INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Perception and production of word-final /ʁ/ in French

Cedric Gendrot
Université Sorbonne-Nouvelle Paris 3 - Laboratoire de Phonétique et Phonologie UMR 7018, FRANCE
cgendrot@univ-paris3.fr

Abstract

Variability of (French) /ʁ/ is a frequently studied phenomenon showing that /ʁ/ can have multiple realizations. In French, all these studies were undertaken using small read corpora and we have reason to believe that these corpora don't allow to look at the full picture. Indeed factors such as local word frequency, as well as speech rate can have almost as much influence as phonemic context in the realization of /ʁ/

According to Ohala's Aerodynamic Voicing principle, /ʁ/ would tend to be either an unvoiced fricative or a voiced approximant. We chose to analyze word final /ʁ/s as they tend to embrace the largest spectrum of variation. The study realized here is two-fold: a perception study in a specific phonemic context, between /a/ and /l/, where /ʁ/ is realized as an approximant, so as to better understand the parameters and their thresholds necessary for /ʁ/ identification, and provide a measure of rhoticity.

In a second step, keeping the rhoticity measurement in mind, we analyzed the realizations of word final /ʁ/s in two continuous speech corpora and modelled the realization of /ʁ/ using predictors such as diphone and digram frequency, phonemic context and speech rate.

Index Terms: phonetics, production, perception, predictors, large corpora.

1. Introduction

1.1. Variation of French /ʁ/

Chafcouloff ([1], [2]), Fougeron [3], Tranel [4], Walter [5], and Meunier [6] for French have mentioned the variability of /ʁ/ that can be realized from voiced ([ʁ̬]) to unvoiced ([ʁ̥]), fricative ([ʁ̞]), or approximant ([ʁ̞]). Numerous studies have also mentioned the variability of /ʁ/ for other languages ([7] and references therein, [8], [9]), with a tendency to derhoticization. From a phonological point of view, /ʁ/ might be considered both as a fricative or as an approximant and evolutions mentioned in the studies above tend to make it go from one category to the other. Considering the latter point, the specification of its voicing may also be reconsidered [10].

Word-final /ʁ/, whether it is followed by a pause or not, seems to be especially subject to variability, which can even lead to its elision. We consider here that the variability of French /ʁ/ is not fully apprehended, since it has been mainly studied in read speech corpora and sometimes even based on impressionistic judgments. In this study, we will make use of a perception test, as well as an acoustic study on two large corpora of continuous speech, so as to better understand the variability of French /ʁ/ in terms of perception and acoustic realization.

The use of large corpora may have two important consequences: (1) finding more variability than in read speech; (2) but also allowing to take into account more natural predictors for the understanding of speech variability. Beyond predictors such as phonemic context and position in syllable/word, we believe that predictors such as local speech rate, lexical frequency, or position within the prosodic phrase are important.

1.2. Perception of French /ʁ/

The criteria used by human listeners for /ʁ/ identification are still unclear, especially if we consider its variability. Are there any stable criteria for /ʁ/ perception whatever its realization? We believe that it is necessary to start from a perception test that will relate the identification of /ʁ/ to its acoustic realization. Very few studies are to be found on /ʁ/ perception, particularly in French. The use of large corpora allows to provide a great number of occurrences by the same speaker and thus constitute a continuum thanks to an appropriate selection of items, this continuum being actually more natural than synthesized stimuli. The details of this selection are developed in section 2.1.

1.3. Physiology of /ʁ/

The variability of /ʁ/ might be explained by its physiological realization. If we consider the aerodynamic voicing constraint [11], the approximant or fricative status of /ʁ/ implies *in fine* a dichotomy between unvoiced fricative and voiced approximant. Especially since French /ʁ/ is a posterior (uvular) phoneme and that the back cavity of the vocal tract is thus much reduced, oral pressure soon reaches the level of sub-glottal pressure and "puts out" voicing. French /ʁ/ could then be realized as an unvoiced fricative or as a voiced approximant according to the degree of constriction. This is what -according to us- explains the great variability of French /ʁ/: the stricture parameter will not only influence the approximant/fricative continuum, but also the voicing feature, these two dimensions being correlated. A study comparing acoustic and EMMA data in [12] validates this hypothesis.

2. Investigation method

2.1. Perception test for /ʁ/ identification

The aim of this experiment is to present natural digrams distinguished by the phonological presence of /ʁ/, (for example "par les" [paʁ#le] vs. "pas les" [pa#le] (*by the* vs. *not the*) to listeners, and to ask them to distinguish between the two. Some items are provided in the multimedia section.

