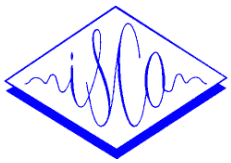


ITRW on Disfluency in
Spontaneous Speech (DiSS'01)
Edinburgh, Scotland, UK
August 29-31, 2001

The Double Function of Disfluency Phenomena in Spontaneous Speech

Mária Gósy

Kempelen Farkas Speech Research Laboratory, Linguistics Institute,
Hungarian Academy of Sciences, Budapest, Hungary
H-1068 Budapest, Benczúr u. 33., Hungary

Abstract

Disfluency in spontaneous speech is the outcome of a speaker's indecision about what to say next. The listener, however, is continuously adapted to both the language signals and the types of disfluency of the heard text. What is in the background of this adaptation process?

This paper analyses the types and characteristics of the disfluency phenomena of a 78-minute spontaneous speech sample (produced by 10 adults). The author's intention is to explain the characteristics of disharmony between speech planning and articulation within the speech production process. In order to explain the hypothesized double function of disfluency in terms of perceptual necessity from the listener's side various experiments have been carried out.

Three different samples of spontaneous speech have been selected for experimental purposes. Three groups of listeners (altogether 60 university students) participated in the experiments. One of the groups had to detect the instances of disfluency in the texts marking them on a paper sheet. The subjects of the other group listened to the same texts and then wrote down their contents. The pauses and hesitations were then eliminated from the texts. The third group of the subjects had the same comprehension task as the previous one had.

Results show that (i) instances of disfluency are consequences of the speaker's speech planning processes, (ii) their reasons and occurrences are unconsciously known by the listener as well, (iii) disfluency phenomena are relatively well predicted, (iv) the listeners need pauses and hesitations in order to comprehend the heard texts successfully.

Keywords: silent pauses, hesitations, speech planning process, comprehension strategies.

1. Introduction

Spontaneous speech is characterized by several phonetic processes like co-articulation or the variability of the phonetic form of words, and by various types of disfluency phenomena (silent and filled pauses, hesitations, prolongations of parts of an utterance, false starts, repetitions). Disfluency in spontaneous speech is the outcome of a speaker's indecision about what to say next and so shows his struggle to achieve control over planning, production, articulation.. All types of disfluency provide a useful approach to study the speech production process and so the complex phenomenon of disfluency has been widened since focusing on the types of hesitations [1]. In order to better understand the nature of disfluency we need to consider also the cognitive processes of planning and performing utterances.

The listener, however, is continuously adapted to both the language signals and the types of disfluency of the heard text. What is in the background of this adaptation process? Does the listener need disfluency in order to comprehend the text better?

Over the last 40 years, scholarly inquiry into the characteristics of various types of pauses (i.e. silent vs. filled ones) has steadily increased while many papers were devoted to other phenomena of disfluency, respectively [3]. Mahl was the first in 1956 who differentiated two types of disfluency labeling one of them *ah*-phenomenon while the other one non-*ah* phenomenon [2]. According to his definition the first category shows the speaker's uncertainty about what to say next while the second one covers other phenomena like repetitions, changes of structures, etc. Despite the widening interest, however, the results have not been enlightening, there has been no agreement on important facts, and the perceptual aspects have frequently been ignored. Among the results very little can be found about the functional aspects of disfluency in speech production and speech perception as well as in comprehension [4].

The most frequently occurring type of disfluency in fluent speech, independently of the language, is silent pause, then filled pause and the combination of these two. There are well-known factors that result in silent pauses: (i) breathing, (ii) intention of interpretation of the text, (iii) pauses determined by syntax, emotion, rhetorical and expressive emphasis, stylistic properties, and (iv) results of various kinds of disharmony between speech planning and articulation. The nature and function of juncture seems to be a matter of disagreement [5]; we understand it as a special case of "cognitive" pauses. The temporal aspects of spoken language raised the question of the duration of pauses. At the beginnings, articulatory pauses shorter than 250 ms were accepted as pauses according to the rationale given by Goldman Eisler [1]. This value was also associated with avoiding the confusion with pauses being parts of a segment (like the silent periods of voiceless stops). Silent pauses as types of disfluency are accepted generally with the durations of 130 and 150 ms [6]; however, the values of 200 and 300 ms can also be found in the literature [5, 6].

