co-text (*i.e.* the tone-unit immediately preceding the subordinate construction), one corresponding to its right co-text (*i.e.* the tone-unit immediately following the subordinate construction). Each extract could therefore be described as a (1)-(2)-(3) sequence. These extracts did not feature any prosodic cue associated with a boundary, and this was validated by a pre-test on 6 native speakers of British English, who were also naive to the experiment and to prosodic studies (see procedure in section 3.1.1). No boundary has been perceived by any participant (agreement rate = 100%).

We only manipulated the visual aspect of the stimuli, *i.e.* gestures. The variables are their alignment with speech and their configuration. Three different conditions were included. Condition 1 includes 6 extracts with the realization of one hand gesture in overlap between the end of the first tone-unit and the second tone-unit, supposedly conveying continuity in discourse (Enfield, 2009). Condition 2 represents 6 extracts with the realization of two identical hand gestures in terms of configuration, one produced in co-occurrence with a lexical item in the first tone-unit, and the other produced in co-occurrence with a lexical item in the second tone-unit (supposedly participating in cohesion through form, space, and direction; McNeill & Levy, 1993; Lascarides & Stone, 2009). The two identical hand gestures are separated with a return to rest position. Finally, Condition 3 is made of 6 extracts with the realization of two different gestures in terms of configuration with one produced in co-occurrence with a lexical item in the first tone-unit, and the other in co-occurrence with a lexical item in the second tone-unit (supposedly creating a boundary through form, space, and direction; Calbris, 2011). These two different hand gestures are separated with a rest position as well.

To design the visual aspect of our stimuli, two persons were filmed (one woman and one man depending on the audio extract) producing one or two gestures depending on the condition while listening to the audio part of the stimuli. These persons are not actors. One of them is the author of the present article while the other is naive to language sciences. They were following instructions regarding the timing and configuration of hand gestures.

The hand gestures produced by these persons belong to the categories of iconics (*i.e.* gestures depicting concrete entities or actions; McNeill, 2005), metaphorics (*i.e.* gestures depicting abstract entities through a metaphoric use of form and space; McNeill, 2005) and pointing gestures (*i.e.* deictic gestures) with a representational value. The gestures share a

semantic link with the lexical affiliate in each stimulus. For the stimuli containing two repeated gestures, we preferred grammatical gestures (for instance marking tense or any kind of modality), since the repetition of a same iconic gesture without the repetition of the lexical affiliate would have given an incongruent result. Each gesture is temporally synchronised with speech, in that the apex of the gesture is temporally aligned with the prominent syllable (specifically, its vowel onset) of the lexical affiliate (Couper-Kuhlen, 1999). This alignment was controlled with a visual representation of prominent syllables in Praat. All gestures show the same duration, amplitude and tension depending on each condition's requirements. A detailed list of each item with a justification for each chosen gesture form is available in the Extra material section.

An example of the stimuli can be seen in Video 1 in the Extra material section, and in Figure 4. Figure 4 is associated with example (5), which illustrates our choices concerning hand gesture types regarding the meaning of each speech extract. In the stimulus represented by example (5) and Figure 4, the speaker is describing the final exam in his architecture course. The lexical item in affiliation with the gesture produced is "work". The gesturing person draws a rectangle on a flat surface, as a sheet of paper on a desk. This stimulus belongs to Condition 1, in which one single gesture is produced in overlap between the first two tone-units.

(5) **Appositive relative clause in a stimulus belonging to Condition 1**

Tim	(1)	you get an assessed [(a) piece of (b) work
	(2)	**which you do on a computer (c)]**
	(3)	using a program called author catway

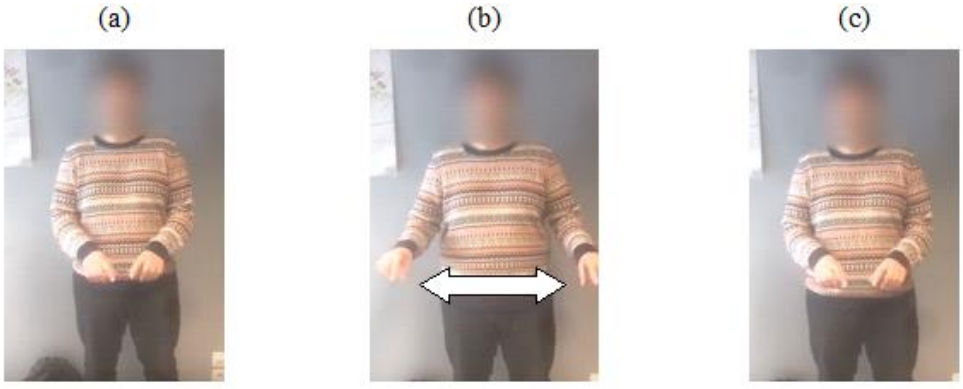

Figure 4. Several moments of the iconic hand gesture produced in Video 1 and example (5). The gesture preparation phase can be seen in image (a), while the realization and retraction phases can respectively be seen in images (b) and (c).

A fixation cross was added to the stimuli two seconds before the beginning of each video in order to focus the participants' attention. The gesturing person's face was blurred, so as to

filter the available visual cues (head and eyebrow movement are made invisible) and to focus the participants' attention on the hands. This also facilitates the interpretation of speech as coming from the gesturing person in the video. The hands are in rest position at the beginning of each video, and return to rest position afterwards. The rest position is the same across the stimuli, as shown in Figure 5. Examples of stimuli belonging to Condition 2 and to Condition 3 can respectively be found as Video 2 and Video 3 in the Extra material section.

Figure 5. Rest position of the gesturing person in videos containing sound extracts produced by a female speaker (a), and in videos containing sound extracts produced by a male speaker (b). These rest positions are adopted before and after the realization of every hand gesture.

Every final version of the stimuli was controlled by 2 persons external to the study and to linguistics in general, to make sure that the audiovisual stimuli were not incongruent. 100% of sequences were validated as such.

3.2.2 Procedure

Our 24 stimuli (18 stimuli + 6 distractors) were randomized and presented sequentially to 25 British participants (aged 19 to 40, mean = 24) via a specifically designed web interface eSurv (2017), permitting to run perception experiments through the internet using a standard web browser. The test lasted approximately 20 minutes.

The recruitment procedure was the same as that of Experiment 1. Most of the participants were recruited online, via mailing lists and social media, while one third of them were recruited by one of the authors in person in Bristol, UK. For the latter third, the test was realised in presence of the experimenter. This latter third participated in both Experiment 1 and Experiment 2. Preliminary questions on the web interface secured the fact that all participants were native speakers of English, had no hearing or visual deficiency, had no experience in

prosodic / gesture annotation, and had headphones plugged in. 1 participant (recruited online) was excluded from the analysis, for not fulfilling one of these criteria (total number of participants after exclusion: 24).

Participants were first presented with a short description of the study, then had to answer the preliminary questions. They were then presented with the stimuli, each extract featuring its orthographic transcription. The transcript did not feature any punctuation, but featured a star to indicate the location of the potential boundary to be identified. A boundary was broadly defined as "anything that acts as a separator between parts of speech". Participants had to tick one of the five boxes of the point scale to rate boundary strength. They could play the video file any number of times they wanted. Figure *6* shows the test interface.

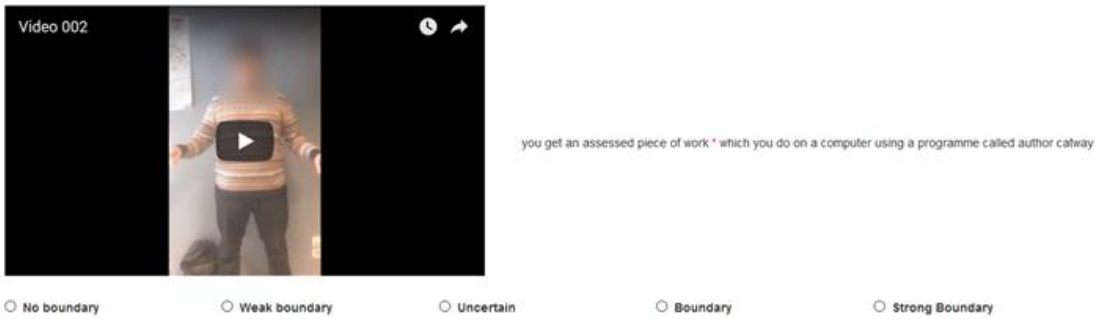

Figure 6. Test interface on eSurv with a video extract and its orthographic transcription, followed by the 5-point scale.

3.2.3 Data analysis

Our hypotheses for Experiment 2 are as follows: 1) speech boundaries between clauses can be visually perceived by naive participants; 2) stronger boundaries are perceived in stimuli containing two different hand gestures in terms of form, trajectory, and space; 3) stronger boundaries are perceived in stimuli containing appositive clauses.

We used Bayesian Cumulative Link Models (CLM) using the R 4.2.1 statistical programming language (R Core Team, 2022) and the *brms* 2.17.0 package (Carpenter *et al.*, 2017; Bürkner, 2021) to account for the degree of influence of each separate variable on perceived boundary strength. All data frames, model scripts, and posterior distributions are available in the Extra material section. Participant and Item were coded as random factors. Type was coded as a three-level, centred predictor (appositive; adverbial; relative), with "relative" as the reference level. All binary predictors were mean-centred. All statistical analyses were conducted using the same packages as in Experiment 1.

4. Results and discussion

4.1 Results of Experiment 1

We noted high inter-rater variability for Experiment 1. A graph showing each participant's mean ratings by syntactic type is available in the Extra material section. Despite this high variability, we used the 95% credible interval to interpret the posterior distributions of each model because of the limited number of participants. The results are given in different sections depending on each tested effect. All data frames with posterior distributions are available in the Extra material section.

4.1.1 Filter

We first explored the potential effect of filter (Pass Hann Band) as a single predictor (fixed factor = Filter; values = yes (1); no (0)) on perceived boundary strength (fixed factor = Ratings; values = strong boundary (5), boundary (4), weak boundary (3), uncertain (2), no boundary (1)). The first model, mb1, was fitted as follows: mb1 <- brm(Rating ~ filter2 + (filter2|Participant) + (1|Item), data = data, family = cumulative, chains = 4, iter = 3000). The list of every fitted model is available in the Extra material section.

Figure 7 shows a plot of the posterior distributions for the filter variable. As a reminder, the posterior distributions correspond to the interval and point estimates of the response (perceived boundary strength) for each variable when the level of the reference variable is set to zero. Zero is marked with a vertical dashed line. Quantile lines correspond to the 95% credible intervals and the mean. The 95% central part of each density ridge is filled with blue, while the tails are filled with pink. The density ridge, corresponding to the filter variable is mostly on the right of the zero vertical dashed line, with a mean estimated coefficient of 1.89.

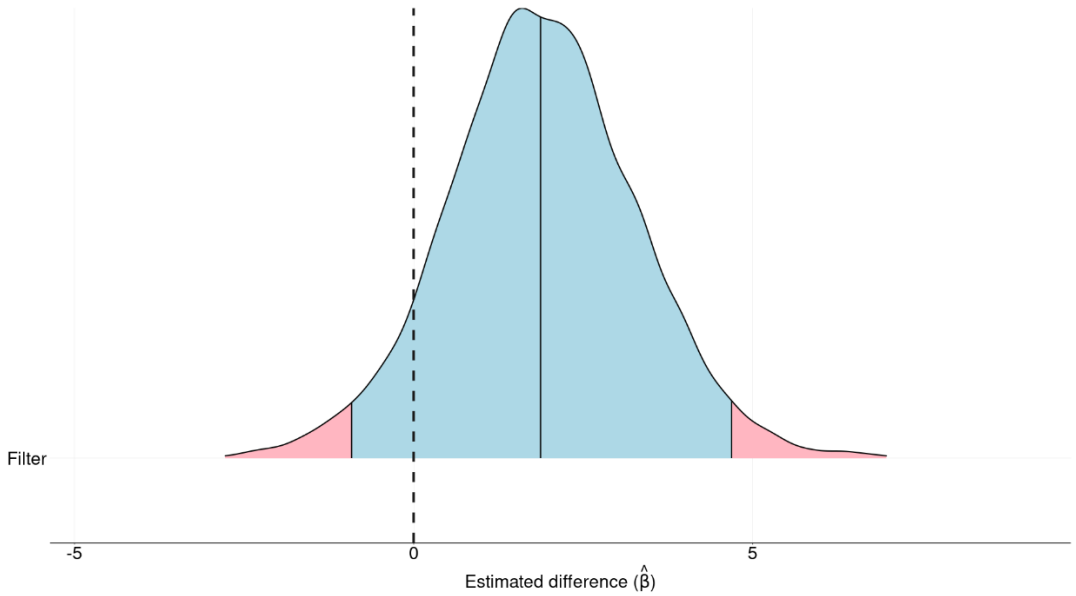

Figure 7. Posterior distribution of the filter variable in mb1.

The posterior distributions calculated in mb1 (available in the Extra material section) strongly supported the existence of an effect of filter on perceived boundary strength, with a stronger perceived boundary for filtered speech (^ß = 1.89, 95% Crl = [-0.91, 4.69], P (ß > 0) = 0.92). P (ß > 0) indicates the probability of filter having an effect on the DV. The data thus support quite well the existence of an effect. This implies that in this model, prosody had an effect on syntactic boundary perception, with stronger boundaries perceived in stimuli only featuring syllabic structure and intonation, and obliterating syntactic and lexical information.

4.1.2 Syntactic type

We then fitted a model, mb2, investigating an overall effect of syntactic type with and without filtering. We first show the ratings for the response variable (perceived boundary strength) before going through the model. A file detailing the total number of answers in each perceptual category (from 1: no boundary to 5: strong boundary) is available in the Extra material section (Answers per Likert category). Figure 8 shows violin plots for ratings in both conditions, where 0 stands for no filter and 1 for filter.