Three types of sequences were chosen for their frequency in our corpora: "par les" [paʁ#le] vs. "pas les" [pa#le], "par la" [paʁ#la] vs. "pas la" [pa#la], "par le" [paʁ#lœ] vs. "pas le" [pa#lœ]. It was judged impossible to use monograms since these items were too short. Four speakers were chosen in the

http://dx.doi.org/10.21437/Interspeech.2017-990

two corpora detailed below on the basis of the largest number of occurrences and their diversity in terms of production. Digrams with a pause between the two words were not selected.

Three men and one woman were chosen: three speakers (two men, one woman) on the broadcast corpus and the remaining male speaker on the spontaneous speech corpus. A minimum of three items (and up to five items) per speaker and per type of digram were kept for the perception test: one occurrence where /ʁ/ is underlyingly absent ("pas les" or "pas la" or "pas le"), one occurrence where /ʁ/ was judged as hyper-articulated and one occurrence where /ʁ/ was judged as reduced ([ʁ̞]). Judgments were provided by three phoneticians (including the author). When that was possible, an occurrence where /ʁ/ was judged as elided by the same phoneticians and another occurrence where /R was reduced were added. The selection was made by gathering items with comparable f0 and duration. A total of 43 items were selected for the perception test.

The test was performed on a laptop with 23 students of 1st to 3rd year linguistics. A PRAAT script was written so as to present the different stimuli according to the following procedure. Five repetitions per item were used in this test, in a semi-random order so as to avoid repetitions and to vary the distribution of speakers. The experiment took place in three distinct steps, one for each type of digram (respectively "par les", "par la", and "par le"). After choosing between two propositions by clicking on the appropriate cell ("pas les" or "par les", "pas la" or "par la", "pas le" or "par le"), participants were asked to indicate on a scale from 1 to 5 the degree of certitude with which they answered, 5 being the maximal score of certitude. A training session was performed with other items in order to familiarize the participants.

2.2. Corpora and measurements

Two kinds of corpora were used for this study: the ESTER corpus [13], a broadcast corpus considered as prepared speech rather than read speech, with few sequences of spontaneous speech; and the NCCF (*Nijmegen Corpus of Casual French*), detailed in Torreira et al. [14] and considered as spontaneous speech. In both cases, orthographic transcription and rough segmentation have been realized by human transcribers, then phonemic and lexical alignment have been realized automatically by the LIMSI alignment tool [15].

/ʁ/ phonemic segmentation has been corrected on both corpora by human transcribers: it shows that if beginning and end phoneme boundaries are usually too long (approximately 15 ms in total), median position of /ʁ/ remains correct. While /ʁ/ elision is rarely predicted by automatic alignment tools, the presence of /ʁ/ is thus fixed by transcribers when necessary. When /ʁ/ is realized as an unvoiced fricative, spectral information contained within the phone itself doesn't cause any issue for its localization and segmentation. However, when it is realized as a voiced approximant, it tends to derhoticize and to be difficult to localize, which is frequently the case in the phonemic context we have chosen here (i.e. preceded by a vowel and followed by a sonorant). These troubles led us to perform a perception test (mentioned in section 2.1) and to elaborate syntagmatic measurements, i.e. relative to the surrounding phonemes (described in the next section).

Acoustic measurements were carried out afterwards on the items used for the perception test. The widest possible panel of measurements was chosen in order to correlate acoustic measurements with the results of the perception test. These were performed automatically with PRAAT and checked manually. Prosodic measurements (duration, f0 and intensity of /ʁ/ and the different phonemes surrounding it) and spectral measurements (frequency, bandwidth and amplitude of formants, spectral moments of /ʁ/ and the different phonemes surrounding it). HNR (harmonic-to-noise ratios) were also taken. Measurements were carried out from the beginning of /a/ (of the word "par" or "pas" to the end of /l/ (of the word "les", "le" or "la") at 20%, 40%, 60%, and 80% of the duration of the whole sequence /aRl/. Ratios between every measurement point allowed us to gather relative measurements between /ʁ/ and its preceding phoneme /a/, but also between /ʁ/ and its following phoneme /l/. The point of these measurements was to check whether variations of values could be more pertinent than absolute values. Relative measurements taken between /ʁ/ and /l/ did not prove significant and won't be developed any further.

