

The Prosody of Discourse Markers *alors* and *et* in French

A Speech Production Study

Ivana Didirková¹, George Christodoulides¹, Anne Catherine Simon¹

¹Centre Valibel, IL&C, Université catholique de Louvain, Belgium

ivana.didirkova@uclouvain.be, george@mycontent.gr, anne-catherine.simon@uclouvain.be

Abstract

We study the prosodic features of two French discourse markers (DMs), *alors* and *et*, in connection with the discourse relation they convey. Twenty adult native speakers of French were asked to prepare and to read aloud 64 sequences (32 per DM) consisting of a first segment, the target DM and a second segment; all first segments were extracted from a natural speech corpus. The sequences were constructed in order to convey one of six pre-defined discourse relations. The prosodic characteristics of the resulting 1280 recorded utterances are analysed based on the DM and the discourse relation. Results suggest that the silent pause duration before the DM, as well as the absolute duration of the DM itself are used by the speaker to differentiate between the core meaning of the DM and its less predictable meanings. Moreover, prosodic cues were not used redundantly and the DMs did not systematically constitute a separate prosodic unit. **Index Terms:** prosody, discourse markers, discourse relations, speech production

1. Introduction

The present study is part of a research project aiming at establishing the role of prosody in the online processes of discourse comprehension, and more specifically the identification of discourse units and the construal of discourse relations. Although the role of prosodic phrasing has been acknowledged as central in language comprehension ([1], [2], [3]), studies in speech production and perception that investigate the role of prosody in identifying discourse relations are still scarce. The overarching research question is whether specific prosodic structures cue specific discourse relations, and whether prosody can facilitate the processing of discourse relations.

We know that prosodic features such as f0, final lengthening and pause duration help the listener reconstruct discourse structure and better understand the message (e.g. [4]). It has also been shown that other linguistic elements play this role and among them discourse markers (henceforth DMs) are particularly important. According to [5], DMs can be defined as “sequentially dependent elements which bracket units of talk”. Discourse markers fulfil various functions, including signalling discourse relations. Words such as *et* (and), *donc* (so), *maintenant* (now) and *alors* (then) can be used as DMs in French.

In this paper, we present a speech production study in which native speakers were asked to read aloud utterances and convey six different discourse relations (addition, temporality, specification, consequence, topic change and concession) between two segments (henceforth S1 and S2) connected with one of two French DMs, *et* (and) and *alors* (then). Previous research on French DMs has shown that these two markers may express several different relations and meanings depending on the context. We investigate whether there are differences in the prosodic features of the utterances produced, depending on the DM and the

intended discourse relation.

1.1. The Prosody of Discourse Markers

Previous studies on the use of discourse markers in speech have investigated their phonetic and prosodic properties. In a study on the DM *hā* (yes) in Hindi, [6] measured the f0 movement on the DM and found that intonation patterns play an important role in cueing different discourse functions: back-channeling, interjection, literal meaning, the beginning of a new topic, or the end of the current topic. Similarly, [7] confirm the importance of intonation in interpreting the Swedish DM *men* (but/and/so), and in choosing between understanding it as a DM and its sentential interpretation. They show that when the token *men* is used as a discourse marker it has a positive f0 reset, with a mean value of 13.8 ST when preceded by a glottalisation, and of 5.7 without glottalisation; whereas in the case of sentential tokens, the mean value of the f0 reset was 2.2 ST.

In English, [8] claimed that the DMs constitute a separate prosodic unit surrounded by brief pauses, and that this configuration helps distinguish between DMs and other uses of the same token. This claim is attenuated by [9] who show that DMs only form a separate intonation unit when opening/closing a conversation or when marking transitions from one topic to another. In this case, the observed melodic pattern is a falling tone; when the DM is used to refer to knowledge shared between the speaker and the listener, other tones are possible depending on the expected listener’s reaction. According to [9] the intonation of DMs depends on the speaker’s perception of how important a particular marker is, and therefore the relationship between the function of a DM, its prosodic characteristics and its position in the utterance is rather arbitrary. Similar results can be found in [7], who report that only 34% of the discourse markers studied form a separate intonation unit, and that this is never the case for sentential tokens. A different result can be found in [10], who show that the Czech DM *jasně* (of course) has a stable prosodic realisation regardless of its different discursive functions.

