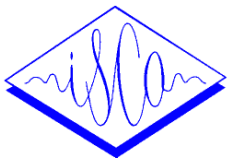

Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop.
10–12 September 2005, Aix-en-Provence, France, pp. 43–46.

Pauses and hesitations in French spontaneous speech

Estelle Campione & Jean Véronis

Equipe DELIC
Université de Provence, Aix-en-Provence, France

Abstract

In traditional terminology, *silent* and *filled pauses* are grouped together, whereas hesitation lengthening is put into a separate category. However, while these various phenomena are very often associated, there have been few studies on how they interact. We analyzed an hour of spontaneous speech to show that silent and filled pauses operate in a totally different way, and that contrary to common belief, silent pauses by themselves never serve as hesitation markers, but only do so when coupled with other markers—mostly syllabic lengthening and filled pauses. These last two hesitation markers have similar acoustic and articulatory characteristics; they are also distributed and function alike.

1. Introduction

There are traditionally two types of pauses (see for example Duez [10]): *silent* pauses, in which all vocal production ceases, except for possible respiratory noises, and *filled* pauses (or pauses containing sound), consisting of quasi-lexical tokens (*euh* in French, *er* and its nasal variant *erm* in English).

By grouping both of these very different acoustic and articulatory phenomena under the same term of *pause*, one makes the implicit hypothesis that they have the same function. However, the likelihood of this is very low. It has been known for a long time (see for example Boomer [2]) that silent pauses have a double role. Some silent pauses are *demarcative*, and appear at the junction of speech segments, which they help structure and parse. Others are *hesitation* pauses, caused by the sporadic difficulties the speaker encounters during "searching and encoding" mental operations (Barik [1]) or the "formulating work" (Morel & Danon-Boileau [17]) inherent to speech production. Conversely, it seems that filled pauses are only used for this second role: they are used as a conventional signal by the speaker to signal that he/she is not done speaking and to prevent interruptions during the time required for building the next part of the speech (see Clark & Clark [7], etc.).

It is worthy to note that the role of hesitation signals, corresponding to filled pauses, is also that of some types of syllabic lengthening (generally affecting a vowel at the end of a word), that we will call *hesitation lengthening*. Their properties are similar to those of filled pauses (see Guaïtella [14], Candea [6]), to the point where they can be detected using the same algorithms (see Goto, Itou, & Hayamizu [13]). Current terminology thus groups under the same term of *pause* two phenomena that are acoustically and functionally very different, whereas it puts hesitation lengthening and *er/erm* in different categories, even though they share similar properties and function alike. We group hereafter hesitation lengthening and *er/erm* under the same term *filled pauses*, as done by Goto et al. [13].

There are many studies on silent pauses in the literature (see for example Zellner's state of the art [20]). However, as demonstrated by a recent thesis (Candea [6]), there are far fewer studies on sentence planning markers in general, and on filled pauses and on lengthening in particular (which is probably due to the fact that most phonetic studies have been for a long time devoted to "laboratory speech", at the expense of spontaneous oral speech, as noted by Cutler [8], Duez [11] and others). In any case, there are almost no studies on how these phenomena interact.

This paper aims at giving a precise study of how silent pauses, filled pauses and hesitation lengthening interact, based on a study of a corpus of French spontaneous speech. We show that silent pauses by themselves never serve as hesitation markers, and need another sentence planning marker, most often a type of lengthening or an *euh* (or a combination of both) to play that part.

2. Corpus

The corpus used in this study consists of 8500 words and 54 minutes of French spontaneous speech, involving 10 different speakers (5 male and 5 female). It is a subset from the *Corpus de référence de français parlé* (Spoken French Reference Corpus) recently recorded by our team [9][1], which consists in 136 recordings of ca. 15 minutes each, involving speakers from 40 different locations covering the France map (36 hours of speech). The corpus is sampled according to age and education levels, and speech genres (public, private and professional speech). Recordings have been made in a quiet room, using minidisk recorders. Disfluency phenomena (hesitations, repetitions, false starts, *euh*, etc.) have been carefully transcribed with several independent verifications. Syllable lengthening and intonation were not marked in the initial transcription.

