Phonetic variations : Impact of the communicative situation

Sandrine Brognaux[1,2]*, Thomas Drugman*[2]

[1]Cental, ICTEAM (Université catholique de Louvain), Belgium
[2]TCTS Lab (Université de Mons), Belgium
sandrine.brognaux@uclouvain.be, thomas.drugman@umons.ac.be

Abstract

While speech synthesis research is now focussing on the generation of various speaking styles or emotions, very few studies have considered the possibility of including phonetic variations according to the communicative situation of the targeted speech (sports commentaries, TV news, etc.). This paper proposes a phonetic analysis of large French corpora to assess the influence exerted by three situational 'traits': read/spontaneous, media/non-media and expressive/non-expressive. It shows that some variations, like elision, tend to be more frequent in spontaneous and non-media speech, conversely to liaisons which appear more often in read and media speech. Interestingly, no phonetic variation draws a clearcut distinction between expressive and non-expressive speech. Finally, a prosodic analysis indicates that the phonetic variations are not directly correlated with the rhythmic features of their corresponding situational 'trait'.
Index Terms: Phonetics, Rhythm, Speaking Style, Speech Synthesis

1. Introduction

Text-To-Speech (TTS) synthesis has reached in the last decades a fairly good level of quality and intelligibility for the generation of neutral read speech. The interest has now shifted to the production of speech corresponding to various speaking styles and emotions. Most studies focussing on this topic have considered modifications at the prosodic and voice quality levels only (see [1, 2]). Surprisingly, potential phonetic modifications of the sentence to synthesize are generally discarded. One rare exception is presented in [3], which modifies the pronunciation of final schwas according to the communicative situation (radio news, conversation, etc.).

This concern is particularly relevant in French, where words are characterized by a high amount of phonetic variants. Schwa elisions and liaisons are the most frequent phonetic modifications. The first consists in the optional pronunciation of a schwa vowel in the middle or at the end of a word (e.g. *petite* pronounced [ptit]). The second relates to latent consonant at the end of a word which can be pronounced when followed by a vowel or a mute h (e.g. *les enfants* pronounced [le͡zɑ̃fɑ̃]). Many linguistic studies have analyzed the modalities of appearance of these phenomena (see [4] and [5] for liaisons, [6] and [7] for epenthetic schwa, [8] and [9] for elisions). They show that the realization of the different variants can be explained by many factors: morpho-syntax [10], speech rate [11, 8, 5], word frequency [4, 5], word probability [12], degree of articulation [13], origin of the speaker [9, 14], age of the speaker [9], etc. Few linguistic analyses, however, have studied the influence exerted on phonetic variations by the communicative situation (TV news, political speech, text reading, etc.), also sometimes referred to as the 'phonogenre' [15, 16]. Yet, the potential interaction between both levels is widely acknowledged [8, 17] and was shown to be very influential for prosody [3, 17, 16]. Only the phonetic differences between read and spontaneous speech have aroused great interest [6, 4, 18].

Most speech synthesizers integrate basic phonetic variations. However, they are trained to produce a phonetic transcription corresponding to standard read speech. For optional variations, the most likely variant is generally produced, independently of the communicative situation. While research is now targeting the generation of expressive [1, 2] and media-related speech (e.g. sports commentaries [19]), the need for a broad study of the influence of these situational 'traits' (as further defined) on phonetics is striking.

Our study proposes an analysis of the influence exerted by the communicative situation on the phonetic realization. Because they cannot be easily ranked on a single scale, the various communicative situations are defined according to three binary 'traits': media/non-media, expressive/non-expressive and read/spontaneous. They will be referred to as 'situational traits' in the remainder of this paper. The main objective of the study is to offer an insightful description of the phonetic features of each 'trait' to outline what should be considered when synthesizing a certain communicative situation.

