



Acoustico-phonetic Characteristics of Filled Pauses in Spontaneous French Speech: Preliminary Results

Danielle DUEZ

CNRS, UMR 6057, Laboratoire Parole et Langage
Université de Provence, 31621 Aix en Provence
Fax: (33) 04 42 59 50 96; email: duetz@lpl.univ-aix.fr

Abstract

In the current analysis we examined the acoustic and phonetic characteristics of filled pauses in spontaneous French speech and their relationship to the prosody of the surrounding context. Two main results emerged: 1) There was no effect of the duration of filled pauses or their sentence location on their F₀ patterns or on the differences between the highest and lowest values. 2) There was no relationship between peak-F₀ values and the F₀ values of filled-pause onsets, but the F₀ values of filled-pause onsets and the F₀-values of non-marked breath-group onsets were highly similar. The F₀ values of filled-pause onsets seem to be stable within the same speaker's speech. They are speaker-dependent and strongly linked to the physiological, absolute aspects of speech production. It is assumed that filled-pause onset may be used by listeners as a reference for evaluating the speaker's pitch range.

I Introduction

This paper reports an ongoing study on the acoustic-phonetic characteristics of filled pauses in spontaneous French speech. In the last few decades, there has been a growing interest in the investigation of disfluencies. There are several important reasons for this, including the following: 1) a renewed interest in so-called spontaneous speech and the manifestations of spontaneity, 2) the investigation of speech encoding processes in various communication situations, and 3) the necessity to improve speech technologies, in particular the naturalness of speech synthesis and the reliability of speech-recognition.

A wide range of studies have concerned with filled pauses. Filled pauses are among the most common disfluencies in spontaneous English and French speech [1]. They are widely used in the speech produced by adults [2] and [3] and young people [4] and [5]. Traditionally there are two main explanations given for their occurrence. The first is based on the idea that filled pauses reflect the speaker's level of anxiety [6] and [7]. The other suggests that filled pauses are linked to the complexity of the message [8]. More recently, another explanation was proposed, based on the notion that filled pauses reflect instead speakers' attention to their speech [9]. Thus, filled pauses appear as symptoms of the attention paid to the speech production process. They occur when the speaker detects an error (or an impending error) in the encoding process, and stops to correct it.

The analysis of the distribution of filled pauses in a sentence is a source of information on the different processes involved in speech production. For example, filled pauses occurring

before lexical words play a role in the lexical-selection process [10], [2], [3] and [11] while filled pauses located at the beginning of a phrase play a role in the programming of the upcoming phrase [12], [5], [13], [14], [2], [3] and [11].

The acoustic-phonetic characteristics of filled pauses and their relationship to the prosodic context have also been the subject of a certain number of studies aimed mainly at examining how filled pauses are integrated into the intonative structure of sentences and/or at defining a certain number of cues relevant to filled-pause recognition. For example, filled pauses in French were shown to be produced at a given, speaker-specific F₀ value regardless of location [15]: the value was roughly equal to the average of the F₀ values of unaccented syllables and major-phrase onsets. Three main patterns have been reported for filled pauses in spontaneous French speech: (1) a slow decline, an irregular decline and a decline followed by a sudden rise [16]. Filled pauses at major syntactic boundaries as well as those within syntactic units have also been shown to have falling or flat F₀ pattern, at relatively low F₀ levels with the lowest F₀ at the end [17]. However, filled pauses at syntactic boundaries tended to start higher in F₀ and then fall, whereas filled pauses internal to a syntactic unit had lower F₀ patterns. Filled pauses at major syntactic boundaries were also found to be longer than those within syntactic units. The analysis of clause-internal filled pauses and the fundamental frequency (F₀) values of preceding peaks in dialogues and conversations in British English and American English have been shown to have higher peaks systematically associated with higher filled-pause values [18]. This was interpreted as an indication of some form of systematic relationship between the peak and corresponding filled-pause F₀ values.