We have selected our sub-corpus in order to balance sexes, age groups and education levels. Five minutes segments were extracted from the original recordings, in which the interviewed speaker was speaking without interruption. A more detailed description and the transcribed corpus itself are available in Campione [3][2].

Disfluency phenomena have been once more verified in the sub-corpus, which was augmented with syllable lengthening and intonation markup. Syllable lengthening was entirely done manually (with the help of a signal editor), and intonation markup was obtained semi-automatically with careful manual verification, according to the method described in Campione & Véronis [5].

3. Silent pauses

[1] Available on-line at:
http://www.up.univ-mrs.fr/veronis/pdf/2004- presentation-crfp.pdf
[2] Available on-line at:
http://www.up.univ-mrs.fr/delic/theses/resume-campione.html

3.1. Tagging

Silent pause transcription is a very difficult exercise when done entirely manually, and we have noticed that most linguists, even highly competent ones, tend to miss many silent pauses, especially when they are coupled with other phenomena (such as hesitation or syllable lengthening). Despite the multiple verification of the original corpus, some silent pauses were still missing (this difficulty confirms Candea's observations [6]). Since, in addition, we needed carefully-measured duration times, silent pauses were detected using a program that calculates the fundamental frequency[3] and isolates voiceless segments. We applied a threshold of 200 ms, in line with past studies (Candea [6]). Shorter silent pauses, whose existence and importance have been underlined in the literature (Hieke, Kowal & O'Connell [15]), were added manually afterwards, with no lower limit (we have found silent pauses as short as 60ms). We have stressed elsewhere the dangers of applying arbitrary thresholds when studying silent pauses (Campione & Veronis [4]), since they can lead to considerable baises in the results.

We then corrected all of the silent pauses using a signal editor: pauses that were not correctly detected (which corresponded in general to voiceless stops) were deleted, and those that had not been detected were added (including those below the initial threshold of 200 ms), and the boundaries of those that were correctly detected were adjusted if needed.

The corpus contained 1375 detected potential silent pauses and 1163 actual silent pauses after correction. The distribution of silent pause lengths is highly skewed, approximately following a log-normal law, with a geometric mean at 496 ms (Campione & Veronis [4]). We categorized silent pauses in three groups, according to the tri-modal behavior described in the same study:

- short (< 200 ms)
- medium (200-1000 ms)
- long (> 1000 ms)

These three types of silent pauses are respectively marked ^, + and ++ in the examples throughout this paper.

3.2. Demarcative role

It is largely accepted by psycholinguists that speech production is based on a planification-execution cycle that results in a series of relatively short units (named discourse segments hereafter), separated by silent pauses (Fromkin [12] ; Levelt [16] ; etc.). Silent pauses are needed by speakers for them to plan their wording, and by listeners for processing the speech. Contrary to hesitation pauses, demarcative pauses play an important part in speech structure, and are probably an important factor in the correct parsing of utterances by listeners.

We tagged all silent pauses that were of a demarcative nature in the corpus, as perceived by two independent experts. These pauses are easy to detect because there are many different converging cues (intonation, vowel duration and quality, syntax, etc.). The presence of a rising or falling intonation (detected automatically and then corrected manually—see above) was a determining factor. The example below shows the type of segmentation that was obtained (speech segments are separated by ||). We provide a literal translation below each example:

ben je travaille dans un pressing ↗ ++ || on fait pas que le pressing on fait aussi la blanchisserie ↗ + || plus la blanchisserie d'ailleurs ↘ + || les draps les nappes la restauration ↗ ++*

well I work at a dry-cleaner's ↗ ++ || we don't do just dry-cleaning we also do laundry ↗ + || and laundry as a matter of case ↘ + || sheets tablecloths catering ↗ ++

Table 1 shows how silent pauses are distributed. 71% percent of silent pauses are demarcative.

