Prosodic correlates of perceived quality and fluency in simultaneous interpreting

George Christodoulides[1], *Cédric Lenglet*[2]

[1] Centre Valibel, Institute for Language & Communication, University of Louvain, Belgium
[2] Department for Specialized Translation and Terminology, FTI-EII, University of Mons, Belgium
george@mycontent.gr, cedric.lenglet@umons.ac.be

Abstract

This study explores the relationship between prosodic features specific to simultaneous interpreting and the listeners' perception of the fluency and accuracy of interpreting, as well as their comprehension of the source speech. Two groups of participants (47 subject experts and 40 non-experts) listened to a 20-minute lecture in German, along with its interpretation into French under two conditions (the actual interpretation, or a read-aloud rendition of the same text by the same interpreter) and answered comprehension and rating questions. The prosodic features of the two conditions were analysed, and differences regarding the temporal organisation of speech, disfluencies, pitch register and the interface between prosody and syntax emerged. Our results suggest that interpreting-specific prosodic features affect the perception of fluency, which in turn affects the perception of accuracy. However the impact on listeners who enjoy relevant contextual knowledge is less pronounced.

Index Terms: speech perception, quality of simultaneous interpreting, fluency

1. Introduction

Simultaneous conference interpreters facilitate multilingual communication in political, technical and other meetings. Typically, they work in soundproof booths, translating speech in real-time so that the participants can follow the debate in their language, without interruption. They are expected to "communicate the speaker's intended messages as *accurately, faithfully, and completely* as possible (…) and to *be clear and lively in [their] delivery*" [1, original emphasis]. Since prosody conveys important information in human communication (e.g. information status, focus, intent, emotion), the interpreter is expected to decode such information from the prosodic structure of the source language (SL) and encode it in parallel constructions in the target language (TL). This 'translation' of prosodic information, along with the linguistic information, from SL to TL inspired studies on the correspondence, or alignment, of prosodic patterns across languages (e.g. [2] where an algorithm is proposed to find equivalences between clusters of intonation patterns in a parallel bilingual corpus).

Beyond these conscious *choices* made by the interpreter, however, the prosodic features of SI are influenced by two *constraints*: the reformulation process of translation and the high cognitive load induced by the task itself [e.g. 3]. SI strategies such as stalling (i.e. waiting for enough input before producing a translation or committing to a syntactic structure) and anticipation (i.e. predicting part of the speaker's input) [4: 201] affect the temporal structure of interpreters' speech. Speech produced under high cognitive load presents specific characteristics, including a lower articulation rate, longer silent pauses, an increased number of filled pauses, corrections and restarts, as well as alterations in voice quality [5, 6, 7].

Surveys among users of simultaneous interpreting suggest that they consider accuracy (or fidelity) to be a crucial quality criterion, whereas prosodic features, such as intonation or accent, are not deemed paramount [8]. However, many users cannot assess the interpreters' accuracy because of their lack of knowledge of the source language. Instead, the users' perception of SI quality may depend on the prosodic features of the interpreters' speech, including intonation, hesitations and pauses. Moreover, the interpreters' liveliness appears to influence the listeners' understanding of the speech content [9, 10]. In other words, the core objective of SI, namely to "produce the same effect on [the listeners] as the original [speech] does on the speaker's audience" [10: 155], might depend not only on what the interpreters say, but also on *how* they say it.

The link between quality perception and prosody is all the more important in SI, as previous research indicates that SI has a distinctive prosodic profile. It results from the interplay of the choices and constraints outlined above, as a speaking style determined both by the situational context and by individual characteristics [c.f. 11, 12]. A particular prosodic profile for SI has been observed in studies for at least the following language combinations: Hebrew to/from English, with a large number of "low-rise non-final pitch movements" [13: 231]; English to German, with "long pauses [and a] high proportion of final pitch movements that indicate a continuation" [14: 72]; and English to French, with less numerous and longer silent pauses, and a narrower pitch range compared to the source speech [15].

