

Pausing Strategies in Children

Anna Esposito^{1,2}

¹ Seconda Università di Napoli, Dipartimento di Psicologia, Via Vivaldi 43, Caserta Italy
anna.esposito@unina2.it, iass.annaesp@tin.it

² IIASS, Via Pellegrino 19, 84019, Vietri sul Mare, Italy, INFN Salerno, Italy

Abstract. This study reports on the cross-modal analysis (video and audio) of spontaneous narratives produced by children (9 plus-minus 3 months years old) and is aimed to test the role of speech pauses (filled and empty) in children discourse organization. Video analysis was necessary to assess the association between utterance's meaning and pauses. Empty speech pauses were divided into three categories according to their duration: a) short - from 0.150 up to 0.500 s long; b) medium - from 0.501 up to 0.900 s long; c) long - more than 0.900 s long. Results show that each category plays a different role in the discourse organization, with short pauses and medium pauses preceding utterances containing new added information and long pauses identifying changes in scene, time and event structures, and the functional role of delimitating paragraphs. Moreover, different pause durations also depend on the amount of information contained in the utterances and signal the cognitive effort required to convey new, or given information. The present data may be relevant in assessing the system of rules that underlie pausing means, and identify predictive schemes of the speech temporal structure useful to improve text-to-speech synthesis systems.

1 Introduction

A characteristic of spontaneous speech, as well as of other types of speech, is the presence of silent intervals (empty pauses) and vocalizations (filled pauses) that do not have a lexical meaning. Such pauses seem to play a role in controlling the speech flow. Several studies have been conducted to investigate the system of rules that underlie speaker pausing strategies and their psychological bases. Research in this field had shown that pauses may play several communicative functions, such as build up tension or generate the listener's expectations about the rest of the story, assist the listener in his task of understanding the speaker, signal anxiety, emphasis, syntactic complexity, degree of spontaneity, gender, and educational and socio-economical information (Bernstein, 1962; Goldman-Eisler, 1968; Abram & Bever, 1969; Kowal et al., 1975; Green, 1977; O'Connell & Kowal, 1983).

Studies on empty pauses distribution in language production showed a relationship between pausing and discourse structure. Empty pauses are more likely to coincide with boundaries, realized as a silent interval of varying length, at clause and paragraph level (Brotherton, 1979; Gee & Grosjen, 1984; Rosenfied, 1987; Grosz & Hir-

shberg, 1992). This is particularly true for narrative structures where it has been shown that pausing marks the boundaries of narrative units (Chafe, 1980; Rosenfield, 1987; Chafe, 1987; O'Shaughnessy, 1995; Oliveria, 2000, 2002).

Several cognitive psychologists have suggested that pausing strategies reflect the complexity of neural information processing. Pauses will surface in the speech stream as the end product of a "planning" process that cannot be carried out during speech articulation and the amount and length of pausing reflects the cognitive effort related to lexical choices and semantic difficulties for generating new information (Goldman-Eisler, 1968; Butterworth, 1980; Chafe, 1987).

Along the above guidelines, the aim of the reported experiments was to investigate how different kind of pausing strategies, such as empty and filled pauses, and phoneme lengthening are used by children to shape the discourse structure in Italian and furthermore, to identify how many of the silent intervals can be attributed to the amount of "given" and "added" information the speaker is conveying in the speech flow.

2 Definitions

The present report interprets the concepts of "given" and "added" according to the definition proposed by Chafe (1974), which considered as "added" any verbal material that produces a modification in the listener's conscious knowledge, and therefore "given" verbal material was intended as not to produce such a modification. Moreover, the label *not classified* is attributed to speech material such as filled pauses, and/or short interruptions (such as "ap*") that follow empty pauses. Together with the above definitions it is necessary to introduce the concept of "changes" that in the present report are labels attributed to empty pauses identifying *changes in scene, time and event structures*. In this context, *changes* could be attributed to any pause independently of the kind of speech material (*given, added, and not classified*) that precedes them.