In the current analysis we examined the acoustic and phonetic characteristics of filled pauses and their relationship to the prosody of the surrounding context. Our objective was to test whether filled pauses are produced at an absolute speaker-specific value or whether they are dependent on the prosodic context. Filled pauses were examined as a function of their location in sentences (within a prosodic word, and between two prosodic words or two breath groups, as defined by Vaissière [19]) in the conversations of four French speakers (two males and two females). As a measure of prosodic context, the F₀ value of the breath-group onset, the closest preceding or following peak (if any). As a measure of filled-pause F₀, the F₀ value of the beginning, middle (or the turning point if any), and end were used. The duration of filled pauses was also measured.

2. Methods

2.1. Subjects

The conversational speech produced by two male and two female French speakers was used for the experiment (one hour in all). The speakers were of the same sociocultural background, without strong regional accents and with normal speech and hearing. Their age ranged from 30 to 50. The subjects, who were unaware of the purpose of the recording, had to reply freely in a relaxed way to questions regarding their life, childhood, work, travels, future plans and current events. The conversations were recorded on a Sony tape-recorder at 9 cm/sec, in a quiet room at the Phonetics Laboratory of Aix en Provence.

2.2. Procedure

The conversations were transcribed orthographically by the author. The presence of filled pauses and the boundary breaks was checked perceptually. Two break levels were defined : prosodic words and breath groups [19]. Filled pauses were also checked to see whether they were accompanied by another disfluency such as a correction, an interruption or a lengthened syllable.

2.3. Filled pauses

A filled pause was any occurrence of "euh" [ø] (a filler like hm or uh in English). It could occur before or after a vowel as in "mais euh" [mæø] or attached to a consonant as in "donc euh" [dɔ̃kø], "visite euh" [vizitø]. The cases where an optional schwa was realised and lengthened as in "que" [kø:], "le"[lø:] and "de" [dø:] were considered as lengthened syllables and not as filled pauses.

2.4. Location of filled pauses

Two main locations were considered : within a prosodic word and between two prosodic words or two breath groups phrases. The former occurred mostly before a lexical word or at the beginning of a phrase just after a conjunction or an adverb. There were different possibilities for the latter depending on the presence of a silent pause. In the absence of a silent pause or when preceded and followed by a silent pause, a filled pause was simply a between-phrase filled pause; a filled pause followed by a silent pause was a final-phrase filled pause; a filled pause preceded by a silent pause was an initial-phrase filled pause.

2.5. Reference points

There are three key points in French breath-group onset, pretonic syllable, and tonic syllable [20] and [21]. Onset- F_0 values can be very high in interrogative and exclamatory sentences, and in phrases with new information, but they are generally quite stable in declarative sentences and are very close to those of pretonic syllables. Since there is no lexical stress in French, prominent syllables are mostly phrase-final syllables [19]. They may also be word-initial syllables since French also possesses an optional initial prominence on lexical words whose realisation depends on the style and speaker [22]. Peaks on a syllable mostly correspond to the realisation of initial and final prominence, but can also be the mark of an emphatic accent. Peak- F_0 values and onset- F_0 values are fundamental in defining sentence structure and in integrating the speaker's pitch range. They were taken as reference values.

2.6. Measurements

The sentences were digitised at a sampling rate of 16 kHz with a Sun computer. Measurements were carried out on spectrograms, oscillograms and F_0 curves displayed on the screen, and by listening to selected segments of the waveform in the region of interest. Filled pauses were identified both perceptually and acoustically. The duration of each filled pause was measured as the duration of the waveform between the first and the last [ø] periods. For each filled pause, five F_0 values were recorded : the first, the middle, and the last F_0 value in the filled pause, F_0 at phrase onset, F_0 at phrase-peak, and F_0 at the syllable following the filled pause (if any).

2.7. Analysis

There were three main patterns for filled pauses (flat, upward for rising F_0 and downward for falling F_0). The combination of the three basic patterns gave nine patterns in all (e.g. downward, downward-flat, downward-upward). The patterns were examined as a function of the location of the filled pauses in the phrases. The initial- F_0 values of the filled pauses were also compared with the mean values of neutral-phrase onsets (default values) and with the onset- F_0 value of the phrases in which the filled pauses were located. The same was done with peak- F_0 values. The initial- F_0 values of the filled pauses were compared with the mean peak- F_0 values and with the peak- F_0 values of the phrases in which the filled pauses were located.

