

THE PROSODIC VARIABILITY OF SPEECH IN A NOISY CONTEXT

N. Kadiri, N. Vigouroux, G. Pérennou

Université Paul Sabatier, IRIT, 118, Route de Narbonne, F- 31062 Toulouse Cedex

ABSTRACT

This paper reports on the prosodic variability (phoneme and word duration, fundamental, vocalic release frequency) which occurs in quiet and noisy conditions, with the presence or not of feedback. The study revealed that: 1) the type of noisy condition and the presence or not of feedback changes the prosodic parameters, 2) these variations are also speaker-dependent, and 3) the vocalic release frequency observed is also function of the nature of the noise/feedback.

INTRODUCTION

In the presence of noise, speech is masked and its production is modified by what is called the Lombard effect [1]. Lane and Tranel [2] showed that the speaker modifies his speech production while speaking in a noisy environment. In particular this study illustrated:

- phonetic cues and features for normal and Lombard speech change,
- prosodic parameters: not only does the loudness (intensity) of the speech increase but also the fundamental frequency and the speech rate.

Such observations are also reported by different authors [3], [4], [5], [6] and [7].

The Lombard effect has been neglected in speech recognition systems until a recent date [4]. In several experimental studies Furui [7] reported that the Lombard effects have a greater effect on speech recognition than does the direct influence of noise entering by microphones, for example.

This paper briefly reports our analysis results observed on the French database BD-BRUIT [9]. Firstly, we will show how the type of noisy conditions and the presence or not of auditory feedback influences the prosodic parameters (phonemes and word duration, pitch)

changes occurring in Lombard speech. We will also discuss the vocalic release frequency observed on the corpora as a function of noise condition, speaker and type of phonological endings.

EXPERIMENTAL DATA

The corpora of our study are issued from BD_BRUIT data base [9]. The recordings have been done under five conditions of noise:

- the reference condition without noise, noted REF,
- white noise without audio feedback, noted BB,
- white noise and audio feedback, noted BB_ret,
- cocktail party noise without feedback, noted Amb,
- cocktail party noise and audio feedback, noted Amb_ret.

Five men and five women have pronounced five times the corpus of ten isolated digits from zero to nine, in the five conditions described above. We have carried out two hand-labelling levels: one in phoneme units (1) and the other in word units (2), both on a temporal and a spectral representation of the speech signal.

The labelling (1) was carried out for two repetitions per speaker and per condition, e.g. a total of 4700 phonemes (47 phonemes per repetition).

The labelling (2) has been made for the complete speech data base files (e.g. 2500). At the end of words, a vocalic release appears sometimes. It can be interpreted as the realization of an underlying schwa such as in the word "quatre" [katrə]. In other cases, it is a release appearing with a final obstruent like in the word "six" [sis]. These releases have not been considered as being a part of the final obstruent (plosive or fricative) which precedes them for the duration of phonemes.

METHODS: VARIANCE ANALYSIS

In order to better explain the set of available observations, we have carried out an analysis of the variance which consists in explaining a quantitative variable by a set of qualitative variables called factors. The phoneme and word duration, the fundamental frequency and vocalic release frequency were analyzed by separate repeated analysis of variances (ANOVAs). Noise condition (noted Noise), the speakers (noted Loc) and the auditory feedback (noted Ret) were the factors. This analysis of variances (ANOVA) has been done with "PROC GLM" of the SAS software [10].

The test consists in observing the probability that a "law of Fisher" with the appropriated degrees of freedom exceeds the statistic of Fisher "F value". If we decide to do the test at 5% (respectively at 1%), we say that the variable has an effect if the probability is inferior to 0.05 (respectively to 0.01).

We have preferred to explain the increase percentage of this parameter in comparison with the mean value of it in the reference condition.

The size of this article does not permit to present all results. So, we will report the most significant results.

DURATION OF PHONEMES AND WORDS

PHONEMES' DURATIONS

As suggested by the results of previous works [3], [5] to measure the noise effect on the duration of speech segments, we define two phonetic classes: vocalic and fricatives. The plosives were not considered because the difficulty to define a robust rule about the status of the consonantic release at the end of words like in /sɛk/ or /set/.

Variation of vowels' duration

The variance analysis shows:

- a significant increase of vowels duration in the presence of noise on average with 21% (F Value = 12.45, $p < 0.0001$),
- the significant effect of the speaker (F Value = 9.56, $p < 0.0001$)
- on the other hand, the type of noise and the auditory feedback of speech are not significant.