In this context, an *empty pause* (EP) is a *silent interval of more than 0.150 s*. Normally *filled pauses* (FP) are used to "hold the floor" i.e. preventing interruption by the listener while the speaker searches for a specific word (Erbaugh, 1987) and different fillers may serve different functions, such as marking a successful memory search ("ah") or signaling the selection of an example ("oh"). However, in this context, filled pauses generally appear as "*hum, eh*" because the task and the lack of an interlocutor preclude their use for other functions. Even though *phoneme lengthening* can appropriately be considered as a filled pause, such potential filled pauses were measured and analyzed on a separate ground. Moreover a "clause" is assumed to be "a sequence of words grouped together on semantic or functional basis" and a "paragraph" was considered as "a sequence of several clauses connected together by the same subject or scene"

3 Materials and Methods

The video recordings on which our analysis is based are of narrations by 10 female children (9 plus-minus 3 months' years old). The children told the story of a 7-minute animated color cartoon they had just seen. The cartoon is of a type familiar to Italian children, involving a cat and a bird. The listener in each case was the child's teacher and other children. This kept out stranger-experimenter inhibitions from the elicitation setting; i.e., factors that could result in stress and anxiety. Limiting these factors allows us to rule out the "socio-psychological"-type of pause (Beaugrande, 1984). The cartoon has an episodic structure, each episode characterized by a "cat tries to get bird/is foiled" narrative arc. Because of the cartoon's episodic structure, typically children will forget entire episodes. Therefore, only two episodes were analyzed, the ones that all the children remembered. The data were recorded at the International Institute for Advanced Scientific Study, Vietri, Italy. None of the participants was aware that speech pauses were of interest. The video was analyzed using commercial video analysis software (VirtualDub™). The program allows viewing of video-shots in 3-D, and movement forward and backward through the shots. The speech waves, extracted from the video, were sampled at 16 kHz and digitalized at 16 bits. The audio was analyzed using Speechstation2™ from Sensimetrics. For the audio measurements the waveform, energy, spectrogram, and spectrum were considered together, in order to identify the beginnings and endings of utterances, filled and empty speech pauses and phoneme lengthening. The details of the criteria applied to identify the boundaries in the speech waveform are accurately described in Esposito and Stevens (1995). Both the video and audio data were analyzed perceptually as well, the former frame-by-frame and the latter *clause-by-clause*.

4 Preliminary Results

Table 1 reports, for each child, the absolute number of occurrences of various pausing means and their percentage (between brackets) over the two episodes. Table 1 also reports the percentage of empty pauses and the percentage of filled pauses and phoneme lengthening for each child.

It can be observed that, on overall, empty pauses are considerably more frequent (49%) than filled pauses (28%) and phoneme lengthening (22%) taken separately. Among children, S6 and S8 are those that use a higher percentage of empty pauses, which is compensated by a reduced number of filled pauses. Moreover, among empty pauses, short pauses (33%) are largely more frequent than medium ones (10%), which in turn are more frequent than long pauses (6%) suggesting that the three duration ranges play a different role in structuring the discourse.

Since this is generally true also for each subject (except S5), it also suggests that children use a similar pause duration strategy to highlight different discourse units.

Table 1. Absolute number of occurrence of empty (short, medium, long) pauses, filled pauses, and phoneme lengthening (% between brackets) broken down for per child

Sub-jects	Short EP	Me-dium EP	Long EP	Filled	Length.	To-tal	% of EP	% of FP and Length.
S1	21 (22)	15 (16)	8 (9)	29 (31)	21 (22)	94	47	53
S2	25 (24)	18 (17)	10 (10)	33 (31)	18 (17)	104	53	48
S3	18 (22)	4 (5)	7 (9)	31 (38)	21 (26)	81	36	64
S4	61 (52)	2 (1)	1 (1)	33 (28)	21 (18)	118	54	46
S5	4 (12)	5 (15)	10 (30)	11 (34)	3 (9)	33	57	43
S6	39 (60)	2 (3)	3 (5)	10 (16)	10 (16)	64	68	32
S7	20 (26)	7 (10)	1 (1)	21 (27)	28 (36)	77	37	63
S8	10 (44)	6 (26)	0	6 (26)	1 (4)	23	70	30
S9	38 (36)	9 (8)	4 (4)	18 (17)	37 (35)	106	48	52
S10	19 (29)	7 (11)	3 (5)	24 (37)	12 (18)	65	45	55
AV Tot.	255 (33)	75 (10)	47 (6)	216 (28)	172 (22)	765	51	49

On overall, filled pauses are frequent as much as phoneme lengthening, even though there is a large intra-speaker and inter-speaker variability. Moreover, Table 1 shows that pausing means are differently used by different children with some child that make use of more filled pauses and vowel lengthening than empty pauses (as S6 and S8), other that does the opposite (as S3 and S7), yet other, along their speech, equally distribute empty and filled pauses.

