

Functional, acoustical and perceptual analysis of vocal hesitations in spontaneous speech

Isabelle Guaitella
 Institut de Phonétique d'Aix-en-Provence
 Laboratoire Parole et Langage URA CNRS 261

ABSTRACT

The aim of this study is to reconsider and to show the interest of the analysis of a phenomenon which is specific to spontaneous speech: vocal hesitations. These latter are often assimilated to silent pauses. Acoustical investigations led us to shed some light on the perceptual and communicative functions of vocal hesitations. This study is based on a general description of the rhythmic organisation of spontaneous speech. We demonstrate that both functions and acoustic manifestations of vocal hesitation are specific.

INTRODUCTION

Among all processes of expression which can be used by speakers in order to plan their discourse (repetitions, silent pauses, etc...; see Butterworth & Goldman-Eisler, 1979; Butterworth, 1980; Siegman 1979; Guaitella 1991), only vocal hesitation cannot signify something else than the perceptible trace of the speaker's planning activity. We think that the vocal manifestation of hesitation plays a specific role in the rhythmic organisation of communication

HYPOTHESIS

Our hypothesis is that when a segmental duration cue is used simultaneously with a f0 cue, the syllable is always accented. On the contrary, when a duration cue alone is used, it can be a vocal hesitation. As a consequence, the whole rhythmic organisation has to be analysed in order to bring to light the specificities of a functional element (for the rhythmic model, see Guaitella 1991).

APPLICATION: STUDY OF INTERVIEW MATERIAL

This study deals with an extract of an interview, recorded in a sound-proof chamber. The speaker is a female, speaking a standard French. The topic is about a journey. Data are obtained from this material and completed by some examples taken from various other materials (other situations and other speakers) which are here to confirm that the results cannot be explained by individual specificities.

F0 configurations and contexts of apparition of hesitation

We observed the totality of the hesitations according to the context of appearance and disappearance (i.e. preceded or followed by text or silence), their duration and their f0 pattern.

We observed that hesitations could appear in all contexts, i.e.:

- preceded and followed by silence (22.5%)
- preceded by speech and followed by silence (50%)
- preceded by silence and followed by speech (5%)
- preceded and followed by speech (22.5%)

Four fundamental frequency patterns were observed:

- a) - Drop in pitch only (or eventually flat contour), see figure 1.
- b) - Drop in pitch including or ended by creaky voice.
- c) - "Step-drop" of pitch, i.e. minor modulations of pitch which cannot be confused with a rising contour, followed by greater falls.

d)- Decreasing following or ended by a "start", i.e. major rising contour.

For the last configurations (d) we consider that a single vowel is divided into two parts: the hesitation following or preceding an accented syllable. This point of view can be justified by perception and by the "linguistic logic" of the phenomenon: we can easily imagine that the speaker could hesitate on a vocalic segment, then accent the same segment to initialise what he is going to say (the opposite strategy is also possible).

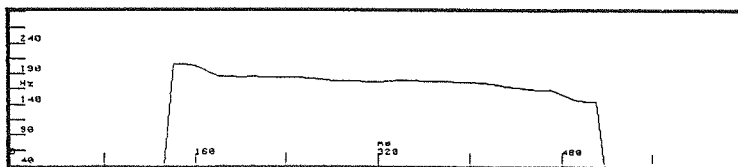


Figure 1. Example of drop in (text: "et", and).

Hesitation and duration variation

The durations of the hesitations are much much greater than those of the other syllables (Tables 1 et 2). The shortest duration is 197 ms and the longest is 1157 ms.

legend:

- T : presence of text
- # : presence of pause
- # - # : hesitation preceded and followed by pauses
- \ : progressive drop in pitch
- k : presence of creaky voice
- m : presence of modulations when decreasing
- Am : presence of a setting of a rising contour

Table 1. Durations of vocal hesitations according to the context (in ms). moy = mean; min = minimum, max = maximum.

	moy	min	max
T - *	465.4	197	712
* - T	574	420	728
T - T	413.9	237	579
* - *	657.7	393	1157

Table 2. Durations of vocal hesitations according to the prosodic configurations (in ms). moy = mean; min = minimum, max = maximum.

	moy	min	max
\	475.4	197	930
k	567.2	237	1157
m	527	420	712
Am	582.7	468	756

We can observe that duration is much longer for the hesitations preceded and followed by silent pauses.