Our analysis has the advantage of relying on a very large corpus in French of about 300 minutes from 32 speakers and 10 communicative situations (sports commentaries, TV news, political speech, etc.). The study of the phonetic realization is based on a strategy making use of natural language processing (NLP) techniques. In a second stage, rhythm is considered as it was shown to be one of the prosodic correlates of phonetic variations [11, 5]. The potential correlation between phonetic and rhythmic features is then evaluated.

The paper is organized as follows. Section 2 presents the corpus and its annotation. The methodology exploited to carry out our study is detailed in Section 3. The main analysis, both phonetic and rhythmic, of the corpus is described and discussed in Section 4. Finally, Section 5 concludes the paper and discusses further works.

2. Corpus design

Our corpus is an extended version of C-PROM [20] including additional sub-corpora exploited for speech synthesis. A special focus is made on sports commentaries [21] which have also been added to the corpus. The phonetization of the speech files was done automatically and further corrected manually. The entire corpus was then automatically phonetically aligned

with EasyAlign [22] and Train&Align [23].

The corpus consists of 300 minutes from 32 French-speaking speakers (French, Belgian and Swiss) and ten sub-corpora corresponding to different communicative situations (interview, political speech, etc.). Each situation contains 2 to 7 speakers and is defined according to three binary situational 'traits': media, read and expressive. Expressive is here defined as an audible emotive implication of the speaker (e.g. excitation, anger, happiness, etc.), be it acted or not. It should be noted that this 'trait' gathers different kinds of expressivity. Emotion valence, for example, can be positive (e.g. happy) or negative (e.g. sad). This could lead to averaged effects in our analysis, and hinder the interpretation of the role played by the various aspects of expressivity.

A summary of the different sub-corpora is shown in Table 1. The situational 'traits' being continuums, some corpora were not classified (NC) if their nature regarding a dimension was ambiguous. The continuum between read and spontaneous speech, for example, goes through 'prepared', which could be assigned to conferences. Because text to speech corpora were not broadcast as such, but could be used for public announcements, they were not classified for the media 'trait'. For interviews, only the parts of the interviewee were kept. The number of speakers per 'trait' is rather balanced and ranges from 13 to 17 with an average length of about 2 hours of speech per trait.

Table 1: *Distribution of the sub-corpora according to the three studied situational 'traits' (TTS corpora were recorded for speech synthesis purposes).*

Communicative situation	Read	Media	Expressive
Sports commentaries	-	+	+
Conference	NC	NC	-
Political discourse	+	+	NC
Interview	-	+	+
Itinerary explanations	-	-	-
TV news	+	+	NC
Expressive speech TTS	+	NC	+
Neutral speech TTS	+	NC	-
Neutral reading	+	-	-
Narration	-	-	+

3. Methodology

For the phonetic analysis, we developed a specific methodology integrating NLP techniques. For each sub-corpus, the orthographic transcription was exploited to produce its automatic phonetization with the NLP tool ELite [24] designed for TTS purposes. It produced a 'standard' phonetization of the text, corresponding to neutral read speech. This transcription was then automatically aligned with the phonetic transcription really pronounced by the speaker (and which was manually checked).

This alignment relies on a slightly modified version of Levenshtein's edit distance [25]. Several adaptations were made:

- Each phoneme is represented by only one character,
- Some phonemes substitutions are not penalized (ø→œ, e → ɛ, etc.) as they might result from a subjective perception of the annotator,
- Insertions and deletions of silences are not penalized.

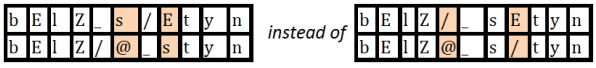
To retrieve the alignment, the matrix obtained by the algorithm is backtracked. Finally, all modifications are stored according to their type (insertion, deletion or substitution).

To avoid potential phonetization errors of the NLP, all sound files containing numbers (written as ciphers) were deleted. Proper names being very frequent in sports commentaries, their deletion would have resulted in a highly reduced sub-corpus. For that reason, we decided to consider only modifications of phonemes not belonging to proper names or to syllables just before and after a proper name.