3. Perception results

Results of the perception test are detailed in table 1. All in all, sequences containing "pas les" were identified as such, i.e. with no presence of /ʁ/, at 96.25%. Sequences "par les" where /ʁ/ was judged as fully realized are very frequently identified as such (97.1%), 87.25% for cases where /ʁ/ was judged approximant by the phoneticians, and more surprisingly up to 74.4% for /ʁ/ that phoneticians had judged as elided. These results, and especially the latter, indicate that acoustic cues remain in the signal, even though the presence of /ʁ/ is judged as questionable by the phoneticians. It has to be mentioned that only one item "par" (one /ʁ/ that was judged approximant by the phoneticians) was identified as "par" by less than 50% of the participants. The second other item that gathered the least "par" identifications is an item where /ʁ/ was judged as elided by the phoneticians, and its identification rate goes as high as 62.5%. These two items were produced by the same male speaker - from the broadcast corpus - which gives first hints about speaker variability. The scores of degrees of certitude indicate a stable mean (cf. table 1) and close to the maximum (approximately 4.5 out of 5) for answers that correspond to the majority of votes (in bold in table 1).

Table 1. *Summary of the identification rates of "par" (and its different realizations) and "pas"*

	pas les /pas le / pas la	par (ø) les / par (ø) le / par (ø) la	par (ʁ̞) les / par (ʁ̞) le / par (ʁ̞) la	par (ʁ) les / par (ʁ) / le par (ʁ) la
"par" identification	3.75% (par) **96.25% (pas)**	**74.4% (par)** 25.6% (pas)	**87.25% (par)** 12.75% (pas)	**97.1% (par)** 2.9% (pas)
certitude (from 1 to 5)	2.95 (par) **4.48 (pas)**	**4.14 (par)** 3.19 (pas)	**4.3 (par)** 2.87 (pas)	**4.5 (par)** 2.8 (pas)

In a first step, we performed a logistic regression between the word form "par" (judged as fricative) and the word form "pas" so as to identify the acoustic criteria between both underlying forms. The /ʁ/ identification is characterized for all speakers by a F2 decrease (p=0.012) and a decrease in duration of the preceding vowel (p=0.01) and of the sequence /aRl/ (p=0.04) for the item "par" compared to "pas". An increase in f0 and F1 values, as well as a decrease in HNR ('harmonic-to-noise-ratio') are also observed for three out of four speakers.

In a second step, the different kinds of /ʁ/ judged by the phoneticians (not realized/elided, approximant, realized as a fricative) were compared in order to identify the criteria for the variation of realization of /ʁ/. A linear regression has been performed and revealed the acoustic parameters implied when /ʁ/ is more frequently identified as present in the perception test. A decrease in F2 (p=0.005) and a decrease in duration of the sequence /aRl/ (p=0.013) firstly allow to predict significantly a higher rate of detections of /ʁ/. The HNR decrease also reaches the level of significance with p=0.045.

4. Acoustic Corpus analysis

Since acoustic parameters implied in the detection of /ʁ/ in both preceding subsections are the decrease in F2 (or more precisely F1 and F2 coming closer to each other from the vowel into the /ʁ/), the decrease in duration of the preceding vowel with the duration of the /aRl/ sequence, and the decrease in HNR, we gathered these parameters in a rhoticity measurement calculated as followed:

$$Rhoticity = \frac{1}{\mathrm{HNR} \times (\mathrm{F2} - \mathrm{F1}) \times \mathrm{duration\ of\ prec.vowel}}$$

The obtained value is thus normalized between 0 and 1, with 1 being the most rhotic.

A clustering tree (cf. figure 2) calculated in R software (version 2.15.1) with the *ctree* function allowed to determine the thresholds for identification of /ʁ/ at 0.5 for our rhoticity measurement (/ʁ/ being detected higher than this threshold), with a variation of F2 estimated at 73 Hz and a decrease in vocalic duration of around 30 ms and a decrease in HNR of 4.

Acoustic measurements on sequences / aR / (in word final position and for the same four speakers, 5000 occurrences in total) were performed on the whole corpora in the same way as measurements realized on stimuli from the perception test, i.e. by analyzing variations between the phoneme preceding /ʁ/ and /ʁ/ itself.

