

THE EFFECT OF SURROUNDING PHRASE LENGTHS ON PAUSE DURATION.

Elena Zvonik and Fred Cummins

Department of Computer Science
University College Dublin
Belfield, Dublin 4, Ireland

{elena.zvonik, fred.cummins}@ucd.ie

ABSTRACT

Little is known about the determining influences on the length of silent intervals at IP boundaries and no current models accurately predict their duration. The contribution of independent factors with different characteristic properties to pause duration needs to be explored. The present study seeks to investigate if pause duration is correlated with the length of sentences or phrases preceding and following a pause. We find that two independent factors—the length of an IP (intonational phrase) preceding a pause and the length of an IP following a pause combine superadditively. The probability of a pause being short (<300 ms) rises greatly if both the preceding and the following phrases are short (≤ 10 syllables).

1. INTRODUCTION

Modeling speech timing in general and generating natural-sounding speech in particular require a close study of pause phenomena and their temporal characteristics. Pauses are known to be a multi-determined and highly variable phenomena which depend on many factors, such as speech rate, speaking style and discourse, to name just a few. This paper presents a quantitative study of the relation between pause duration and the length of the flanking phrases.

Most published studies exploring pause behavior deal with pauses and speaking rate, where speaking rate is conventionally understood to include pauses and articulated speech, while articulation rate excludes pauses. For example, defining pauses as being silent intervals of at least 200 ms, Fletcher [1] finds that most speakers vary the number of pauses produced, but some vary pause length to alter speech rate, especially when instructed to speak rapidly. Grosjean and Lane [2] looked at the relative contributions to perceived speech rate of both articulation rate and the number of pauses. They found that both pauses and speed of articulation contributed to perceived rate, but that the effect of the latter was much stronger than that of the former.

In an analysis of horse race commentaries, Trouvain and Barry [3] showed that as the race progresses, the commen-

tator's breath rate increases, and the duration of interpause intervals decreases, but pause durations do not uniformly shorten. The result is more pauses per time unit. Neither speech rate (syllables/sec, including pauses) nor articulation rate (excluding pauses) is consistently increased towards race end. The resulting percept of increased 'tempo' is persuasive, showing that pause behavior is intimately linked to perceived rate of speech, but the relationship is not a simple one.

In a study of clear speech, Uchanski et al. [4] inserted pauses into conversational speech, and deleted pauses from clear speech (which has a slower speech rate). The resulting speech was less intelligible in both cases, again pointing at a complex relationship between pauses and underlying speech timing.

Investigating pause behavior in relation to discourse has also received much attention. Having analyzed a number of spontaneous monologues (Dutch), Swerts et al. [5] come to the conclusion that filled pauses may carry information about discourse structure, i.e. major discourse boundaries tend to co-occur with filled pauses. They also note that filled pauses at stronger boundaries often have preceding and following silent pauses.

Gustafson-Čapkova and Megyesi [6] reported the effect of speaking style and discourse structure on pause distribution in Swedish. They mainly compared syntactic and discourse context in which speakers tend to make pauses in professional and non-professional readings and in spontaneous dialogs. Their results showed that in professional readings all the pauses appeared at strong boundaries, whereas in non-professional readings speakers tended to locate pauses at sentence and clause boundaries and in front of conjunctions. In spontaneous dialogs, however, silent intervals also appeared at weak boundary positions.

The above studies have looked at the function of pauses, and their distributional characteristics. Quantitative modeling of appropriate pause durations, however, has not received as much attention. Having analyzed a large corpus of read and spontaneous speech in five languages, Campione and Véronis [7] claimed that pause durations also

vary across languages. They found the average duration of pauses to be lower in Italian and higher in Spanish. The authors also described a trimodal distribution of pauses, categorizing them as brief (<200 ms), medium (200–1000 ms) and long (>1000 ms). All the data was log-transformed.

An interesting attempt to relate prosody to syntax in Dutch sentences was made by Terken and Collier [8]. They reported that the silent interval duration was gradually increasing at the NP-VP boundary when the length of the NP was increased. They reported two factors influencing the duration structure - syntactic complexity of the stretch of speech preceding as well as following the boundary and the length of the included words, and, moreover, that the effect of both factors is additive.