To investigate if different duration ranges play a different role in structuring discourse units and if the need of pausing is due to the cognitive effort to recall from memory and lexicalize concepts that are not yet known by the listener, we evaluated the amount of *given*, *added* and *not classified* information that precedes each empty pause. Table 2 reports for each speaker the number of short, medium, and long pauses that follow *given*, *added*, and *not classified* speech material. Here the label *not classified* is attributed to speech material such as filled pauses, and/or short interruptions (such as “ap*”). We also report the number of short, medium and long empty pauses that are associated to a change (*changes*) of scene or paragraph structure. Note that *changes* could happen independently of the kind of speech material (*given*, *added*, and *not classified*) that precedes them and therefore, should not be counted in the total percentage.

Results in Table 2 reveal an interesting patterning. Most of the pauses, independently of their length, follow new added information, except for a few short (4%) and medium (3%) pauses that follow given information. Moreover, most of the long pauses (96%) are associated to a change of scene suggesting that children follow a pausing strategy similar to adults in signaling discourse boundaries.

Table2. Absolute number of occurrences of short, medium, and long empty pauses associated with a given, added, not classified information and with changes of scene in the discourse structure. The percentage (between brackets) is computed over the number of pauses in each duration range.

Subjects	Short_EP	Medium_EP	Long_EP
given_S1	2 (10)	1 (7)	0 (0)
added_ S 1	17 (81)	13 (87)	7 (88)
not classified_S1	2 (9)	1 (6)	1 (12)
changes_ S1	1 (5)	11 (73)	8 (100)
given_S2	0 (0)	1 (6)	0 (0)
added_ S2	22 (88)	17 (94)	7 (70)
not classified_ S2	3 (12)	0 (0)	3 (30)
changes_ S2	1 (4)	16 (89)	10 (100)
given_S3	0 (0)	0 (0)	0 (0)
added_ S3	18 (100)	4 (100)	5 (71)
not classified_S3	0 (0)	0 (0)	2 (29)
changes_ S3	8 (44)	4 (100)	7 (100)
given_S4	3 (5)	0 (0)	0 (0)
added_ S4	52 (85)	2 (100)	1 (100)
not classified_S4	6 (10)	0 (0)	0 (0)
changes_ S4	16 (26)	1 (50)	1 (100)
given_S5	0 (0)	0 (0)	0 (0)
added_ S5	4 (100)	5 (100)	6 (60)
not classified_S5	0 (0)	0 (0)	4 (40)
changes_ S5	1 (25)	3 (60)	9 (90)
given_S6	0 (0)	0 (0)	0 (0)
added_ S6	38 (97)	2 (100)	3 (100)
not classified_S6	1 (3)	0 (0)	0 (0)
changes_ S6	8 (21)	2 (100)	3 (100)
given_S7	0 (0)	0 (0)	0 (0)
added_ S7	20 (100)	7 (100)	1 (100)
not classified_S7	0 (0)	0 (0)	0 (0)
changes_ S7	5 (25)	6 (86)	1 (100)
given_S8	0 (0)	0 (0)	0 (0)
added_ S8	9 (90)	5 (83)	0 (100)
not classified_S8	1 (10)	1 (17)	0 (0)
changes_ S8	4 (40)	6 (100)	0 (0)
given_S9	3 (8)	0 (0)	0 (0)
added_ S9	29 (76)	7 (78)	4 (100)
not classified_S9	6 (16)	2 (22)	0 (0)
changes_ S9	5 (13)	6 (67)	3 (75)
given_S10	1 (5)	0 (0)	0 (0)
added_ S10	16 (84)	5 (71)	2 (67)
not classified_S10	2 (11)	2 (29)	1 (33)
changes_ S10	2 (11)	6 (67)	3 (100)
total_given	9 (4)	2 (3)	0 (0)
total_added	225 (88)	67 (89)	36 (77)
total_not_classified	21 (8)	6 (8)	11 (23)

total changes	51 (20)	61 (81)	45 (96)
----------------------	----------------	----------------	----------------

Children pause, like adults, to recover from their memory the new information they try to convey to the listeners, showing that higher is the recovery effort, longer is the pausing time. As it could be seen in Table 2, longer are the pauses, lower is the probability that they can be associated to a given information. The relationship with the cognitive effort can be easily seen examining the amount of long pauses associated with changes (of scene, time and event structure). In fact, an high percentage of medium (81%) and long (96%) pauses are made to signal these changes, whereas only a low percentage of short (20%) pauses serve this purpose.

The above data also suggest a predictive scheme of the speech temporal structure and the alternating pattern of cognitive rhythm in the production of spontaneous narratives. If we want to identify, in this alternating pattern, paragraphs or changes (of scene, time, and structure) we should check for long pauses first, that account for the highest percentage (96%) of such changes and then for medium pauses (81%). Pauses shorter than 0.500 s have the lowest probability to correspond to a change of scene (20%) in the flow of the narration, even though they are more frequent than medium and long pauses.