Table 1: Types of silent pauses

Length	Demarcatives	Non-demarcative	Total
Short	0	24	24
Medium	673 (70%)	293 (30%)	966
Long	148 (86%)	25 (14%)	173
Total	821 (71%)	342 (29%)	1163

Short silent pauses are never demarcative, which seems to justify the thresholds of 200 ms used in some studies. We agree with Morel & Danon-Boileau [17] for whom silent pauses that are below this threshold do not have a defined iconic value. They often are of a respiratory nature and appear where the intonative cohesion is blatant:

donc c'est ^ la calandre qui travaille aussi

thus it's ^ the drying machine that works as well

4. Filled pauses

4.1. Tagging

Hesitation lengthening and *euh* (that we refer both to as filled pauses in this paper, as said before) are characterized by a continuous vowel lasting much more than the norm and with a constant vocalic quality, and are associated with a flat or slightly falling fundamental frequency (F_0) curve (Guaïtella [14]). These characteristics seem to be common to many languages (Quimbo, Kawahara & Doshita [18]; Goto et al. [13]).

Aside from the time required for the task, tagging filled pauses in the corpus is relatively easy provided the latter is listened to carefully with the help of a signal editor. Hesitation lengthening is easily set apart from types of lengthening pertaining to syntactic structure, to stress, etc., which generally have a rising or falling intonation contour, at least in French (see Vaissière [19]). Hesitation lengthening often occurs on function words, or at positions that are not syntactic or discourse breaking points. *Euh* is an easy to recognize *quasi-lexical* token (it is listed in dictionaries), and human annotators almost never disagree on its presence. Only a few cases of words ending with a schwa are borderline: lengthening of the schwa or progressive change to an *euh*? Neither listening, nor intonation curves, nor sonagrams seem to provide any decisive answer to this question. One might wonder whether this "hesitation" on hesitation is not indicative that hesitation lengthening and *euh* have the same function. We use a colon to note hesitations in the form of syllabic lengthening:

*voilà alors hein ^ on **ne:** ++ il est il est difficile d'aborder **la:** + la question du métier sans parler un petit peu des origines*

[3] Developed by Robert Espesser, as well as various tools including the signal editor used in this study.

*Well then eh ^ we do **not:** ++ it's difficult to tackle **the:** + the question of one's job without talking a little about the origins*

4.2. Filled pause sequences

In 12% of cases, filled pauses are part of a complex sequence containing several types of lengthening, several *euh*'s, or a combination of both (possibly mixed with one or more silent pauses). The following excerpt provides an example of a particularly long combination (it is interesting to note that the next sequence is a word fragment, followed by another filled pause:

*euh Beaune est une **euh la : la : la : euh le :** cé- le : cépage de : euh la ville de Beaune je veux dire*

*er Beaune is **er : the : the : the : er the :** ty- the : type of vine of : er the city of Beaune I mean*

There are 679 filled pauses in all, 591 of which are separate sequences. Table 2 shows the number of pauses of each type.

Table 2. Types of filled pauses

Type	N	sub-type	N
lengthening	230	*simple*	216
		complex	14
euh	323	*simple*	305
		complex	18
combination	38		
Total	**591**		

5. Interaction study

Silent and filled pauses can hardly be studied separately:

- 380 of the 1163 silent pauses of our corpus occur next to a filled pause (33%).
- Conversely, 344 of the 591 sequences of filled pauses are next to a silent pause or contain a silent pause (58%).

This shows the importance of the interaction between these two types of pauses.

5.1. Two types of filled pauses

We think it is important to separate filled pauses (or sequences) into two functionally different types. Filled pauses that are within a speech segment are the most common type (478 out of 591 sequences, or 81%). They mark an interruption that may or may not be followed by a repetition and/or a repair:

*euh Beaune est une **euh la : la : la : euh le :** cé- le : cépage de : euh la ville de Beaune je veux dire*

*er Beaune is **er : the : the : the : er the :** ty- the : type of vine of : er the city of Beaune I mean*

However, in 113 out of the 591 sequences (19%), filled pauses occurred at the beginning of a segment:

*enfin bon voilà euh le dé*but de mon voyage ça a été ça ++ || **euh** et après bon ben après les choses se sont mises en place hein*

*well then er the start of my trip went well ++ || **er** and after well after things got into place eh*

In such cases, the interruption does not result from a difficulty in setting up the end of the segment in a lexical or syntactic sense, but is rather a way of "filling in" during the time it takes to make up the rest of the speech, and of preventing others from interrupting. This phenomenon is quite frequent since it affects 133 out of 822 segments (14%). These 113 cases are split into two groups:

- 97 sequences starting with *euh*;
- 16 sequences starting with a type of lengthening.