Besides f0, other prosodic features have been studied, specifically aiming to distinguish between the sentential and the discourse use of discourse particles. For example, [7] found that the absolute duration of tokens used as DMs is longer compared to their sentential use. Their lexical analysis of the word class of items following the DM also indicates a difference between the two uses: while discourse tokens are often followed by conjunctions, pause fillers and particles, the sentential tokens are mainly followed by pronouns.

1.2. The Prosody of Discourse Relations

Other studies focus on the prosodic marking of discourse relations, trying to find prosodic features that would help clarify and/or classify various discourse or rhetorical relations. For example, [11] conclude that prosody alone is sufficient to indicate

the rhetorical relations between dialogue acts. They demonstrate a classification of rhetorical relations based on an analysis of pitch contour, pitch variance, relative energy, speech rate, number of silent frames and duration. For example, in the case of a causal rhetorical relationship, the nucleus has a long duration, numerous silences and flat intonation, and the satellite has a long duration, low speech rate and a high maximum f0; while in the case of a contrastive rhetorical relationship, there is little or no pause between the nucleus and the satellite, and the satellite has a high speech rate and an increased energy.

In their study on correlation between the discourse structure of a spoken monologue and its prosody, [12] analysed f0, intensity, speech rate, pause duration and inter- and intra-segmental features of discourse relations, following the typology of relations of Rhetorical Structure Theory [13]. This study is more nuanced than [11], claiming that only some rhetorical relations can be identified solely on the basis of prosodic features. Finally, [14] shows that speakers read aloud discourse segments that have a causal relationship with shorter pause between them and at faster rate, than segments that are not related.

1.3. Objective and Hypotheses

The main purpose of this study is to measure to what extent prosodic cues are used by native speakers of French to convey discourse relations. Results will be used for a subsequent perception experiment where those cues will be manipulated in order to measure to what extent they influence or facilitate the choice of one specific discourse relation over another.

Based on previous studies, we expect differences in pause duration, f0 contour and f0 reset, and speech rate, according to the different discourse relation under investigation. More precisely, we hypothesise that:

- discourse relations strongly corresponding to the core meaning of the DM (for example addition / temporal relation for *et* (and), or consequence for *alors* (then)) will be made less prominent by prosody, i.e. shorter silent pause, regular speech rate without decelerations, and no melodic reset on the DM;
- discourse relations that are less often introduced by the DM in question (like concession with *et* and *alors*) or that are of a more global nature, in the sense that they indicate textual structure (like topic shift with *alors*), will be prosodically marked, to help the hearer understand them in context. In these cases we expect a longer silent pause before (or after) the DM, a pitch reset on the DM, a more prominent realization of the DM and a slower articulation rate.

2. Method

2.1. Corpus and Stimuli Preparation

The instances of the target words *alors* and *et* were embedded in utterances containing a first discourse segment (S1), a DM (*alors* or *et*) and a second segment (S2). All S1s and some of the S2s were extracted from natural spoken data (from the LOCAS-F corpus, [15]). In this corpus, discourse relations have been previously annotated as follows: two experts annotated original sequences containing the S1, the corresponding DM and the S2 for discourse relation. Experts reached consensus after discussion on disagreements. Two DMs and six different discourse relations were selected for this study (see [16] for more details on the definition of these discourse relations): consequence and

specification (introduced by both *et* and *alors*), topic shift and concession (for *alors* only), and addition and temporality (for *et* only); each DM can convey four discourse relations.

Table 1: Pairs of discourse relations for DMs *et* and *alors*

DM	Discourse Relation
<i>et</i> (and)	temporality – addition consequence – specification
<i>alors</i> (then)	concession – specification consequence – topic shift

For each original S1, three additional S2s were then constructed in order to represent the three additional discourse relations; the constructed S2s were controlled for syntactic structure. This procedure resulted in 176 utterances (S1+DM+S2) which were then submitted for annotation to 44 naïve subjects, following the procedure reported in [17]. Participants were asked to identify the discourse relation for every sequence using a multiple-choice questionnaire. Based on the results of this annotation experiment, we paired discourse relations that most frequently overlapped. More precisely, we extracted regularities in the identification of discourse relations when a disagreement had been observed between the naïve and the expert annotators, and then paired discourse relations accordingly. For *et* the temporality relation was paired with addition, and the consequence relation with specification; for *alors*, the concession relation was paired with specification and the consequence relation with topic shift.