Following Terken and Collier, Strangert [9] investigated how the silent interval marking the boundary between an NP and a VP depended on the complexity of NP, the complexity of VP and on the length of the final word of the NP. The silent interval here seems to be a prolonged consonantal closure (full information on the sentential material is not provided), rather than a pause, but the durations reported clearly extend into the domain normally associated with pauses, rather than stops (max of 531 ms). The silent interval was longer as NP complexity increased, as VP complexity increased (a smaller effect), and as word length increased. Furthermore, the effects of word length and VP complexity combined additively (a single speaker was employed). These results are suggestive of a relationship between constituent length and/or complexity and juncture marking by silence.

All the published studies on silent pause durations demonstrate clearly that pause behavior is highly variable, depending in a complex fashion on both speaker and discourse situation. Synchronous Speech (SS) introduced by Cummins [10] may provide a means for reducing variability somewhat. SS is speech elicited when two speakers are asked to read a text together. In [10], pause placement was found to be considerably more predictable in synchronous speech than in unaccompanied readings. The results of the study described in [11] demonstrate not only that pause location is more consistent in synchronous speech, but that the pause duration is also less variable. In general, speakers produced pauses of comparable duration in both solo and synchronous conditions. In the case of one pause, however, durations in the solo condition were highly variable, while this variability was much reduced in the synchronous condition.

This particular pause also exhibited by far the longest mean duration, and it was suggested that this may have resulted from the greater than average length and complexity of the preceding sentence. Again, this suggests a relationship between silent duration and complexity. The nature and strength of this relationship needs to be explored, and

the results of such an investigation might be of use in the quantitative modeling of pause duration in speech synthesis.

The present study seeks to investigate if and how silent pause duration is correlated with the length of the phrase the phrases on either side. Synchronous speech is used as means of reducing inter-speaker variability.

2. METHODS

Six speakers (3 subject pairs) participated in the experiment. All were from the area of Dublin, Ireland. Each recording session provided a series of text readings which were made in the following order: one speaker from the pair first read the text alone, then both speakers read the text in synchrony and finally the second speaker read the text alone. Each speaker therefore read each text twice – once alone (solo condition) and once in synchrony. Each speaker was the first solo reader for every other text. Speakers wore head-mounted microphones; recordings were made onto the right and left channels of a stereo file. No control for familiarity of speakers within a pair was made.

The texts recorded were seven randomly chosen entries from "Bridget Jones's Diary" by Helen Fielding [12]. Each diary entry provided a separate reading. A set of predictors possibly affecting the duration of a pause was chosen: number of intonational phrases (IPs), number of lexical words and syllables in the sentence preceding a pause and number of lexical words and syllables in the IP preceding a pause. The same set of predictors was used to investigate the relationship between pause duration and the following sentence/IP. Only inter-sentential pauses were considered for analysis. The text was marked up with expected syllable and IP boundaries. A control reading from a phonetician, native speaker of English (Dublin, Ireland) was obtained, and predicted IP boundaries compared with those of the control reading. Syllable boundaries were marked according to dictionary rules. The electronic addition of Merriam-Webster Dictionary was used for this purpose. A total of 1212 pauses (101 per speaker (6) and condition (2)) of 50 milliseconds or longer were measured.

Based on cursory examination of the data, the present analysis will be restricted to consideration of the relationship between pause duration and the number of syllables in the IP preceding and following the pause. Syllables within an IP, rather than words within an IP or words/IPs within a sentence seemed to provide the cleanest results.

3. RESULTS

3.1. Pause Variability

For each pause in the text an index of variability for that pause was calculated as the standard deviation of pause duration produced by the 6 speakers. This provided matched indices for the solo and synchronous conditions. A sign test on the difference between matched indices showed inter-speaker variability to be smaller in the synchronous condition (sign test, $z=4.37$, $p < 0.001$).

Pause durations are highly variable, as found both here and in all previous studies. The reduction in inter-speaker variability seen in the synchronous condition is clearly a potential bonus for a quantitative analysis. In what follows, therefore, we present only the data from the synchronous condition.