Fluency has been regularly used as a quality criterion in expectation surveys on SI [8], although it is a polysemous concept. In ordinary language, fluency refers to general (often foreign) language proficiency, whereas a more technical definition associates fluency with speech flow and absence of disfluencies such as pauses, hesitations and repetitions [16: 537]. Experimental research has shown that these temporal features do not only influence the perception of the interpreter's fluency, but also that of its intonation [8: 67]. Listeners asked to rate "fluency" seem to blend temporal features with intonation (e.g. pitch variation). Consequently, it seems appropriate to merge these parameters in a perceptual study.

Does the particular prosody of SI have an impact on its perception? To date, most studies on prosody and quality in SI are based on carefully doctored speeches [8, 9]. Consequently, their findings cannot be linked directly to the perception of the *authentic* prosody of SI. Shlesinger [13] conducted a small-scale experiment on the impact of authentic SI prosody on 15 listeners' understanding of speech content, comparing excerpts of speeches produced under two different conditions: read aloud from a script and interpreted simultaneously. In a

listening test, the subjects' scores in the "read-aloud" condition were approximately twice as good. She concluded that SI intonation affected meaning and perception, but she argued that this effect would be counterbalanced in the case of authentic conference participants with the relevant contextual knowledge [13: 234]. Our experiment aims to answer the following research questions: Does simultaneous interpreting from German into French have particular prosodic features? If yes, do these features influence the listeners' objective and subjective understanding of the speech content, and their perception of the interpreter's fluency and accuracy?

2. Method

2.1. Perceptual Experiment

Our goal was to create a situation as close to authentic SI as possible. The original speech is an abridged presentation on investment strategy by a German fund manager; its duration is 20 minutes. German was selected as the source language in order to increase the likelihood that French-speaking listeners rely on the interpreter only. A professional conference interpreter (male, French native speaker, 6 years of experience) interpreted the German presentation into French in a state-of-the-art interpreting booth. The recording of this interpretation was transcribed; punctuation was added at syntactically-complete clause boundaries; discourse markers and interjections were included in the transcription (only filled pauses, e.g. 'euh', were omitted). The same interpreter read the transcript, after rehearsing it, and was recorded in a booth. We thus obtained two different prosodic profiles by the same speaker: under authentic SI conditions; and prepared reading, without the cognitive constraints of SI. The two versions were synchronized to the video of the original presentation using *Praat, Audacity* and *AviDemux*.

The experimental design is a conference simulation adapted from [9]. The subjects watch the video of the German presentation and listen to an interpretation into French. An interpreter pretends to work in a booth at the back of the room. This creates the impression of a live interpretation, whereas actually, the subjects are listening to one of the recordings, according to the experimental condition they were assigned to.

The subjects were 87 French-speaking university students: 47 students of economics and 40 translation students. Students in economics were chosen because of their specialized knowledge and their greater availability than professional economists. Translation students were chosen in order to control the influence of prior thematic knowledge. The subjects were matched for academic performance (based on grade records) to control memory and prior knowledge. The resulting pairs were randomly distributed between two experimental conditions: interpreted and read-aloud speech.

We use a listening comprehension test and an assessment questionnaire, which we both pretested extensively. The listening test consists of 3 multiple-choice and 4 half-open questions and assesses the comprehensibility of the speech with a listening score (interval scale). In the assessment questionnaire, the subjects are asked to rate on a 7-point ordinal scale how fluent the interpreter's delivery was (Fluency), how well they think they understood the lecture (Subjective Comprehension) and how accurately they reckon the interpreter rendered the speech (Accuracy).

2.2. Linguistic and Prosodic Analysis

The two recordings (SI and Read) were orthographically transcribed in Praat [17]. We obtained a phonetic transcription as well as an automatic segmentation of words, syllables, phones and pauses, automatically using EasyAlign [18]; the alignment was corrected manually. A 'delivery' tier was added to annotate articulation-related (schwa, creaky voice, liaison and elision) and paralinguistic phenomena (audible breath, noises). Part-of-speech tagging and multi-word unit detection were obtained automatically using DisMo [19] and subsequently verified manually. We applied an annotation scheme for disfluencies based on [20]: i) single-token disfluencies: filled pauses, hesitation-related lengthening, lexical false starts and intra-word pauses; ii) structured disfluencies: repetitions (of one or more words), deletions, substitutions, insertions, and complex combinations of the above.