The distribution of disfluency phenomena seems to show similar tendencies across languages [7]; most of them are various kinds of pauses. In dialogues pauses appeared in 32% of all speaking time where 25% was silent while only 7% was filled pause. Another study revealed that disfluency appeared also in dialogues in every 42 seconds [8]. More research is required to learn more about disfluency being a universal or a language-specific phenomenon. There are many other questions that need answers. Should tongue tip be considered as a specific type of disfluency? Are pauses and hesitations

specific signals for the listener to perform his decoding strategy better? Or, on the contrary, do they disturb the decoding process? The present paper aims at answering these questions by carrying out several experiments with Hungarian-speaking speakers and listeners.

2. Method, material, subjects

A 78-minute spontaneous speech sample produced by 10 Hungarian speaking adults (5 females and 5 males) served as speech material. No hearing or speech defects were reported with any of the subjects. After a short period of introductory communication the subjects were asked to speak about their work. None of them were aware of the aim of the requested task, however, everybody knew that their speech was being recorded. The whole material was analyzed with respect to the occurrences and types of disfluency phenomena. Acoustic-phonetic analysis was carried out by means of the Kay Elemetrics CSL 4300B digital system.

Three different samples of spontaneous speech have been selected for experimental purposes. The data of tempo and pauses of the three texts is shown in Table 1.

Table 1: *Pauses of the three text samples (AT=articulation tempo in sounds/s, ST1=speech tempo in sounds/s, ST2= speech tempo in words/minute)*

Texts	AT	ST1	ST2	Pause time (ms/item)
1st	15.29	11.83	125.7	13.9/24
2nd	15.58	12.26	168.1	8.05/29.77
3rd	11.47	9.56	112	11.95/26

Three groups of listeners (altogether 60 university students, females and males) participated in the listening experiments. One of the groups (with 20 subjects) had to detect the instances of disfluency in the texts marking them on a paper sheet. They heard each text two times, and were asked to draw a vertical line where they thought they found the place of disfluency. The subjects of the other group (another 20 subjects) listened to the same texts and then wrote down their contents. The pauses and hesitations were then carefully eliminated from the texts (by means of the same Kay Elemetrics CSL system). The third group of the subjects (the last 20 subjects) had the same comprehension task as the previous one had. The written narratives of these two groups were compared. Analyzing the narratives the number of words used and the main ideas of the text samples were taken into consideration. To test statistical significance and statistically relevant interrelations, a two-way ANOVA was used (SPSS 8.0 for Windows statistics package).

This paper aims at investigating (i) the frequency, phonetic types and acoustic properties of disfluency phenomena of Hungarian spontaneous speech and (ii) the function of pauses from the aspects of speech production and speech comprehension. Our hypothesis is that pauses bear a double function in speech; they solve the disharmony between speech planning and articulation on the one hand and provide a better strategy for the listeners in comprehension, on the other hand.

3. Results

3.1 Disfluency phenomena. The phonetic analysis of our material showed that the following 6 types of disfluency phenomena could be detected covering 37% of the texts on average: silent pauses; hesitations; repetitions at various levels of speech, e.g. vowels, syllables, words (both function and content words), word combinations; prolongations (particularly at the end of phrases and utterances); alterations and changes (both structural and semantic); and false starts. Slip of the tongue was considered as a disfluency category. Almost all categories of disfluency could be detected with all speakers though the occurrences showed enormous individual differences ($p<0.001$). There was one (male) who had no filled pauses, three others (two females, one male) with whom neither prolongations nor false starts existed while 7 subjects did not struggle with slips of the tongue. Silent pause has a marked occurrence in all speakers' speech production. Taking disfluency phenomenon as an 'umbrella term' to be 100%, the occurrence of the 7 types shows this distribution across all speakers (on average), cf. Table 2.