The rhoticity measurement was used as a dependent variable and predictors were local speech rate (measured as the number of phonemes per second, including pauses), phonemic context, position within the word or the prosodic phrase, the number of syllables in the word, its grammatical category, the lexical frequency of the word calculated in the same corpus. The position of /ʁ/ within the prosodic phrase was calculated after a specific automatic annotation based on a previous work [16] with two positions retained only, phrase final or phrase median position. Linear regressions were performed and showed that the best prediction of rhoticity is obtained with the three following predictors: lexical frequency of the word containing /ʁ/ (p=0.001), position in prosodic phrase (p=0.012) and speech rate (p=0.04). Altogether they predict a variation of /ʁ/ with a R^2 of 30.2%. Occurrences measured with a low rhoticity index (cf. figure 4) have been checked in the corpora and appear difficult to identify perceptively, but these will have to be tested in another perception test in a next step.

In a second step of this acoustic analysis, we extended the analysis to all speakers and extracted 6300 /ʁ/ in word-final positions preceded by any vowel and followed by a pause [6]. In these contexts, /ʁ/ can also be unvoiced and formants cannot be measured, HNR alone was thus used for rhoticity measurement as duration is also too dependent on the vowel. Recall that a lower HNR value theoretically stands for a higher (with less voicing) friction and vice-versa. F0 difference was measured on the preceding vowel so at to detect intonational phrase final (when f0 is rising) versus sentence final /ʁ/ (when f0 is falling).

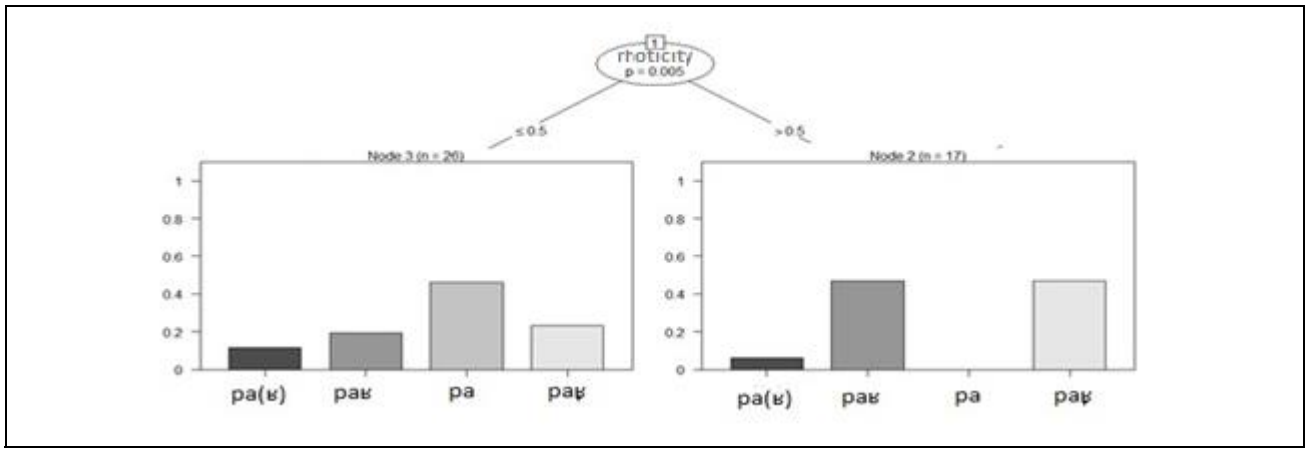

Figure 2. *Clustering tree indicating the detection threshold in our perception test.*

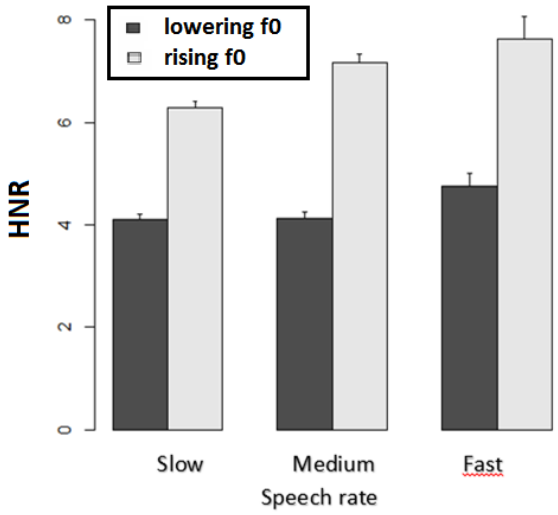

Figure 3: *HNR as a function of speech rate and preceding f0 contour: word final /ʁ/ preceded by a vowel and followed by a pause*

Figure 3 shows that HNR is lower when the f0 contour on the preceding vowel is lowering ($p<0.01$) and when speech rate is slower ($p<0.05$), thus indicating a more fricative and less voiced /ʁ/ in these contexts (also cf. figure 5, for an example).