To process our data we used *Praaline* [21], a toolkit that interfaces with *Praat* and runs a cascade of scripts and/or external analysis tools, each of which may add features to an annotation level (e.g. syllables, words etc.), stored in a relational database. We applied *Prosogram's* [22] two-step algorithm for pitch stylisation: for each syllable, vocalic nuclei are detected based on intensity and voicing, and then the f_0 curve on the nucleus is stylised into a static or dynamic tone, based on a perceptual glissando approach. Syllabic prominence was estimated with *ProsoProm* [23], *Analor* [24], and a manual perceptual annotation was also performed (cf. 2.3). Segmentation into accentual and intonational phrases was performed by an expert annotator (taking into account all prominence scores); furthermore, perceptually-motivated prosodic boundaries were calculated based on the approach proposed in [25]. Several aggregate measures were calculated using *ProsoReport* [12]. In order to study the interface between prosody and syntax, a three-level syntactic annotation was added. First, an annotation into minimal chunks based on the phrasal tag-set of the French Treebank [26] was added manually. We also applied the model for syntactic annotation into functional sequences and dependency clauses detailed in [27] and obtained segmentation into Basic Discourse Units whenever the major prosodic boundaries and the dependency clause boundaries coincide. In total, the two recordings are 42-minutes long (1256 seconds each), and contain 8760 syllables, 1335 silent pauses, and 6143 tokens (words).

2.3. Evaluation of automatic tools

As a corollary study, we evaluate the performance of the above-mentioned automatic tools. Results for prominence detection are shown in Table 2; in line with [28] there was a fair agreement between the human annotator and the tools, and between the tools themselves (consistently lower for the SI).

Table 1. *Evaluation of prominent syllable detection*

Both conditions	**ProsoProm vs. Analor**	**ProsoProm vs. Manual**	**Analor vs. Manual**
Precision	97.1%	81.9%	59.3%
Recall	39.9%	49.1%	86.4%
Correct	77.4%	84.4%	81.6%
F-measure	56.6%	61.4%	70.3%
Cohen's kappa	**0.447**	**0.524**	**0.576**
Interpreting κ	0.394	0.456	0.561
Reading κ	0.478	0.568	0.581

A comparison between the (manually corrected) segmentation into accentual and intonation phrases (AP/IPs), and the perceptually-motivated prosodic boundaries (PBs) proposed by [25] can be seen on Table 2. As expected, the perceptual PBs are coarser than the hierarchical segmentation into AP/IPs (which, for French, is based on the prominence of the last syllable of an each unit).

Table 2. *Comparison of prosodic boundary detection*

Sylls with PBs vs. IP/APs	IP boundary		AP boundary	
	Yes	No	Yes	No
No PB	427	**6773**	1262	**5938**
Minor PB	244	264	381	127
Intermediate PB	55	72	73	54
Major PB	**921**	1	**922**	0

The precision of the POS annotation (DisMo) was 96.3 %, while the syntactic annotation was performed manually.

3. Results

3.1. Global Prosodic Features

A selection of global prosodic features is shown in Table 3.

Table 3. *Global prosodic measures.*

Measure	SI	Read
Articulation ratio (%)	72.6	62.7
Articulation rate (syll/s)	4.91	5.27
Speech rate (syll/s)	3.59	3.33
Speech segments (runs)	493	858
… with average length (syll)	9.2	4.9
Var. coefficient of vowel duration	0.089	0.022
Var. coefficient of syllable duration	0.079	0.042
Median pitch (Hz)	127	152
Pitch range (semitones)	7.8	14.2
Pitch trajectory (semitones/s)	15.13	22.47

