# A Physical Pause as a Sequence of Special Articulation Gestures

**Pavel Skrelin**

Department of Phonetics, St-Petersburg State University,
Universitetskaya emb. 11, Saint-Petersburg, Russia
e-mail: paul@phonet.lang.pu.ru

## Abstract

The paper deals with a new method of calculating duration of pauses. The method is based on the assumption that a physical pause is a kind of the articulation gesture (or their sequence) , which the speaker applies for segmenting speech flow which he produces. Mean duration of such gesture can be measured using the average syllable duration values of a particular speaker. The results of the experiment described show that the use of the average syllable duration as a unit for measuring pause duration allows to calculate duration of physical pauses between minor and major units with reliable accuracy.

## 1 Introduction.

Duration of pauses is traditionally described either in absolute values, milliseconds, showing their physical duration, or in some relative terms, like "long", "medium", "short" [1], depending on their possible duration values and functional load in the intonation structure of an utterance, as it happens in intonation modeling . In the latter case the definition merely implies that pauses of the first type are longer (or shorter) than the other. This is in agreement with data of many observations on pauses, but tells you nothing about their physical values in the text read by a particular speaker. The aim of the experiment described was to define the inner unit of temporal organization of the text which the speaker applies in producing pauses of required duration. Absolute values of such unit may vary from speaker to speaker and may change from one speech segment to another depending on, for example, speech tempo.

## 2 The choice of the measuring unit

Studying the data obtained by N.Volskaya [2] as a result of the analysis of the intonation phrasing of the phonetically balanced text read by 8 speakers, has led us to a very interesting observation: mean duration of pauses can be equal to mean syllable duration multiplied by 2 or 3. The number of syllables derived from the ideal phonemic transcription in the phonetically balanced text, including closed syllables at the end of the intonation units, is 984. Having this data, we obtain mean syllable duration for each speaker. Table 1 presents data on mean syllable duration for each speaker, mean duration of pauses in the read text, calculated pause duration which is equal to the average syllable duration multiplied by 2 or 3 , and the percentage of relative error, which is calculated as the relation of the absolute difference between real and calculated pause duration to the real (mean) duration of pauses. Analysis of the data presented in Table 1 shows that the relative difference of the calculated and real values does not exceed 5% ( for four speakers), and in the case of one speaker only it exceed 10% (11,6%). The reasons for this deviation of the real values from the calculated ones can be the following:

1. the data presented in the study performed by N.Volskaya reflect only the pauses marked by the listeners, their real number could be greater;
2. the number of actually realized syllables could be fewer compared to the number reflected in the ideal transcription.

As a result, the averages of physical duration of pauses and syllables can be slightly different from the data given in Table 1.

To minimize possible calculation errors and define the primary set of factors which influence pause duration, special experimental material was prepared.

## 3 Research material and methods

A digitized recording of the phonetically balanced read by a professional actor (aged 30) and two unprofessional speakers, lecturers of Moscow and St. Petersburg Universities was used as a material for



ISBN 1-876346-48-5 © 2003 UAB

| Speaker | Average syllable duration | Average pause duration | Calculated pause duration | Rel. diff. between calculated and real values |
|---|---|---|---|---|
| S. 1 | 146 | 264 | 2·146=292 | 11,6 % |
| S. 2 | 160 | 319 | 2·160=320 | 0,3 % |
| S. 3 | 112 | 351 | 3·112=336 | 4,5 % |
| S. 4 | 143 | 389 | 3·143=429 | 9,3 % |
| S. 5 | 134 | 371 | 3·134=402 | 7,7 % |
| S. 6 | 150 | 319 | 2·150=300 | 6,3 % |
| S. 7 | 152 | 295 | 2·152=304 | 3,0 % |
| S. 8 | 152 | 464 | 3·152=456 | 1,8 % |

**Table 1:** Average pause duration calculated on the basis of mean duration of syllable values for each speaker (ms); relative difference between real and calculated duration values

the study. The text was manually segmented into phonemes (sounds) and syllables. Special attention was devoted to the definition of boundaries of pauses: in selecting the boundary place we took into consideration a complete – and not perceived – end and beginning of the articulation of the sound. The segmentation was performed using the EDSW program , version 1.05. The segmentation file was then exported to the structured text file, which was later treated using the EXCEL for MS Office 97. The study of the segmentation results showed that 54 physical pauses were realized corresponding to the boundary punctuation marks (full stops and interrogation marks); 37 (for speaker 1), 38 (for speaker 2) and 47 (for speaker3) physical pauses, corresponding to the minor units boundaries were realized, in many cases they were marked in the text by commas, dashes and colons.