3.2. Pause Duration

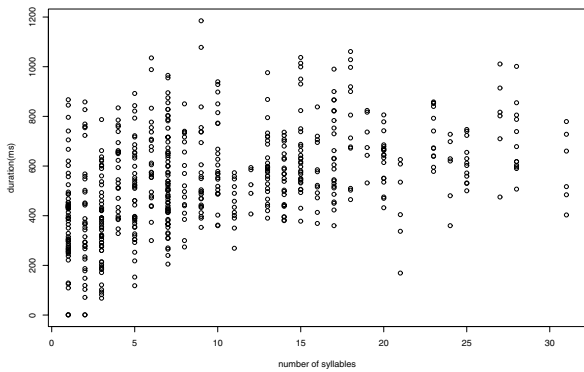


Fig. 1. Pause duration by number of syllables in a preceding IP.

Figure 1 shows the pause duration data as a function of the number of syllables in the preceding IP ($n=606$). No attempt to extract a measure of central tendency has been made, as the interesting feature of these data lies in the details seen for IPs of less than 10 syllables. These data are not very well described using a simple linear relationship ($R^2 = 0.17$), such predictability as there is comes largely from those pauses of duration less than about 300 ms, which occur almost exclusively with syllable counts of less than 10. If we exclude pauses of less than 300 ms duration from the analysis, the proportion of variance accounted for by a linear relationship all but vanishes ($R^2 = 0.09$).

Figure 2 shows a similar plot of pause duration as a function of the number of syllables in the following IP ($n=606$). Again, pauses of less than 300 ms almost only occur when the following IP has less than 10 syllables, al-

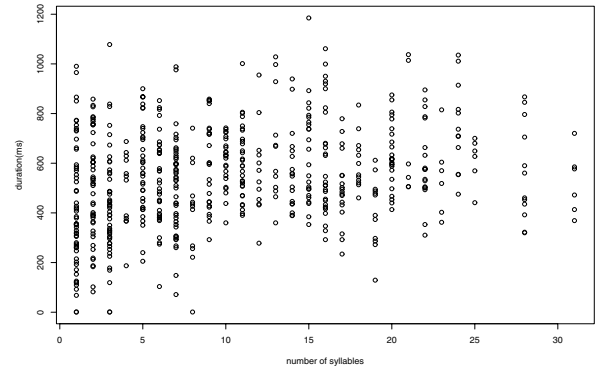


Fig. 2. Pause duration by number of syllables in a following IP.

though this relationship is not quite as clear as in the previous case. The proportion of variance (R^2) in pause duration accounted for by syllable count in the following IP is 0.06, and if we exclude pauses of less than 300 ms, this drops to 0.015.

These data suggest that pauses of less than 300 ms (approximately) may be special, in that they only occur when either the number of syllables in the preceding or following IP is less than about 10. These two predictors could themselves be correlated. This turns out, however, not to be the case. No significant correlation existed between the number of syllables preceding a pause and those following ($p = 0.19$, n.s.). There are therefore two independent factors which affect the probability of a pause being less than about 300 ms.

To look at how these factors interact, we counted the number of pauses with duration less than 300 ms for all combination of long and short IPs preceding and following the pause. As can clearly be seen from Table 1, these factors combine superadditively. The probability of a pause being short (<300 ms) rises greatly if both the preceding and the following IPs are short (≤ 10 syllables).

preceding IP	following IP	number of short pauses
short	long	8
short	short	55
long	short	2
long	long	0

Table 1. Number of puses of duration less than 300 ms as a function of the length of the surrounding IPs. For IPs, 'short' means having less than or equal to 10 syllables.