These results are illustrated in the figure 1 which shows the duration

increase in comparison with reference according to the noise conditions for all speakers. We will notify that the type of noise has an important influence for two speakers (130 for the Amb condition for one of speakers 29 on average for the other speakers).

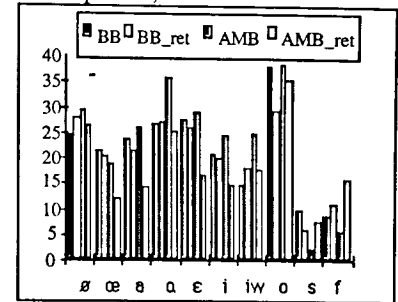


Figure 1: Increase average of the phonemes' duration according to the noise conditions (phonemes in API, the code "iw" represents a labialized /i/).

Variation of fricatives' duration

For fricatives' duration, no effect was significant for speech in noise. On the other hand, the increases of duration are significantly affected only by the speaker (F Value = 6.12, $p < 0.0001$). This result is not totally in agreement with other observations, where the consonant durations are decreasing by the effect of noise [7].

The type of noise has some significant effect on the duration of phonemes but these effects are not confirmed for consonantic phonemes in isolated speech. It seems that the speaker factor has also an influence on the phoneme variation.

WORDS' DURATION

The results of variance analysis

The computation of words' duration raises a question: how to determine the end boundary of the word. We have decided to take into account all the words of the corpora, whether they end with a plosive (like /set/) or not. In all cases the final release, voiceless or voiced, after a plosive or after a fricative, has been counted as making part of the word. However, we have separated the results into two groups in order to distinguish the case where the vocalic

release does not appear (words 0, 1, 2 and 3) and where it may appear (words 4, 5, 6, 7, 8 and 9).

We ran variance analyses to explain the variation of the word duration. We added another factor to explain this duration variation: the type of the word. The variance analysis shows:

- significant effects of the noise (F Value = 167.91, $p < 0.0001$, around 14% on average of duration increase), the speakers (F Value = 81.98, $p < 0.0001$ and the type of words (F Value = 15.51, $p < 0.0001$).

- on the other hand, the type of noise and the presence or not of the auditory feedback have no significant effect on the increase of the word duration.

The results will be presented according the two types of word endings: vocalic and consonantic.

Words with vocalic endings

The words have on average, a duration increase of about 16% in the presence of noise. The type of the noise and the presence of feedback are not significant. The variation of the word duration is also function of the word type. This can be explained by the fact that they do not contain the same number of consonants — these ones do not show a tendency to lengthen under the influence of noise. The speaker factor has sometimes an effect on the variation.

Words with consonantic endings

The results are identical as for the vocalic endings:

- the noise has a significant influence (12,5% on average of increase for all noises merged),
- the type of noise (white noise or cocktail party noise) does not influence significantly the increase of the duration,
- the speaker is also a significant factor.

The analysis of the word duration confirms the study done on phonemes. It shows that speakers have, according to noise conditions, a variable behaviour (but significant, speaker has a typical behaviour).

CONSEQUENCE OF NOISE ON THE VOCALIC RELEASE AT THE END OF WORDS

Here, we will demonstrate if the conditions of speech production have an

influence on the phonological strategy of speakers in noisy and/or feedback conditions. In a previous study [5] have observed:

- an increase of the number of omissions of consonants located at the end of the word, more particularly of phonemes /t/, /p/, and /f/

- and an insertion of phonemes in the shape of schwa appearing at the end of words in a ratio of two more times in noised speech than in reference speech.

In this study, we are interested in the observation of the schwa realizations and in its effects on the syllabic organization of the utterance. The vocalic releases at the word end, are characterized by voiced segments similar to weak central vowels. These segments appear in words like /katrə/ as well as in words like /sek/ or /sis/ ... without a final /ə/.

The results of the variance analysis

Two factors are added as explicative factors:

- the sex of the speaker
- and the phonological type of the word ending. We define three final types: T: for words ending with a plosive: /sɛk/ /set/, and /ɥit/,

S: for words ending with a fricative: /sis/ and /nɛf/

tre: for words ending with the set of phonemes /trə/ in the word /katrə/.

The variance analysis shows the significant influence of the noise condition, the speaker, the speaker sex (216 vocalic releases for men and 133 for women), the phonological type of the words' ending and also the effect of the auditory feedback (131 vocalic release with feedback and 216 without feedback).