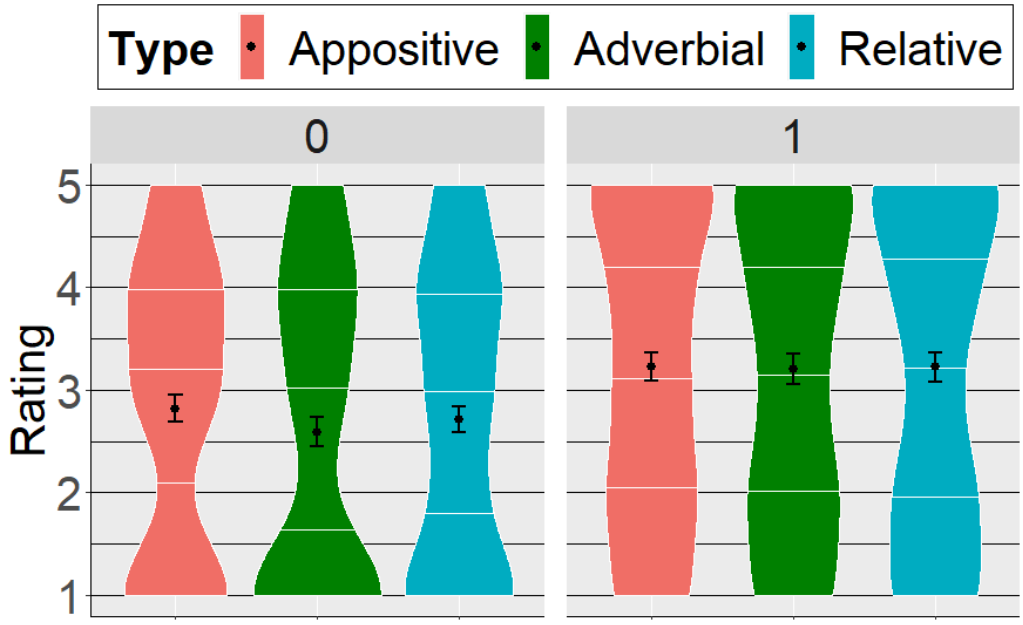

Figure 8. Ratings for the response variable, perceived boundary strength, represented in violin plots in the no filter (0) and filter (1) conditions.

Ratings are presented on the vertical axis, ranging from 1 (no boundary) to 5 (strong boundary). The shape of each violin plot varies in function of the distribution of answers in each perceptual category, *i.e.* each rating. The quantiles are shown with horizontal white lines in each violin plot. Each black point represents the mean rating score for every syntactic type in both conditions, along with its standard error bar. The filter condition (1) shows higher ratings in perceived boundary strength (mean ratings = 3.23 for appositives, 3.2 for adverbials, 3.22 for relative clauses). The no filter condition (0) shows more contrasts in ratings (mean ratings = 2.82 for appositives, 2.59 for adverbials, 2.71 for relative clauses). In both conditions, appositive clauses feature a slightly higher perceived boundary strength than adverbial and relative clauses.

The fitted model, mb2, did not show much of an effect of syntactic types on perceived boundary strength (^ß = 0.13, 95% Crl = [-0.38, 0.63], P (ß > 0) = 0.7 for appositives; ^ß = -0.08, 95% Crl = [-0.68, 0.46], P (ß < 0) = 0.61 for adverbials), while the same main effect of filter persisted.

4.1.3 Silent pause

We next investigated the effect of the presence of a silent pause and syntactic type on perceived boundary strength, with and without filtering. The model was fitted as mb4. Figure 9 shows the ratings for each syntactic type in the no filter (0) and filter (1) conditions.

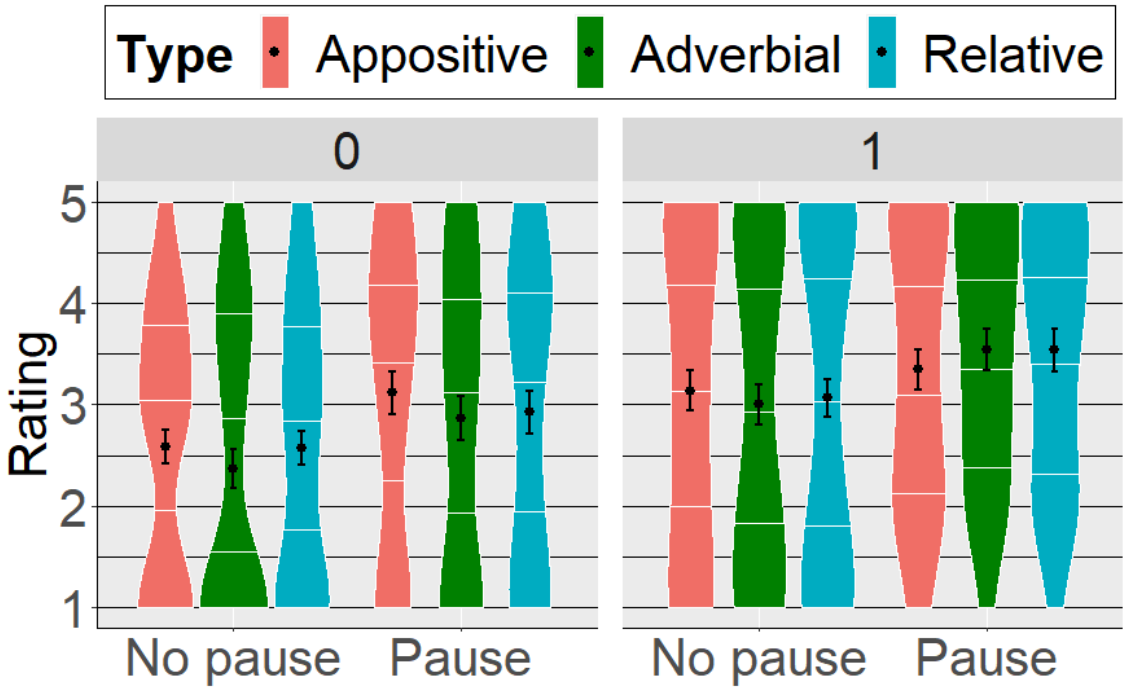

Figure 9. Violin plots of perceived boundary strength according to the presence or absence of a silent pause in the stimuli, in the no filter (0) and filter (1) conditions.

In the no filter condition (0), the overall ratings are higher with the presence of a pause, with a higher contrast in between the perception of syntactic types (mean ratings for types without a silent pause = appositives 2.58, adverbials 2.37, relatives 2.57; mean ratings for types with a silent pause = appositives 3.12; adverbials 2.87, relatives 2.92). Appositives show the highest ratings, followed by relative and adverbial clauses. Perceived boundary strength increases by 0.5 with a silent pause for appositive and adverbial clauses, while it increases by 0.35 for relative clauses. In the filter condition (1), the overall ratings are also higher with the presence of a pause. These ratings are much higher than in the non-filter condition. Adverbial and relative clauses show the highest ratings with the presence of a silent pause (mean ratings for types with a silent pause = appositives 3.35, adverbials 3.54, relatives 3.54).

The effect of the presence of a silent pause was probable, with a stronger perceived boundary for stimuli containing a silent pause (^ß = 1.04, 95% Crl = [-0.6, 2.64], P (ß > 0) = 0.91). We also found an interaction between silent pauses and filtering, in that this effect was different for filter and non-filter conditions, with higher ratings overall in the filter condition, for both no pause and pause.

Figure 10 shows a plot of the posterior distributions for each syntactic type, the pause variable, the filter variable, as well as the interactions between each variable.

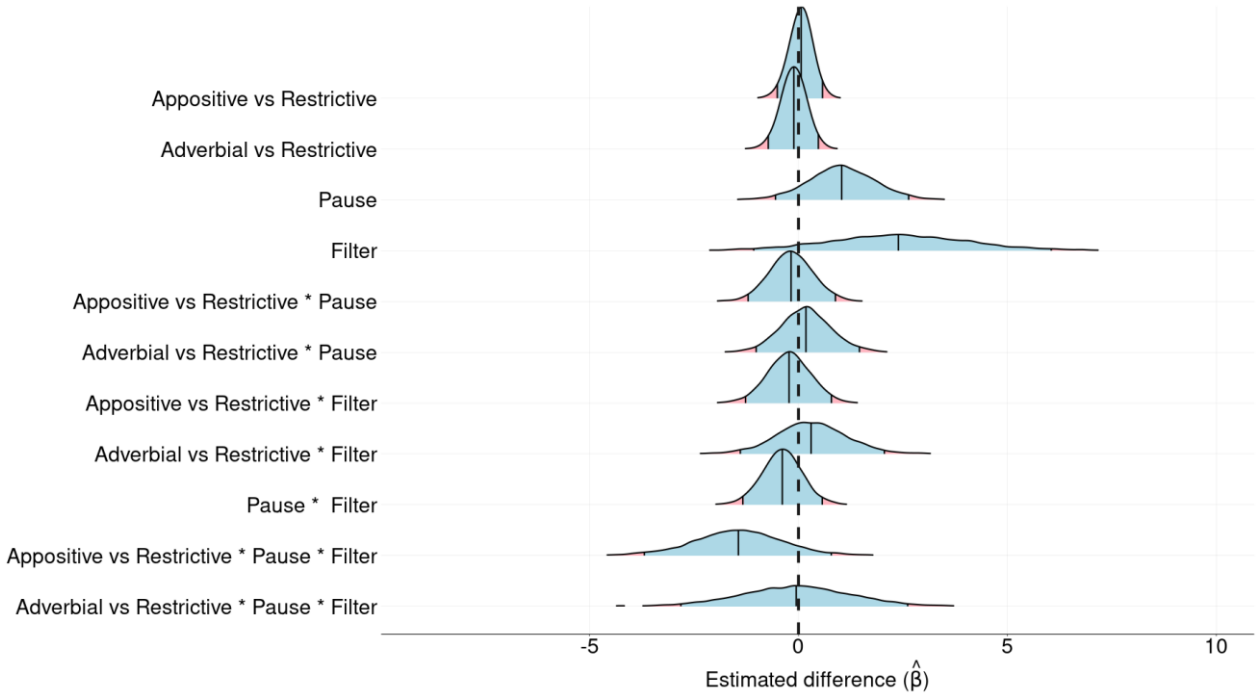

Figure 10. Posterior distributions of each syntactic type, along with those of the pause and filter variables and their interactions in mb4.

The third density ridge from the top, corresponding to pause, is on the right of the zero vertical dashed line, with a mean estimated coefficient of 1.04.

4.1.4 Final syllable lengthening

We also tested the effect of final syllable lengthening on perceived boundary strength. The model was fitted as mb5. Figure 11 shows the ratings for perceived boundary strength with no lengthening (0) and lengthening (1).

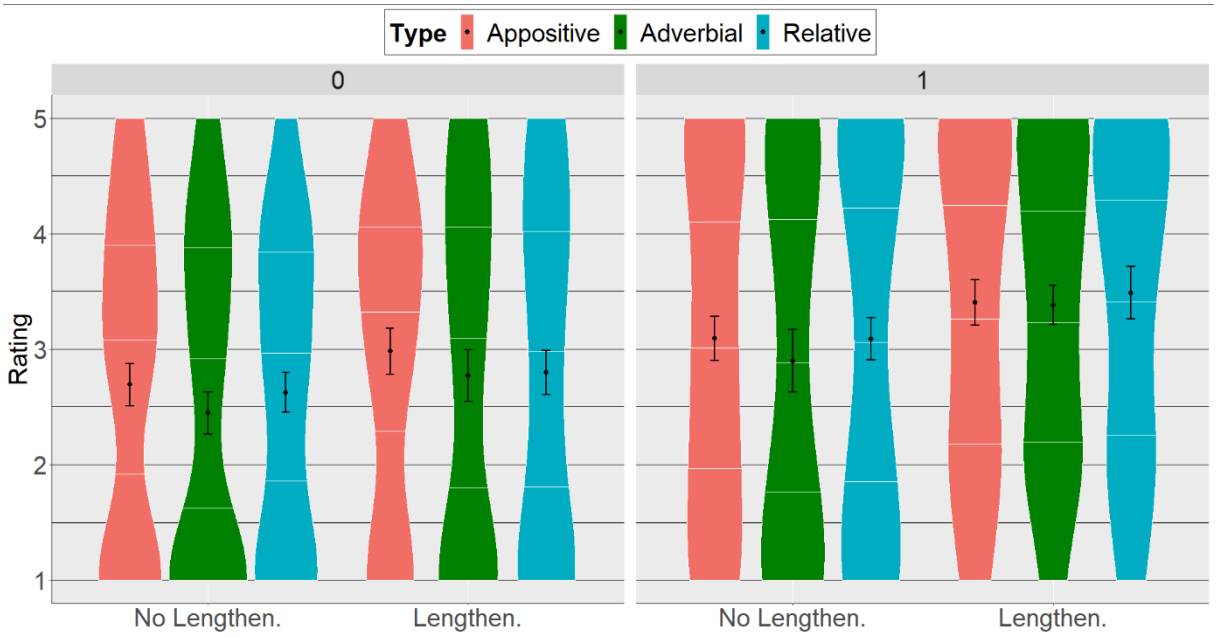

Figure 11. Violin plots of perceived boundary strength according to the presence or absence of final syllable lengthening in the stimuli, in the no filter (0) and filter (1) conditions.

All syntactic types show an increase in ratings with final syllable lengthening. In the no filter condition (0), appositive clauses show the highest ratings in both non-lengthened (mean rating = 2.69) and lengthened stimuli (mean rating = 2.98). In the no filter condition, ratings increase by 0.3 for appositive clauses and by 0.5 for adverbial clauses, while those for relative clauses increase by about 0.2.

The posterior distributions in mb5 supported the existence of an effect of final syllable lengthening on perceived boundary strength, with a stronger boundary perceived for speech containing final syllabic lengthening (^ß = 0.72, 95% CrI = [-0.21, 1.62], P (ß > 0) = 0.94).

Although the posterior distributions in mb5 showed the same main effect of filtering, there was no visible interaction between lengthening and filtering, as there was no difference between the filter and no filter conditions. This suggests that lengthening is more robust across conditions as an effect than silent pauses in our models, which were modulated by filtering.

Figure 12 shows a plot of the posterior distributions for each variable as well as their interactions. The third density ridge from the top, corresponding to final syllabic lengthening (length), is on the right of the zero vertical dashed line, with a mean estimated coefficient of 0.72.