A first analysis of the alignment highlighted recurrent errors made by the algorithm when two modifications occurred in a small phonetic context. This led to some refinements:

- The deletion and insertion of schwas being more frequent, they were favored and assigned a reduced cost.
- The cost was also reduced for substitution of [i] by [j], which is rather frequent.

This avoided alignments such as:



The main advantage of using NLP-produced phonetization is that it allows for an easy comparison of the pronunciation of the corpus with a so-called 'standard' pronunciation. The latter already considers most mandatory phonetic variations such as liaisons or elisions dictated by the linguistic context. It provides a more precise analysis than a comparison with a phonetized dictionary (see e.g. [4]) while being fully automatic.

Throughout this study, the statistical significance of the results is calculated via unilateral t-tests or Wilcoxon tests depending on the normality of the variable distribution. Correlations are evaluated using Spearman's coefficient.

4. Phonetic and prosodic analysis

4.1. Analysis of the phonetic variations

In this section, we first consider the overall proportion of phonetic variations for each situational 'trait' (4.1.1). We then focus on the analysis of four phonetic variations that were qualitatively assessed to be the most frequent in our corpus: schwa elision (4.1.2), epenthetic schwa (4.1.3), final consonant elisions (4.1.4) and liaisons (4.1.5).

4.1.1. Overall proportion of phonetic changes

Phonetic variations are analyzed by comparing the NLP-produced standard phonetization with the real pronunciation by the speaker. The overall proportion of phonetic changes is computed as the total amount of modifications (deletions, insertions or substitutions) divided by the maximal number of characters, i.e. the number of characters of the longest of both strings.

Table 2 shows significant differences in the amount of phonetic changes for both media and read dimensions (with respectively p=0.043 and p=1.2e-05). This indicates that spontaneous and non-media speech differ rather strongly from what produces a generic NLP. Conversely, read speech corpora exploited for speech synthesis display, on average, only 1.31% of phonetic changes. This finding partly implies that, while the NLP produces suitable phonetizations for neutral read speech, it requires non-negligible modifications to produce non-media

or spontaneous speech. Finally it is worth noting that no significant differences are found along the 'expressive' dimension. This might be due to the heterogeneity of this 'trait', which notably gathers emotions with different valences.

The next sections investigate typical phonetic phenomena for each 'trait'.

4.1.2. Schwa elision

Schwa elision is one of the most intricate phonetic variations in French. It relates to schwas which can be pronounced or not at the middle or the end of a word. Our analysis excludes final schwas which may rather be linked to liaisons and are further investigated in the next subsection. The percentage of elided schwa is here computed as the number of schwa deletions, inside words, divided by the total amount of schwas inside words. Figure 1 shows the significant role played by the distinctions spontaneous/read and media/non-media (with respectively p=0.005 and p=1.7e-04). It shows that more schwas are elided in spontaneous speech, corroborating earlier studies [4, 9, 6]. An interesting finding is that more schwas are also elided in non-media speech. This may be explained by the fact that media speech tends to belong to a higher level of language which has been said to be correlated with lower elision rates [26]. As in [8], we observe rather high inter-speaker variability.

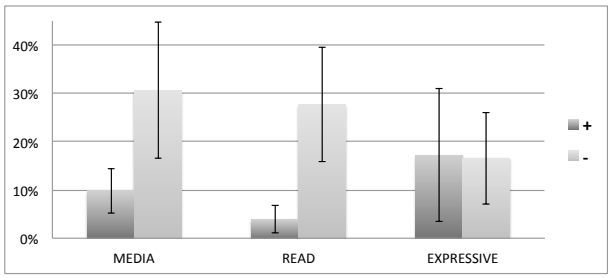


Figure 1: *Percentage of elided schwas in middle word position, together with their 95% confidence intervals.*

4.1.3. Epenthetic schwa

The epenthetic schwa is seen as the insertion of a final schwa on a word ending or not in -e (e.g. *match* pronounced [matʃə̲]) [6, 27, 28]. Candea [7] shows that its frequency has increased in the last decades and that it can occur independently of the phonetic or rhythmic context. While it was previously seen as a sign of informality, the sociolinguistic aspect is now fading out.