Table 2 presents data which were used for further calculations aimed at modeling the duration of pauses between minor and major units in the read text

At the next step of the study we tried to model physical duration of each pause and compare calculated duration with the realized one. In modeling pause duration we assumed that a pause is a special articulation unit, or gesture, rather than a delayed sound articulation or its cessation. If this is the case, we can represent its duration as the integer number of such articulation gestures. Calculation of physical pause duration is based on the average syllable duration, which can be adjusted to meet particular circumstances, such as tempo and the number of syllables in the intonation unit: the greater the number of syllables in the intonation unit, the longer the pause after it is supposed to be and vice versa. In the general case the formula for calculating pause duration looks like this:

| Speakers: | 1 (actor) | 2 (prof.) | 3 (prof.) |
|---|---|---|---|
| Number of realized sounds | 2237 | 2228 | 2208 |
| Average sound duration (ms) | 78,8 | 74,8 | 75,6 |
| Number of realized syllables | 975 | 961 | 965 |
| Average syllable duration (ms) | 181 | 173,4 | 173 |
| Average number of sounds in the syllable | 2,3 | 2,3 | 2,3 |
| Number of physical pauses | 91 | 92 | 101 |
| Average pause duration (ms) | 544 | 877,4 | 516 |
| Average duration of intonation unit (ms) | 1665 | 1735,7 | 1525 |
| Average number of syllable in the intonation unit | 9 | 10 | 8,8 |

**Table 2:** Principal characteristics of the research material

$$T_{pause} = N \cdot T_{syll} \cdot K_t \cdot K_{nsyll}$$

where N - the number of syllables in the pause;

$T_{syll}$ - average syllable duration in the text (ms);

$K_t$ - tempo coefficient in the intonation unit before the pause;

$K_{nsyll}$ - "number of syllables" factor in the intonation unit before the pause.

## 4 Tempo coefficient

A tempo coefficient is the relation of the mean sound duration in the intonation unit preceding the pause to the average sound duration in the text. For example the coefficient is <1 indicates speeding up of the tempo (the decrease of the syllable duration compared to its average duration in the text). This coefficient is necessary for all calculations. By applying this coefficient to the average syllable duration in the text allows us to get a normalized (for current conditions) average syllable duration which depends only on the tempo in this particular intonation unit and is not connected with the average number of sounds in the syllable (in this study it ranges between 1,7 and 3).

## 5 "Number of syllables" factor

This coefficient is the relation of the number of syllables in the intonation unit before the pause to the average number of syllables in the intonation unit in the text, or the average number of syllables calculated in the text to the average number of syllables in the intonation unit before the pause. Thus, this coefficient can be used to increase pause duration at the end of the sentences, marked by punctuation signs; and to reduce pause duration in the case of inner prosodic breaks. This is motivated by the fact that the longer the realized sound sequence is the more time is required for speaker to fill up his reserve of air, the shorter the sound sequence, the shorter the required pause, especially if he has to pronounce several intonation units before the end of the sentence.

## 6 Calculation results

Calculations of pause duration were performed for all types of pauses described above. Table 3 presents data (%) on average deviation of the calculated values from the real ones. In the **Appendix** an example of calculation of one boundary pause duration for Speaker 2 is shown.

| Speakers: | 1 | 2 | 3 |
|---|---|---|---|
| Final pauses | 3,5% | 3,4 % (2,1 %) | 4,6 % (2,8 %) |
| Non-final (inner) pauses | 4,4 % | 6,7 % (5,2 %) | 7,1 % (5,9 %) |

**Table 3:** Mean deviation of calculated pause duration from the real ones

The results of the pilot study (speaker 1 in Table 3) revealed, that

- the difference between the values of calculated and real pause duration is mostly due to an additional break in the intonation unit signaled by temporal and melodic changes;
- taking into account the inner segmentation of the final intonation unit and its function (some pauses corresponding to boundary punctuation marks were realized as inner pauses) allows us to increase the precision of calculation, particularly when we do not use the "number of syllables" factor (this is absolutely correct because of the inner segmentation of the intonation unit into minor ones). In half of such cases it is better not to use the tempo coefficient either, because tempo serves as an additional factor for the segmentation of final intonation units.
- accordingly, taking into account the inner segmentation of minor units allows us to increase the accuracy of calculation of pause duration [3].