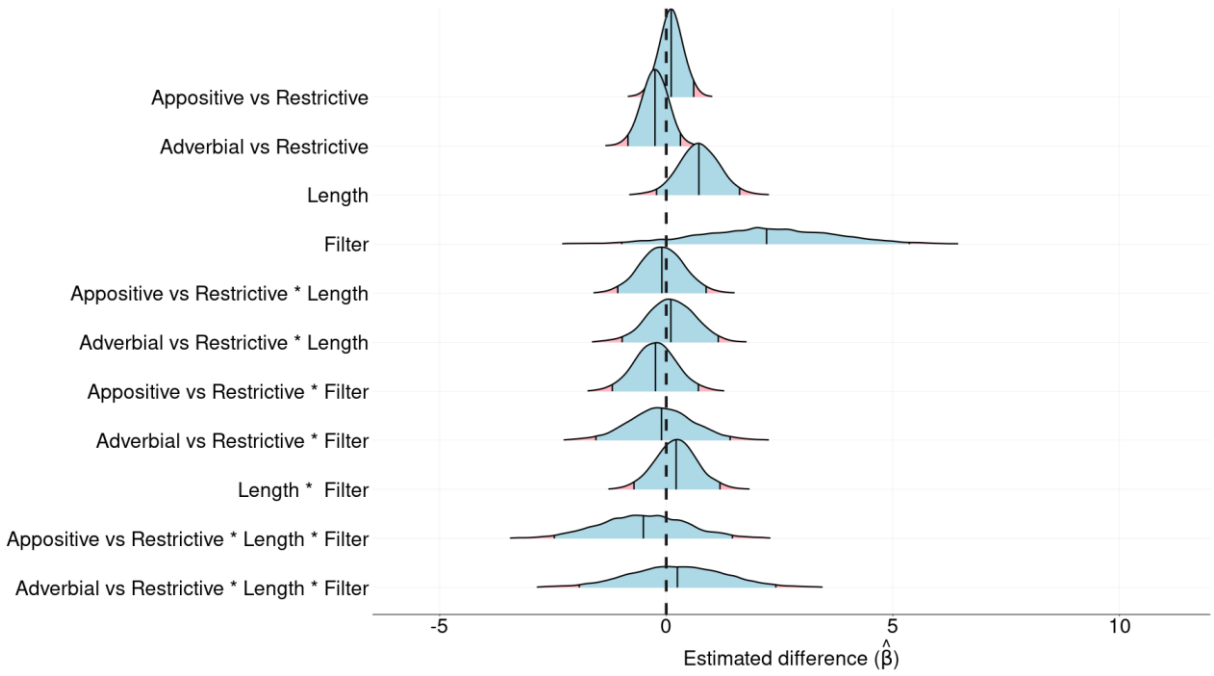

Figure 12. Posterior distributions of each syntactic type, along with those of the final syllabic lengthening and filter variables and their interactions in mb5.

The posterior distributions in mb5 also indicated a probable interaction between appositives and lengthening, with a weaker perceived boundary for filtered appositive clauses with final syllable lengthening (^ß = -0.5, 95% CrI = [-2.47, 1.46], P (ß < 0) = 0.7) than that for filtered relative clauses. However, the probability (based on our number of responses) led us to consider this effect unreliable. This can nonetheless be seen in Figure 11 in the filter condition (1), where appositive clauses without any final syllable lengthening feature a mean rating of 3.09, while relative clauses with final syllable lengthening show a mean rating of 3.48.

4.1.5 Pitch reset

The next model, mb6, tested the effect of pitch reset on perceived boundary strength. Figure 13 shows the ratings in the no filter (0) and filter conditions (1).

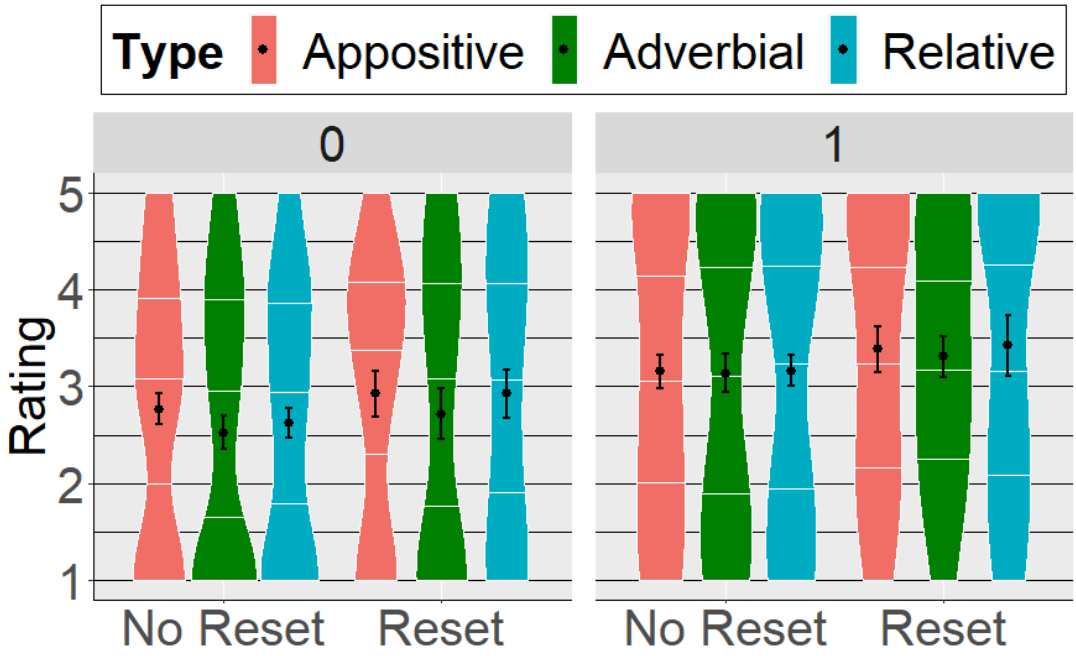

Figure 13. Violin plots of perceived boundary strength according to the presence or absence of pitch reset in the stimuli, in the no filter (0) and filter (1) conditions.

In both no filter and filter conditions, pitch reset shows slightly higher ratings across syntactic types. Ratings are higher for speech containing a pitch reset in the filter (1) condition (mean rating = appositives 3.39, adverbials 3.31, relatives 3.42) than in the no-filter (0) condition (mean rating = appositives 2.92, adverbials 2.72, relatives 2.92). In the no filter condition, ratings increase by 0.2 for all three syntactic types with the presence of a pitch reset.

Along with a consistent main effect of filtering, we found a main effect of pitch reset on perceived boundary strength, with a stronger perceived boundary for speech containing a pitch reset (^ß = 0.46, 95% CrI = [-0.2, 1.07], P (ß > 0) = 0.93). Figure 14 shows a plot of the posterior distributions for each variable, as well as their interactions.

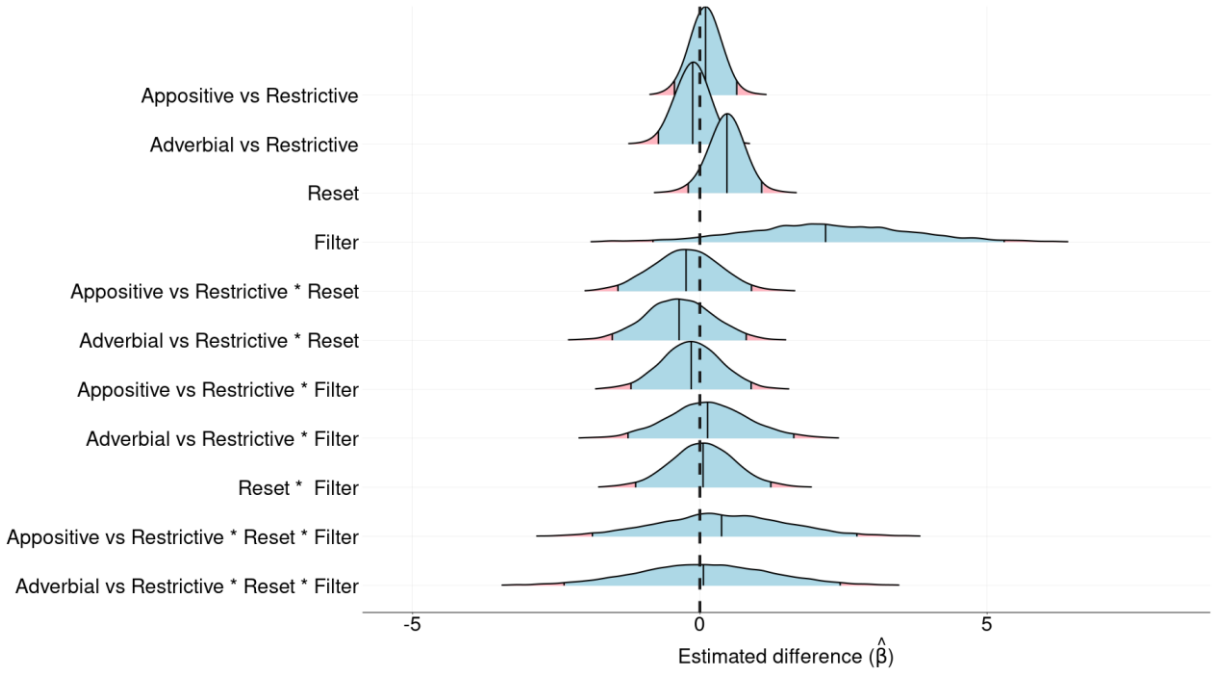

Figure 14. Posterior distributions of each syntactic type, along with those of the pitch reset and filter variables and their interactions in mb6.

The third density ridge from the top, corresponding to pitch reset, is on the right of the zero vertical dashed line, with a mean estimated coefficient of 0.46.

4.1.6 Number of prosodic cues

We then explored the effect of number of prosodic cues (0; 1; 2; 3) on perceived boundary strength, with and without filtering. The model was fitted as mb3. Figure 15 shows violin plots for ratings in the no filter and filter conditions, where 0 stands for no filter and 1 for filter on a vertical axis. It also features the ratings for the different number of cues simultaneously present in the stimuli, with 0, 1, 2, and 3 on a horizontal axis.

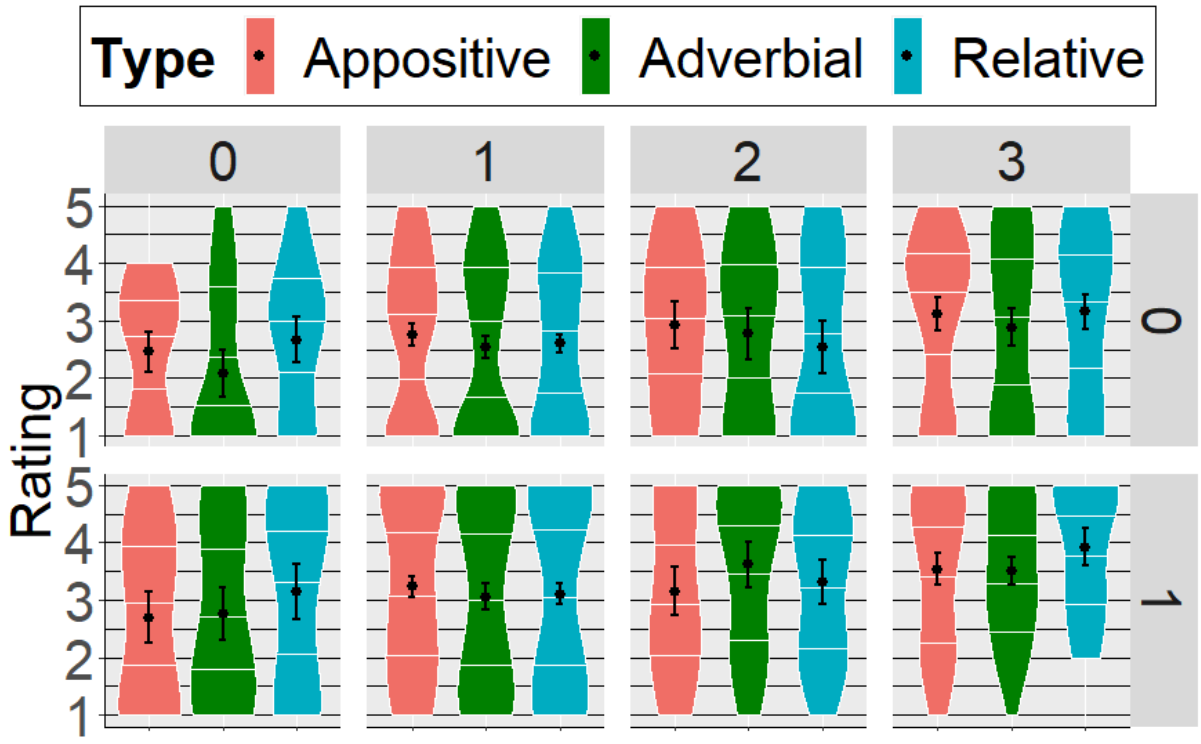

Figure 15. Violin plots of perceived boundary strength according to the number of simultaneous boundary cues in the speech stimuli (horizontally), in the no filter (0) and filter (1) conditions (vertically).

In both non-filter and filter conditions, ratings are higher as the number of simultaneous cues increases. In the filter condition, relative clauses show stronger ratings overall (3.15 for 0 cues; 3.1 for 1 cue; 3.31 for 2 cues; 3.92 for 3 cues). In the no filter condition, with a cluster of three prosodic cues as opposed to no cue, ratings increase by 0.65 for appositive clauses, 0.81 for adverbial clauses, and 0.48 for relative clauses.

The posterior distributions calculated in mb3 supported the existence of an effect of number of cues on perceived boundary strength, with a stronger perceived boundary for a higher number of cues (^ß = 0.45, 95% Crl = [-0.12, 1.03], P (ß > 0) = 0.95). The posterior distributions still supported a main effect of filter on perceived boundary strength, with a stronger perceived boundary for filtered speech. Interestingly however, no interaction was found between number of cues and filter, meaning that there was no difference in boundary ratings between filter and no filter, with regards to the impact of the number of cues (^ß for filter*number = -0.02, 95% Crl = [-0.51, 0.55], P (ß < 0) = 0.55).

Figure 16 shows a plot of the posterior distributions for the number variable, the filter variable, and the interaction between number and filter.