5 Conclusions

This study investigates the system of rules that underlie children pausing strategy in discourse structure and their psychological bases. We found that, as in adults (reported by several authors - see introduction) pausing is an indicator of the cognitive effort made for planning speech. A considerable amount of pausing made by children is used to bring new concepts to the listener's conscious knowledge. Only a few among the short (4%), and medium (3%) empty pauses mark given information. More data are necessary to assess the role of pauses that mark not classified information, since such speech material consists of filled pauses and interruptions that can also be considered as indicators of cognitive planning.

Children use pausing to mark clauses and paragraph consistently. We would expect children, being less skilled in the use of the language's lexicon to make more pauses at word level than at the clause and paragraph level, unless we hypothesize an innate model for discourse structures. This hypothesis is further on supported by the fact that in 56% of the cases children pauses occur right after the first word in a clause, i.e. right after a filler conjunction that signals a major transition in the speech flow and serves to plan the content of the continuation of the discourse. Another interesting result of this analysis is the fact that a considerable number of long pauses and a consistent number of medium ones is used to mark paragraph boundaries whereas short pauses rarely served for this function. The consistency among the subjects in the use of pausing means seems to suggest a very coarse and general timing model, that speakers use to regulate speech flow and discourse organization. It would be interesting to conduct an analysis on more data to make sense of how this model works, since it would be of great utility in the field of human-machine interaction,

favoring the implementation of more natural speech synthesis and interactive dialog systems.

References

- Abrams, K. & Bever, T. G. (1969) Syntactic structure modifies attention during speech perception and recognition, *Quarterly Journal of Experimental Psychology*, 21, 280-290.
- Bernstein, A. (1962) Linguistic codes, hesitation phenomena, and intelligence, *Language and Speech*, 5, 31-46.
- Beaugrande, R. (1984). Text production. Norwood NJ: Text Publishing Corporation.
- Brotherton, P. (1979) Speaking and not speaking; process for translating ideas into speech. In *Of Time and Speech* (edited by A. Siegman and S. Feldestein), Hillsdale, Lawrence Erlbaum, N.J., 179-209.
- Butterworth, B. L. (1980). Evidence for pauses in speech. In *Language Production: Volume 1. Speech and Talk* (edited by Butterworth B. L.), London Academic Press, 155-176.
- Chafe, W. (1974) Language and consciousness, *Language*, 50, 111-133.
- Chafe, W. (1980) The deployment of consciousness in the production of a narrative. In *The Pear Stories*.(edited by W.L. Chafe), Norwood N.J., Ablex, 9-50.
- Chafe, W. (1987) Cognitive constraint on information flow. In *Coherence and Grounding in Discourse* (edited by R. Tomlin) John Benjamins, 20-51.
- Esposito, A. & Stevens, K. N. (1995) Notes on Italian vowels: An acoustical study (part I), *RLE Speech Communication Working Papers*, 10, 1-42.
- Erbaugh, M. S. (1987) A uniform pause and error strategy for native and non-native speakers. In *Coherence and Grounding in Discourse* (edited by R. Tomlin) John Benjamins, 109-130.
- Kowal, S., O'Connell, D. C. & Sabin, E. J. (1975) Development of temporal patterning and vocal hesitations in spontaneous narratives, *Journal of Psycholinguistic Research*, 4, 195-207.
- Gee, J. P. & Grosjean, F. (1984) Empirical evidence for narrative structure. *Cognitive Science* 8, 59-85.
- Green, D. W. (1977) The immediate processing of sentence, *Quarterly Journal of Experimental Psychology*, 29, 135-146.
- Goldmar Eisler F. (1968) *Psycholinguistic: experiments in spontaneous speech*, London: New York, Academic press.
- Grosz, B. & Hirschberg, J. (1992) Some intentional characteristics of discourse structure. *Proceedings of International Conference on Spoken Language Processing, Banff*, 429-432.
- O'Connell, D. C. & Kowal, S. (1983) Pausology, *Computers in Language Research* 2, 19, 221-301.
- Oliveira, M. (2000) *Prosodic Features in Spontaneous Narratives*, Ph.D. Thesis, Simon Fraser University.
- Oliveira, M. (2002) Pausing strategies as means of information processing narratives. *Proceeding of the International Conference on Speech Prosody, Ain-en-Provence*, 539-542.
- O'Shaughnessy D. (1995). Timing patterns in fluent and disfluent spontaneous speech. *Proceedings of ICASSP Conference, Detroit*, 600-603.
- Rosenfield, B. (1987). *Pauses in Oral and Written Narratives*. Boston, Boston University.