A high increase is to be noted at the end of noise. In the absence of auditory feedback, this percentage is multiplied by 4. The feedback reduces in a sensible way this influence. The type of noise does not seem to have a real effect.

The figure 2 shows the effect of noise conditions and of phonological type on the frequency of vocalic release realization. For the whole noise conditions, the phonological type of the ending has an important influence on the rise of the realization percentage: the increase of type 'tre' is the double of the type S. The type 'tre' gives the most important

increase since it is almost four times more important than that of type S. The percentage of vocalic release of the words ending with a plosive is more important than those ending with a fricative.

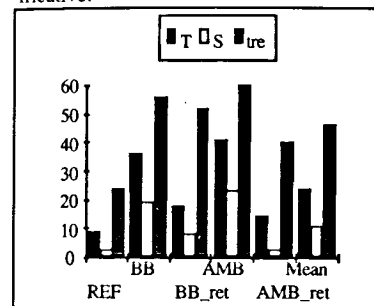


Figure 2 - Percentage of vocalic release.

The feedback has a constant effect on the rise of this percentage. The presence of vocalic release at the end of the word is significantly dependent of the speaker, of the noise conditions, of the auditory feedback, of the speaker sex and of the phonological type of the end of the word. On the other hand, the nature of noise (white noise or cocktail party noise) has no influence on this presence.

NOISE CONSEQUENCE ON FUNDAMENTAL FREQUENCY F_0

[5], [6], and [7] indicate an increase of the fundamental frequency (often correlated to an intensity increase) under the noise. The maximal value of the fundamental frequency (noted F_0) for each occurrence of the word was computed. The study of the F_0 variation to reference conditions shows:

- the significant influence of the speaker factor (F Value = 337.21, $p < 0.0001$), of the noise factor (F Value 588.58, $p < 0.0001$, on average 20,25% of fundamental increase for the noise merged), of the feedback (F Value 75.86, $p < 0.0001$).

- the no effect of the type of the word. Like observed by [6] our analyses of fundamental revealed no further significant differences between conditions (white noise and cocktail party noise) for isolated words. The type of noise is more significant without feedback than with auditory feedback. The variation is also dependent of the

speaker.

CONCLUSION

To gain an understanding about the Lombard effect we analyzed the change of prosodic and phonologic parameters occurring in the five conditions of noise. The analyses made on French speech corpora., to confirm previous studies on English and American, demonstrated the effect of the phonologic strategy of speakers due to the noisy/feedback conditions.

REFERENCES

- [1] Lombard E. (1911), "Le signe de l'élévation de la voix", Ann. Maladiers Oreille, Larynx, Nez, Pharynx, 37 pp.101-19.
- [2] Lane H., Tranel, B. (1971), "The lombard Sign and the Role of Hearing in Speech" in Speech and Hearing, 14.
- [3] Summers, W., Pisoni D.B., Bernacki R.H., Pedlow R., Stokes M., (1988) "Effects of Noise on Speech Production : Acoustic and Perceptual Analyses", in J. Acoust. Soc. Am., 84(3) : pp. 917-28.
- [4] Hassen J. and Bria O. (1990), "Lombard Effect Compensation for Robust Automatic Speech Recognition in Noise" in Proc. of ICSLP-90, pp. 1125-8.
- [5] Junqua J.C., Anglade Y. (1990), "Acoustic and Perceptual Studies of Lombard Speech: Application to Isolated-Word Automatic Speech Recognition", in Proc. of ICASSP-90, pp. 841-4.
- [6] Howell P., Young K. and Sackin B. (1992), "Acoustical changes to speech in noisy and echoey environments", in Proc. of ESCA on Speech Processing in Adverse Conditions, Antibes, pp. 43-51.
- [7] Junqua J.C. (1992), "The variability of Speech Produced in Noise", in Proc. of ESCA on Speech Processing in Adverse Conditions, Antibes, pp. 223-5.
- [8] Furui S. (1992), "Toward Robust Speech Recognition Under Adverse Conditions" in Proc. of ESCA on Speech Processing in Adverse Conditions, Antibes, pp. 31-42.
- [9] BD_BRUIT (1993), "Base de Données Parole, Locuteurs soumis au Bruit", GDR-PRC-CHM et ICP, Février.
- [10] SAS/STT (1990), Users Guide, Volume 2 GLM-VARCOMP, version 6, SAS Istitute Inc.