Table 2: *Types and occurrences of disfluency phenomena across all speakers*

Disfluency type	Occurrence (%)	Range (%)	sd
silent pause	44.57	22.7-77.9	17.5
hesitation	12.37	0-25.7	7.84
repetition	3.25	1.3-5.33	1.51
prolongation	1.77	0-4	1.44
alteration, change	3.05	1.06-5.2	1.43
false start	1.34	0-2.85	1.06
slip of the tongue	0.36	0-1.2	0.48

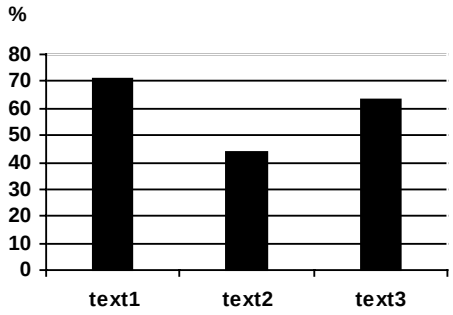
Comparing our results to those obtained with other (analyzed) languages, some interesting differences should be mentioned. Prolongations appear frequently on definite articles (*a*, *az*), false starts occur quite often when definite articles should be replaced (by content words), and alterations/changes concern both the words (stems) and various inflections that come from the strongly agglutinative character of Hungarian (e.g. *mentek .. mentünk* or *filmet .. filmeket*). Analyzing the correlations between paired types of disfluency, significant correlation was found only between false starts and changes ($p<0.002$).

3.2 Experiment I. For the sake of this experiment the duration of silent pauses had to be defined. Silent periods were taken to be pauses when 65% of all the listeners identified them between two lexical items. In these terms, the duration of the shortest pause was 80 ms as a perceptual agreement among the listeners.

All the hesitations – in Hungarian with the vowel [ø] and sometimes with the consonant [m] – were detected 100% correctly. Subjects were able to detect unfilled pauses in 60% of all occurrences that shows better performance than that reported in the literature (cf. 28,2%) [2]. The identification of silent pauses, however, was not independent of the actual text sample (Fig. 1), and the difference proved to be significant ($F(2, 54) = 17.238$, $p<0.0001$). The faster the speech the less

correct the perception of pauses with all listeners. However, it seems to be contradicted by the results of text1. Text1 was faster than text3 but the identification scores of pauses for text1 were slightly better than for text3. Acoustic analysis revealed that the average duration of pauses was different in the two texts: 551 ms on average in text1 while 411.7 ms in text3. The speaker of text3 produced pauses independently of structural boundaries in about 40% of all occurrences while the speaker of text1 only in 15%. This means that not only the actual speech tempo but also the average duration of pauses and their place influence the listener's correct perception.

Figure 1



Correct identification of silent and filled pauses of the three text samples

There was no significant difference in correct identification of pauses between females and males but there was a slight tendency indicating that females show more sensitivity toward correct pause perception. However, there was a significant difference concerning the false markings of pauses between females and males in text1 and text2 ($F(1, 54 = 10.205$, $p < 0.002$). The quantity of false markings was bigger with females than with males which indicates that females felt the texts to be more dissected (Fig. 2).

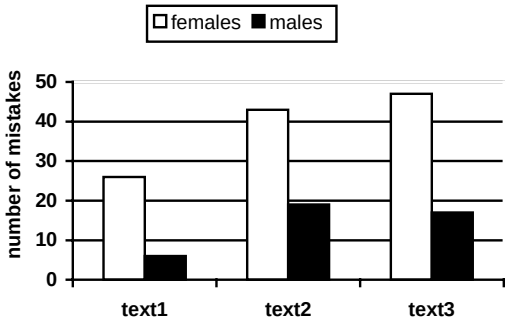


Figure 2
Number of false identification of pauses in the three texts

There was a strong correlation between the duration of pauses and their correct identification as it had been hypothesized ($r_p = 0.714$, $p < 0.0001$). The longer the pause the more correct its identification (Fig. 3). Again, factors other than actual duration also influenced the correct identification of pauses like their place within an utterance. Listeners were able to identify pauses that appeared at a structural boundary or phrase boundary significantly better (75% correct responses) as opposed to those that appeared elsewhere (36.8% correct identification).