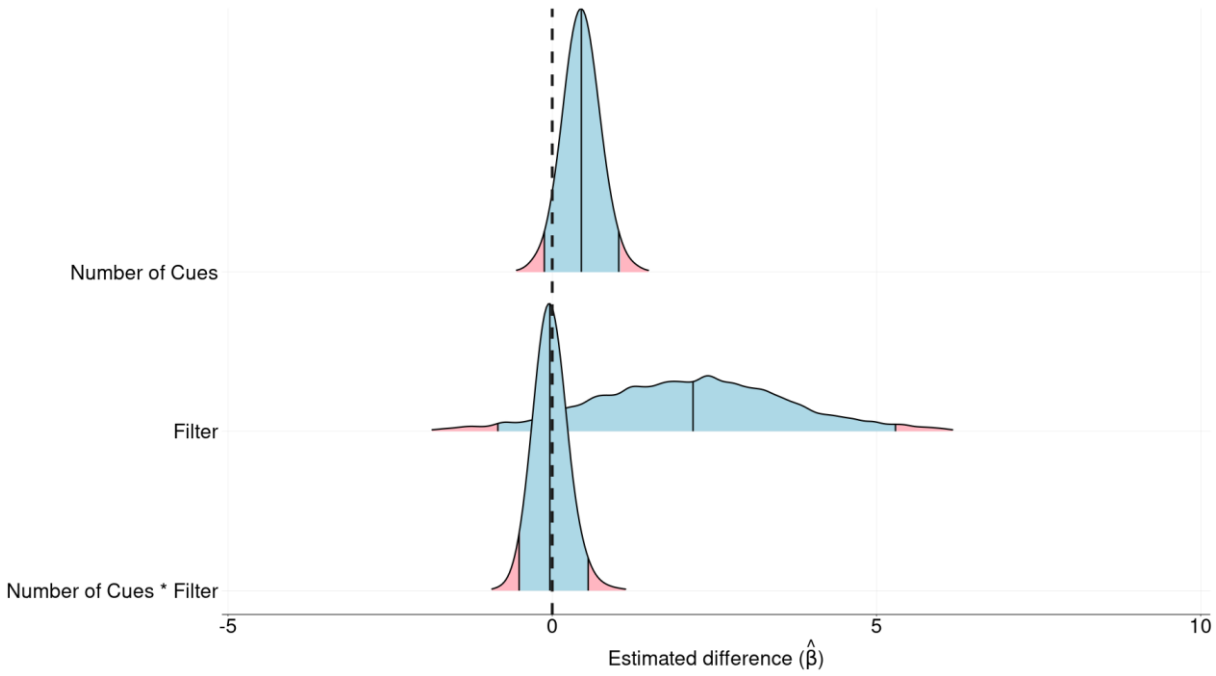

Figure 16. Posterior distributions of the number and filter variables along with that of their interaction in mb3.

The first density ridge, corresponding to the number of cues variable, is on the right of the zero vertical dashed line, with a mean estimated coefficient of 0.45.

4.1.7 Rising and falling tones

We finally tested the effect of the presence of a rising tone (mb7) and that of a falling tone (mb8) on perceived boundary strength.

Figure 17 shows the ratings in the no filter (0) and filter conditions (1) with the presence or absence of a rising tone.

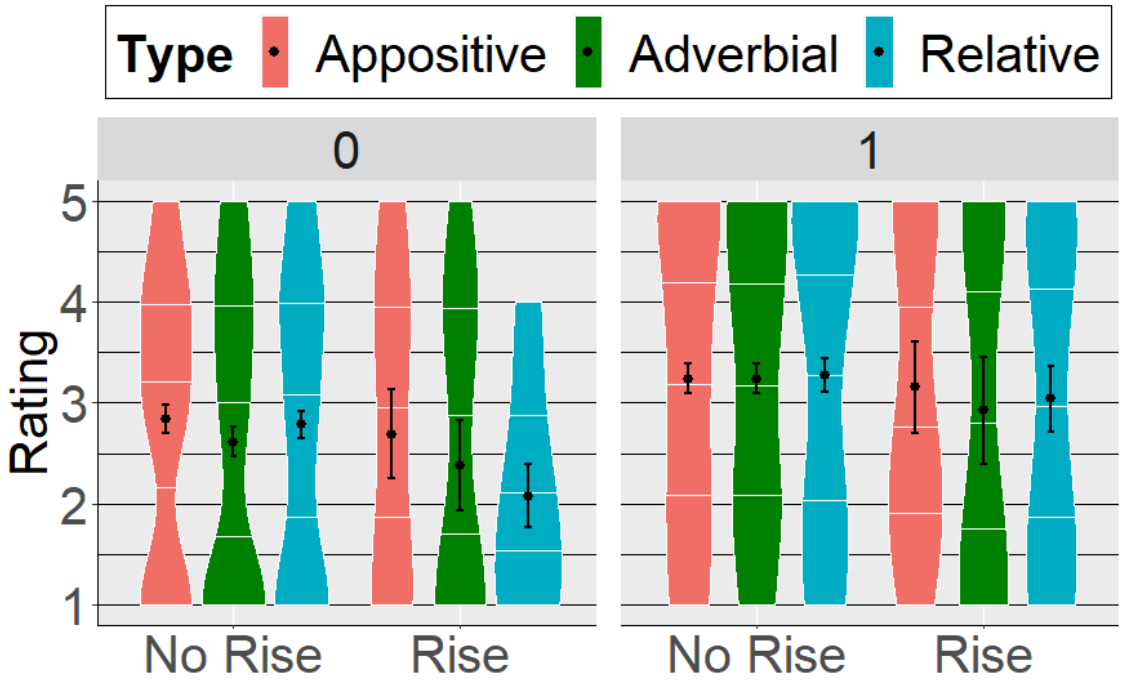

Figure 17. Violin plots of perceived boundary strength according to the presence or absence of a rising tone in the stimuli, in the no filter (0) and filter (1) conditions.

Ratings are weaker across the board for speech containing a rising tone, in both no filter (0) and filter conditions (1). In the no filter condition, ratings decrease by 0.15 (from 2.84 to 2.69) for appositives, by 0.23 from (2.61 to 2.38) for adverbials, and by 0.71 (from 2.78 to 2.07) for relatives. Ratings are stronger in the filter condition (1), and the contrast in ratings between syntactic types is clearer in speech containing a rising tone.

The posterior distributions of mb7 indicated a main effect of the presence of a rising tone on perceived boundary strength, with a weaker perceived boundary for speech containing a rising tone (^ß = -0.67, 95% CrI = [-1.65, 0.27], P (ß < 0) = 0.93). Figure 18 shows a plot of the posterior distributions for each variable and each of their interactions.

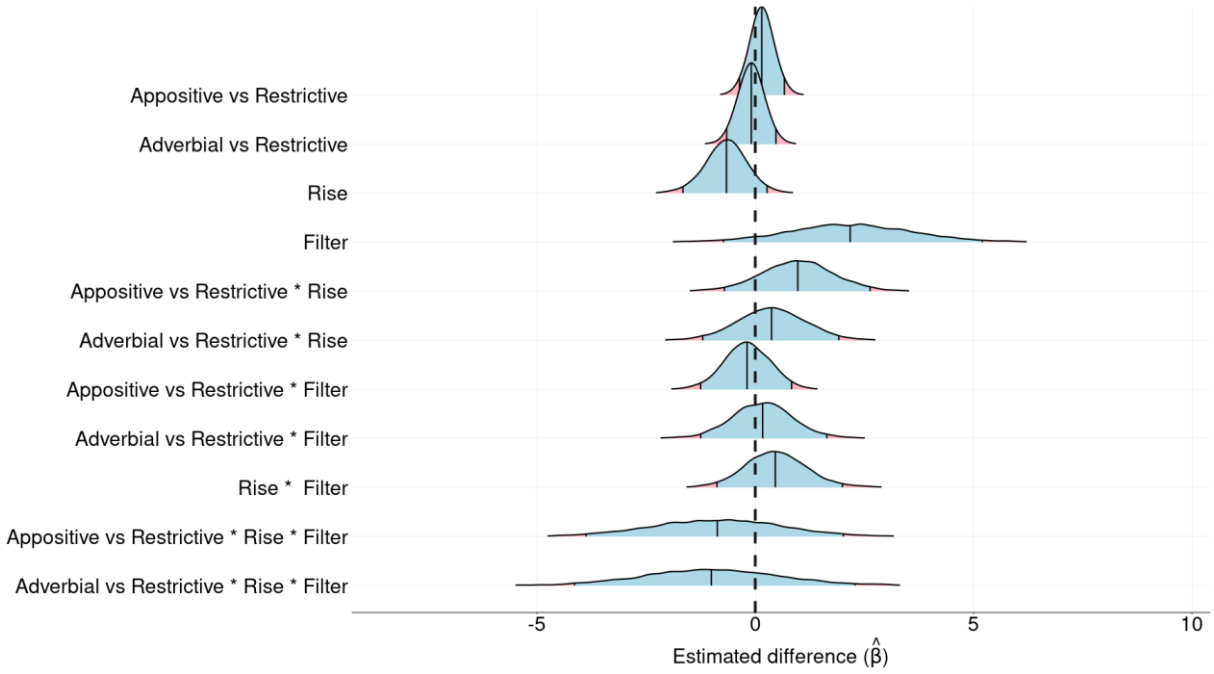

Figure 18. Posterior distributions of each syntactic type, along with those of the rising tone and filter variables and their interactions in mb7.

The third density ridge from the top, corresponding to rising tones (rise), is on the left of the zero vertical dashed line, with a mean estimated coefficient of -0.67.

Concerning the presence of a falling tone, Figure 19 shows the ratings in the no filter (0) and filter conditions (1).

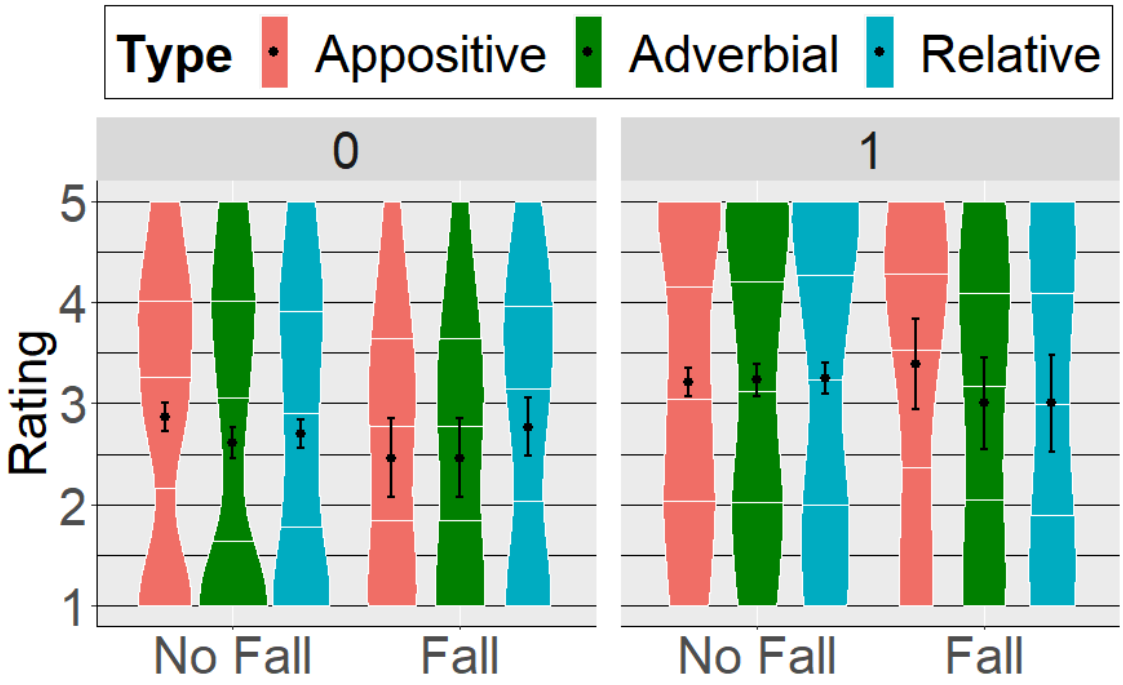

Figure 19. Violin plots of perceived boundary strength according to the presence or absence of a falling tone in the stimuli, in the no filter (0) and filter (1) conditions.

Ratings are overall weaker for speech containing a falling tone, in both no filter (0) and filter conditions (1). In the no filter condition, ratings decrease by 0.41 (from 2.87 to 2.46) for appositives and by 0.15 (from 2.61 to 2.46) for adverbials. They slightly increase by 0.07 (from 2.69 to 2.76) for relatives. An increasing pattern is also found for appositive clauses in the filter (1) condition.

With mb8, we found a main effect of the presence of a falling tone on perceived boundary strength, with a weaker perceived boundary for speech containing a falling tone (^ß = -0.39, 95% Crl = [-1.46, 0.72], P (ß < 0) = 0.79). However, the estimated coefficient was not high, and the probability score led us to consider this effect of falling tones as moderate. Figure 20 shows a plot of the posterior distributions for each variable and each interaction.

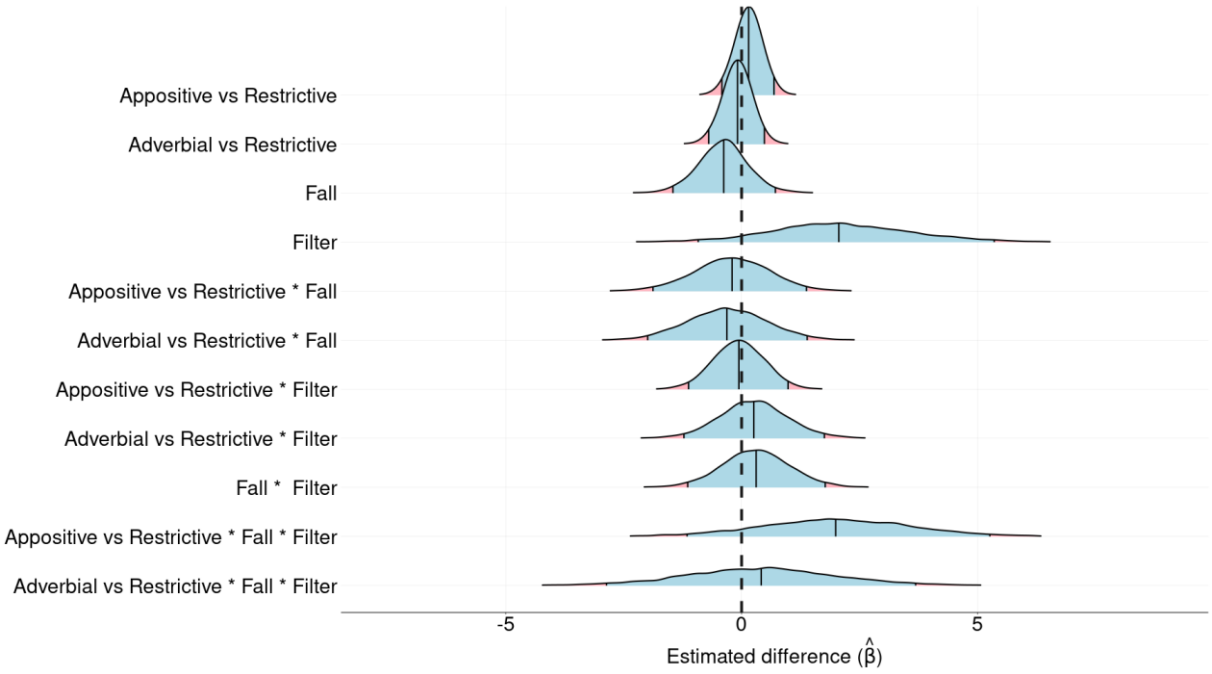

Figure 20. Posterior distributions of each syntactic type, along with those of the falling tone and filter variables and their interactions in mb8.