At least in our corpus, these types of lengthening are all monosyllabic words, mainly connectives such as *and*, *but*, *then* (11 out of 16 cases). The rest is/are function words that introduce phrases (*of*, *where*, *the*).

5.2. Role of silent pauses

As shown above, short silent pauses are never demarcative. Among the 318 non-demarcative silent pauses left, 289 (91%) are associated with a filled pause. In a large majority of cases (257 out of 289 or 89% of pauses), a filled pause comes before the silent pause:

*on l'ap- au départ on faisait euh le **le:** + euh le: ma*cérer le: le: **le poulet:** + le poulet bien sûr qui est issu aussi de la Bourgogne puisque **euh** + le poulet de Bresse*

*at the beginning we made er the **the:** + er the: macerate the: the: **the chicken**: + the chicken: of course that comes from Burgundy since **er** + the chicken from Bresse*

In only 32 cases (11%), there is a sudden interruption and the filled pause occurs immediately after:

et il a apprécié + euh ce Corton blanc

*and he appreciated + **er** this white Corton*

These cases, which seem to occur after a major intonation change and/or a stressed syllable, should be studied in detail.

Let us take a deeper look at the 29 cases of non-demarcative silent pauses that are not associated with a filled pause. They can be broken down as such:

1. In 13 cases, the speaker hesitates, but the silent pause is in fact associated with another sentence planning marker:

- word fragment (8 cases)

  *et puis là on vit **au jour le j- + au jour** le jour*

  *and then now we live **one d- + one day** at a time*

- repetition (5 cases)

  *donc s- nous avons un rôle **de + de** soutien de marché*

  *thus we have a role **of + of** supporting the market*

2. In 11 other cases, the silent pause is associated with a *discourse marker*:

- onomatopoeia (*pff*, etc.) (2 cases)

alors que moi ça me dérange pas du tout au contraire: ***pssff*** *+ je trouve que chacun a ses limites*

whereas that does not bother me at all on the contrary: ***pssff*** *+ I think everyone has their limits*

- particle (*ben*, *hein*, etc.) (3 cases)

  elle apprendra comme moi sur le tas ***hein +*** *de toute façon*

  she'll learn as I have on the fly ***eh*** *+ anyway*

- focal stress (6 cases)

  j'avais ***+ te*llement*** *soif de cette liberté-là*

  I longed + ***so much*** *for this type of freedom*

This last phenomenon is quite interesting and is worthy of a more in-depth study using a larger set of data. We hypothesize that it helps reinforce how well the stress is perceived and allows the speaker to fully reload his/her pulmonary capacity before or even after the stress (we mark the stress with a star after the stressed syllable):

3. Only 5 out of the 1163 silent pauses of the corpus remain unassociated with any other cue:

elle va se sentir dans ***un + endroit*** *de confiance donc*

she'll feel in ***a*** *+* ***secure*** *place then*

In these 5 cases there is a syntactic cohesion of the segment (there even is a liaison in the example above). There is no intonative discontinuity at the point of the silent pause. It would be worthwhile to find more examples and make a minute analysis of this phenomenon from an acoustic and syntactic aspect. Nevertheless, one conclusion that can be drawn is that there is no perceived hesitation in the cases we observed. These pauses act exactly like short pauses, and in fact they do not last long since their duration is below the (geometric) mean (200 to 470 ms).