Our analysis focussed only on epenthetic schwas in words with no final -e. Interestingly, this variation is significantly more frequent in media compared to non-media speech, as shown in Figure 2 (p=0.013). This goes in line with [3] which assessed a higher rate of ending schwa pronunciations (all words considered) in radio news and political speech compared to conversational speech. This rate is also significantly higher in spontaneous and expressive speech (p=0.019 and p=0.008), most likely due to their high frequency in sports commentaries. High inter-speaker variability is however observed. While they are often studied on Parisian French, no difference was witnessed in our corpus between French, Belgian and Swiss speakers.

4.1.4. Final consonant elisions

When analyzing the alignment of both predicted and real phonetizations, we observed that the 'il' pronoun (meaning 'he' or 'it') is often pronounced [i], with elision of the final 'l'. This phenomenon occurring quasi-exclusively in front of phonetic consonants, we analyzed its appearance in that specific context (see Table 2). Only sub-corpora containing at least 3 occurrences of the pronoun in that context were kept. Both media and read 'traits' are significant, with higher elision rates in spontaneous and non-media speech (respectively p=6.1529e-05 and p=0.002) which goes in line with results obtained for the elision of schwa in Section 4.1.2.

Another type of consonant elision regards the elision of the final liquid when preceded by an obstruent (e.g. 'peut-être" pronounced [pøtɛt]) [29]. A first qualitative analysis shows that this phonetic variation highly depends on the phonetic context. The liquid is nearly always pronounced when followed by a vowel. Conversely, it may be dropped when followed by a consonant. Our analysis was carried out in this latter phonetic context only, on corpora containing at least 4 occurrences of such phonetic context. Table 2 shows that, here again, spontaneous and non-media corpora display significantly more elisions of the liquid than read and non-expressive speech (with respectively p=1.1e-05 and p=0.04).

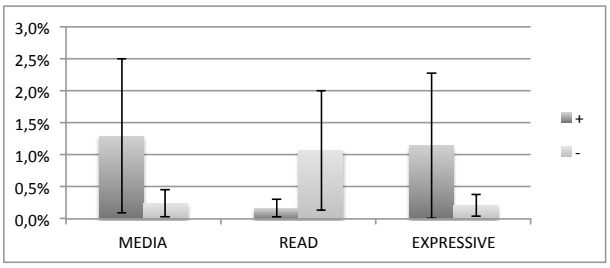


Figure 2: *Percentage of words, not ending in -e, pronounced with a final schwa, together with their 95% confidence intervals.*

4.1.5. Liaisons

We defined potential liaisons contexts as words ending in a French liaison consonant /t, n, z, R, p/ and followed by a vowel, as in [4, 5, 10]. It should be noted that these potential liaisons do not refer to so-called 'optional liaisons', all liaisons being considered, be they mandatory, facultative or prohibited. Table 2 shows that read speech displays a significantly higher rate of liaisons (p=4.6396e-05) which confirms findings in [4, 5, 18]. Interestingly, media speech also shows more liaisons, even if the difference is not significant. This might be due to the fact that media and read speech tend to be more formal, 'sustained' speech being more inclined to high liaison rates [30, 26].

4.2. Prosody: Any correlations between phonetic and rhythmic features?

Rhythm, and speaking rate in particular, has been shown to be correlated with schwa elisions, more elisions appearing in fast speech [11]. This section analyzes various rhythmic features to assess their correspondence with the situational 'traits' and evaluate their correlation with the aforementioned phonetic variations. It focusses on three rhythmic measures: the articulation rate, the mean duration of inter-pausal units (IPU) and the proportion of prominent syllables.