The third density ridge from the top, corresponding to falling tones (fall), is mostly on the left of the zero vertical dashed line, with a mean estimated coefficient of -0.39.

4.1.8 Summary of results for Experiment 1

To sum up, the exploration of interactions between isolated prosodic cues and prosodic boundary ratings shows that some cues were used by naive listeners to identify boundaries across our three syntactic types. These cues were silent pauses, final syllabic lengthening, and pitch reset. The presence of rising and falling tones in the stimuli led participants to assign lower ratings for perceived boundary strength. Table 1 provides a summary of effects for each variable in the form of their estimated coefficient, along with the probability the exact coefficient is greater or smaller (rising tone, falling tone) than 0. The results of Experiment 1 are discussed in section 5.

Table 1. Summary of main effects for each variable tested.

variable	^ß	P (ß > 0)
filter	1.89	0.92
silent pause	1.04	0.91
final syllable lengthening	0.72	0.94
pitch reset	0.46	0.93
combination of cues	0.45	0.95
rising tone	-0.67	0.93 (<)
falling tone	-0.39	0.79 (<)
appositive clauses (reference level: relative clauses)	0.13	0.7
adverbial clauses (reference level: relative clauses)	-0.08	0.61(<)

4.2 Results of Experiment 2

Just as in Experiment 1, we noted high inter-rater variability for Experiment 2. A graph showing each participant's mean ratings by syntactic type is available in the Extra material section.

We investigated the effects of syntactic type and gesture type on perceived boundary strength. Figure 21 shows violin plots for the overall ratings of each syntactic type.

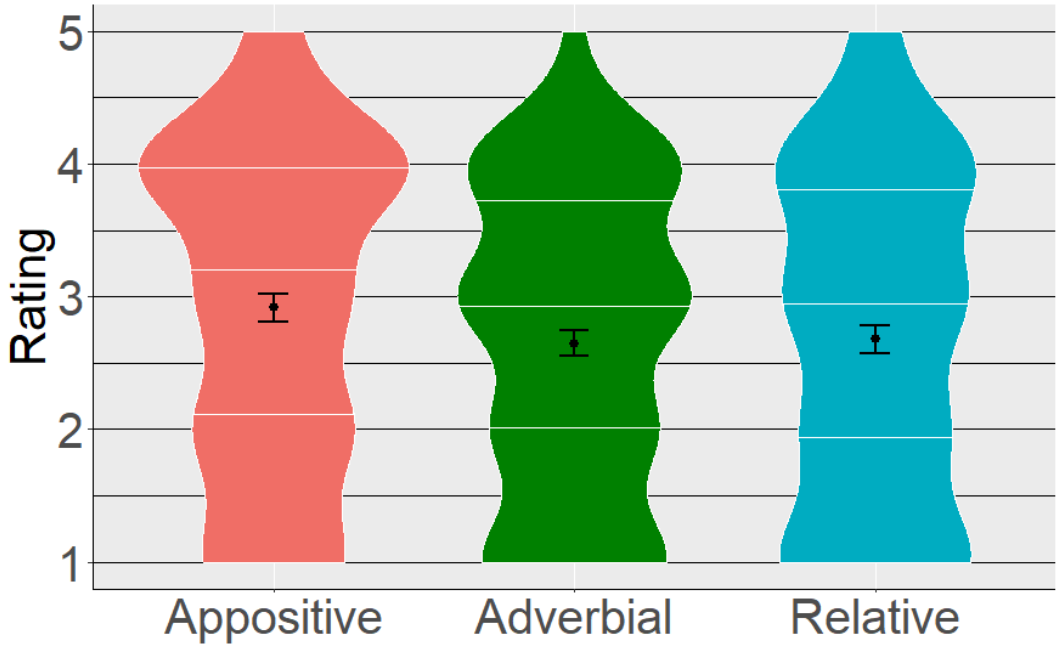

Figure 21. Violin plots of perceived boundary strength by syntactic type.

Appositives display a stronger boundary score (mean rating = 2.92) than adverbial (mean rating = 2.65) and relative clauses (mean rating = 2.68). All mean ratings are comprised in an interval from 2.5 to 3.

Figure 22 shows violin plots for the overall ratings for each gesture condition.

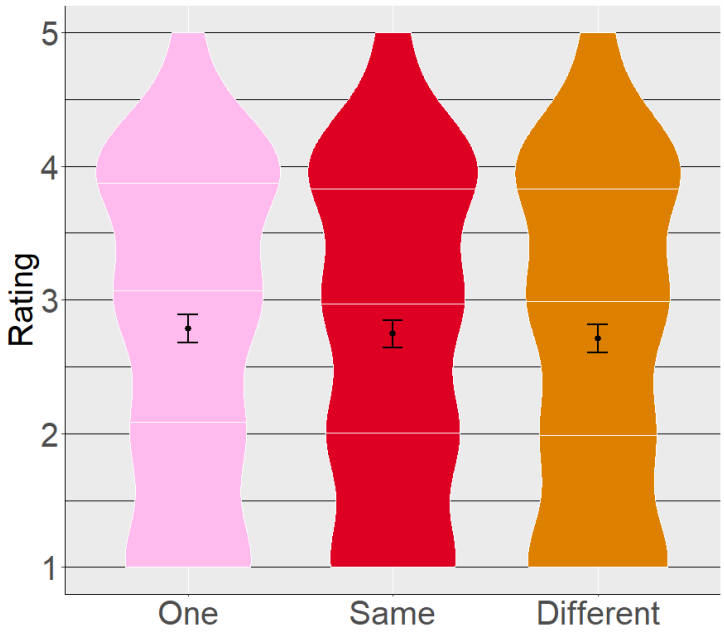

Figure 22. Violin plots of perceived boundary strength by gesture condition.

The three gesture conditions lead to fairly similar ratings, with a mean rating of 2.79 for one hand gesture, 2.75 for two identical hand gestures, and 2.71 for two different hand gestures.

As a single predictor, gesture condition does not appear to have an effect on perceived boundary strength.

Figure 23 shows violin plots of perceived boundary strength for syntactic types in each gesture condition.

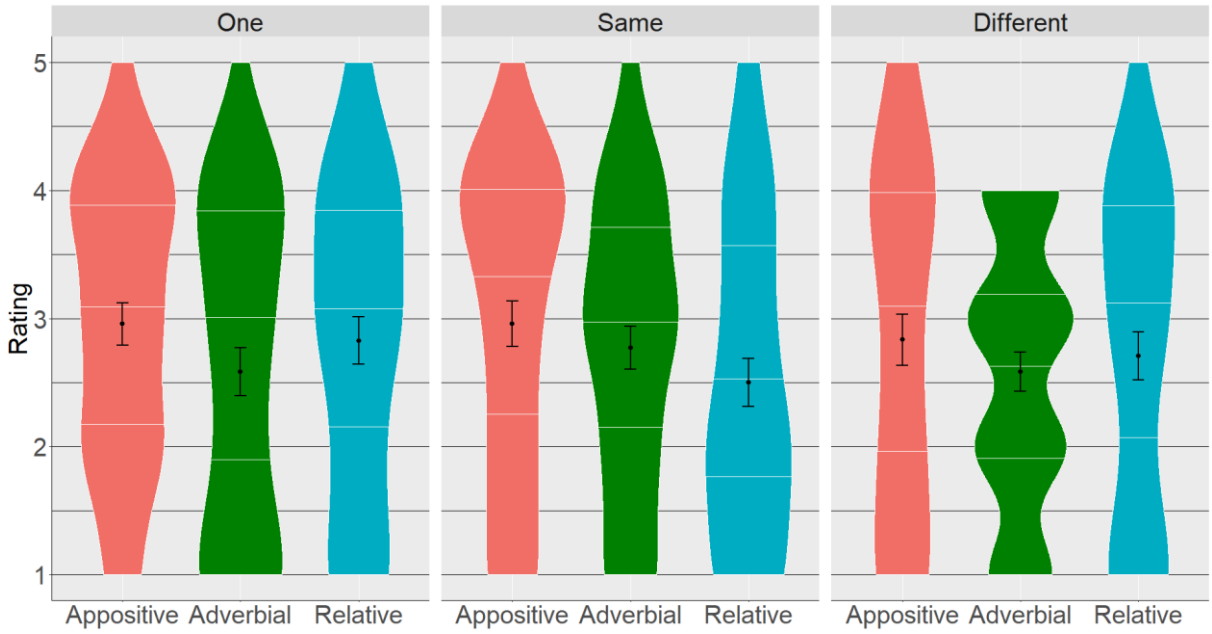

Figure 23. Violin plots of perceived boundary strength by syntactic type and by gesture condition.

When considering subsets for each gesture condition (one gesture overlapping tone-units; two identical gestures; two different gestures), ratings for syntactic types show more variety. Appositives still show higher mean ratings across all conditions (mean rating = 2.96 for one gesture; 2.96 for two identical gestures; 2.83 for two different hand gestures) than the other two syntactic types. The contrast between each syntactic type increases in the "Same" hand gesture conditions (mean rating = 2.96 for appositives; 2.77 for adverbials; 2.5 for relatives). This is the only condition in which relative clauses show weaker ratings than adverbials. Mean ratings for adverbial clauses across conditions are 2.58 for one gesture, 2.77 for two identical gestures, and 2.58 for two different hand gestures. Mean ratings for relative clauses are 2.83 for one gesture, 2.5 for two identical gestures, and 2.71 for two different hand gestures.

We modelled the "rating" dependent variable as influenced by the interaction of two independent variables: type and coding. The model was fitted as mbG. The posterior distributions calculated in mbG did not support the existence of an effect of syntactic type or gesture type on perceived boundary strength across the board. However, they moderately

supported the presence of an effect of syntactic type. Appositive clauses were linked to a stronger perceived boundary than relative clauses (^ß = 0.53, 95% CrI = [-0.51, 1.56], P (ß > 0) = 0.86). This discrepancy increased in the "same gesture" condition (*i.e.* in stimuli containing two identical hand gestures; ^ß = 0.6, 95% CrI = [-1.88, 3.06], P (ß > 0) = 0.7). The discrepancy between adverbial and relative clauses also increased in this condition (^ß = 1.03, 95% CrI = [-1.3, 3.49], P (ß > 0) = 0.82). It has to be noted that these results represent tendencies in data that moderately support the presence of an effect. More data for this test is needed to verify generalizability.

Figure 24 shows a plot of the posterior distributions for the syntactic type and gesture type variables.

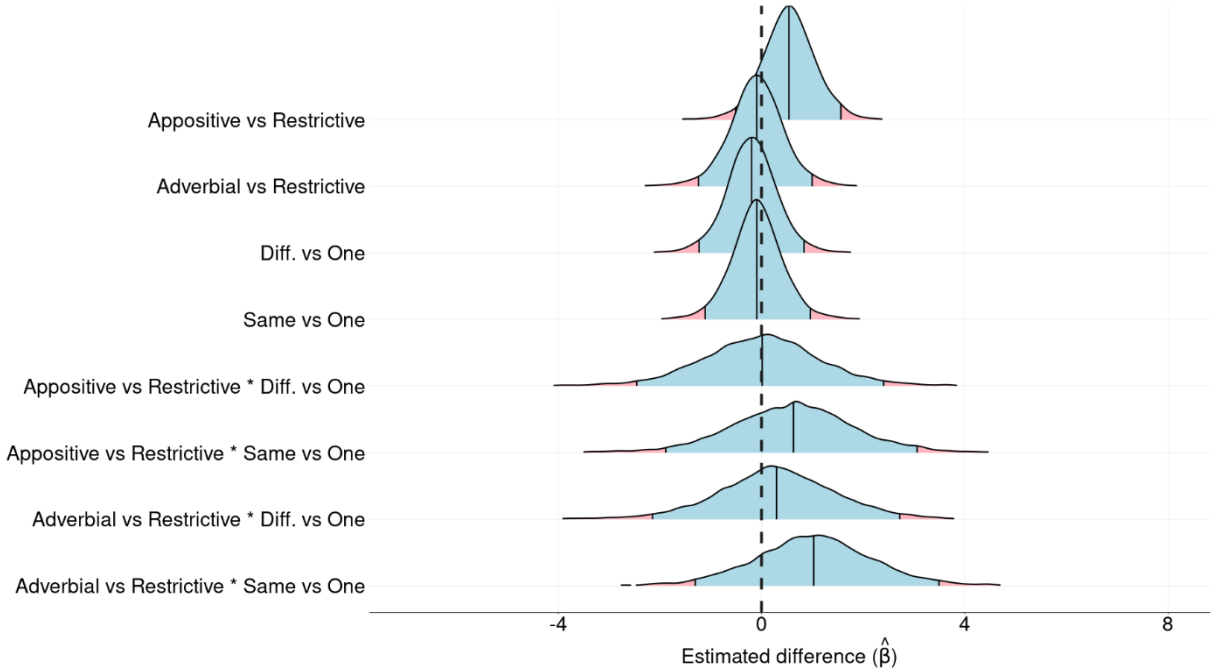

Figure 24. Posterior distributions of each syntactic type, along with those of the gesture conditions and their interactions in mbG.

The first density ridge, corresponding to the difference between the ratings for appositive and relative clauses, is on the right of the zero vertical dashed line, with a mean estimated coefficient of 0.53.

To sum up, as a single predictor, gesture condition did not have any effect on perceived boundary strength. Our data moderately supported the presence of an effect of syntactic type (increased in the "same" gesture condition), which suggests a trend whose generalizability needs to be verified.

5. General discussion and conclusion

5.1 Discussion of Experiment 1

Experiment 1 evaluates the perception of boundaries by naive listeners, in extracts of spontaneous speech containing syntactic subordinate structures. Research question (1) can be answered as follows: prosody modulates the strength of perceived syntactic boundaries. The cues involved in the modulation of perceived syntactic boundaries are silent pauses, final syllable lengthening, pitch reset, and the simultaneous presence of several cues on one stimulus.

These results cannot be explained in terms of constituent length, sequential position, discourse status, or prosodic cues, since these parameters were controlled. However, constituent complexity and weight remain potential factors. Likewise, the degree of (potential) syntactic closure varies in between segments. Specifically, some of the adverbial clauses are preceded by a (potential) full syntactic closure, while others are preceded by a nominal clause projecting a verbal clause.