3. Results

3.1. F_0 patterns of filled pauses

Table 1. Number of filled pauses as a function of sentence location (B: between two prosodic words or two breath groups ; W: within a prosodic word) and F_0 pattern (D: Downward, DF: downward-flat, DU: downward-upward, F: flat, FD: flat-downward, FU: flat-upward, U:upward, UD: upward-downward and UF: upward-flat). T means total.

	D	DF	DU	F	FD	FU	U	UD	UF	T
B	35	14	9	10	1	4	2	5	1	81
W	29	22	10	13	4	2	3	5	1	89
T	64	36	19	23	5	6	5	10	2	170

The results reported in Table 1 indicate a number somewhat similar for the different patterns of the filled pauses located within a prosodic word and between-two prosodic words or two breath groups. There was no significant interaction between the distribution of filled pauses and the F_0 patterns (χ^2 : 0.7, df(8), n.s.).

Decreasing F_0 patterns were the most frequent: 105 out of 170 in all. There was a continuous declining F_0 for 64 of them, 36 with a descent followed by a plateau, and 5 with a plateau followed by a descent. There were 23 filled pauses with a plateau. For the remaining filled pauses, there was a combination of a rise with a plateau or a rise with a descent (and vice versa)

In general, the differences between the initial F_0 value and the final F_0 value for the filled pauses were less than 20%, which is in agreement with previous results reported in the literature [16] and [17]. However, 36 out of the 170 filled pauses were found to exhibit larger differences (ranging from 25% to 50%). The majority of them had sharp falls (16 out of

36) or downward-flat patterns (8 out of 36). For the remaining ones the distribution was as follows : downward-upward (2), flat-downward (1), flat-upward (2), upward (3), upward-downward (4). A χ^2 test performed on the results revealed a significant interaction between F_0 patterns and F_0 differences ($\chi^2 = 19.8$, $p=0.01$). The filled pauses with differences above 25% were located within a prosodic word (22) as well as between prosodic words or groups (3 at the beginning and 11 at the end). For 14 out of 36, there was also a silent pause. Some abrupt falls or steep rises (reaching or attaining F_0 -values close to 60 Hz) for male speakers (9 cases) and female speakers (8 cases) were linked to a creaky voice.

3.2. Filled-pause duration

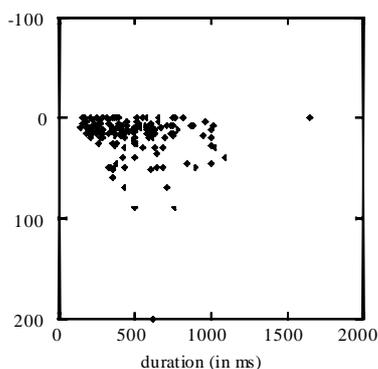
Table 2. Filled-pause duration (in ms) as a function of sentence location (B: between group; W: within a prosodic word) and F_0 pattern (D:Downward, DF: downward-flat, DU: downward-upward, F: flat, FD: flat-downward, FU: flat-upward, U: upward, UD: upward-downward and UF: upward-flat). T means total.

	D	DF	DU	F	FD	FU	U	UD	UF	T
B	432	669	468	468	449	428	500	666	202	476
W	450	501	577	478	376	491	485	313	1000	519
T	441	585	522	473	412	459	492	489	601	

An ANOVA on the duration of filled pauses did not reveal a significant effect of F_0 pattern [$F(152,8)=1.2$, $p=0.2$], or location [$F(152,1)=0.5$, $p=0.4$] but a significant interaction between the two factors [$F(152,8)=2.1$, $p=0.03$].

3.3. Effect of duration on F_0 -differences between the lowest and highest values of filled pauses

Fig1. F_0 differences (expressed as a %) as a function of duration



As can be seen in Figure 1, there was no effect of filled-pause duration on the differences (expressed as a %) between the highest value (most often the initial value) and the lowest one (most often the final value). The r^2 value (0.018) confirmed the absence of a correlation between duration and F_0 differences. Thus it appears that F_0 values and patterns are strongly independent on filled-pause duration. This independence is a characteristic of filled pauses.