6. Conclusion

Based on a subset of the *Corpus de référence du français parlé* lasting about one hour and involving 10 speakers (5 male and 5 female), the present study shows that a silent pause by itself never serves as a hesitation or sentence planning marker. It has that function only when coupled with other markers, mainly filled pauses (syllabic lengthening and the quasi-lexical item *euh*). Other cues are also associated with silent or filled pauses, such as word fragments, repetitions, or quasi-lexical items like *well*, *eh*, *pff*, etc. An extensive study on the phenomena associated with sentence planning should be conducted.

7. References

[1] Barik, H. C. 1968. On defining juncture pauses: a note on Boomer's "Hesitation and grammatical encoding". *Language and Speech, 11*, 156:159.

[2] Boomer, D.S. 1965. Hesitation and grammatical encoding. *Language and Speech, 8*, 148:158.

[3] Campione, E. 2001. *Étiquetage semi-automatique de la prosodie dans les corpus oraux: algorithmes et méthodologie.* Thèse de doctorat, Université de Provence, Aix-en-Provence.

[4] Campione, E., & Véronis , J. 2002. A Large-Scale Multilingual Study of Silent Pause Duration. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 conference* (pp. 199-202). Aix-en-Provence: Laboratoire Parole et Langage.

[5] Campione, E., & Véronis, J. 2004. Semi-automatic tagging of intonation. In G. Sampson & D. McCarthy (Eds.), *Readings in Corpus Linguistics* (pp 462-473). London: Continuum.

[6] Candea, M. 2002. *Contribution à l'étude des pauses silencieuses et des phénomènes dits "d'hésitation" en français oral spontané.* Thèse de doctorat, Université Paris III.

[7] Clark, H. & Clark, E. 1977. *Psychology and Language*. New York : Harcourt, Brace, Jovanovich.

[8] Cutler, A. 1998. The recognition of spoken words with variable representations, *Proceedings of the ESCA Workshop on the Sound Patterns of Spontaneous Speech* (pp. 83-92). Aix-en-Provence, France.

[9] DELIC. 2004. Présentation du Corpus de référence du français parlé. *Recherches sur le français parlé*, 18, 11-42.

[10] Duez, D. 1982. Salient pauses and non salient pauses in three speech style. *Language and Speeh*, 25(7), 11:28.

[11] Duez, D. 1998. The aim of SPoSS, *Proceedings of the ESCA Workshop on the Sound Patterns of Spontaneous Speech* (pp.VII-IX). Aix-en-Provence, France.

[12] Fromkin, V. A. 1971. The non-anomalous nature of anomalous utterances. *Language,* 47, 27:52.

[13] Goto, M., Itou, K., & Hayamizu, S. 1999. A Real-time Filled Pause Detection System for Spontaneous Speech Recognition, *Proceedings of the 6th European Conference on Speech Communication and Technology (Eurospeech '99)* (pp. 227-230). Budapest.

[14] Guaïtella, I.1991. *Rythme et parole: comparaison critique du rythme de la lecture oralisée et de la parole spontanée.* Thèse de doctorat, Université de Provence, Aix-en-Provence.

[15] Hieke, Kowal, & O'Connell, D. C. 1983. The trouble with "articulatory" pauses. *Language and Speech, 26*, 203:214.

[16] Levelt, W. J. M. 1989. Speaking: From Intention to Articulation. Cambridge MA : MIT Press.

[17] Morel, M. A., & Danon-Boileau, L. 1998. *Grammaire de l'intonation. L'exemple du français.* Paris : Ophrys.

[18] Quimbo, F. C. M., Kawahara, T., & Doshita, S. 1998. Prosodic analysis of fillers and self-repair in Japanese speech, *Proceedings of the International Conference on Spoken Language Processing (ICSLP).* Sydney, Australia.

[19] Vaissière, J. 1991. Rhythm, accentuation and final lenghtening in French. In J. Sundberg & L. Nord & R. Carlson (Eds.), *Music, Language, Speech and Brain*. Macmillan Press.

[20] Zellner, B. 1998. *Caractérisation et prédiction du débit de parole en français. Une étude de cas.* Thèse de doctorat, Université de Lausanne, Lausanne.