# Assessment of Prosody Disturbances in Stutterers by means of phonetic Indices

*Audrey Leclercq, Kathy Huet, Myriam Piccaluga, Bernard Harmegnies*

Speech sciences laboratory (AUWB)
University of Mons, Belgium

audrey.leclercq@umons.ac.be, kathy.huet@umons.ac.be, myriam.piccaluga@umons.ac.be, bernard.harmegnies@umons.ac.be

## Abstract

This paper is focused on the contribution of phonetic tools used in prosody analyzes for the study of stuttering. Three contrasted subjects (1 Non Stuttering Person, NPS and 2 Person Who Stutter, PWS with contrasted stuttering profiles) have been recorded while performing a map task under varying auditory feedback conditions (NAF, DAF 80, DAF 120 & DAF 160). The fluency index used (ISI) reveals differences between subjects and conditions, and therefore appears as a possible tool for use in further research aiming at a better understanding of stuttering and/or at the development of reliable assessment methods is discussed.

**Index Terms**: speech rate, ISI, stuttering, fluency

## 1. Introduction

Persons who stutter (PWS) encounter various difficulties in delivering the speech sounds corresponding to the phonetic realizations of the messages they intend to emit. Those difficulties result in objective perturbations of the flow of speech, with, e.g., unexpected repetitions of phonic events, sudden blockings, collateral verbal tics, etc. Rhythm disturbances and various other consequences for other aspects of their prosody may appear. They result, for the listener, in a global sensation of dysfluency The precise etiology and the nature of processes involved in stuttering are still largely debated. Nevertheless, the phenomenon has been widely studied, but very often as a pathology that speech therapists, physicians, psychologists, etc. aim at curing. Usually, the descriptions of the phenomena remain quite macroscopic: they may consist, e.g., in counting the number of words or stuttered syllables [1], or trying to assess the degree of severity of the pathology [2]. They are very often based on subjective expertise, sometimes relying on scales aiming at objectivising the judgment [3], or even based upon self assessment [4]. Curiously, very few scientific work has been devoted to this pathology by phoneticians and phonologists, though the phenomena involved in stuttering precisely fall into their field of expertise. Moreover, apart from some very interesting studies focused on English-speaking stutterers [e.g., 5] the phonic aspects of stuttering have very rarely been studied in languages of the Romance group [6]. In order to tend to a better understanding of stuttering, and also for the purpose of better assessing the results of various therapeutic approaches, objective and reliable descriptions are nevertheless needed. This paper is meant as an exploratory essay in this framework, aiming at testing the interest of a phonetic index (the Inter Syllabic Interval: ISI), originally devised to evaluate interpreters' fluency [8, 9]. The language involved is French. In order to maximize the variability of the phenomena to be observed, and therefore better assess the interest of the variables under study, we will 1. maximize the inter-suject variability by studying 2 PWSs with contrasted stuttering profiles, alongside with a NSP and, 2. maximize the intra-subject variability by using various conditions of auditory feedback, since this technique has proved to be a source of improvement in stutters [6].

## 2. Method

### 2.1. Subjects

Our 3 contrasted subjects are Belgian males aged 25-30 years. They all are monolingual and French is their mother tongue. We collected data from two persons known as stutterers since early childhood (PWS1 and PWS2). From a clinical point of view, PWS1 stuttering would rather be classified as "clonic", since it is characterized by a large amount of repetitions (of phonemes, syllables, words and sentences) whereas PWS2 stuttering would rather be considered as "tonic", given the frequent blockings that appear in his speech flow. We also recorded a subject which has never been said to be a stutterer, and does not show any sign of stuttering (NSP).