4. DISCUSSION

The present study, along with earlier studies described in [10] and [11] demonstrates that inter-speaker variability in pause duration is significantly reduced in SS, as a result of speakers' ability to coordinate pause timing when reading a given text together. The basis for this ability to produce this type of reduced speech is still unclear. Two hypotheses may be considered. The first one suggests that the speakers share an unconscious common knowledge of default timing values for speech. According to the second hypothesis, reduced variability is a result of eliciting speech while being engaged in a concurrent task, in our study- reading in tight synchrony with another person. If this is the case, it should be possible to get reduced variability by having subjects read a text in solo while being engaged in a concurrent task different to the task of reading together with another speaker. In order to test that, a small experiment was carried out. Subjects were asked to read the texts presented on a computer screen. They were instructed to hit quickly different keys on the keyboard if the background color of the screen flickered from white to red or from white to black when they were reading. Speakers wore headphones and were played cocktail party noise at a comfortable volume. Duration of inter-sentential pauses was measured and an analysis of variability was done as described in 3.1., providing matching indices for the solo-concurrent and synchronous-concurrent conditions. Results showed that inter-speaker variability was significantly lower in both the solo(sign test, $z=2.19$, $p < 0.03$) and synchronous(sign test $z=6.37$, $p < 0.001$) conditions, compared to concurrent. In other words, performing another non-speech test concurrently with reading did not reduce inter-speaker variability. This finding suggests that speakers have shared knowledge of speech timing, which they have to exploit to achieve tight synchrony when reading together. Resulting speech is similar across the speakers because they revert to these common shared values in synchronous condition.

Our study sheds some light on the determining influences on pause duration not previously described in the timing literature. Using an innovative paradigm of Synchronous Speech we have been able to capture data that contains reliably less inter-subject variability than conventional techniques. This adds a level of robustness to our results which allows us to conclude that short pause durations (<300 ms) are determined by the length of preceding and following IPs. These two factors are independent and are found to operate superadditively.

Much work remains to be done in further examination of the nature and strength of independent contributions to pause duration. The study of these effects will contribute significantly to future quantitative modeling efforts.

5. REFERENCES

- [1] Janet Fletcher, "Some micro and macro effects of tempo change on timing in French," *Linguistics*, vol. 25, pp. 951–967, 1987.
- [2] François Grosjean and Harlan Lane, "Effects of two temporal variables on the listener's perception of reading rate," *Journal of Experimental Psychology*, vol. 102, no. 5, pp. 893–896, 1974.
- [3] Jürgen Trouvain and William J. Barry, "The prosody of excitement in horse race commentaries," in *Proceedings ISCA-Workshop on "Speech and Emotion", Belfast (Northern Ireland)*, 2000, pp. 86–91.
- [4] Uchanski Rosalie M., Sunkyoung S. Choi, Louis D. and Reed Charlotte M. Braid, and Nathaniel I. Durlach, "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," *Journal of Speech and Hearing Research*, vol. 39, pp. 494–509, 1996.
- [5] Marc Swerts, Anne Wichmann, and Robert-Jan Beun, "Filled pauses as markers of discourse structure," in *Proceedings ICSLP96, Fourth International Conference on Spoken Language Processing*, 1996, vol. 2, pp. 1033–1036.
- [6] Sofia Gustavson-Čapkova and Beáta Megyesi, "Silence and discourse context in read speech and dialogues in Swedish," in *Proceedings of Prosody 2002*, 2002.
- [7] Estelle Campione and Jean Véronis, "A large-scale multilingual study of silent pause duration," in *Proceedings of Prosody 2002*, 2002.
- [8] Terken J. and Collier R., "Syntactic influences in prosody," in *Speech Perception, Production and Linguistic Structure*, Tokhura Y., Vatikiotis-Bateson E., and Sagisaki Y., Eds., pp. 427–438. IOS Press, Amsterdam, Washington, Oxford, 1992.
- [9] Eva Strangert, "Relating prosody to syntax: boundary signalling in Swedish," in *Proceedings of the 5th European Conference on Speech Communication and Technology*, 1997, vol. 1, pp. 239–242.
- [10] Fred Cummins, "On synchronous speech," *Acoustic Research Letters Online*, vol. 3, no. 1, pp. 7–11, 2002.
- [11] Elena Zvonik and Fred Cummins, "Pause duration and variability in read texts," in *Proceedings ICSLP02, Seventh International Conference on Spoken Language Processing*, 2002, pp. 1109–1112.
- [12] Helen Fielding, *Bridget Jones's diary*, Pickador, 1997.