We note a higher **articulation rate** (syllables per second *excluding* pauses) under the Reading condition. The interpreter made more silent pauses in the Reading condition (cf. 3.2.); this is reflected in the lower articulation ratio, and the lower speech rate (sylls/s *including* pauses). Speech segments (continuous stretches of speech separated by silent pauses >250ms) are considerably more numerous and shorter under the Reading condition than in SI. These measures indicate that the interpreter **over-segmented** his speech in the Reading condition (short utterances and extensive use of pauses). The observed difference between the variance coefficients of vowel duration and of syllable duration indicate that under the SI condition, the interpreter accelerated and decelerated his articulation more frequently than under the Reading condition. Finally, **pitch range and pitch trajectory** are smaller under the SI condition, compared to Reading; this indicates that the latter was a livelier rendition of the text.

3.2. Silent Pauses and Disfluencies

A Mann-Whitney U test on average **silent pause length** indicates that it is longer under the SI condition ($p < 0.001$). We modelled silent pause length as a mixture of log-normal distributions, following the methodology in [29] and [30]:

$$f(x) = \sum_{i=1}^{N} \pi_i \Lambda_i(\mu_i, \sigma_i^2, x)$$

Three component distributions are identified (using a Bayesian Information Criterion), and their parameters estimated using the Expectation-Maximisation algorithm (Table 4, Figure 1). Cut-off values *t* (local maxima of the model's uncertainty function) are used as thresholds to categorise pauses as 'short', 'medium' or 'long' (instead of *ex-nihilo* fixed thresholds).

Table 4. *Log-normal mixture model of silent pause length (in ms), 1.3 < σ < 1.8, N = 3.*

Pause type	SI			Read		
	π	M	t	π	μ	t
Short	44%	195	283	39%	136	203
Medium	32%	568	1037	48%	581	602
Long	24%	1570		13%	1221	

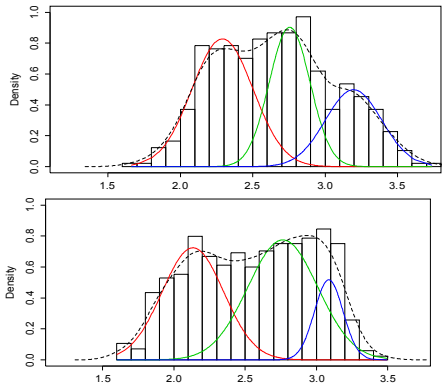

Figure 1. *Density plots of log (silent pause length) and component distributions (SI: top, Reading: bottom).*

Regarding **disfluencies**, under the SI condition the interpreter produced 272 filled pauses vs. only 8 under the Reading condition. Other types of disfluencies were almost inexistent in Reading. Under the SI condition false starts, repetitions and deletions (in order of frequency) were observed. In total, **9.8%** of the tokens were disfluent in SI, compared to **0.4%** in Reading.

3.3. Prosody-Syntax Interface

Basic Discourse Units (BDUs) in [27] are proposed as "the segments that speakers and listeners use to interpret the discourse they are engaged in". Based on the observation that listeners use both prosody and syntax as cues to information structure, BDUs are defined as segments that run between the points where major prosodic boundaries and dependency clause boundaries **coincide**. In a congruent BDU, one intonation unit (IU) contains one dependency unit (DU); in intonation-bound BDUs, one IU contains several DUs; in syntax-bound BDUs, one DU packs several IUs. Regulative BDUs contain only discourse markers or adjuncts. A mixed-boundary BDU contains more than one DU and more than one IU, and is the product of a lack of synchrony between prosodic and syntactic boundaries. Table 5 shows the distribution of BDUs of different types under the two conditions.

Table 5. *Number and average duration of BDUs per type and condition.*

Condition / BDU type	SI		Read	
	%	Avg dur (s)	%	Avg dur (s)
Congruent	21.3	2.57	20.2	2.04
Regulative	21.9	1.24	24.4	0.85
Intonation-bound	5.3	6.43	0.4	1.98
Syntax-bound	30.2	8.20	55.0	5.60
Mixed-boundary	21.3	11.79	0	

We observe that in the SI condition, there was frequently a **mismatch between prosodic and syntactic boundaries**. These are typically cases in which the interpreter constructs a phrase incrementally, pausing inside syntactic units. Figure 2 shows how the three different types of silent pauses are distributed between and within syntactic units (chunks and functional sequences) and BDUs.