### 2.2. Task

We used the same map task as in [6]: the subjects were asked to explain to a remote, silent French-speaking interlocutor how to get from a given city to another. For this purpose, they were given a detailed map covering a large area of Northern France and of the French-speaking part of Belgium. We asked the subjects to plan travels fulfilling varying constraints (e.g., avoid large cities, prefer touristic areas, etc.), in order to have them speak as much as possible. The subjects were given time to silently prepare their itinerary and were allowed to put graphic marks on the maps, but not to write any word or sentence. They were told to begin speaking only when they would have entirely prepared the itinerary. Each task was carried out under normal auditory feedback (NAF) and delayed auditory feedback (DAF). In DAF conditions, three delays were used: 80 ms, 120 ms and 160 ms. For each condition, the subject had to describe a specific journey, with specific points of departure and arrival and specific constraints. We therefore collected 4 corpora by subject.

### 2.3. Acoustical treatment

The delayed feedback was obtained by means of the Kay CSL 4300 DAF routine. All the recordings were performed in a sound proof room at the Phonetics Laboratory of the University of Mons, by means of a Neumann U87 P 48 microphone, connected to a Sony 501 ES PCM coder, connected to a Panasonic VHS video recorder. The analyses were performed using the Praat software (Version 5.1.05).

### 2.4. Events

The unit phenomenon we focus on is basically the acoustic result (phonic segment or group of phonic segments) of the realization of a syllable in the classical phonological sense, i.e., each event can be linked with a part of a word (or an entire monosyllabic word) belonging to the French lexicon. Nevertheless, in spontaneous speech (and especially in spontaneous stutterers' speech), phonic events that cannot be covered by this definition frequently appear (isolated sounds or groups of sounds corresponding to a syllable constituting a portion of word of the language used, isolated speech sounds recognizable or not as syllables onsets, filled pauses etc.). These events, all resulting from a 'unit of pronunciation' [9, p. 338], therefore corresponding to the rather wide acceptation of the term 'syllable' in a phonetic view will also be taken into account, for they are highly relevant to the topic under study. Thus the whole set of phenomena we study is divided into two groups, which we will refer to, respectively, as the one of the 'α events' (sound or group of sounds resulting from the normative pronunciation of a syllable of a word belonging to the French lexicon) and the one of the 'β events' (any other isolated sound or group of sounds recognizable as a pronunciation unit and not contributing to the normative pronunciation of a word of the French language, e.g., repetition of an isolated syllable, production of a syllable onset, production of a filled pause, etc.). In fig. 1, for the purpose of illustration, we show an excerpt of PWS1 production under NAF. The subject says *puis là, je vais continuer* ("then there, I will continue"), with repetitive productions of the initial /p/ in *puis* and of initial /k/ in *continuer*, and with intrusion of *ouais* (familiar *yes*) among the group of reiterated [p] and a filled pause ([ɔ:]) before *continuer* ([p p p p p p p p p p p p p p wɛ p p pɥi la ɔ: ʒə ve k k kɔ̃ ti nɥe]). The table 1, aligned with the spectrogram, just above it in figure 1 examples of the way in which we label the phenomena.

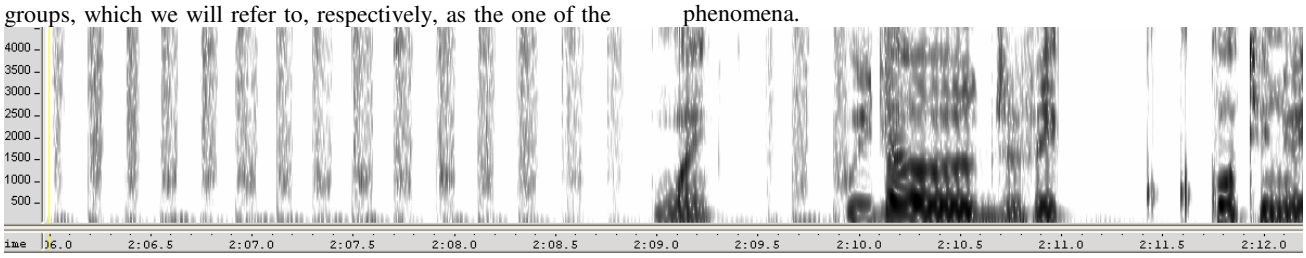


**Figure 1:** *spectrogram of an excerpt of PWS1 production under NAF*

| | p | p | p | p | p | p | p | p | p | p | p | p | p | p | p | wɛ | p | p | pɥi | la | ɔ: | ʒə | ve | k | k | kɔ̃ | ti | nɥe | **Tot** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **α** | | | | | | | | | | | | | | | | 1 | | | 1 | 1 | | 1 | 1 | | | 1 | 1 | 1 | **8** |
| **β** | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | | | 1 | | | 1 | 1 | | | | **20** |