Initial- F_0 values of filled pauses and prosodic context

An ANOVA yielded significant differences between the F_0 values of filled pauses and the onsets of the phrases in which they were located [$F(1, 306)=11$, $p=0.001$], a significant effect

of speakers [$F(3,306)=235$, $p=0.0001$], and a significant interaction [$F(3,306)=3.2$, $p=0.02$]. The differences were probably due to the fact that some phrase onsets were semantically or syntactically marked and exhibited high values (for example, the highest values for Spk2 and Spk3 were 220 Hz and 300 Hz, respectively).

The comparison of the F_0 values of filled pauses and those of non-marked onset phrases did not reveal significant differences [$F(1,343)=3$, $p=0.06$], which supports the above hypothesis (Mean F_0 values can be seen in Table 3). Moreover, there was no statistically significant correlation between the initial F_0 -values of filled pauses and peak F_0 -values ($r^2=0.04$).

The initial values of filled pauses clustered around the mean and were very stable. However, there were a few exceptions with low values (60 Hz, 80 Hz, and 60 Hz for Spk1, Spk2 and Spk3, respectively) due to physiological effects (creaky voice) These low values were characteristic of some filled pauses that started with a steep rise and of filled pauses located between two silent pauses.

Table 3. Mean F_0 values of filled-pause onset (FPO) and phrase-onsets. There were two different values for phrase onset: those obtained for the phrases in which the filled pauses were located (PO) and those obtained for non-marked onsets (NMFP). SD values are in parentheses.

Speaker	Male		Female	
	Spk1	Spk2	Spk3	Spk4
FPO	110 (15)	114 (14)	170 (32)	176 (14)
PO	116 (12)	116 (20)	190 (39)	186 (20)
NMFP	116 (15)	115 (18)	179 (30)	179 (11)

4. Conclusions

Two main results emerge from the present analysis: 1) There was no effect of the duration of filled pauses and their location within sentences on their F_0 patterns, or their highest and lowest values. 2) There was no relationship between peak- F_0 values and the F_0 values of filled-pause onsets, but great similarity between the F_0 values of filled-pause onsets and the F_0 values of non-marked breath-group onsets.

In French, four levels are usually used to represent intonation contours [23]. Level 2 is the level of non-marked onsets, level 3 is the level of final syllables of minor-continuation phrases, level 4 is the level of final syllables of major-continuation phrases, and level 1 is the finality level. A fifth level can be used for emphasis and exclamatory sentences [15].

As already mentioned [20] and [21], level 2 is a key point in the production and perception of intonation units. It is a reference value for evaluating the speaker's pitch range, and consequently, a baseline for the integration of the various key points. The fact that filled pauses are anchored in level 2 suggests that they may also constitute a reference for integrating the various levels. This interpretation is quite in line with Léon's proposal [15] to take the F_0 values of filled pauses as a reference for defining level 2. In turn, the F_0 values of phrase-onsets can be used as criteria in the recognition of filled pauses.

The present results contradict the common idea that filled pauses are rest times. It is well known that vocal onsets bring into play the breath and the glottal muscles in order to produce the first vibrations of a sound. Physiologically, this consists of four different phases: 1) Tensing of the laryngeal and respiratory muscles, which makes the larynx go from its rest position to its phonatory position, 2) beginning of expiration, 3) end of vocal-fold adduction, and 4) initiation of vocal-fold vibration. The lack of adjustment between phases 2 and 3 may result in a creaky voice which is a frequent characteristic of filled pauses. There appears to be a correspondence between physiological and cognitive activity in the production of filled pauses : at the same time, the speech organs are being positioned and the phonetic program is being prepared for execution.

The F₀ values of filled-pause onsets seem to be stable within the same speaker. They are speaker-dependent and strongly linked to the absolute, physiological aspects of speech production [24].

However, some filled pauses may also serve pragmatic aims and be used for specific communicative goals. In this sense, the initial F₀ values of filled pauses might also be manifestations of the relative, functional aspects of speech. Some cases observed here suggest that some F₀ values may be higher than those of non-marked breath-group onsets. This assumption has to be tested with a greater number of speakers and in various spontaneous speech styles.

