

# Children's Organization of Discourse Structure Through Pausing Means

Anna Esposito<sup>1,2</sup>

<sup>1</sup>Seconda Università di Napoli, Dipartimento di Psicologia, Via Vivaldi 43, Caserta, Italy  
anna.esposito@unina2.it, iiass.annaesp@tin.it

<sup>2</sup>IIASS, Via Pellegrino 19, 84019, Vietri sul Mare, INFN Salerno, Italy

**Abstract.** This study aims to investigate on how different kinds of pausing strategies, such as empty and filled pauses, and phoneme lengthening are used by children to shape the discourse structure in Italian, and 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. To this aim a cross-modal analysis (video and audio) of spontaneous narratives produced by male and female children (9 plus-minus 3 months years old) was performed. 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. The analysis showed that each of the above categories seems to play a different role in the children discourse organization. Children pause, like adults, to recover from their memory the new information they try to convey. Higher is the recovery effort, longer is the pausing time. Longer are the pauses, lower is the probability that they can be associated to a given information. Most of the long pauses (96% for female and 94% for male) are associated to a change of scene suggesting that long pauses are favored by children in signaling discourse boundaries. The consistency, among subjects, in the distribution of speech pauses seems to suggest that, at least in Italian, there is an intrinsic model of timing, probably a very coarse model, that speakers use to regulate speech flow and discourse organization.

## 1 Introduction

Spontaneous speech, as well as other types of speech, is characterized by the presence of silent intervals (empty pauses) and vocalizations (filled pauses) that do not have a lexical meaning. These pausing means play several communicative functions. In fact, it has been shown that their occurrence is determined by several factors 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 [1, 2, 13, 15, 16, 18].

Pauses are not only generate by psychological motivations but also as a linguistic mean for discourse segmentation. Speakers systematically signal changes in scene, time, and event structures using speech pauses. Empty pauses are more likely to coincide with boundaries, realized as a silent interval of varying length, at clause and paragraph level

[4, 14, 17, 22]. This is particularly true for narrative structures where it has been shown that pausing marks the boundaries of narrative units [7, 8, 19, 20, 21, 22].

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 [5, 8, 16].

Moreover, there is a practical interest in studying pause distribution along speech for application on automatic speech synthesis and recognition. Most of the current automatic speech systems ignore the effects of pausing, temporal timing, utterance-final vowel lengthening. Thus, a better knowledge of how pausing strategies affect spontaneous speech should be of support in the design of more natural speech synthesis systems and improve the performance of automatic speech recognition systems.

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 work interprets the concepts of "given" and "added" according to the definition proposed by Chafe [6], 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 work 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 [12] 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 and 4 male 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 [3]. 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 [9]. 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

Tables 1 and 2 report, for each male and female child respectively, the absolute number of occurrences of various pausing means and their percentage (between brackets) over the two episodes. Tables 1 and 2 also report the percentage of empty pauses and the percentage of filled pauses and phoneme lengthening for each child.

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.** *Female children:* absolute number of occurrence of empty (short, medium, long) pauses, filled pauses, and phoneme lengthening (% between brackets) broken down for per child

Female children	Short EP	Medium EP	Long EP	Filled	Vowel Length.	Total	% of EP	% of FP and Vowel 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
<b>Tot.</b>	<b>255 (33)</b>	<b>75 (10)</b>	<b>47 (6)</b>	<b>216 (28)</b>	<b>172 (22)</b>	<b>765</b>	<b>51</b>	<b>49</b>

**Table 2.** *Male children:* absolute number of occurrence of empty (short, medium, long) pauses, filled pauses, and phoneme lengthening (% between brackets) broken down for per child

Male children	Short EP	Medium EP	Long EP	Filled	Vowel Length.	Total	% of EP	% of FP and Vowel Length.
S1	13 (28)	10 (21)	4 (8)	13 (28)	7 (15)	47	57	43
S2	20 (39)	9 (17)	2 (4)	8 (15)	13 (25)	52	60	40
S3	25 (45)	10 (18)	3 (5)	7 (12)	11 (20)	56	68	32
S4	18 (24)	13 (19)	7 (9)	21 (29)	15 (20)	74	52	49
<b>Tot.</b>	<b>76 (33)</b>	<b>42 (18)</b>	<b>16 (7)</b>	<b>49 (22)</b>	<b>46 (20)</b>	<b>229</b>	<b>59</b>	<b>41</b>

On overall, filled pauses are frequent as much as phoneme lengthening, even though there is a large intra-speaker and inter-speaker variability. Moreover, Tables 1 and 2 show 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 S3 and S7 for female children), other that does the opposite (as S6 and S8 for female children, and S3 for male children), yet other, along their speech, equally distribute empty and filled pauses.