The duration of the hesitations (the overall mean duration is 533 ms) is greater than the mean of syllable durations (184.5 ms) and less than the mean duration of silent pauses (817.1 ms). This result can be considered the consequence of the presence of three classes of segmental duration: syllable duration, hesitation duration, pause duration.

Hesitation and pitch

A falling contour is always observed on hesitations (a few cases of flat contours were also found, but never rising ones). However, the amplitude of the fall can vary (Tables 3 and 4).

Table 3. Pitch fall according to the context (in Hz). moy = mean; min = minimum, max = maximum.

	moy	min	max
T - *	- 28.6	- 10	- 67
* - T	- 46.5	- 42	- 55
T - T	- 22.75	0	- 50
* - *	- 45	- 25	- 65

Table 4. Pitch fall according to the prosodic configurations (in Hz). (For the "settings" we have taken into account the last value before the rise or the first value before the fall; for the "modulations" we have considered the global decreasing of pitch). moy = mean; min = minimum, max = maximum.

	moy	min	max
\	- 30.75	0	- 65
k	- 32.65	- 10	- 50
m	- 34.8	- 20	- 67
Am	- 39	- 27	- 60

When hesitation is preceded and followed by text, it shows a smaller fall. In the same way this lowering is often shorter, which seems logical. Hesitations preceded by a pause (and followed by text or pause) show a deeper fall and longer duration. The context of appearance of hesitations seems to influence the quantity of the fall. On the other hand, this phenomenon does not seem to depend on the prosodic configurations.

PERSPECTIVES AND DISCUSSION: CONSEQUENCES FOR THE PERCEPTION OF RHYTHM.

Hesitation corresponds to the desire of the speaker to "keep the floor" while preparing what he is going to say (Maclay, Osgood, 1959). What is specific to vocal hesitation is its vocal but non-verbal nature. They are certainly necessary for maintaining contact by keeping talking during the period of time which the speaker saves for conceptualisation. During vocal hesitations voice sounds like a sustained note. However, while this sustained note shows, at acoustic level, a decreasing of pitch, it corresponds to a diminution of sub-glottal pressure, i.e. the physiological dimension of declination. The acoustical realisations of vocal hesitations are of interest because they seem to prove their physiological origin.

This could explain why hesitations are hardly "detectable" by listeners in a standard situation of communication. It is well known that listeners only notice hesitations if they are extremely frequent. But, if we consider that the vocal reality of hesitations is only physiological, the non-verbal nature of hesitations has also to be admitted. However, it does not mean that hesitations have no communicative functions.

Moreover, the specific duration scale of hesitations could facilitate a specific perceptual treatment. This interpretation is logical if we consider that hesitations are not dependant (in most cases) on the speaker's will.

We can admit that there exists a subconscious agreement among speakers in order to not consciously detect hesitations. If we maintain that hesitations are inevitable and essential in communication, speakers often consider them as failures of reasoning.

Their role in communication (and especially in perception) is probably to create a "time of rest" during which both speaker and listener can review what has been said and extrapolate what is going to be said.

CONCLUSIONS

Lengthening of syllable duration, without break of pitch, cannot be considered as a cue for accentuation but - at least for the cases of extreme increasing - an hesitation cue considered as another segmental category. Determining a duration threshold could permit the discrimination of non-accented syllables (with increase in duration) from vocal hesitation. The parameter of duration appears to be an accentuation cue but remain a secondary cue associated with a pitch break. The existence of hesitation phenomena confirm the prime role of the localisation of direction of slope on the syllable in the perception of perception.

REFERENCES:

- Butterworth B., 1980, "Evidence from Pauses in Speech", in Butterworth *Language Production*, vol.1 "*Speech and Talk*", Academic Press, pp. 155-76.
- Butterworth B., Goldman-Eisler F., 1979, "Recent Studies in Cognitive Rhythm", in Siegman, Feldstein, *Of speech and time Temporal Speech Patterns in Interpersonal Contexts*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 211-24.
- Guaïtella I., 1991, *Rythme et parole: comparaison critique du rythme de la lecture oralisée et de la parole spontanée*, Doctoral Thesis, Aix-en-Provence.
- Maclay H., Osgood C.E., 1959, "Hesitation phenomena in spontaneous English speech", *Word*, 15, pp. 19-44.
- Siegman A.W., 1979, "Cognition and Hesitation in Speech", in: Siegman, Feldstein, *Of Speech and Time*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 151-78.