The analysis of the *articulation rate*[1] highlights significant

[1] We focus here on the articulation rate, i.e. the speaking rate (num-

Table 2: *Summary of the phonetic changes for the three situational 'trait' dimensions.*

	All changes		Elision of [l] in 'il"		Elision of liquid in obstruent+liquid		Liaisons	
Situational 'trait'	+	-	+	-	+	-	+	-
Read	1.78%	4.25%	8.33%	76.56%	3.65%	50.34%	59.33%	42.15%
Media	2.93%	4.11%	51.87%	96.67%	18.53%	49.99%	54.52%	44.77%
Expressive	3.57%	2.92%	52.53%	34.72%	33.54%	21.74%	44.53%	50.27%

differences for the read 'trait' only, with lower speaking rates in spontaneous speech (p=0.019). This can be explained by the presence of lengthened syllables, due to hesitations. As in [17], we also observe a significantly lower percentage of articulation for spontaneous speech (p=0.049). Conversely to existing studies (e.g. [11]), however, no significant correlation is found between speaking rate and schwa elision ($|Rho| < 0.09$), or any other phonetic variation. We have shown, on the contrary, that elisions are more frequent in spontaneous speech which displays a lower articulation rate. This difference might be due to the fact that most existing studies focus on one specific task (e.g. text reading) in which only speaking rate is modified to observe the frequency of elisions. In our corpus, too many factors are influencing the speaking rate, e.g. hesitations, communicative situation, speaker idiosyncrasies, etc.

The mean duration of IPU turns out to be significantly longer in spontaneous speech compared to read speech (p=0.04). This seems to indicate that spontaneous speech displays less but longer silences. A possible explanation is that short pauses are rarely silent in spontaneous speech and rather tend to be filled by disfluency markers. Finally, the *percentage of prominent syllables* is assessed with Prosoprom [31], an automatic algorithm for detecting prominent syllables on an acoustic basis. Only the media dimension seems to be influential for that measure, media speech containing more prominent syllables (see Figure 3). This distinction also stands out when looking at initial stresses only, media displaying 19.1% of prominent initial syllables against 16.4% in non-media speech. This confirms findings in [15]. However, the inter-speaker variability is rather high which makes this difference not significant (p=0.07).

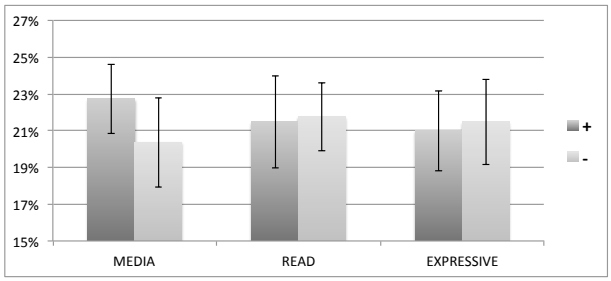


Figure 3: *Percentage of prominent syllables, together with their 95% confidence intervals.*

IPU duration is moderately correlated with schwa elisions and 'l' elisions of the 'il' pronoun (respectively Rho=0.59 and Rho=0.43). This may imply that more elisions are realized when longer sections of speech are uttered without pauses. This, however, does not hold for liquid elisions after obstruents. Interestingly, no correlation can be found between the percentage of prominences and phonetic variations.

ber of syllables per second) with silences excluded, the sub-corpora displaying different silence densities.

On the whole, correlations between rhythmic and phonetic variations are rather weak. This seems to indicate that phonetic variation is more directly dependent on the situational 'trait' itself than on the rhythmic features of the 'trait'. It should be noted that, as for phonetic variation, the expressive 'trait' is not characterized by any specific rhythmic feature.