5. Discussion and conclusion

For the voiced approximant /ʁ/ analyzed in the perception test, we were able to provide a rhoticity measurement for French uvulars, characterized by a decrease, in HNR, F2 and of the preceding vowel duration, altogether favoring /ʁ/ perceptual identification. This measurement is in turn influenced by speech rate, position in prosodic phrase and lexical frequency, factors not mentioned in the previous French literature, which suggests that observed variability for /ʁ/ is very similar to that observed for other phonemes. However it seems more noticeable perceptually because of aerodynamic constraints as mentioned in 1.3.

We extended this analysis using HNR measurements only and confirmed the results obtained in the specific context of the perception test: /ʁ/ acoustic realization is dependent on f0 contour of the preceding phoneme as well as local speech rate. These results suggest that /ʁ/ is to be considered as a fricative when realized in its full form (longer with a complete articulatory gesture), and its approximant realization is seemingly a shorter/reduced form.

HNR measurement for /ʁ/ is currently used in a follow-up study as a primary acoustic cue for didactic purposes in a real-time analysis tool ReperMe ! allowing for the user to practice the French /ʁ/ with a microphone (http://www.univ-paris3.fr/anr-reper-231657.kjsp).

6. Acknowledgements

This study was funded by the ANR project jeunes chercheurs ("junior researcher") REPER (2013 edition) ANR-13-JSH2-0005-01 and was partly funded by Labex EFL CGI

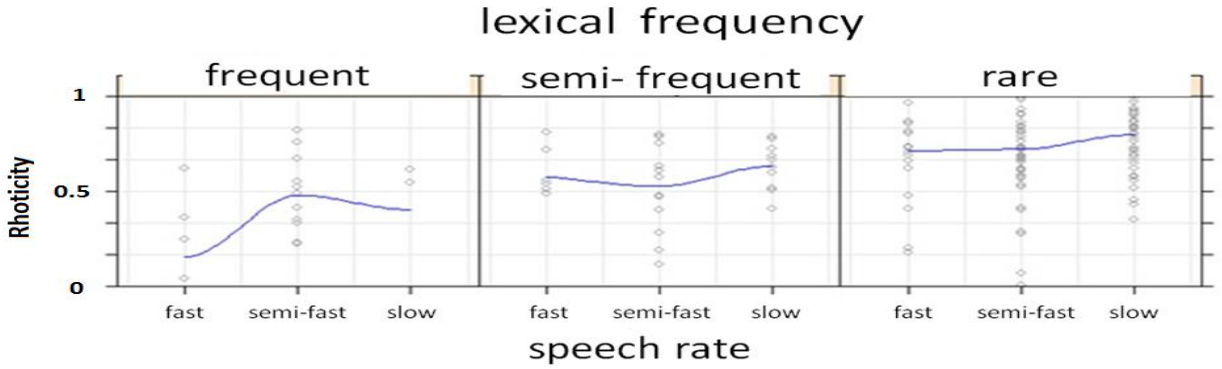

Figure 4. */ʁ/ variation according to our rhoticity measurement as a function of speech rate and lexical frequency.*