References

- [1] Grosjean, F. & Deschamps, A., Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes :phénomènes d'hésitation, *Phonetica*, **31**, 1975, pp. 144-84.
- [2] Grosjean, F. & Deschamps, A., Analyse des variables temporelles du français spontané, *Phonetica*, **26**, 1972, pp. 129-149.
- [3] Grosjean, F. & Deschamps, A., Analyse des variables temporelles du français spontané, *Phonetica*, **28**, 1973, pp. 191-226.
- [4] Sabin, E.J., Clemmer, E.J., O'Connell, D.C. Kowal, S., A pausological approach to speech development, in *Of speech and time : temporal speech patterns in interpersonal context*, A. Siegman and S. Feldstein, Eds , pp.35-55, New Jersey : Hillsdale, 1979.
- [5] Candea, M., *Contribution à l'étude des pauses silencieuses et des phénomènes dits "d'hésitation" en français oral spontané*, thèse de doctorat nouveau régime, Paris III, unpublished manuscript, 2000.
- [6] Jurich, A.P. and Polson, C.J. Nonverbal assesment of anxiety as a function of intimacy of sexual attitude questions, *Psychological reports*, **57**, 1985, 1247-1253.
- [7] Mahl, G. F., Disturbances in the patient's speech in psychotherapy, *Journal of Abnormal and Social Psychology*, **42**, 1956, pp. 3-32.
- [8] Siegman, A.W. and Pope, B. Ambiguity and verbal fluency in the TAT, *Journal of Consulting Psychology*, **30**, 1966, 239-245.
- [9] Christenfeld, N., Effects of a metronome on the filled pauses of fluent speakers, *Journal of Speech and Hearing Research*, **39**, 1996, pp. 1232-1238.
- [10]Goldman-Eisler, F., *Psycholinguistics. Experiments in spontaneous speech*. London and New York: The Academic Press, 1968.
- [11]Maclay, H. & Osgood, C.E., Hesitation Phenomena in Spontaneous English Speech, *Word*, **1**, 1959, pp. 19-43.
- [12] Boomer, D.S., Hesitation and grammatical encoding, *Language and Speech*, **8**, 1965, pp. 148-158.
- [13]Duez, D., *Contribution à l'étude de la structuration temporelle de la parole en français*, Thèse de Doctorat d'état, Université de Provence, unpublished manuscript, 1987.
- [14] Duez, D., *La pause dans la parole de l'homme politique*, Editions du CNRS, collection sons et parole, Paris, 1991.
- [15] Léon, P., *Prolégomènes à l'étude des structures intonatives*, Studia Phonetica 2 (P. Léon and P. Martin, Eds) Montréal, Paris and Bruxelles: Didier, 1969.
- [16] Guaïtella, I, *Rythme et parole : comparaison critique du rythme de la lecture oralisée et de la parole spontanée*, thèse de doctorat, Université de Provence, unpublished manuscript, 1991.
- [17] O'Shaughnessy, D., Recognition of hesitations in spontaneous speech, *IEEE*, **1**, pp. 521-524, 1992.
- [18] Shriberg, E. & Lickley, R., Intonation of clause-internal filled pauses, *Phonetica*, **50**, 1993, pp.172-179.
- [19] Vaissière, J., Rhythm, accentuation and Final lengthening in French, in *Music, Language and Brain*, J. Sundberg, L. Nord, L. and R. Carlson, Eds. pp 108-120, Macmillan, Houndsmills, 1991.
- [20]Di Cristo, A., Des traits acoustiques aux indices perceptuels. Application d'un modèle d'analyse prosodique à l'étude du vocatif en français, *Travaux de l'Institut de Phonétique*, **3**, 1976, 213-358.
- [21] Di Cristo, A., *De la microprosodie a l'intonosyntaxe*, Thèse de doctorat d'etat, Université de Provence, 1978.
- [22]Fonagy, I., L'accent français : accent probabilitaire, in *L'accent en Français contemporain*, I. Fonagy and P. Léon, Eds; Montréal, Paris, Bruxelles : Studia Phonetica **8**, 1980, 53-97.
- [23] Delattre, P., (1966) Les dix intonations de base du français, *The French Review*, **40(1)**, 1-14.
- [24] Crystal, D. (1971). Relative and absolute in intonation analysis. *Journal of the International Phonetic Association*, **1 (1)** : 17-28.