5. Conclusion and perspectives

While speech synthesis of various speaking styles and emotions is now the focus of much research, very few studies consider potential phonetic variations according to the communicative situation (expressive, spontaneous, etc.). This paper proposed a phonetic analysis of a large corpus in French to assess the influence played by three situational 'traits': read/spontaneous, media/non-media and expressive/non-expressive. It first showed that spontaneous and non-media speech exhibit a significantly higher percentage of phonetic variations compared to standard read speech as produced by a conventional NLP for TTS. Regarding the various phenomena, we showed that spontaneous speech is characterized by a higher elision rate and less liaisons, which confirms results of earlier studies. Media speech usually follows the same phonetic tendencies as read speech, which may be due to the overall higher level of language compared to spontaneous and non-media speech. However, it displays much more epenthetic schwas, which seems to be particularly characteristic of sports commentaries. Interestingly, the expressive 'trait' is not associated with any specific phonetic feature. The diversity of the corpora in that 'trait' (e.g. emotions with different valences) should be further investigated to evaluate the role played by the different aspects of expressivity.

Rhythm was analyzed in a second stage and showed higher speaking rates and shorter inter-pausal units for read speech. Higher proportions of prominences were observed for media speech. Here again, no specific rhythmic feature was associated with the expressive 'trait'. Finally, low levels of correlation were found between the rhythmic parameters and the phonetic variations, except for a moderate correlation between the duration of the inter-pausal unit and some types of elisions. This implies that phonetic variations depend primarily on the situational 'trait' (read/spontaneous and media/non-media) and not on the rhythmic features of that 'trait'.

Further studies will focus on the perceptive analysis of those phonetic changes to assess whether they only constitute possible variants or if they influence the credibility of the message. Required phonetic variations will then be integrated in HMM-based synthesis according to the targeted communicative situation.

6. Acknowledgements

Authors are supported by FNRS. The project is partly funded by the Walloon Region Wist 3 SPORTIC. Authors are grateful to J.-P. Goldman for his insightful advice.

7. References

[1] J. Yamagishi, K. Onishi, T. Musuko, and T. Kobayashi, "Acoustic modeling of speaking styles and emotional expressions in HMM-based speech synthesis," *IECE Transactions on Information and Systems*, vol. E88-D(3), pp. 502–509, 2005.

[2] R. Tsuzuki, H. Zen, K. Tokuda, T. Kitamura, M. Bulut, and S. Narayanan, "Constructing emotional speech synthesizers with limited speech database," in *ICSLP*, 2004, pp. 1185–1188.

[3] S. Roekhaut, J.-P. Goldman, and A. C. Simon, "A model for varying speaking style in TTS systems," in *Speech Prosody*, 2010.

[4] C. Fougeron, J.-P. Goldman, and U. H. Frauenfelder, "Liaison and schwa deletion in French: an effect of lexical frequency and competition?" in *Interspeech*, 2001, pp. 639–642.

[5] C. Fougeron, J.-P. Goldman, A. Dart, L. Gulat, and C. Jeager, "Influence de facteurs stylistiques, syntaxiques et lexicaux sur la réalisation de la liaison en français," in *Actes of TALN*, 2001.

[6] A. Hansen, "The covariation of [schwa] with style in parisian french: an empirical study of 'e caduc' and prepausal [schwa]," in *ESCA Workshop on Phonetics and Phonology of Speaking Styles*, 1991.

[7] M. Candea, "Le e d'appui parisien : statut actuel et progression," in *XXIVe Journées d'Etudes sur la Parole, Université de Nancy II*, 2002, pp. 185–188.

[8] A. Burki, M. Ernestus, C. Gendrot, C. Fougeron, and U. H. Frauenfelder, "What affects the presence versus absence of schwa and its duration: A corpus analysis of French connected speech," *The journal of the Acoustical Society of America*, vol. 130 (6), pp. 3980–3991, 2011.

[9] F. Hambye, "La prononciation du français contemporain en Belgique. Variations, normes et identités." Ph.D. dissertation, Université catholique de Louvain, Belgique, 2005.

[10] P. Boula de Mareuil, M. Adda-Decker, and V. Gendner, "Liaisons in French: a corpus-based study using morpho-syntactic information," in *ICPhS*, 2003.

[11] A. Lacheret-Dujour, "Phonological variations in read speech, reduction phenomena and speaker classes: Do allophonic choices represent speaking style?" in *ESCA Workshop on Phonetics and Phonology of Speaking Styles*, Barcelona (Spain), 1991.