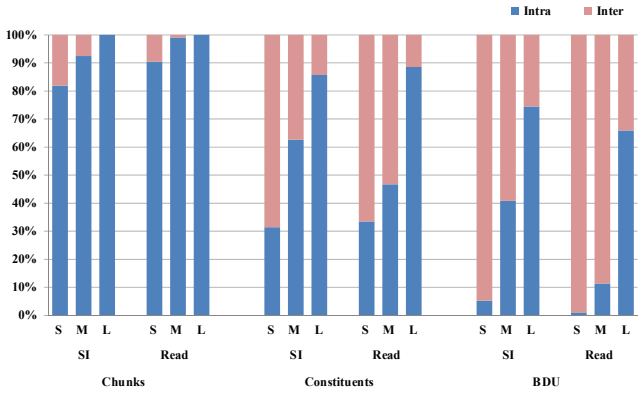

Figure 2. *Silent pauses of different types (S: short, M: medium, L: long), within (intra) and between (inter) syntactic units.*

We observe that in the SI condition, medium-length pauses *within* constituents and *within* BDUs occur more frequently than in the Reading condition (+15% and +30% respectively). This is in line with the high percentage of mixed-boundary BDUs observed. The high proportion of syntax-bound BDUs with a relatively short duration under the Reading condition is another indication of over-segmentation.

3.4. Perception of Quality and Fluency

The questionnaire data were coded and processed with *IBM SPSS Statistics*. Missing observations were excluded. The mean Listening Score, which measures Objective Comprehension (the sum of the correct answers in comprehension questions; max = 17), and the median Accuracy, Fluency and Subjective Comprehension ratings (1 = best), broken down by subject groups and experimental conditions are shown in Tables 6 and 7. Highest scores and best ratings are in boldface.

Table 6. *Mean Listening Score per group (17 = max)*

Group	SI	Read	Both
Translation students (TRAN)	8.47	**9.39**	8.94
Economics students (ECON)	7.95	**8.39**	8.18
Both groups	8.18	**8.83**	8.52

Table 7. *Median quality and subjective comprehension ratings per group (1 = best)*

Group	Rating	SI	Read	Both
TRAN	Accuracy	2	2	2
	Fluency	3	**2**	2
	Subjective Compr.	5	**4.5**	5
ECON	Accuracy	3	3	3
	Fluency	3	3	3
	Subjective Compr.	3	3	3
Both	Accuracy	2	2	2
	Fluency	3	**2**	3
	Subjective Compr.	4	4	4

The translation students' median ratings of fluency and subjective comprehension are better in the Reading condition. Across all groups, there is a moderate correlation between the experimental condition and the fluency ratings, which turns out to be significant in a Spearman correlation test (Translation: $r = -0.506$; $p = 0.001$; Economics: $r = -0.323$; $p = 0.027$; Translation + Economics: $r = -0.393$; $p < 0.001$; two-tailed). In other words, **the subjects who listened to the read-aloud speech tended to rate fluency better**.

Concerning the fluency ratings without regard to the experimental condition, there is a moderate and significant correlation between fluency ratings and subjective comprehension among the students of economics ($r = 0.480$; $p = 0.001$, two-tailed). There is a slightly stronger significant correlation between fluency and accuracy ratings across all groups (Translation: $r = 0.490$; $p = 0.002$; Economics: $r = 0.496$; $p = 0.001$; Translation + Economics: $r = 0.520$; $p < 0.001$; two-tailed).

4. Conclusions and Perspectives

In this paper we presented a perceptual study based on a conference simulation, striving to create a situation as close to authentic conditions as possible.