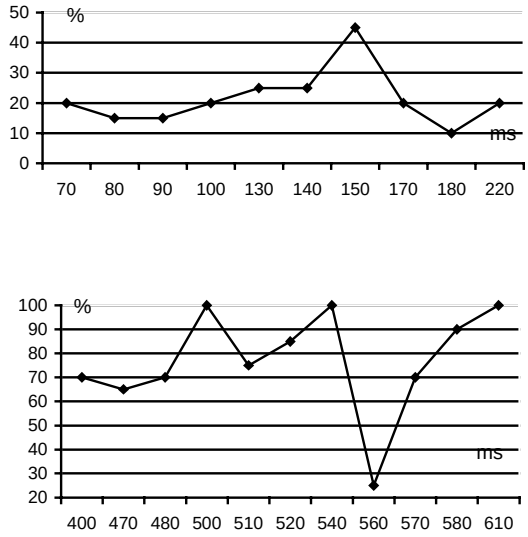


Figure 3
Interrelation of the duration of pauses and their identification (70-220 ms and 400-610 ms)

3.3 Experiment II. Recall accuracy was analyzed in the subjects' written narratives where the mean number of words and the main idea-units were taken into consideration. The number of words refers also to the quantity of narratives; the more the listener comprehended the more he could write down. Results show a significant difference depending on the presence vs. absence of pauses in the texts (Fig. 4). (Statistical analysis for words: $t(19) = 5.936$, $p < 0.0001$.)

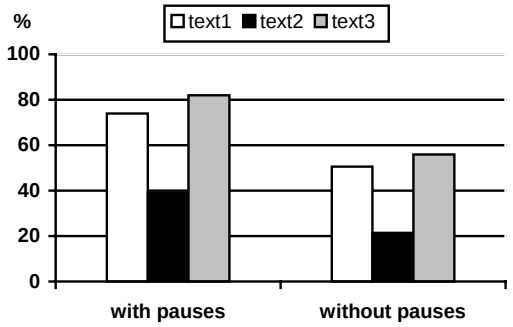


Figure 4
The differences of the number of words in the narratives

The number of words and the number of main idea-units of a narrative based on verbal comprehension predict the performance level of correct comprehension quite well [9]. A lot of extra information was found concerning the details with

those subjects who listened to the texts with their original pauses. No such information could be traced, however, in those subjects' narratives who had no pauses in their heard texts. Figure 5 shows the differences of the main ideas in percentages across the three texts (statistical analysis for the main idea-units: $t(19) = 8.320$, $p < 0.0001$).

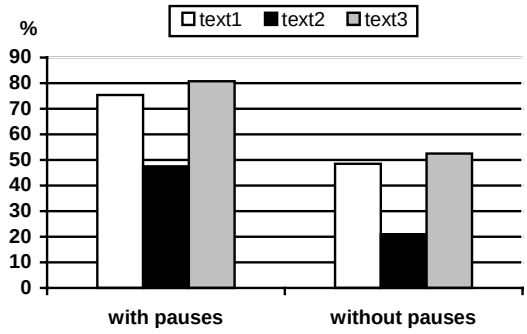


Figure 5
Differences of the main idea-units recalled in the narratives

The analysis of false statements revealed that subjects made three times more mistakes if there were no pauses in the heard texts. All these results indicate that the listener needs pauses and consequently other types of disfluency phenomena to a certain extent in order to fully comprehend the heard texts. It can be assumed, however, that too many interruptions will affect the comprehension negatively, as well.