Empty pauses, taken separately, are considerably more frequent than filled pauses and phoneme lengthening both for male (59% against 21% and 20%) and female (51% against 28% and 22%). Moreover, short pauses (33% both for male and female) are more frequent than medium ones (18% for male and 10% for female), which in turn are more frequent than long pauses (7% for male and 6% female). This is generally true for each subject (except female S5) suggesting that children use a similar strategy in distributing short, medium and long pauses in their narrations.

**Table 3.** Absolute number of occurrences of short, medium, and long empty pauses associated with a given, added, and not classified information, and changes of scene in the discourse structure. The percentage (between brackets) is computed over the number of pauses in each duration range. (Results are reported both for male and female children)

<i>10 Female children</i>	<i>Short EP</i>	<i>Medium EP</i>	<i>Long_EP</i>
<b>Given</b>	9 (4%)	2 (3%)	0 (0)
<b>Added</b>	225 (88%)	67 (89%)	36 (77%)
<b>Not classified</b>	21 (8%)	6 (8%)	11 (23%)
<b>Changes</b>	51 (20%)	61 (81%)	45 (96%)
<b>4 Male children</b>			
<b>Given</b>	5 (6%)	0	0
<b>Added</b>	65 (86%)	37 (88%)	13 (81%)
<b>Not classified</b>	6 (8%)	5 (12%)	3 (19%)
<b>Changes</b>	6 (8%)	14 (33%)	15 (94%)

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 3 reports 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\**”).

The number of short, medium and long empty pauses that are associated to a change (*changes*) of scene or paragraph structure is also reported. 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.

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 3, most of the pauses follow new added information, except for a few short (4% for female and 6% for male) and medium (3% only for female) pauses. Most of the long pauses (96% for female and 94% for male) are associated to a change of scene suggesting that long pauses are favoured by children in signalling discourse boundaries.

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% for female, 33% for male children) and long (96% for female and 94% for male children) pauses are made to signal these changes, whereas only a low percentage of short (20% for female and 8% for male children) pauses serve this purpose.

The above data also suggest a predictive scheme for the alternating pattern of cognitive rhythm in the production of spontaneous narratives. In this alternating pattern, long pauses account for the highest percentage (96% and 94% for female and male

children respectively) of paragraphs or changes followed by medium pauses (81% and 33% for female and male children respectively). Even though they are more frequent than medium and long pauses, short pauses (no longer 0.500 s) have a low probability to signal a change of scene (20%) in the flow of the narration.

Tables 4 and 5 gives, for each child, the number of words (Wds), clauses (Cls) and paragraphs (Phs) marked by a pause (#Ps). In this case, filled and empty pauses are grouped together. Note that pauses that mark a word boundary can also mark a clause, a filler conjunction and a paragraph boundary.

The results in Tables 4 and 5 show that, on the average, 19% of the word boundaries are marked by a pauses independently of the number of words used. The pattern is still more reliable at clause and paragraph level where children mark with a pause 70% of the clause boundaries and 100% of the paragraph boundaries. There is a variability in the above pattern mostly observed among the female children, since the female children S5, S6, S8, and S10 pause less than the others at word boundaries, and S7 and S9 use also other means (such as vowel lengthening) to signal paragraphs boundaries.