Among these pairs, we selected 64 sequences (32 for *alors* and 32 for *et*) so that every discourse relation was represented 8 times for each DM. For every discourse relation, two sequences would have a high inter-annotator agreement expressed in percentages ranging from 72.72% up to 100%. For two other sequences, there was no consensus on the annotation of the discourse relation. The last four occurrences for the same discourse relation were not controlled for agreement, as they were part of the pair of discourse relations as described above. The aim of this selection was to cover, at the same time, utterances where the discourse relation seemed to be easy to identify based on written stimuli and utterances where the discourse relation was not easy to identify. This will allow us to explore to what extent prosodic cues contribute to the disambiguation of discourse relation in these stimuli.

2.2. Speech Elicitation

The 64 selected sequences were read aloud by 20 adult native speakers of French (13 female and 7 male). Participants were recruited via the Participants' Pool of the Faculty of Psychology of Université catholique de Louvain in Belgium (via Facebook), and were remunerated for their participation.

Prior to the recording session, participants were handed a document containing the 64 sequences along with a document explaining and illustrating the six discourse relations. They were asked to prepare for the recording by silently reading the sequences, in order to be able to produce a fluent, natural-like rendition conveying the targeted discourse relation. Recordings took place in a quiet room. Sequences to read were presented on a computer screen, grouped by DM and discourse relation. Each discourse relation was explained again at the beginning of the corresponding group of sequences, and then the utterances were presented on the screen, one by one. Participants were

free to prepare each utterance without any time pressure. When they felt ready, they had to press a button to start the recording of the presented utterance. They were free to re-record each utterance as many times as necessary, to their satisfaction. A new recording was mandatory in cases where speakers produced disfluencies, omitted or added an element, or modified the original sequence in any other way. Speech was recorded using a DPA 4066-B34 omnidirectional headset microphone, connected to a Boss BR-800 digital multi-track recorder.

2.3. Data Analysis

The 1280 recorded utterances were analysed using *Praaline* [18]. They were added to a corpus, along with the corresponding speaker and recording metadata, and each utterance was linked to its transcription using a script (the custom software used for the recording followed a coding scheme to facilitate this process). The transcription was automatically aligned to the phone, syllable and token level, using the Forced Alignment feature of *Praaline* (short silent pauses were also automatically detected as part of the forced alignment process; the precision of this detection was satisfactory because the recordings were of high quality). The phonetic alignment around the DM (approximately one intonation unit before and after the target DM) was then manually verified. Pitch stylisation was performed using *Prosogram* [19]. An automatic annotation of prosodic prominence and prosodic boundaries was performed using *Promise* [20]. Features extracted using these plug-ins are stored in an SQL database, and include durations (of pauses, segments, syllables etc), pitch information (e.g. intonation contour descriptors), and symbolic annotations (e.g. prominences and boundaries). The database from *Praaline* was linked to the R statistical software [21] for analysis.

3. Results

The differences in the prosodic features of utterances where the same DM is used to signal different discourse relations were less clear-cut and less pronounced than our initial hypotheses. However, there are interesting results related to pause duration and pitch reset that will be presented in the following sections.

3.1. Silent Pauses, Syllable Duration, Articulation Rate

In our data, the DM is almost never followed by a silent pause: there is no silent pause between the DM and the S2 it introduced. However, the duration of the silent pause before the DM varies according to the DM and the discourse relation it conveys. In general, the silent pause duration before *alors* is longer than before *et* (see Figure 1).

Furthermore, silent pause duration increases for non-typical discourse relations (i.e. not semantically coded by the DM). With regards to the DM *alors* (then), we observe that subjects insert a longer silent pause before the DM when it introduces a segment opening a new topic (topic change, CHN) and, to a lesser extent, when the DM introduces a specification (SPE). This is not the case when *alors* introduces a consequence (CSQ) which is the typical relation for this DM since it corresponds to its core meaning, or a concession (CCS, negative consequence).

With regards to the DM *et* (and), we observe a slightly longer pause before a specification (SPE) or a temporal (TMP) S2 segment, and it is not the case when S2 is interpreted as a consequence (CSQ) or as an addition (ADD).