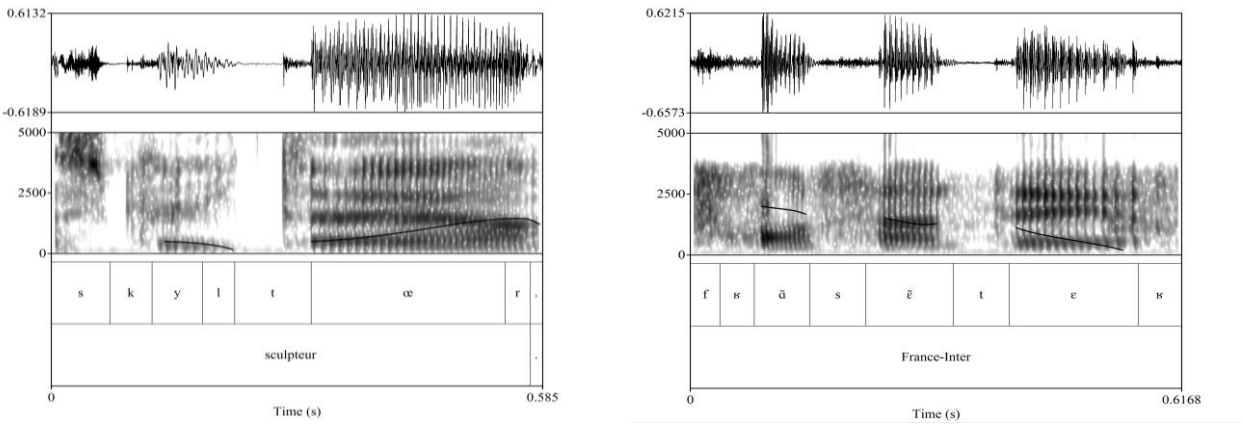

Figure 5. *Examples of /ʁ/ variation according to f0, voiced on the left and unvoiced on the right.*

7. References

[1] Chafcouloff, M. Les caractéristiques acoustiques de [j, ɥ, w, l, r] en français. In *Travaux de l'Institut de Phonétique d'Aix,* 1-52: 1980.

[2] Chafcouloff, M. A propos des indices de distinction /l-R/ en français. *Speech Communication.* 2:137–139, 1983.

[3] Fougeron, C. Word boundaries and contrast neutralization in the case of enchaînement dans French. In Cole, J. & Hualde J.I. (eds) *Papers in Laboratory Phonology IX: Change in Phonology*, Berlin: Mouton de Gruyter, 609-642, 2007.

[4] Tranel, B. *The Sounds of French: An Introduction*. Cambridge: Cambridge University Press, 1987.

[5] Walter, H. *Le français dans tous les sens*. Paris : Robert Laffont, 1988.

[6] Meunier, C. Les groupes de consonnes : problématique de la segmentation et variabilité acoustique. Thèse de Doctorat. Université d'Aix-en-Provence, 1994.

[7] Maddieson, I. *Patterns of Sounds*. Cambridge: Cambridge University Press, 1984.

[8] Lawson, E., Stuart-Smith, J., Scobbie, J. Articulatory Insights into Language Variation and Change: Preliminary Findings from an Ultrasound Study of Derhoticization in Scottish English. *University of Pennsylvania Working Papers in Linguistics.* 14(2), 2008.

[9] Mielke, J, A. Baker, and D. Archangeli. Variability and homogeneity in American English allophony and retraction. Variation, Detail, and Representation. *LabPhon 10*. Berlin: Mouton de Gruyter. 699-719, 2010.

[10] Russell Webb, E. Explanatory adequacy in phonology: a deductive approach to /ʁ/. *Linguistica Atlantica*, 25: 77–94, 2004a.

[11] Ohala, J. J. The origin of sound patterns in vocal tract constraints. In: MacNeilage, Peter (ed.): *The production of speech*. New York: 189-216, 1983.

[12] Gendrot, C. Kühnert, B. Demolin, D. (2015). Aerodynamic, articulatory and acoustic realization of French /ʁ/. 18th International Congress of Phonetic Sciences (ICPhS'15), Aug 2015, Glasgow, United Kingdom. Proceedings of the 18th ICPHS 2015

[13] Galliano S., Geoffrois E., Mostefa D., Choukri K., Bonastre J.-P., Gravier G. The ESTER Phase II evaluation campaign for the rich transcription of French broadcast news, *Proceedings of Eurospeech-Interspeech*, Lisbonne, 1149-1152, 2005.

[14] Torreira, F., Adda-Decker, M., & Ernestus, M. The Nijmegen corpus of casual French. *Speech Communication*, 52: 201-212, 2010.

[15] Gauvain, J.L., Lamel, L. and Adda, G. The Limsi Broadcast News Transcription System, *Speech Communication*, 37(1-2): 89-108, 2002.

[16] Gendrot, C. and Gerdes, K. "Prosodic hierarchy and spectral realization of vowels in French", In : Yoo, H-Y & Delais-Roussarie, E. (eds), Actes d'Interfaces Discours Prosodie, http://makino.linguist.jussieu.fr/idp09/docs/IDP_actes/Articles/gendrot.pdf, 191-205, 2011.

[17] Severin, F., Bozkurt, B., Dutoit, T. HNR extraction in voiced speech, oriented towards voice quality analysis. Proceedings of EUSIPCO'05, Antalya,Turkey, 2005.