4. Conclusion

Acoustic-phonetic analysis of spontaneously uttered texts revealed that disfluency phenomena are characteristic of the speaker and are relatively independent of the actual articulation tempo. The types and occurrences of disfluency seem to be (i) speaker-dependent and (ii) language-dependent to a certain extent. This latter means that language structure affects speech planning and speech production processes that result in specific occurrences of disfluency phenomena. In our material the most frequent types of disfluency were filled and unfilled pauses with the duration range of 70 and 1200 ms.

Listeners identified the majority of silent and filled pauses independently of their frequency, of the topic of the text, or of the speaker's fundamental frequency characteristics. There is a close correlation between the duration of pauses and their correct identification, however, a lot of other factors influence the listeners' final perceptual judgments like grammatical, syntactical predictions, semantic presuppositions, effects of suprasegmental patterns, actual articulation, etc. False pause identifications originated in (i) their unnoticed existence and (ii) their identification where no pauses existed. Sex differences were significant in this respect.

The listener, decoding verbal messages, activates all his language (and other stored) knowledge in order to follow the acoustic flow of heard speech. In order to construct meaning from a continuous acoustic stream he unconsciously controls the process of speech comprehension by his own inner speech production. The listener behaves as speaker when he/she processes speech. Objective data confirm the activation also of the Broca-area in the left hemisphere when comprehending speech. Successful comprehension requires that the listener should identify words, detect syntactic structures, and extract meaning from individual sentences, and finally build relations among the various parts of the text. Pauses are meant to provide time for such operations on the one hand, and for correction processes if they are needed, on the other hand. If the context does not enable the listeners to "fill in" the gaps created during the comprehension process, pauses or other types of disfluency can take over the role. This does not mean that speech is incomprehensible without pauses but it does mean that comprehension performance is significantly restricted in their absence. In this case comprehension is focused on key words and/or on a limited number of ideas the text contains. This would lead to an uncertain decoding process that results in a semantically restricted outcome with the possibility of false statements. It seems to be a paradox that two contradictory facts support the same explanation, i.e. pauses are not all identified, however, pauses are needed for correct decoding operations. If they are needed why cannot they be perceived absolutely correctly? Pauses can be ignored both perceptually and functionally during comprehension when they are not useful on the one hand but they are assumed to be used as parts of the decoding strategy even if they are consciously unnoticed, on the other hand.

Our results show that (i) instances of disfluency are consequences of the speaker's speech planning processes indicating their language-specific nature, (ii) they are relatively well predicted, (iii) their reasons and occurrences are unconsciously known by the listener as well, and (iv) the listeners need pauses and hesitations in order to comprehend the heard texts better. On the basis of these results the hypothesis of the double function of disfluency in spontaneous speech is confirmed.

5. References

[1] Goldman Eisler, F.: Psycholinguistics. (Experiments in Spontaneous Speech.) Academic Press. London 1968.

[2] Mahl, G.: Disturbances and silences in the patient's speech in psychotherapy. Journal of Abnormal and Social Psychology 53. 1956, 1-15.

[3] Hieke, A.E.–Kowal, S.–O'Connel, D.C.: The trouble with "articulatory" pauses. Language and Speech 26. 1983, 203-219.

[4] Duez, D.: Silent and non-silent pauses in three speech styles. Language and Speech 25. 1982, 11-25.

[5] Laver, J.: Principles of Phonetics. Cambridge University Press. Cambridge 1995.

[6] Shapley, M.: Prosodic variation and audience response. Papers in Pragmatics Vol. 1. No. 2. 1987, 66-80.

[7] Misono, Y.–Kiritani, S.: The distribution pattern of pauses in lecture-style speech. Logopedics and Phoniatrics 2. 1990, 110-113.

[8] Horga, D.: Samoispravljanje u govornoj proizvodnju. Suvremena linguistika 23/1-2. 1997, 91-105.

[9] Wingfield, A.– Tun, P.A.–Rosen, M.J.: Age differences in veridical and reconstructive recall of syntactically and randomly segmented speech. Journal of Gerontology Vol. 50. 1995, 257-266.