**Table 2**: *phonetic transcription and classification of the phonic events in figure 1.*

## 3. Results

### 3.1. Distribution of the phonic events

Figure 2 shows the evolution of the global amount of phonic material produced by minute (α- and β events together) for each subject under the four experimental conditions. As can be seen, the NSP has a very stable rate under the 4 situations (159-166 events by minute).

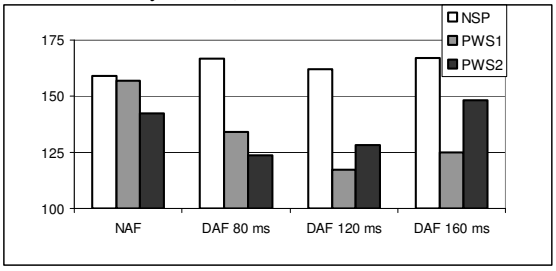


**Figure 2:** *global amount of phonic material by minute (ordinate), for each subject, in each experimental condition.*

On the contrary, both PWS exhibit changes in event rates under the DAF conditions. In general, the effect observed consists in lowering the rate, relative to the NAF condition (with the exception of PWS2 under DAF 160, with a slight 4% increase re NAF). The maximal decreases are found under the DAF 120 condition in PWS1 (25%) and DAF 80 condition in PWS2 (13%). Those observations suggest that the DAF conditions exert substantial, but varying effects on the PWS, but few if any on the NSP. However, they do not allow appreciating in any way whether the change consists in degradation or improvement. We have therefore computed the contribution of the β events to the total number of events. The resulting index is the ratio of the rate of β events by minute to the global rate of events (α and β) by minute.

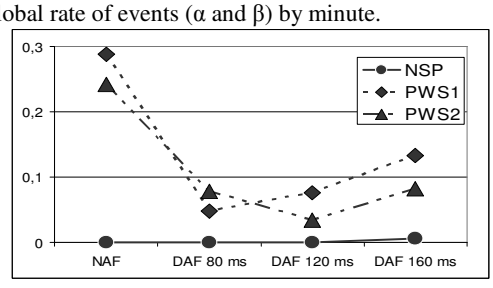


**Figure 3:** *proportion of β events relative to the global amount of phonic material by minute (ordinate), for each subject in each experimental condition.*

As shown in figure 3, the ratio is zero for the NSP (with the exception of DAF 160, for which it is very small: 2 β events on the whole set). On the contrary, the ratio is fairly high for both PWS in the NAF condition (between 25% and 30% of the phonic material uttered consist of β events). In other words, both PWS emit, when speaking in normal conditions, large amounts of linguistically non-relevant phonic material; nevertheless, under the DAF conditions, a dramatic decrease of the ratio can be observed, with values as low as .05 (PWS1)

and .03 (PWS2). It has to be noticed, too, that the maximal effects are observed at delay values of 80 ms (PWS1) or 120 ms (PWS2); under the DAF 160 condition, both PWS are characterized by values of the ratio increased though still lower than those corresponding with NAF.