The models showed that naive participants are able to perceive different degrees of boundary strength. This study therefore adds to the existing evidence of naive listeners' sensitivity to prosodic cues (Mo, Cole, and Lee, 2008; Cole, Mahrt, and Hualde, 2014). However, inter-rater variability was high for Experiment 1. We assume this perceptual variability is directly linked to the recruitment method for this study, since two thirds of the participants were recruited online and completed the experiment from home. We also presume that the use of different, personal material (headphones, laptop or PC) for all participants during the procedure increased variability. Standard factors for high variability include performance factors (attention and fatigue) and interference from environmental noise or activity. Variability in prosody perception can also arise due to differences in participants' linguistic experience (Cole, Mo, and Baek, 2010), for instance unfamiliarity with a speaker’s voice or their phonetic expression of prosody. These differences may well be heightened with stimuli containing non-elicited spontaneous speech.

Slightly stronger boundaries are perceived in appositive relative clauses. Although syntax and prosody are not necessarily isomorphic, this result is in accordance with macro-syntactic production results and with their description in the literature, since appositive clauses are considered macro-syntactically detached from the matrix clause (De Vries, 2006; Krifka, 2007; Loock, 2007). Our data is also in agreement with the results found by Auran & Loock

(2011), who showed that silent pauses and final syllabic lengthening are typical features of discourse sequences containing an appositive clause.

One of the potential explanations for the fact that adverbial clauses are in most cases associated with weaker boundaries than appositive and relative clauses comes from the Discourse Analysis literature, in which adverbial clauses are defined as featuring a flexible macro-syntax (*e.g.* Thompson, 2002), mirroring their variety of discourse functions. They also show more variety in their degree of potential syntactic closure.

The fact that relative clauses show slightly higher ratings than adverbial clauses contradicts both the description of relative clauses in the syntactic literature and our production results. The literature generally agrees with the fact that restrictive relative clauses are mostly fully integrated to their left co-text in macro-syntactic terms (Matthiessen & Thompson, 1988; De Vries, 2006). In our production study, restrictive relative clauses featured very few prosodic boundaries, all of them being produced after the realization of the restrictive relative clause (Lelandais & Ferré, 2016). In our ENVID corpus, restrictive relative clauses are in majority produced in the same tone-unit as the preceding one, while the rest is typically preceded with a continuation contour (*i.e.* final rise). The final syllable of (1) and the initial syllable of (2) are usually realized at the same F0 height. The response is therefore not based on any production-based learning effect which would imply a strict correspondence between a syntactic type and prosodic features.

Given the fact that subjects labelling prosodic boundaries in their own language have been described as usually perceiving more boundaries than subjects annotating delexicalized speech (Mettouchi *et al.*, 2007), the fact that participants perceived stronger boundaries in delexicalized speech in this study suggests the presence of subordinate clauses in stimuli, whatever their type, downgrades perceived boundary strength. This remark has yet to be hypothesized and probed with another experimental setup, with stimuli containing main clauses and subordinate clauses. These subordinate clauses are all verb phrases that display traditional prosodic phrasing cues at their right edge, as described in Cole, Mo, and Baek (2010). This study adds to the evidence that the boundaries at their left edge, however, are perceived differently.

Concerning the interactions between isolated prosodic cues and prosodic boundary perception, silent pauses represent the strongest cue for boundary perception. This falls in line with a vast number of studies in various experimental settings with read, radio, or narrative

speech (Carlson & Swerts, 2003; Yoon, Cole, & Hasegawa-Johnson, 2007; Wagner & Watson, 2010; Roy, Cole, & Mahrt, 2017; Kuang, Pik Yu Chan, & Rhee, 2022).

The second strongest cue in our models is final syllabic lengthening, without the simultaneous presence of any silent pause or pitch reset. These findings differ from those of Wightman *et al.* (1992), who had observed that major boundaries were better distinguished with a combination of silent pauses, lengthening, and intonational cues. More recently, Kuang, Pik Yu Chan, & Rhee (2022) as well as Pintér, Mizuguchi, and Tateishi (2014) found final syllabic lengthening more reliable when coupled with a silent pause, with a more global effect of durational cues in general. Cole, Mo, and Baek (2010) had pointed out, however, that lengthening might well be sufficient to drive the perceptual processes underlying speech segmentation for major boundaries. Hilton *et al.* (2019) provided backing evidence for the conspicuousness of durational cues and of lengthening, but specified that the scope, amplitude and targets for these durational features to cue a boundary were not yet precisely delimited. This main effect of final syllabic lengthening takes place despite the fact that acoustic effects of prosody vary according to the phonological content of the word or syllable. Lee *et al.* (2006) and Mo (2008) both reported variability in final lengthening effects on vowels as a function of vowel phoneme.

Pitch reset, in our models, has a similar effect on the response variable as an isolated cue to the presence of several cues at once. This is unexpected given the fact that the literature has yielded mixed results about the effect of pitch reset as a single cue in English (Biron et al., 2021; Kuang, Pik Yu Chan, and Rhee, 2022) and in other languages (see Simon & Christodoulides, 2016 for French). Other studies have pointed out a weak effect of pitch cues in general (Roy, Cole, & Mahrt, 2017). These mixed results are usually accounted for by unnaturalness. Indeed, pitch reset is rarely found as a single cue in natural speech, as it is usually produced with a pause and final lengthening (de Pijper & Sanderman, 1994). Furthermore, native speakers of English have been described as more sensitive to unit-final boundary cues than unit-initial cues (Kuang, Pik Yu Chan, and Rhee, 2022). The steady effect of pitch reset in our models is even more unexpected since it is the only unit-initial cue we included in the stimuli.

The number of prosodic cues also has an effect on perceived boundary strength. Ratings increase substantially when three prosodic cues (silent pause, final syllabic lengthening and pitch reset) are clustered. These results are in line with the majority of perceptual studies in

English and in other languages, observing higher perceived boundary strength when cues are combined (*e.g.* Hilton *et al.*, 2019; Kuang, Pik Yu Chan, and Rhee, 2022).

Rising tones were also found to have a main effect on perceived boundary strength, in that participants reliably perceived weaker boundaries in stimuli with a rising tone (when filter = 0). The effect is attenuated in filtered speech, suggesting congruency between a syntactic subordinate clause and the presence of a rising tone. Because of the controlled maximum pitch height for all rises, participants are likely to have interpreted these rises as continuation (*i.e.* nonfinal) contours. Relative clauses are particularly affected by the difference in ratings. This is consistent with their production pattern, since continuation contours are frequent features of relative clauses (Lelandais & Ferré, 2016).

Falling tones caused participants to perceive weaker boundaries as well. The direction of this effect was unexpected, since falling tones ending with lower F0 values are generally mentioned as part of boundary cues (Cole, Mo, and Baek, 2010). We think the reason for this effect lies in the fact that the tone does not end low enough to be perceived as conveying finality (Cole, Mo, and Baek, 2010), because of the way the stimuli were manipulated to only feature a falling tone without any pitch reset. A competing explanation has to do with the position of the tone itself. A falling tone heard in the middle of two tone-units might not carry any meaning of finality as a single cue, because of the existence of an immediate co-text. In conversational speech, adverbial clauses are commonly preceded with a falling contour on (1), which is often in relation with the management of interpretative frames (Chafe, 1984). Adverbial clauses can be analysed as a means of grouping several discourse segments together, linked by the fact that they have to be interpreted by a same criterion, which is delivered in an adverbial clause (Brown & Yule, 1983). This implies that co-speakers have to keep this criterion in mind for the treatment of the host segment and the segments that follow, until the production of a cue indicating the end of its range. A falling tone on the preceding segment can be such a cue.

In the context of relative clauses, falling tones caused slightly higher ratings. We suspect the presence of a falling tone introducing an embedded clause caused incongruency, and triggered a higher response from participants. Kim *et al.* (2006) reported falling F0 contours and other cues having a greater incidence on the perception of major rather than lower-level boundaries. Cole, Mo, and Baek (2010) pointed out that the differences between lower- and higher-level prosodic boundaries in terms of syntactic factors had not been explored, but they expected that the lower-level boundaries were more commonly associated with the edges of lower-level syntactic constituents, or were used in contexts of syntactic embedding.

Lastly, we address the fact that some of the prosodic cues impacted perceived boundary strength differently across syntactic types in the filter condition. This phenomenon could stem from the fact that speech rate and vowel intensity were not controlled. Vowel intensity has been shown to play a role in the perception of prosodic boundaries (Mo, 2008), and differences in the stimuli despite the Pass Hann band and the amplification might have led our participants to assign higher scores. Speech rate has not been shown to play a role in the perception of boundaries, but in that of prominence (Priva, 2017). Some of the stimuli might have been perceived as prominent, and interfered with participants' judgement about perceived boundary strength.

In a nutshell, our results show that silent pauses, final syllabic lengthening and pitch reset are used as prosodic cues by naive listeners to identify boundaries in speech and assess their weight, in sequences containing subordinate clauses. These results match our production results in British English, since more prosodic cues are produced in appositive clauses than in adverbial and restrictive relative clauses (Lelandais & Ferré, 2016). These results are also in line with recent studies (Kuang, Pik Yu Chan, and Rhee, 2022) showing the importance of silent pauses and final syllabic lengthening in the perception of boundaries in speech.

5.2 Discussion of Experiment 2

Experiment 2 evaluates the perception of boundaries by naive listeners, in extracts of spontaneous speech containing syntactic subordinate structures, with speakers producing different iconic, metaphoric, or pointing gestures. Research question (2) can be answered as follows: gesture, as a single predictor, does not modulate the strength of perceived syntactic boundaries. However, the discrepancies in ratings for each syntactic type increase in a specific condition (when two identical hand gestures are produced sequentially). The fact that the posterior distributions moderately support the existence of an effect for this condition suggests a tendency that has to be probed with more data.

As in Experiment 1, these results cannot be explained in terms of constituent length, sequential position, and discourse status, as these parameters were controlled in the study. However, constituent complexity, weight, and degree of syntactic closure remain potential factors.

A trend can be observed in Condition 2 only, when two identical gestures are produced. Condition 2 caused slightly higher ratings for appositive and adverbial clauses compared to those in the other conditions. Gesture repetition has never been studied from the perspective of

boundary perception in speech, but has yielded conflicting results from the point of view of their informational input (Hoetjes *et al.*, 2015). The concept differs in that in most studies, gesture repetition is accompanied with referential repetition, which is not the case in our data. Hoetjes *et al.* (2015), who studied the extent to which gesture reduction was comparable to other forms of linguistic reduction, found that participants judged gestures from repeated references as less precise than those from initial ones. However, they also found that gestures from repeated and initial references were equally successful in communicating information.

The results reported by Hilton *et al*. (2019) sheds some light on the trend observed for Condition 2. Hilton *et al.* (2019) found that visually presented non-speech stimuli in the form of action sequences performed with hands elicited the same electrophysiological correlate (a Closure Positive Shift) as prosodic boundary processing. The shared features of the boundary cues included in their study were all durational. We believe that the identical hand gestures in Condition 2 might have prompted participants to focus on the return to rest position in between them (although present in all conditions) and might have increased their sensitivity to the deceleration before this return to rest position.

Another related explanation for this tendency is that the repetition of two identical hand gestures has a priming or bootstrapping effect on the response variable. Whether this effect is specific to perceived boundary strength or to any response variable deserves future investigation.

In this same Condition 2, lower responses were given to relative clauses only. If repeated gestures do create cohesion as claimed by some production studies (Lascarides & Stone, 2009), participants only perceived it in speech stimuli containing relative clauses. Cohesion implemented through gesture might be a subtle cue that can only be reliably perceived by participants when other, more robust cues for cohesion are available, such as syntactic embedding.

The fact that the production of two different hand gestures (condition 3) does not influence perceived boundary strength contradicts our predictions. The production of two successive hand gestures with different configurations and trajectories did not have any consistent effect on perceived boundary strength. Since the capacity of hand gestures to create a boundary in discourse has been mentioned in production studies (Streeck, 2009; Enfield, 2009; Calbris, 2011) but never been probed experimentally, it is plausible that this cue, in isolation, does not affect participants' speech segmentation. The fact that hand gestures

produced in co-occurrence with appositive clauses show different formal configurations from the other hand gestures produced in their co-texts (Lelandais, 2020) is not linked to any pattern in perception. Another possibility is that the nature of the stimuli hampers participants' judgment about boundary strength. The stimuli we used were short, and do not reproduce participants' actual exposure to discourse because of their length. Participants might need more time to perceive such boundaries, especially with stimuli containing non-elicited spontaneous speech. More context and more exposure might modulate participants' performance. Another potential factor is that participants might have been unfamiliar with the gestures that were implemented in the stimuli, and some of them might have been judged incongruous despite our pre-screening.

The absence of effect in our models for Conditions 1 and 3 indicates that Condition 1 and Condition 3 are not predictors for boundary perception, and that they are not predictors for the perception of an absence of boundary either. Our results do not show that the production of a gesture in overlap between two different tone-units (Condition 1) or that the production of two different gestures (Condition 3) are associated with the absence of boundaries, or with cohesion in discourse. The gestures represented by these two conditions cannot be regarded as cues for continuity.

The hierarchy in overall ratings for perceived boundary strength between syntactic types echoes that in Experiment 1, with a slightly higher gap in between the highest- and lowest-rated syntactic types, *i.e.* appositive and adverbial clauses respectively, in both experiments. The potential explanations for the fact that adverbial clauses show the lowest ratings have been discussed in Experiment 1, and remain relevant for Experiment 2. From the side of production, these syntactic constructions were produced with very few gestural boundary markers in the ENVID corpus (Lelandais, 2020).

In production, the quantity and distribution of gestural cues are lower compared to those of vocal cues. Speakers use more prosodic than gestural boundary cues when producing subordinate structures, and gestural cues are rarely used in combination on a same occurrence. Two different reasons could account for this gap. Speakers could give more importance to the vocal modality for the expression of a boundary and use gestural cues as a complement. Alternatively, speakers could consider the few gestural cues they use are sufficient to express a boundary. The fact that head beats and eyebrow movement do not appear together in different discourse situations has for instance been highlighted by House, Beskow, and Granström (2001). These two cues have been analyzed as having a sufficient impact on co-speakers'

perception to be used as isolated markers. The present perception test allows us to refine these possibilities. The production of two successive different hand gestures in configuration and trajectory (Condition 3) is not enough for participants to perceive a boundary in discourse, in the absence of prosodic boundary cues. As a potential follow-up of this test, one might agglomerate several gestural boundary cues in order to determine whether a correlation of cues can significantly impact perceived boundary strength.

The prosodic value of gestures is often referred to while studying speech segmentation, especially that of beat gestures or head and eyebrow movement. In this test, we did not include any variable related to the prosodic value of gestures. On the basis of Experiment 2, we cannot say that speech segmentation can be achieved with other types of hand gestures, but according to our model, two identical hand gestures are the most likely to trigger a response.