With respect to our first research question, our findings confirm previous studies regarding the **particular prosodic characteristics of SI**. In the SI condition, the interpreter produced long silent pauses, frequent filled pauses and several reformulation-related disfluencies. The articulation rate was more variable (i.e. more accelerations and decelerations), and the pitch range and trajectory were both narrower in SI, indicating that the same person rendered the text in a more lively fashion, when freed from the cognitive constraints of interpreting. The main effect of SI was observed in the prosody-syntax interface, with often mismatched prosodic and syntactic boundaries, and more intra-unit pauses. The combination of these prosodic features has had an effect on the perceived fluency rating. As previous research has shown that "some disfluencies may be considered felicitous by listeners" when used for communicative purposes [31], it will be interesting to explore the contribution of each prosodic factor in fluency ratings, in a future study.

With respect to our second research question, the results lend additional support to the claim that **the perception of the interpreters' accuracy is linked to that of their fluency**, thus confirming previous experimental findings [e.g. 8, 9]. The differences in listening scores (objective comprehension) between the experimental conditions are less pronounced among students of economics. This seems to support Shlesinger's claim that the prosody of interpreting has less impact on the listeners who enjoy relevant contextual knowledge. One explanation could be that the translation students processed the speech at a more superficial level and hence, were more affected by perturbations of the prosodic structure of the speech. The students of economics could use their prior knowledge to process the speech content at a deeper level and make inferences to compensate for disturbing prosodic variations. Admittedly, the higher mean listening score of translation students is unexpected. We hypothesize that these students benefited from their capacity to capture the gist of speeches in their notes thanks to an elaborate note-taking technique they develop in introductory courses to conference interpreting. In a future study, perceptual and prosodic data could be correlated to test the effect of each prosodic factor on perceived quality and fluency.

5. Acknowledgements

We would like to thank Prof. Liesbeth Degand (University of Louvain) for her help with the syntactic annotation; Dr. Mathieu Avanzi for the annotation of accentual and intonation units; our anonymous interpreter colleague who volunteered for the recordings; our subjects; and the teaching and technical staff at UMons, who made this experiment possible. The second author is supported by a UMons 100% research grant.

6. References

[1] AIIC, "Practical guide for professional conference interpreters", Online: http://aiic.net/page/628/practical-guide-for-professional-conference-interpreters/lang/1#5, 1990/2004.

[2] Agüero, P.D., Adell, J., and Bonafonte, A., "Prosody generation for speech-to-speech translation", Proceedings of the ICASSP, 2006.

[3] Seeber, K. and Kerzel, D., "Cognitive load in simultaneous interpreting: Model meets data", International Journal of Bilingualism 16(2): 228–242, 2012.

[4] Gile, D., Basic Concepts and Models for Interpreter and Translator Training, Revised edition, Amsterdam and Philadelphia, John Benjamins, 2009.

[5] Jameson, A., Kiefer, J., Müller, C., Grossmann-Hutter, B., Wittig, F. and Rummer, R., "Assessment of a user's time pressure and cognitive load on the basis of features of speech", in M. Crocker and J. Siekmann [Eds] Resource-adaptive cognitive processes, 171–204, Berlin, Springer, 2009.

[6] Tet Fei Yap, "Speech production under cognitive load: Effects and classification", Diss., The University of New South Wales, 2012.

[7] Schuller, B., Batliner, A., "Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing", Wiley, 2013.

[8] Collados Aís, Á, Macarena Pradas Macías, E., Stévaux, E. and García Becerra, O. [Eds], La Evaluación de la Calidad en Interpretación Simultánea: Parámetros de Incidencia. Granada, Comares, 2007.

[9] Holub, E. and Rennert, S., "Fluency and intonation as quality indicators", Paper presented at the Second International Conference on Interpreting Quality, Almuñécar, Spain, 2011.

[10] Déjean le Féal, K., "Some thoughts on the evaluation of simultaneous interpretation", in D. and M. Bowen [Eds], Interpreting: Yesterday, Today, and Tomorrow, 154–160, Binghamton, State University of New York Press, 1990.

[11] Léon, P., Précis de phonostylistique, Parole et expressivité, Nathan Université, Paris, 1993.