## 3.2. Variations in the flow of speech

### *3.2.1. The Inter Syllabic Interval*

The study of temporal discourse structure is often grounded on the analysis of disruptions in the speech signal. The groundbreaking work done by Goldman-Eisler [e.g. 10] has shown various benefits of this approach. Nevertheless, the detection of the pauses and their categorization is a very tricky question (for a larger discussion, see [8]). Broadly speaking, analyzing a phenomenon on the basis of a "*non-event*" (a break in the phonation stream), quite apart from its epistemological limitations, can cause technical problems. In order to avoid these limitations, one of us [7] has set up an index (the Inter Syllabic Interval: ISI), which is obtained using time differentiation of intensity peaks. For each peak, the time interval separating it from the preceding peak is calculated. For a corpus of N events, a chronologically ordered list of N–1 inter-event delays can be obtained. ISI graphs show a more or less regular alternating pattern of peaks and troughs. The peaks correspond to long ISIs, which are characterized by a momentary reduction in speech delivery; the troughs correspond to a period of faster delivery speed.

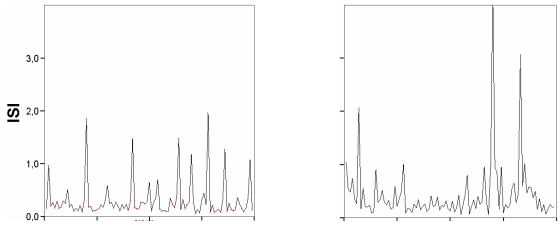


**Figure 4:** *ISI values (s) (ordinate) corresponding to 100 consecutive events (abscissa) in the flow of speech of the NSP under NAF condition (left) and of PWS1 under DAF 80 condition (right).*

For instance, in fig. 4 (left), the sequence of 100 ISI values presented (NSP in NAF) show a fairly good regularity in the subject's production. On the contrary, in fig. 4 (right), values presented (PWS1) suggest both the presence of quick clonic repetitions of short portions of speech (contiguous very low ISI values) and of phonatory blockings (very high ISI values).

### *3.2.2. ISI variability*

An overall descriptive analysis of the ISI values suggests sensitivity to the effect of the experimental condition. In the no-delay condition, the 3 subjects have very close ISI average values (378 ms – 423 ms). Under the effect of the DAF, the NSP exhibits quite stable values (360 ms – 371 ms). On the contrary, both PWS tend to have higher values under the DAF effect. The major values are 514 ms in PWS1 under the DAF 120 condition and 489 ms in PWS2 under the DAF 80 condition. For both PWS, a noticeable decrease is observed under the DAF 160 condition. The analysis of variance confirms these overall observations, exhibiting a significant subject effect ($F_{(2,3354)}$=14.788, p <.001), a significant interaction effect ($F_{(6,3354)}$=2.300, p=.032) and a weakly significant condition effect ($F_{(3,3354)}$=2,470, p= .060). Broadly speaking, it turns out that not all subjects react to DAF in the same way, that mainly PWS are affected by DAF, and that the effect consists of an ISI increase under the effect of the DAF 80 and/or DAF 120.

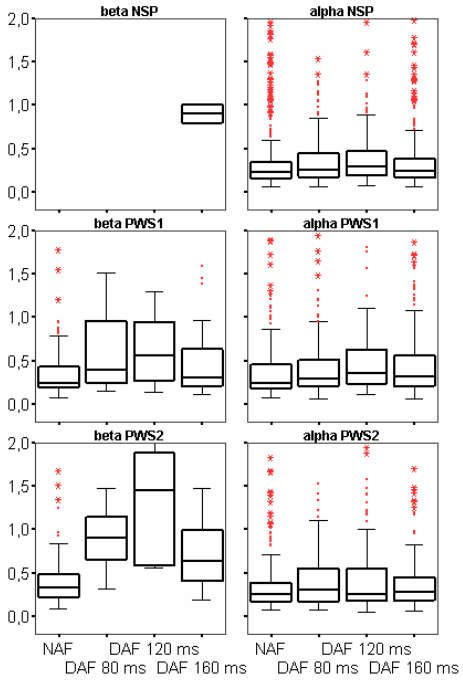


**Figure 5:** *Boxplots (medians, .25 and .75 percentiles and outliers) for the ISI values (s) (ordinate) for each event type, for each subject and for each condition.*