The differences observed as for silent pause duration are confirmed by the duration of the DM itself and by its de-

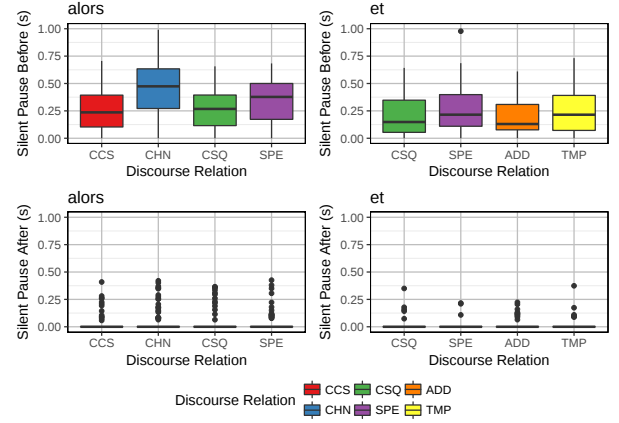


Figure 1: Pause duration before (top) and after (bottom) the DM, for *alors* (left) and *et* (right).

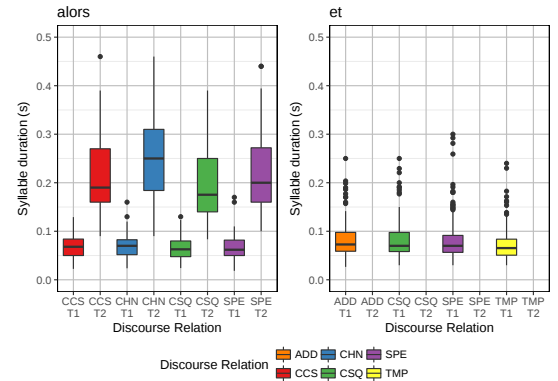


Figure 2: Duration of the DM, for disyllabic *alors* (left) and monosyllabic *et* (right). T1 and T2 are the first and second syllables of the target DM respectively.

gree of prominence (see Figure 2). The DM *alors* is significantly lengthened, and prominent, when signalling topic change (CHN), and slightly lengthened for specification (SPE). On the other hand, we observe no variation for *et* in any of the 4 discourse relations.

Articulation rate does not significantly vary depending on the discourse relation, neither on the S1, nor on the S2 segment. There is a slight trend to reduce the speech rate on S2 for topic change or specification with *alors*, and on consequence or addition with *et*.

3.2. Prosodic Boundary on S1 and Pitch Reset on the DM

Prosodic boundaries and pitch resets are prosodic cues signalling a clear segmentation between the end of S1 and the subsequent DM introducing S2. We measured whether there was a difference in pitch between the last syllable of S1 and the onset of the DM, and between the last syllable of the DM and the onset of S2 (see Figure 3).

As for *alors*, although differences in pitch are reduced, results confirm the trend for specification (SPE) and topic change (CHN) to be marked by a pitch reset on the DM, while concession (CCS) and consequence (CSQ) are not. Interestingly, we also observe a pitch reset on *et* when it introduces a specification (SPE); on the other hand, *et* introducing a consequence

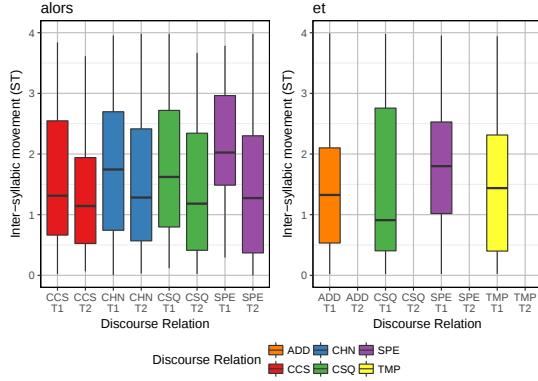


Figure 3: Pitch reset between the end of S1 and the DM *alors* (left) or *et* (right). T1 and T2 are the first and second syllables of the target DM respectively.

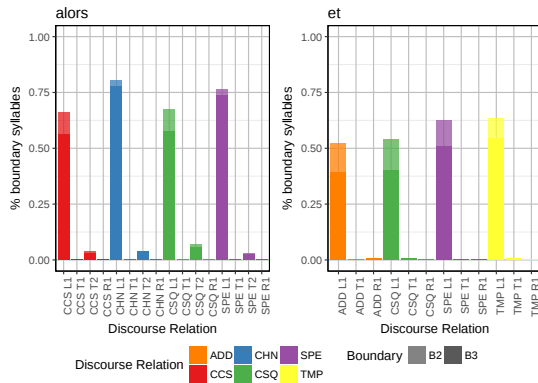


Figure 4: Prosodic boundary (percentage) at the end of S1, before *alors* (left) or *et* (right). B3 = major prosodic boundary and B2 = medium prosodic boundary. L1 is the first syllable to the left of the target DM, R1 is the first syllable to the right of the target DM etc.