[12] D. Jurafsky, A. Bell, M. Gregory, and W. D. Raymond, "Probabilistic relations between words: Evidence from reduction in lexical production," *Typological studies in language*, vol. 45, pp. 229–254, 2001.

[13] B. Picart, T. Drugman, and T. Dutoit, "Analysis and synthesis of hypo and hyperarticulated speech," in *Speech Synthesis Workshop*, 2010.

[14] A. Martinet, *La prononciation du français contemporain*. Librairie Droz, 1971.

[15] J. P. Goldman, A. Auchlin, and A. C. Simon, "Discrimination de styles de parole par analyse prosodique semi-automatique," in *Interface Discours Prosodie (IDP)*, 2009.

[16] T. Prsir, J. P. Goldman, and A. Auchlin, "Variation prosodique situationnelle: étude sur corpus de huit phonogenres en français," in *Interface Discours Prosodie (IDP)*, 2013.

[17] A. C. Simon, A. Auchlin, M. Avanzi, and J. P. Goldman, "Les phonostyles: une description prosodique des styles de parole en français," in *Les voix des Français. En parlant, en écrivant*. Abecassi, M. & G. Ledegen, 2009.

[18] V. Lucci, *Etude phonétique du français contemporain à travers la variation situationnelle (débit, rythme, accent, intonation, e muet, liaisons, phonèmes)*. Publications de l'Université des Langues et Lettres de Grenoble Grenoble, 1983.

[19] B. Picart, S. Brognaux, and T. Drugman, "HMM-based speech synthesis of live sports commentaries: Integration of a two-layer prosody annotation." in *8th ISCA Speech Synthesis Workshop (SSW8)*, 2013.

[20] M. Avanzi, A. C. Simon, J. P. Goldman, and A. Auchlin, "C-PROM. An annotated corpus for French prominence studies." in *Speech Prosody*, 2010.

[21] S. Brognaux, B. Picart, and T. Drugman, "A new prosody annotation protocol for live sports commentaries," in *Interspeech*, 2013.

[22] J.-P. Goldman, "Easyalign: an automatic phonetic alignment tool under Praat," in *Interspeech*, 2011, pp. 3233–3236.

[23] S. Brognaux, S. Roekhaut, T. Drugman, and R. Beaufort, "Train&Align: A new online tool for automatic phonetic alignments," in *IEEE Workshop on Spoken Language Technologies*, 2012. [Online]. Available: http://cental.fltr.ucl.ac.be/train_and_align/

[24] V. Colotte and R. Beaufort, "Linguistic features weighting for a text-to-speech system without prosody model," in *Interspeech*, 2005, pp. 2549–2552.

[25] V. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals." *Soviet Physics Doklady*, vol. 10 (8), pp. 707–710, 1966.

[26] L. Warnant, *Orthographe et prononciation en français. Les 12000 mots qui ne se prononcent pas comme ils s'écrivent.*, Duculot, Ed., 1996.

[27] A. B. Hansen, "Etude du e caduc - stabilisation en cours et variations lexicales," *Journal of French Language Studies*, vol. 4, pp. 25–54, 1994.

[28] A. B. Hansen and M.-J. Hansen, *Structures linguistiques et interactionnelles dans le français parlé. Actes du colloque international*. Museum Tusculanum Press, 2003, ch. Le schwa prpausal et l'interaction.

[29] J. W. de Reuse, "La phonologie du français de la région de Charleroi (Belgique) et ses rapports avec le wallon," *La linguistique*, vol. 23, pp. 99–115, 1987.

[30] F. Argod-Dutard, *Eléments de phonétique appliquée*, Colin, Ed., 1996.

[31] J.-P. Goldman, M. Avanzi, A. Lacheret-Dujour, A. C. Simon, and A. Auchlin, "A methodology for the automatic detection of perceived prominent syllables in spoken French," in *Interspeech*, 2007, pp. 98–101.