5.3 Conclusion

Experiments 1 and 2 confirmed that (1) prosody modulates the strength of perceived syntactic boundaries. The prosodic cues involved in this modulation are silent pauses, final syllabic lengthening, pitch reset, and a combination of these cues. Our two experiments, however, did not confirm that (2) gesture modulates the strength of perceived syntactic boundaries. A trend has been observed with two successive hand gestures showing identical configurations and trajectories, but has to be verified with further data to be meaningful.

Experiment 1 shows a gradation in perceived boundary strength in function of prosodic cues, which mirrors the production patterns observed on an earlier production study. Silent pauses and final syllabic lengthening are used by naive participants in all three types of subordinate constructions to identify boundaries in speech.

Experiment 2 did not reveal any consistent effect of gesture on perceived boundary strength. We used parameters which are proper to the gestural modality, such as articulator configuration and trajectory. We did observe a trend for condition 2 (production of two successive hand gestures that are identical in trajectory and configuration), which suggests a small degree of systematicity in the use of some gestural cues in perception. Although this tendency has to be probed with more data, participants might be sensitive to some (non-prosodic) gestural features in boundary strength perception.

It has to be noted that our two perception tests are different in the number of cues included as variables. We did not include any combination of several gestural cues in

Experiment 2, and we did not test a wide array of gestural cues. The motivations of this specific test were merely exploratory, since no study had documented the segmentation value of gesture without any prosodic quality. We were also bound to having few stimuli because of the online nature of the test and participants' attentional fatigue. We wanted to work with videos showing gesturing persons other than actors, while avoiding video synthesis. It would however be useful, in a follow-up study, to test correlations between prosodic and gestural cues in one same experiment.

The present study reveals differences in perceived boundary strength with stimuli containing different types of subordinate clauses, and contributes in this sense to the exploration of the interface between syntax and prosody (Experiment 1) and of that between syntax and gesture (Experiment 2).

The most immediate and straightforward development of this study would be to increase the number of participants to probe the strength of the effects we obtained and to gain insight on the trends we observed. This study could also be expanded with stimuli containing syntactic coordinate structures. We would expect higher perceived boundary strength overall, but whether we find stronger effects of prosodic or gestural cues would be of particular interest. Since participants perceived stronger boundaries in delexicalized speech in this study, it would be relevant to compare the effect strength of filter as well.

Finally, we aim at improving the spontaneous aspect of the stimuli. Our test does not reflect the online spontaneous decisions made by participants in a situation of conversation because of the isolation of the prosodic and gestural cues, and because of the shortness of stimuli. This test nonetheless exposes naive participants to stimuli that look like short extracts from spontaneous conversation. One of the advantages of the stimuli is that they have been produced in presence of an interlocutor (Petrone *et al.*, 2017). Mo & Cole, in their 2010 study, showed proof of an important perceptual distinction between read and spontaneous speech. While pauses are necessary to the identification of a boundary in read speech, they are important but not essential to the identification of a boundary in spontaneous speech. Including extracts of spontaneous speech in more perceptual studies may facilitate comparisons across speech genres and discourse formats, in that it helps overcome limitations stemming from controlled or virtual environments (*e.g.* reducing the analysis to head or eyebrow movement; Jiménez-Bravo & Marrero-Aguiar, 2020). In our data however, the fact that the gesturing persons follow

instructions remains another important limitation, which plays a role in the relevance of our results.

Author contributions

Manon Lelandais: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, review & editing.

Gabriel Thiberge: Methodology, Software, Visualization, Writing – review & editing.

Acknowledgements

We deeply thank Lauren Fromont and Elizabeth Couper-Kuhlen for their valuable insight and comments on a previous version of this manuscript.

References

Amir, N., Silber-Varod, V., & Izre'el, S. (2004). Characteristics of Intonation Unit Boundaries in Spontaneous Spoken Hebrew—Perception and Acoustic Correlates. *Proceedings of Speech Prosody 2004*, 1–4. Nara, Japan: ISCA.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. London: Cambridge University Press.

Astésano, C., Bertrand, R., Espesser, R., & Nguyen, N. (2012). Perception of boundaries and prominences in French. *Proceedings of JEP-TALN-RECITAL*, 1–5. Grenoble, France.

Atterer, M., Baumann, T., & Schlangen, D. (2008). Towards incremental end-of-utterance detection in dialogue systems. *Proceedings of the International Conference on Computational Linguistics*, 11–14. Manchester, UK: ACM.

Auran, C., Colas, A., Portes, C., & Vion, M. (2005). Perception of breaks and discourse boundaries in spontaneous speech: Developping an on-line technique. *Proceedings of IDP05*, 1–7. Aix-en-Provence, France.

Auran, C., & Loock, R. (2006). Appositive Relative Clauses and their Prosodic Realization in Spoken Discourse: A corpus study of phonetic aspects in British English. *Proceedings of the Workshop on Constraints in Discourse*, 19–26. Maynooth, Ireland: National University of Ireland.

Auran, C., & Loock, R. (2011). The prosody of discourse functions: The case of Appositive Relative Clauses in spoken British English. *Corpus Linguistics and Linguistic Theory*, *7*(2), 181–201.

Barkhuysen, P., Krahmer, E., & Swerts, M. (2008). The interplay between the auditory and visual modality for end-of-utterance detection. *The Journal of the Acoustical Society of America*, *123*(1), 354–365.

Barth-Weingarten, D. (2016). *Intonation Units Revisited. Cesuras in talk-in-interaction*. Amsterdam: John Benjamins.

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

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30*(6), 644–663.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic Typology. The Phonology of Intonation and Phrasing* (pp. 9–54). New York: Oxford University Press.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*(1), 255–309.

Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia*, *44*(2), 178–190. https://doi.org/10.1016/j.neuropsychologia.2005.05.007

Biau, E., Fernandez, L. M., Holle, H., Avila, C., & Soto-Faraco, S. (2016). Hand gestures as visual prosody: BOLD responses to audio-visual alignment are modulated by the communicative nature of the stimuli. *NeuroImage*, *132*, 129–137.

Biau, E., Fromont, L., & Soto-Faraco, S. (2018). Beat Gestures and Syntactic Parsing: An ERP Study. *Language Learning*, *68*(1), 102–126.

Bigi, B. (2012). SPPAS: a tool for the phonetic segmentation of Speech. *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2012)*, 1748–1755. Istanbul, Turkey.

Biron, T., Baum, D., Freche, D., Matalon, N., Ehrmann, N., Weinreb, E., Biron, D., & Moses, E. (2021). Automatic detection of prosodic boundaries in spontaneous speech. *PLoS ONE*, *16*(5), e0250969. https://doi.org/10.1371/journal.pone.0250969

Blaauw, E. (1994). The Contribution of Prosodic Boundary Markers to the Perceptual Difference between Read and Spontaneous Speech. *Speech Communication*, *14*(4), 359–375.

Boersma, P., & Weenink, D. (2021). Praat: Doing Phonetics by Computer. Retrieved 30 January 2021, from http://www.fon.hum.uva.nl/praat/

Borràs-Comes, J., & Prieto, P. (2011). Seeing tunes. The role of visual gestures in tune interpretation. *Laboratory Phonology*, *2*(2), 355–380.

Brehm, L, & Alday, P. (2020). A decade of mixed models: It's past time to set your contrasts. *Proceedings of the 26th Architectures and Mechanisms for Language Processing Conference (AMLAP 2020)*. https://amlap2020.github.io/a/131.pdf

Brown, G., & Yule, G. (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.

Brugos, A., Shattuck-Hufnagel, S., & Veilleux, N. (2006). Transcribing Prosodic Structure of Spoken Utterances with ToBI. Retrieved from Online course: http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/

Buhmann, J., Caspers, J., van Heuven, V., Hoekstra, H., Martens, J.-P., & Swerts, M. (2002). Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the Spoken Dutch Corpus. *Proceedings of LREC 2002*, 779–785. Las Palmas, Spain: ELRA.

Bürkner, P. (2021). Bayesian Item Response Modeling in R with brms and Stan. *Journal of Statistical Software, 100*(5), 1–54. doi: 10.18637/jss.v100.i05.

Buxó-Lugo, A., Watson, D. G., 2016. Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language 90*, 1–13. https://doi.org/10.1016/j.jml.2016.03.001.

Calbris, G. (2011). *Elements of meaning in gesture*. Amsterdam: John Benjamins.

Carlson, R., & Swerts, M. (2003). Perceptually based prediction of upcoming prosodic breaks in spontaneous Swedish speech materials. *Proceedings of the 15th International Congress of Phonetic Sciences*, 507–510. Barcelona, Spain: International Phonetic Association.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Cassell, J., Nakano, Y. I., Bickmore, T. W., Sidner, C. L., & Rich, C. (2001). Non-verbal cues for discourse structure. *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, 114–123. Toulouse, France.

Chafe, W. (1984). How People Use Adverbial Clauses. *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society*, 437–449. Berkeley, CA, USA: Linguistic Society of America.

Chen, X., Alexopoulou, T., & Tsimpli, I. (2021). Automatic extraction of subordinate clauses and its application in second language acquisition research. *Behavior Research Methods, 53*(2), 803–817. https://doi.org/10.3758/s13428-020-01456-7

Cho, H., & Hirst, D. (2006). The contribution of silent pauses to the perception of prosodic boundaries in Korean read speech. *Proceedings of Speech Prosody 2006*. Dresden, Germany. Retrieved from http://www.isca-speech.org/archive

Choi, J.-Y., Hasegawa-Johnson, M., & Cole, J. (2005). Finding intonational boundaries using acoustic cues related to the voice source. *The Journal of the Acoustical Society of America*, *118*(4), 2579–2587.

Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics*, *37*, 871–887.

Cole, J., Mahrt, T., & Hualde, J. I. (2014). Listening for sound, listening for meaning: Task effects on prosodic transcription. *Proceedings of Speech Prosody 2014*, 859–863. Dublin, Ireland: ISCA Archive.

Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, *25*(7–9), 1141–1177.

Collier, R., de Pijper, J. R., & Sanderman, A. (1993). Perceived prosodic boundaries and their phonetic correlates. *Proceedings of the Workshop on Human Language Technology*, 341–345. Stroudsburg, PA, USA: Association for Computational Linguistics.

Couper-Kuhlen, E. (1999). Hearing and notating conversational rhythm. In Auer, P., Couper-Kuhlen, E., & Müller, F. (Eds.) *Language in Time. The Rhythm and Tempo of Spoken Interaction*, 35–55. Oxford: Oxford University Press.

Crowhurst, M. J. (2018). The influence of vowel duration and creak on the perception of internal phrase boundaries. *The Journal of the Acoustical Society of America*, *143*(3), 1–8. https://doi.org/10.1121/1.5025325

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

Cruz, M., Swerts, M., & Frota, S. (2017). The role of intonation and visual cues in the perception of sentence types: Evidence from European Portuguese varieties. *Laboratory Phonology*, *8*(1), 1–24.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

De Kok, I., & Heylen, D. (2009). Multimodal end-of-turn prediction in multi-party meetings. *Proceedings of the 2009 International Conference on Multimodal Interfaces*, 91–98. New York, USA: ACM.

de Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *The Journal of the Acoustical Society of America*, *96*(4), 2037–2047.

De Vries, M. (2006). The syntax of appositive relativization: On specifying coordination, false free relatives, and promotion. *Linguistic Inquiry*, *37*(2), 229–270.

Debreslioska, S., Özyürek, A., Gullberg, M., & Perniss, P. (2013). Gestural viewpoint signals referent accessibility. *Discourse Processes*, *50*(7), 431–456.

Dimitrova, D., Chu, M., Wang, L., Özyürek, A., & Hagoort, P. (2016). Beat that Word: How Listeners Integrate Beat Gesture and Focus in Multimodal Speech Discourse. *Journal of Cognitive Neuroscience*, *28*(9), 1255–1269.

Duez, D. (1985). Perception of silent pauses in continuous speech. *Language and Speech*, *28*, 377–389.

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture and Composite Utterances*. Cambridge: Cambridge University Press.

Engelmann, F., Granlund, S., Kolak, J., Szreder, M., Ambridge, B., Pine, J., Theakston, A., & Lieven, E. (2019). How the input shapes the acquisition of verb morphology: Elicited production and computational modelling in two highly inflected languages. *Cognitive Psychology*, *110*, 30–69.

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, *56*(3), 850–864.

eSurv (2017). http://esurv.org/

Frazier, L., Clifton Jr, C., & Carlson, K. (2004). Don't break, or do: Prosodic boundary preferences. *Lingua*, *114*(1), 3–27.

Frederiksen, A. T. (2016). Hold + Stroke Gesture Sequences as Cohesion Devices: Examples from Danish Narratives. *San Diego Linguistics Papers*, *6*, 2–13.

Fritz, I., Kita, S., Littlemore, J., & A. Krott. (2019). Information packaging in speech shape information packaging in gesture: the role of speech planning units in the coordination of speech-gesture production. *Journal of Memory and Language*, *104*, 56–69.

Fromont, L. A., Soto-Faraco, S., & Biau, E. (2017). Searching high and low: Prosodic breaks disambiguate relative clauses. *Frontiers in Psychology [Online]*, *8*. https://doi.org/10.3389/fpsyg.2017.00096

Granström, B., & House, D. (2005). Audiovisual representation of prosody in expressive speech communication. *Speech Communication*, *46*(3), 473–484.

Granström, B., House, D., & Lundeberg, M. (1999). Prosodic cues in multimodal speech perception. *Proceedings of the International Congress of Phonetic Sciences (ICPhS99)*, 655–658. San Francisco, USA.

Grover, C., Facrell, J., Vereecken, H., Martens, J. P., & Van Coile, B. (1998). Designing prosodic databases for automatic modeling in 6 languages. *Proceedings of the 3rd ESCA/COCOSDA Workshop on Speech Synthesis*, 93–98. Jenolan Caves, Australia: ESCA/COCOSDA.

Hadar, U., Steiner, T. J., Grant, E. C., & Clifford Rose, F. (1983). Head movement correlates of juncture and stress at sentence level. *Language and Speech*, *26*(2), 117–129.

Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.

Henry, L. (2020). ggstance: Horizontal ggplot2 components. https://github.com/lionel-/ggstance

Hilton, M., Räling, R., Wartenburger, I., & Elsner, B. (2019). Parallels in processing boundary cues in speech and action. *Frontiers in Psychology*. 10:1566. https://doi.org/10.3389/fpsyg.2019.01566

Hirst, D. J. (2007). A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation. *Proceedings of the XVIth International Conference of Phonetic Sciences*, 1233–1236. Retrieved from http://fedora.tge-adonis.fr:8090/fedora/get/CRDO-Aix:234079/DEPOT_DESC_2068.pdf

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., & Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language*, *79*, 1–17.

Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., & Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Frontiers in Psychology*, *3*(74), 1–12.

Holler, J., & Bavelas, J. B. (2017). Multi-modal communication of common ground. In *Why Gesture? How the hands function in speaking, thinking, and communicating* (pp. 213–240). Amsterdam: John Benjamins.

House, D., Beskow, J., & Granström, B. (2001). Timing and interaction of visual cues for prominence in audiovisual speech perception. *Proceedings of Eurospeech*. Aalborg, Denmark. Retrieved from http://perso.telecom-paristech.fr/~chollet/Biblio/Congres/Audio/Eurospeech01/CDROM/papers/page387.pdf

Huddleston, R., & Pullum, G. K. (2006). Coordination and Subordination. In B. Aarts & A. McMahon (Eds.), *The Handbook of English Linguistics* (pp. 198–219). Oxford: Blackwell.

Jiménez-Bravo, M., & Marrero-Aguiar, V. (2020). Multimodal perception of prominence in spontaneous speech: a methodological proposal using mixed models and AIC. *Speech Communication*, *124*, 28–45. https://doi.org/10.1016/j.specom.2020.07.006

Kakouros, S., & Räsänen, O. (2016a). Perception of sentence stress in speech correlates with the temporal unpredictability of prosodic features. *Cognitive Science*, *40*(7), 1739–1774.

Kakouros, S., & Räsänen, O. (2016b). Statistical Learning of Prosodic Patterns and Reversal of Perceptual Cues for Sentence Prominence. *Proceedings of the 38th Annual Conference*

of the Cognitive Science Society, 1–6. Retrieved from http://users.spa.aalto.fi/orasanen/papers/cogsci_stat_learning_prosody.pdf

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260–267.

Kendon, A. (1983). Gesture and speech: How they interact. In J. M. Wiemann & R. P. Harrison (Eds.), *Nonverbal interaction* (pp. 13–45). Beverly Hills, USA: Sage Publications.

Kim, H., Yoon, T-J., Cole, J., & Hasegawa-Johnson, M. (2006). Acoustic differentiation of L- and L-L% in switchboard and radio news speech. *Proceedings of Speech Prosody 2006*, paper 214. Dresden, Germany: ISCA. https://www.isca-speech.org/archive_v0/sp2006/sp06_214.html

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32.

Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes*, *22*(8), 1212–1236.

Krahmer, E., & Swerts, M. (2007). Perceiving focus. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and Focus* (pp. 121–137). Dordrecht: Springer.

Kreiman, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics*, *10*(2), 163–175.

Krifka, M. (2007). Basic Notions of Information Structure. *Interdisciplinary Studies on Information Structure*, *6*, 13–55.

Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical Transactions of the Royal Society B.*, *369*(20130397). http://dx.doi.org/10.1098/rstb.2013.0397

Kuang, J., Pik Yu chan, M., & Rhee, N. (2022). The effects of syntactic and acoustic cues on the perception of prosodic boundaries. *Proceedings of Speech Prosody 2022*, 699–703. Lisbon, Portugal: ISCA. doi: 10.21437/SpeechProsody.2022-142

Ladd, R. D. (2008). *Intonational phonology*. Cambridge: Cambridge University Press.

Lascarides, A., & Stone, M. (2009). Discourse coherence and gesture interpretation. *Gesture*, *9*(2), 147–180.

Lee, E. K., Cole, J., & Kim, H. (2006). Additive effects of phrase boundary on English accented vowels. *Proceedings of Speech Prosody 2006*. Dresden, Germany.

Lehiste, I. (1979). Perception of sentence and paragraph boundaries. In B. Lindblom & S. Ohman (Eds.), *Frontiers of Speech Communication Research* (pp. 191–201). New York: Academic Press.

Lelandais, M. (2020). Modelling the interpretative impact of subordinate constructions in spontaneous conversation. *CORELA (Cognition, Representation, Language)*, *18*(2), 12827. https://doi.org/10.4000/corela.12827

Lelandais, M., & Ferré, G. (2016). Prosodic boundaries in subordinate syntactic constructions. *Proceedings of Speech Prosody 2016*, 183–187. Boston, USA: ISCA. http://dx.doi.org/10.21437/SpeechProsody.2016-38

Lewandoswki, W., & Özçalışkan, S. (2018). How event perspective influences speech and co-speech gestures about motion. *Journal of Pragmatics*, *128*, 22–29.

Li, W.-J., & Yang, Y. (2009). Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. *Neuroscience*, *158*(4), 1416–1425.

Loehr, D. P. (2004). *Gesture and intonation*. PhD Thesis, Georgetown University. Retrieved from http://www9.georgetown.edu/faculty/loehrd/pubs_files/loehr04.pdf

Loock, R. (2007). Appositive relative clauses and their functions in discourse. *Journal of Pragmatics*, *39*(2), 336–362.

Lüdecke, D. (2021). sjPlot: Data visualization for statistics in social science. https://cran.r-project.org/package=sjPlot

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2016). Imposing Cognitive Constraints on Reference Production: The Interplay Between Speech and Gesture During Grounding. *Topics in Cognitive Science*, 1–18. https://doi.org/doi:10.1111/tops.12217

Matthiessen, C., & Thompson, S. A. (1988). The structure of discourse and ``subordination''. In J. Haiman & S. A. Thompson (Eds.), *Clause Combining in Grammar and Discourse* (pp. 275–329). Amsterdam: John Benjamins.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

McNeill, D., & Duncan, S. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and Gesture* (Vol. 2, pp. 141–161). Cambridge: Cambridge University Press.

McNeill, D., & Levy, E. T. (1993). Cohesion and gesture. *Discourse Processes*, *16*(4), 363–386.

Mendoza-Denton, N., & Jannedy, S. (2011). Semiotic layering through gesture and intonation: A case study of complementary and supplementary multimodality in political speech. *Journal of English Linguistics*, *39*(3), 265–299.

Mertens, P., & Simon, A.-C. (2013). Exploring acoustic and syntactic cues to prosodic boundaries in French. A multi-genre corpus study. *Proceedings of ICPhS 2013*, 81–87. Glasgow, UK: IPA Archive.

Mettouchi, A., Lacheret-Dujour, A., Silber-Varod, V., & Izre'el, S. (2007). Only Prosody? Perception of speech segmentation in Kabyle and Hebrew. *Nouveaux Cahiers de Linguistique Française*, *28*, 207–218.

Meyer, L., Obleser, J., Kiebel, S. J., & Friederici, A. D. (2012). Spatiotemporal dynamics of argument retrieval and reordering: An fMRI and EEG study on sentence processing. *Frontiers in Psychology*, *3*, [Online]. https://doi.org/10.3389/fpsyg.2012.00523

Mo, Y. (2008). Duration and intensity as perceptual cues for naive listeners' prominence and boundary perception. *Proceedings of Speech Prosody 2008*, 739–742. Campinas, Brazil: ISCA.

Mo, Y., & Cole, J. (2010). Perception of prosodic boundaries in spontaneous speech with and without silent pauses. *Journal of the Acoustical Society of America*, *127*(3), 1956.

Mo, Y., Cole, J., & Lee, E.-K. (2008). Naive listeners' prominence and boundary perception. *Proceedings of Speech Prosody 2008*, 735–738. Campinas, Brazil: ISCA.

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, *23*(1), 103–123. https://doi.org/10.3758/s13423-015-0947-8

Oben, B., & Brône, G. (2015). What you see is what you do: On the relationship between gaze and gesture in multimodal alignment. *Language and Cognition*, *7*(04), 546–562.

Oben, B., & Brône, G. (2016). Explaining interactive alignment: A multimodal and multifactorial account. *Journal of Pragmatics*, *104*, 32–51.

Özyürek, A., Kita, S., Allen, S. E. M., Furman, R., & Brown, A. (2005). How does linguistic framing of events influence co-speech gestures?: Insights from crosslinguistic variations and similarities. *Gesture*, *5*(1–2), 1–2.

Pagel, V., Carbonell, N., Laprie, Y., & Vaissière, J. (1995). Spotting prosodic boundaries in continuous speech in French. *Proceedings of the XIIIth ICPhS*, 308–311. Stockholm, Sweden.

Parrill, F., & Kimbara, I. (2006). Seeing and hearing double: The influence of mimicry in speech and gesture on observers. *Journal of Nonverbal Behavior*, *30*(4), 157–167.

Perniss, P., & Özyürek, A. (2015). Visible Cohesion: A Comparison of Reference Tracking in Sign, Speech, and Co-Speech Gesture. *Topics in Cognitive Science*, *7*(1), 36–60.

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.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. PhD Thesis, Massachusetts Institute of Technology.

Pintér, G., Mizuguchi, S., & Tateishi, K. (2014). Perception of prosodic prominence and boundaries by L1 and L2 speakers of English. *Proceedings of Interspeech 2014*, 544–547. Singapore: ISCA.

Portes, C. (2002). Approche instrumentale et cognitive de la prosodie du discours en français. *Travaux Interdisciplinaires Du Laboratoire Parole et Langage d'Aix-En-Provence (TIPA)*, *21*, 101–119.

Pozniak, C., & Burnett, H. (2021). Failures of Gricean reasoning and the role of stereotypes in the production of gender marking in French. *Glossa: a journal of general linguistics*, *6*(1). https://doi.org/10.5334/gjgl.1310

Priva, U. C. (2017). Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition*, *160*, 27–34.

Pynte, J. (2006). Phrasing effects in comprehending pp constructions. *Journal of Psycholinguistic Research*, *35*(3), 245–265.

R Core Team. (2022). A language and environment for statistical computing. Retrieved 29 June 2022, from R foundation for statistical computing website: http://www.r-project.org/

Rienks, R. J., Poppe, R., & Heylen, D. (2010). Differences in head orientation behavior for speakers and listeners: An experiment in a virtual environment. *Transactions on Applied Perception*, *7*(1). https://doi.org/10.1145/1658349.1658351

Rietveld, A. C., & Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, *13*, 299–308.

Roux, G., Bertrand, R., Ghio, A., & Astésano, C. (2016). Naive listeners' perception of prominence and and boundary in French spontaneous speech. *Proceedings of Speech Prosody*, 912–916. Boston, MA, USA: ISCA.

Roy, J., Cole, J., & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology*, *8*(1), 1–36. https://doi.org/10.5334/labphon.108

Schlangen, D. (2006). From reaction to prediction: Experiments with computational models of turn-taking. *Proceedings of INTERSPEECH 2006*. Pittsburgh, USA: ISCA.

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *The Journal of the Acoustical Society of America, 71*(4), 996–1007.

Silverman, K. E. A., Beckman, M. B., Pitrelli, J. F., Ostendorf, M., Wightman, C. W., Price, P., Hirschberg, J. (1992). TOBI: a standard for labeling English prosody. *Proceedings of the 2nd International Conference on Spoken Language Processing (ICSLP 92)*, *2*, 867–870.

Simon, A.-C., & Christodoulides, G. (2016). Perception of Prosodic Boundaries by Naive Listeners in French. *Proceedings of Speech Prosody*, 1158–1162. Boston, MA, USA: ISCA.

Smith, C. L. (2009). Naive listeners' perceptions of French prosody compared to the predictions of theoretical models. In H.-Y. Yoo & E. Delais-Roussarie (Eds.), *Proceedings of IDP09* (pp. 335–349).

Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology*, *12*(3), 175–200. http://dx.doi.org/10.20982/tqmp.12.3.p175

Stan Development Team. (2017). shinystan: Interactive visual and numerical diagnostics and posterior analysis for Bayesian models. Retrieved 29 June 2022 from http://mc-stan.org/

Streeck, J. (2009). Gesturecraft. The manu-facture of meaning. Amsterdam: John Benjamins.

Streefkerk, B. M., Pols, L. C., & ten Bosch, L. F. (1997). Prominence In Read Aloud Sentences, As Marked By Listeners And Classified Automatically. *Proceedings of IFA 21*, 101–116. Amsterdam, Netherlands: Institute of Phonetic Science.

Svartvik, J., & Quirk, R. (1980). *A Corpus of English Conversation*. Lund, Sweden: Lund University Press.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America, 101*(1), 514–521.

Swerts, M., & Krahmer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics*, *36*(2), 219–238.

't Hart, J. (1981). Differential sensitivity to pitch distance, particularly in speech. *The Journal of the Acoustical Society of America*, *69*, 811–821.

't Hart, J., Collier, R., & Cohen, A. (1990). *A Perceptual study of Intonation*. Cambridge: Cambridge University Press.

Thompson, S. A. (2002). ``Object complements'' and conversation: Towards a realistic account. *Studies in Language*, *26*(1), 125–163.

Turk, A. & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, *35*(4), 445–472.

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

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, *57*, 209–232. https://doi.org/10.1016/j.specom.2013.09.008

Watson, D., & Gibson, E. (2005). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, *19*(6), 713–755.

Watson, D., & Gibson, E. (2004). Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica*, *59*(2-3), 279–300.

Wells, J. C. (2006). *English Intonation: An Introduction*. Cambridge: Cambridge University Press.

Wickham, H., François, R., Henry, L., & Müller, K. (2022). dplyr: A grammar of data manipulation. https://dplyr.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. New York: Springer Verlag. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org/

Wightman, C., Shattuck-Hufnagel, S., Otsendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of Acoustical Society of America*, *91*, 1707–1717.

Wilke, O. (2021). ggridges: Ridgeline plots in ggplot2. https://cloud.r-project.org/package=ggridges

Winter, B., Duffy, S. E., & Littlemore, J. (2020). Power, gender and individual differences in spatial metaphor: The role of perceptual stereotypes and language statistics. *Metaphor and Symbol*, *35*(3), 188–205. https://doi.org/10.1080/10926488.2020.1794319

Yang, Y., & Wang, B. (2002). Acoustic correlates of hierarchical prosodic boundary in Mandarin. *Proceedings of Speech Prosody*. Aix-en-Provence, France: ISCA Archive.

Yoon, T.-J., Chavarria, S., Cole, J., & Hasegawa-Johnson, M. (2004). Intertranscriber Reliability of Prosodic Labeling on Telephone Conversation using ToBI. *Proceedings of INTERSPEECH 2004*, 2722–2732. Jeju, South Korea.

Yoon, T.-J., Cole, J., & Hasegawa-Johnson, M. (2007). On the edge: Acoustic cues to layered prosodic domains. *Proceedings of ICPhS XVI*, 1264–1267. Saarbrücken, Germany.