(CSQ) is melodically strongly connected to the preceding S1 segment. The pitch profile for additive (ADD) and temporal (TMP) relations is in-between.

We do not observe any pitch reset between the DM and S2, i.e. the DM is prosodically integrated to the utterance it introduces. This is confirmed by the absence of a prosodic boundary before the DM (see Figure 4, where T1 and T2 are the first and second syllables of the DM respectively).

However, we observe that the DM is frequently preceded by a strong prosodic boundary, and this is even more frequent before *alors* than before *et*. As expected, the highest frequency of a major prosodic boundary occurs before *alors* introducing a topic change (CHN) or a specification (SPE), and the lower frequency of a prosodic boundary occurs before *et* introducing addition (ADD) or consequence (CSQ).

4. Discussion and Conclusion

4.1. Prosodic Cues of Discourse Relations

We can synthesise the observations on the prosodic cues presented in the previous sections, and interpret them in terms of “prosodic segmentation” (long silent pause, prosodic boundary,

lengthened DM, pitch reset, etc.) vs “prosodic integration” (no pause, less frequent prosodic boundary, no lengthening, etc.). The result is that an S2 segment introduced by *alors* is more detached from the preceding S1 segment, than an S2 segment introduced by *et*. Table 2 shows the conclusions for each combination of DM and discourse relation.

Table 2: Synthesis of produced prosodic cues depending on the DM and on the discourse relation

<i>alors</i> (then)	CHN > SPE CSQ = CCS	Prosodic segmentation (marked) In-between
<i>et</i> (and)	SPE = TMP CSQ = ADD	In-between Prosodic integration (unmarked)

In sum, our experiment and the speech elicited shows that the strongest prosodic segmentation, the less predictable the discourse relation signalled by the DM. This is mostly the case with *alors* (then) indicating topic change and specification, and to a lesser extent with *et* (and) introducing specification and temporal relation. On the other hand, *alors* introducing consequence or specification, and *et* introducing consequence or addition, are not prosodically salient; the introduced discourse segments remain prosodically integrated. The fact that a DM introducing consequence is never “marked” by salient prosodic cues suggests that consequence is a basic, even “default”, discourse relation.

Moreover, our results suggest that prosodic cues are not used redundantly (silent pause *and* lengthening *and* melodic reset) but rather combine purposefully (silent pause *or* lengthening *or* melodic reset). A more detailed analysis cannot be provided due to lack of space, but the interested reader may refer to the supplementary material to this article.

4.2. Methodological Issues of Speech Elicitation

Finally, some observations regarding the methodological constraints of speech elicitation studies are in order. We have attempted to keep the utterances produced as close to natural speech as possible by selecting material from a natural speech corpus (S1 segments and some of the S2 segments). However, given that potentially confounding variables need to be controlled in an experimental protocol, the resulting utterances were produced out of context and in somewhat artificial conditions. Confronted with this problem, we will take the position of [22] that we need “a stronger methodological awareness in investigations of speech phenomena and more cautious interpretations of the findings that we make [as well as] a much better understanding of the extent to which our methods and our ways of collecting speech data influence our results”.

5. Acknowledgements

This research is supported by the Belgian F.R.S./FNRS (MIS Project PhraDiCo: Prosodic phrasing for discourse comprehension. Perception of prosodic boundaries and their contribution to the construal of discourse coherence, convention No. F 4520.16, PI: Anne Catherine Simon).