This inferential treatment should nevertheless be considered with caution, since, as shown by figure 5, ISI distributions tend to be fairly dissymmetrical, with concentration in the lower part of the axis and a great amount of outliers in the upper part (in fig. 5, the axis has been truncated to 2, but the greatest values are around 6.5). If medians rather than means are taken into account, the observation of a gradual ISI increase from NAF to DAF 120 holds true for PWS1 (median test: Chi Square=35,347 , df=2, p<.001) but not to PWS2 (Chi Square=1.699 , df=2, p =.428). Moreover, a significant effect is observed in NSP (Chi Square=20.338, df=2, p<.001). Nevertheless, the amplitude of the ISI increase is fairly greater in PWS1 (medians from .241 up to .372) than in NSP (medians from .222 up to .294).

On the other hand, figure 5 emphasizes that the DAF effect strongly acts on the β events, for which the increase is dramatically greater than in α events. In order to better understand this observation, it is important to have a look at the β events ISI distributions. In figure 6, distributions of the ISI values corresponding to β events only have been drawn for each PWS under each experimental condition (remembering that our NSP does not produce any phenomenon of this kind, except 2, only under the DAF 160 condition). It turns out that in NAF, both PWS exhibit concentration in very low values. These values may correspond to brief phenomena, like, e.g., syllables onsets, verbal tics, and, obviously, clonic repetitions (it is interesting to notice that, for PWS1, referred to as a "clonic stutterer", the local mode, around 200 ms, is lower than for PWS2, around 300 ms, therefore suggesting the presence of a larger amount of clonic repetitions in PWS1). When the subjects are exposed to the first two DAF

conditions, a dramatic rarefaction of the small ISI values is observed. The phenomenon is very strong, for PWS1 in the DAF 80 condition (where only one event with duration less than 500 ms is observed), and for PWS2 in the DAF 120 condition (where no phenomenon with duration less than 500 ms is observed).

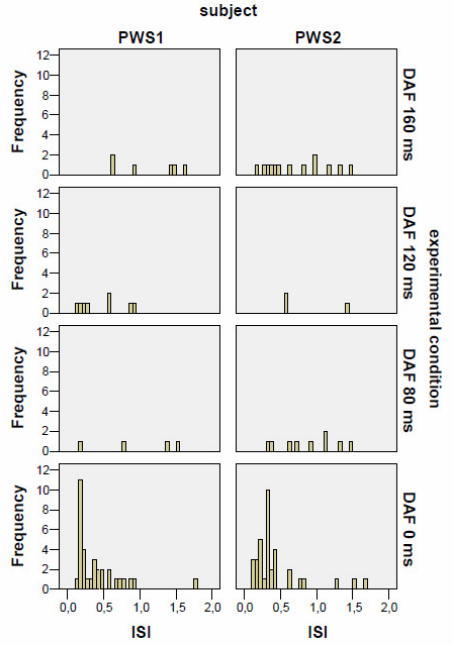


**Figure 6:** *ISI distributions for the β events (first minute of speech in PWS1 and PWS2 under the 4 experimental conditions).*

Thus, one important factor of the ISI means increase in PWS under DAF condition is the vanishing of the very short β events, which are quite numerous for both PWS under NAF, and absent in the productions of the NSP.

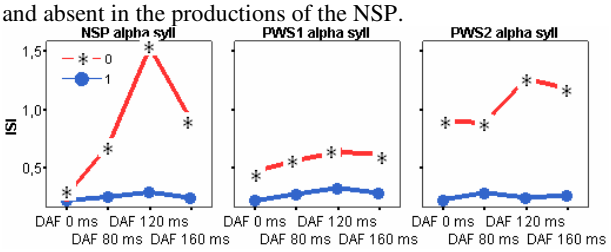


**Figure 7**: *ISI average values for each subject in each condition: α events only, in words with (circles) or without (stars) direct semantic relation with the itinerary described by the speaker.*