**Table 4.** Absolute number and percentage (between brackets) of words (Wds), clauses (Cls), and paragraphs (Phs) marked by a pause (#Ps) for female children

<b>10 Female children</b>	<b>Wds</b>	<b>#Ps</b>	<b>(%)</b>	<b>Cls</b>	<b>#Ps</b>	<b>(%)</b>	<b>Phs</b>	<b>#Ps</b>	<b>(%)</b>
<b>S1</b>	217	54	(25)	43	31	(72)	10	10	(100)
<b>S2</b>	249	49	(20)	56	37	(66)	11	11	(100)
<b>S3</b>	224	43	(19)	39	28	(72)	13	13	(100)
<b>S4</b>	307	74	(24)	49	36	(73)	15	15	(100)
<b>S5</b>	118	18	(15)	23	13	(57)	8	8	(100)
<b>S6</b>	324	47	(15)	68	36	(53)	15	15	(100)
<b>S7</b>	206	40	(19)	34	26	(76)	9	8	(89)
<b>S8</b>	149	15	(10)	25	14	(56)	6	6	(100)
<b>S9</b>	217	45	(21)	41	31	(76)	13	10	(77)
<b>S10</b>	187	29	(16)	34	23	(68)	11	11	(100)
<b>Averaged total</b>	<b>2198</b>	<b>414</b>	<b>(19)</b>	<b>412</b>	<b>275</b>	<b>(68)</b>	<b>111</b>	<b>107</b>	<b>(96)</b>

**Table 5.** Absolute number and percentage (between brackets) of words (Wds), clauses (Cls), and paragraphs (Phs) marked by a pause (#Ps) for male children

<b>4 Male Children</b>	<b>Wds</b>	<b>#Ps</b>	<b>(%)</b>	<b>Cls</b>	<b>#Ps</b>	<b>(%)</b>	<b>Phs</b>	<b>#Ps</b>	<b>(%)</b>
<b>S1</b>	142	29	(20)	21	15	(71)	9	9	(100)
<b>S2</b>	182	33	(18)	28	20	(71)	6	6	(100)
<b>S3</b>	218	35	(16)	34	23	(68)	8	8	(100)
<b>S4</b>	208	49	(24)	37	27	(73)	12	12	(100)
<b>Averaged total</b>	<b>750</b>	<b>146</b>	<b>(19)</b>	<b>120</b>	<b>85</b>	<b>(71)</b>	<b>35</b>	<b>35</b>	<b>(100)</b>

## 5 Conclusions

This study was devoted to investigate on the system of rules that underlie children pausing strategy and their psychological bases. The reported data show that empty pauses of short, medium, and long duration are largely used by children to signal new information to the listeners' conscious knowledge and only a few among the short (4% for female and 6% for male), and medium (3% only for female) empty pauses mark given information. This suggests that children pause, like adults [16, 20, 21, 22], to recover from their memory the new information they try to convey. Higher is the recovery effort, longer is the pausing time. Moreover, longer are the pauses, lower is the probability that they can be associated to a given information, supporting the hypothesis that pausing plays the functional role of indicating the cognitive effort needed for planning speech.

Pauses are not only generate by psychological motivations but also as a linguistic mean for discourse segmentation. Pauses are used by children to mark words, clause, and paragraph boundaries. The results show that a similar percentage of word (19% both for male and female children), clause (68% for female and 71% of male children) and paragraph (96% of female and 100% of male children) boundaries is marked by a pause. Short pauses (67.6% for females and 56.7 for males) than medium (19.9% for females and 31.4% for males), that in turn are more frequent than long pauses (12.5% for females and 11.9% for males). However, children systematically signal changes in scene, time, and event structures using medium (81% for female and 33% for male) and long pauses (96% for female and 94% for male) suggesting that only pauses longer than 0.5 s are favored by children to mark paragraph boundaries whereas short pauses rarely served for this function.

This result favors the hypothesis of an universal model for discourse structure, otherwise 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. This hypothesis is further on supported by the fact in a previous work [10-11] it has been shown that in 56% of the cases children's 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 message content for the continuation of the discourse. 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. More data are needed 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.

## Acknowledgements

This work has been partially funded by the American project *IT's Enabled Intelligent and Ubiquitous Access to Educational Opportunities for Blind Student*, in collaboration with professor Nicholas Bourbakis, Information Technology Research Institute (ITRI), Wright State University, Dayton, OHIO, USA. The author would like to thank professor Maria Marinaro for her useful comments and suggestions, and Luisa Del Prete and Giulia Palombo who collected the video recordings. Miss Tina Marcella Nappi is acknowledged for her editorial help.