[12] Goldman, J.-Ph., Auchlin, A. and Simon, A.C., "Description prosodique semi-automatique et discrimination des styles de parole" in H.-Y. Yoo and E. Delais-Roussarie [Eds], Actes d'IDP 2009, 207-221, Paris, September, 2009.

[13] Shlesinger, M., "Intonation in the production and perception of simultaneous interpretation", in S. Lambert and B. Moser-Mercer [Eds], Bridging the Gap: Empirical Research in Simultaneous Interpretation, 225–236, Amsterdam and Philadelphia, John Benjamins, 1994.

[14] Ahrens, B., "Prosodic phenomena in simultaneous interpreting: A corpus-based analysis", Interpreting 7(1): 51–76, 2005.

[15] Christodoulides, G., "Prosodic features of simultaneous interpreting", in P. Mertens and A.C. Simon [Eds], Proceedings of the Prosody – Discourse Interface Conference, 33–37, Leuven, Belgium, 2013.

[16] Chambers, F. "What do we mean by fluency?", System 25(4): 535-544, 1997.

[17] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer", online at http://www.praat.org

[18] Goldman, J.-Ph. EasyAlign: an automatic phonetic alignment tool under Praat, Proceedings of InterSpeech, Florence, Italy, 2011.

[19] Christodoulides, G., Avanzi, M. and Goldman, J.-Ph., "DisMo: A morphosyntactic, disfluency and multi-word unit annotator: An evaluation on a corpus of French spontaneous and read speech", Paper submitted to LREC, 2014.

[20] Shriberg, E., "To 'errrr' is human: ecology and acoustics of speech disfluencies", Journal of the International Phonetic Association 31(1): 153–169, 2001.

[21] Christodoulides, G., "Praaline: integrating tools for speech corpus research", Paper submitted to LREC, 2014.

[22] Mertens, P., "The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model" in B. Bel and I. Marlien [Eds], Proceedings of Speech Prosody 2004, Nara, Japan, 23–26 March, 2004.

[23] Goldman, J.-Ph., Avanzi, M., Simon, A.C. and Auchlin, A., "A continuous prominence score based on acoustic features", Proceedings of InterSpeech 2012, 9–13 September, 2012.

[24] Avanzi, M., Lacheret, A., and Victorri, B. "ANALOR. A Tool for Semi-Automatic Annotation of French Prosodic Structure", Proceedings of Speech Prosody 2008, 119–122, 2008.

[25] Mertens, P. and Simon, A.C., "Towards automatic detection of prosodic boundaries in spoken French" in P. Mertens and A.C. Simon [Eds], Proceedings of the Prosody – Discourse Interface Conference, 81–87, Leuven, Belgium, 2013.

[26] Abeillé, A., Clément, L. and Toussenel, F., "Building a treebank for French" in A. Abeillé [Ed] Treebanks: Building and using parsed corpora, 165–188, Kluwer, Dordrecht, 2003.

[27] Degand, L., Simon, A.C., "On identifying basic discourse units in speech: theoretical and empirical issues", Discours (4), Online: http://discours.revues.org/5852, 2009.

[28] Avanzi, M., Rousier-Vercruyssen, L., Schwab, S., Gonzalez, S., and Fossard, M., "C-PROM-Task: A New Annotated Dataset for the Study of French Speech Prosody", Proceedings of TRASP, Aix-en-Provence, 27–30, 2013.

[29] Goldman, J.-Ph., François, T., Roekhaut, S., Simon, A. C., "Étude statistique de la durée pausale dans différents styles de parole", Actes des 28èmes Journées d'Étude sur la Parole (JEP), Association Francophone de la Communication Parlée, Mons, Belgium, 25-28 May, 2010.

[30] Little, D., Oehmen, R., Dunn, J., Hird, K., Kirsner, K., "Fluency Profiling System: An automated system for analyzing the temporal properties of speech", Behavior Research Methods 45 (1): 191–202, 2012.

[31] Moniz, H., Trancoso, I., Mata da Silva, A.I., "Classification of disfluent phenomena as fluent communicative devices in specific prosodic contexts", Proceedings of Interspeech 2009, 1719-1722, ISCA, Brighton, UK, September 2009.