6. References

- [1] A. Cutler, D. Dahan, and W. van Donselaar, "Prosody in the comprehension of spoken language: A literature review," *Language and Speech*, vol. 40, no. 2, pp. 141–201, January 1997.
- [2] C. Féry, *Intonation and prosodic structure*, ser. Key topics in phonology. Cambridge: Cambridge University Press, 2017.
- [3] L. Frazier, K. Carlson, and J. Clifton, "Prosodic phrasing is central to language comprehension," *Trends in Cognitive Sciences*, vol. 10, no. 6, pp. 244–249, 2006.
- [4] M. Horne, P. Hansson, G. Bruce, J. Frid, and M. Filipsson, "Cue words and the topic structure of spoken discourse: The case of Swedish men 'but'," *Journal of Pragmatics*, vol. 33, no. 7, pp. 1061–1081, 2001.
- [5] D. Schiffrin, *Discourse markers*. Cambridge: Cambridge University Press, 1987.
- [6] K. Bali, "F0 cues for the discourse functions of 'hā' in Hindi," in *Interspeech 2009 – 10th Annual Conference of the International Speech Communication Association, September 6-10, Brighton, United Kingdom, Proceedings*, 2009.
- [7] M. Horne, P. Hansson, G. Bruce, J. Frid, and M. Filipsson, "Discourse markers and the segmentation of spontaneous speech: The case of Swedish men 'but/and/so'," *Working Papers, Lund University, Dept. of Linguistics*, vol. 47, pp. 123–139, 1999.
- [8] R. Carter and M. McCarthy, *Cambridge grammar of English: A comprehensive guide: spoken and written English grammar and usage*. Cambridge: Cambridge University Press, 2006.
- [9] S. Komar, "The interface between intonation and function of discourse markers in English," *English Language Overseas Perspectives and Enquiries (ELOPE)*, vol. 4, no. 1-2, p. 43, 2007.
- [10] J. Volin, L. Weingartová, and O. Niebuhr, "The prosody of the Czech discourse marker 'jasně': An analysis of forms and functions," *Phonetica*, vol. 73, no. 3-4, pp. 314–337, 2016.
- [11] G. Murray, S. Renals, J. Carletta, and J. Moore, "Incorporating speaker and discourse features into speech summarization," R. C. Moore, Ed. Morristown, NJ, USA: Association for Computational Linguistics, 2006, pp. 367–374.
- [12] J. Kleinhans, M. Farrús, A. Gravano, J. M. Pérez, C. Lai, and L. Wanner, "Using prosody to classify discourse relations," in *Interspeech 2017 – 18th Annual Conference of the International Speech Communication Association, August 20–24, Stockholm, Sweden, Proceedings*, 2017, pp. 3201–3205.
- [13] W. C. Mann and S. A. Thompson, "Rhetorical structure theory: Toward a functional theory of text organization," *Text - Interdisciplinary Journal for the Study of Discourse*, vol. 8, no. 3, 1988.
- [14] H. d. Ouden, L. Noordman, and J. Terken, "Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports," *Speech Communication*, vol. 51, no. 2, pp. 116–129, 2009.
- [15] L. Degand, L. J. Martin, and A. C. Simon, "Unités discursives de base et leur priphérie gauche dans LOCAS-F, un corpus oral multigenres annoté," in *Actes du 4^{ème} Congrès Mondial de Linguistique Française*. Berlin, Germany: EDP SciencesC, 2014.
- [16] L. Degand and A. C. Simon, "Variation of discourse markers across a multi-genre corpus of spoken French: Annotating function and meaning," in *Presentation at Variation et changement pragmatique-discursif 3 (DiPVaC), Toronto, 4-6 May*, 2016.
- [17] I. Didirková, A. C. Simon, and G. Christodoulides, "Same discourse markers, different discourse relations: What do we learn from naïve annotation experiments?" in *Presentation at the Dis-Rom 2017 Conference, Louvain-la-Neuve, Belgium*, 2017.
- [18] G. Christodoulides, "Praaline: integrating tools for speech corpus research," in *LREC 2014 – 9th International Conference on Language Resources and Evaluation, May 26–31, Reykjavik, Iceland, Proceedings*, 2014, pp. 31–34, <http://www.praaline.org>.
- [19] P. Mertens, "The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model," in *Proc. of Speech Prosody 2004, March 23–26, Nara, Japan*, 2004, pp. 549–552.
- [20] G. Christodoulides, M. Avanzi, and A. C. Simon, "Automatic labelling of prosodic prominence, phrasing and disfluencies in French speech by simulating the perception of naïve and expert listeners," in *Interspeech 2017 – 18th Annual Conference of the International Speech Communication Association, August 20–24, Stockholm, Sweden, Proceedings*, 2017, pp. 3936–3940.
- [21] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. [Online]. Available: <https://www.R-project.org/>
- [22] P. Wagner, J. Trouvain, and F. Zimmerer, "In defense of stylistic diversity in speech research," *Journal of Phonetics*, vol. 48, pp. 1–12, 2015.