No similar observation can be made for α events, but the increase of their ISI values could be linked with other factors. We have re-analyzed the whole corpus at the word-level and wondered, for each α event, whether it was or not contained in an isolated word, without direct semantic relation with the itinerary described, e.g., words like *alors* ("then"), *bon* ("well"), *ouais* ("yeah"), etc. Broadly speaking, those items seem to act as fillers, but, as they belong to the lexicon of the language used, they may be linked with the momentary arousal of incidental, padding discourse.. As shown in figure 7, the events belonging to items of that kind are more sensitive to the DAF effect, included in the NSP, whose values in the DAF 120 conditions are the greatest in the sample. It is important, nevertheless, to keep in mind that those phenomena are not quite frequent (under NAF, respectively 4%, 24% and 8% of the α events, respectively for NSP, PWS1 and PWS2). Under DAF, the figures may drop under 1% for NSP and PWS2, but they remain around 10% in PWS1.

## 4. Conclusion

In this paper, focused on French-speaking PWS, we have shown that the use of a very simple, assumption free, phonetically based index (the Inter-Syllabic Interval, ISI) allows to point out striking differences in the speakers behaviours, depending upon their speaker profile (NSP vs. PWS), the conditions they have to cope with while speaking (NAF vs. DAF) and the kind of stuttering (clonic vs. tonic), if any. We have also underlined that the observed ISI distributions are far to be Gaussian; beyond the implications in terms of statistical procedures to avoid or favour, this suggests an underlying complex structure of causality, with specific groups of ISI values being related to specific phenomena (e.g., very short ISIs corresponding to clonic behaviour or very long ISIs associated with irrelevant or parallel discourse, etc.), presumably linked with varying psycholinguistic sub-processes and/or levels of language production. In future research, experiments should therefore be conducted with the aim of setting up a semiology of the ISI values, with the double aim of 1. better describing and assessing the stutterers performance under varying conditions, 2. better understanding the processes involved in coping with feedback alteration, in stutterers, as well as in non-stutterers.

## 5. Acknowledgments

Thanks are due to Profs. Dolors Poch (UAB) and Marielle Bruyninckx (UMONS) for their early cooperation and advice.

## 6. References

[1] Armson, J.and Stuart, A., "Effect of extended exposure to frequency altered feedback on stuttering during reading and monologue", JSLHR, 41, 479-490, 1988.

[2] Sparks, G., Grant, D., Millay, K., Walker-Baston, D., and Hynan, L. (2002), "The effect of fast speech rate on stuttering frequency during delayed auditory feedback", Journal of Fluency Disorders, 27, 187-201, 2002.

[3] Stuart, A. and Kalinowski, J., "The perception of speech naturalness of post-therapeutic and altered auditory feedback speech of adults with mild and severe stuttering", Folia Phoniatrica et Logopaedica, 56, 347-357, 2004.

[4] Kalinowski, J., Guntupalli, V. K., Stuart, A. and Saltuklaroglu, T., "Self-reported efficacy of an ear-level prosthetic device that delivers altered auditory feedback for the management of stuttering", International Journal of Rehabilitation Research, 27, 167-170, 2004.

[5] Subramanian, A., Yairi, E. and Amir, O., "Second formant transitions in fluent speech of persistent and recovered preschool children who stutter", Journal of Communication disorders, 36, 1, 59-75, 2003.

[6] Harmegnies, B., Bruyninckx, M. and Poch, D., "Delayed Auditory Feedback effects on stutterers' voice qualities and vowels systems", *Proc. of the 13th ICPhS*, 1, 400-403, 1995.

[7] Piccaluga, M., "Approches psycholinguistiques de l'inter-prétation", Ph D thesis, Université de Mons-Hainaut, 2004.

[8] Piccaluga, M., Nespoulous, J.-L. and Harmegnies, B., "Disfluency surface markers and cognitive processing; the case of simultaneous interpreting", *Proc. of the 16th ICPhS*, 1317-1320, 2007.

[9] Crystal, D., "A dictionary of Linguistics and Phonetics", Cambridge Ma, Basil Blackwell, 1992.

[10] Goldman-Eisler, F., "Segmentation of input in simultaneous translation, *Journal of Psycholinguistic Research*, 1, 2, 127-140, 1972.