## References

1. Abrams, K., Bever, T. G.: Syntactic Structure Modifies Attention during Speech Perception and Recognition. *Quarterly Journal of Experimental Psychology*, Vol. 21 (1969) 280-290
2. Bernstein, A.: Linguistic Codes, Hesitation Phenomena, and Intelligence. *Language and Speech*, Vol. 5 (1962) 31-46
3. Beaugrande, R.: Text Production. Norwood NJ: Text Publishing Corporation (1984)
4. Brotherton, P.: Speaking and not Speaking; Process for Translating Ideas into Speech. In A. Siegman, A., and Feldestein, S. (eds): *Of Time and Speech*. Hillsdale, Lawrence Erlbaum, N.J. (1979) 79-209
5. Butterworth, B. L.: Evidence for Pauses in Speech. In Butterworth B. L. (ed): *Language Production: Speech and Talk*, London Academic Press, Vol. 1 (1980) 155-176
6. Chafe, W. L.: Language and Consciousness. *Language*, Vol. 50 (1974) 111-133
7. Chafe, W. L.: The Deployment of Consciousness in the Production of a Narrative. In Chafe W.L.(ed): *The Pear Stories*, Norwood N.J., Ablex (1980) 9-50
8. Chafe, W. L.: Cognitive constraint on information flow. In Tomlin, R. (ed): *Coherence and Grounding in Discourse*, John Benjamins (1987) 20-51
9. Esposito, A., Stevens, K. N.: Notes on Italian Vowels: An Acoustical Study (Part I), RLE Speech Communication Working Papers, Vol. 10 (1995) 1-42
10. Esposito, A.: Pausing Strategies in Children. In *Proceedings of the International Conference in Nonlinear Speech Processing, Cargraphics, Barcelona, SPAIN, 19-22 April (2005)* 42-48
11. Esposito, A., Marinaro, M., Palombo, G.: Children Speech Pauses as Markers of Different Discourse Structures and Utterance Information Content. *Proceedings of the International Conference From Sound to Sense: +50 years of discoveries in Speech Communication, MIT, Cambridge, June 10-13 (2004)*
12. Erbaugh, M. S. (1987) A Uniform Pause and Error Strategy for Native and Non-Native Speakers. In Tomlin, R. (ed): *Coherence and Grounding in Discourse*, John Benjamins (1987) 109-130
13. Kowal, S., O'Connell, D. C., Sabin, E. J.: Development of Temporal Patterning and Vocal Hesitations in Spontaneous Narratives, *Journal of Psycholinguistic Research*, Vol. 4 (1975) 195-207
14. Gee, J. P., Grosjean, F.: Empirical Evidence for Narrative Structure. *Cognitive Science*, Vol. 8 (1984) 59-85.
15. Green, D. W.: The Immediate Processing of Sentence. *Quarterly Journal of Experimental Psychology*, Vol. 29 (1977) 135-146
16. Goldman-Eisler, F. *Psycholinguistic: Experiments in Spontaneous Speech*. London: New York, Academic press (1968)
17. Grosz, B., Hirschberg, J.: Some Intentional Characteristics of Discourse Structure. *Proceedings of International Conference on Spoken Language Processing, Banff (1992)* 429-432
18. O'Connell, D. C., Kowal, S.: Pausology. *Computers in Language Research* 2, Vol.19 (1983) 221-301
19. Oliveira, M.: *Prosodic Features in Spontaneous Narratives*, Ph.D. Thesis, Simon Fraser University (2000)
20. Oliveira, M.: Pausing Strategies as Means of Information Processing Narratives. *Proceeding of the International Conference on Speech Prosody, Ain-en-Provence (2002)* 539-542
21. O'Shaughnessy D.: Timing Patterns in Fluent and Disfluent Spontaneous Speech. *Proceedings of ICASSP Conference, Detroit (1995)* 600-603
22. Rosenfield, B.: *Pauses in Oral and Written Narratives*. Boston, Boston University (1987)