Thus if we use more accurate methods of calculation proposed in the pilot study we can decrease the average error in calculated pauses (see data in brackets in Table 3). But even by using the simplest methods of calculation proposed in this paper we obtain satisfactorily low error in pause duration prediction.

## 7 Conclusion

The study is based on the assumption that a physical pause is a kind of the articulation gesture (or a succession of such gestures), which the speaker conscientiously employs for structuring the speech flow he produces. The use of the average syllable duration as a unit for measuring pause duration allows us to describe its values for different types of intonation units with acceptable accuracy.

New data obtained as a result of the analysis of spontaneous speech and reading also show a close relation between the average duration of syllables and pauses. Table 4 illustrates the results of the study based on the recordings made from 4 speakers of different age and gender. 586 pauses in reading style and 594 pauses in spontaneous speech were analyzed. (For a more detailed description of the material and method see [4]). Mean deviation of calculated pause duration from the observed one is 2,4 % for text reading and 2,9 % for spontaneous speech.

| Speaker | Average syllable duration (ms) | Average pause duration (ms) | Calculated pause duration (ms) | Rel. diff. between calculated and real values |
|---|---|---|---|---|
| f18R | 143,3 | 445 | $3 \cdot 143 = 430$ | 3,4 % |
| f18S | 161 | 502 | $3 \cdot 161 = 483$ | 3,8 % |
| m18R | 145 | 432 | $3 \cdot 145 = 435$ | 0,7 % |
| m18S | 155 | 489 | $3 \cdot 155 = 465$ | 4,9 % |
| m43R | 162 | 509 | $3 \cdot 162 = 486$ | 4,5 % |
| m43S | 180 | 543 | $3 \cdot 180 = 540$ | 0,6 % |
| m52R | 180 | 545 | $3 \cdot 180 = 540$ | 0,9 % |
| m52S | 187 | 731 | $4 \cdot 187 = 748$ | 2,3 % |

**Table 4:** Average pause duration calculated on the basis of mean duration of syllable values for 4 speakers; relative difference between real and calculated duration values (f - female, m - male; digits - age; R - text reading, S - spontaneous speech)

The average syllable duration can also be used in mod-


ISBN 1-876346-48-5 © 2003 UAB

eling pauses between major and minor units in the text produced by a particular speaker. The choice of average syllable duration of the speaker as a measuring unit for calculating pause duration proved to be fairly reliable. At the same time, as our observations show, it is quite possible to use for this purpose the average sound duration of a particular speaker.

In the speech synthesis systems developed nowadays pauses, which are used for structuring the text, are of fixed duration: 200 and 400 ms are pauses separating minor units, and 600 ms are pauses for major units [5]. The application of the proposed method for modeling pause duration in speech synthesis models can help to improve naturalness of the synthesized text, due to a variety of physical values of pauses it offers for separating intonation phrases and sentences.

In the present study we did not set the task to define the exact number of units (mean syllable duration specific for a particular speaker) each particular pause should correspond to, since this is the task of another, quite different study, which will be based on the analysis of formal characteristics of sentences and intonation phrases as well as the text as a whole. At the same time the application of such abstract units instead of absolute physical values makes the fulfillment of the task much more comfortable.

## Acknowledgments

This study is supported by the Intas grant 2000-00915 "Spontaneous Speech of Typologically Unrelated Languages (Russian, Finnish and Dutch): Comparison of Phonetic Properties".

The author would like to thank Ekaterina Miasnikova for her help in data processing.

## References

[1] Zharkov I.V., Slobodyanyuk S.L., Svetozarova N.D. "The Automatic Transcriber of the Russian Text," *Bulletin of Phonetic Fund of the Russian Language*, N 5, pp. 58-71, 1994 (in Russian).

[2] Volskaya N.B. "About real and virtual pauses," in *Problems and methods of experimental–phonetic studies*, SPb, pp. 165-170, 2002 (in Russian).

[3] Skrelin P.A. "Pause duration: in search of the measuring unit," *Problems and methods of experimental–phonetic studies*, SPb, pp. 224-229, 2002 (in Russian).

[4] De Silva V. "Universal and Language Dependent Phonetic Differences between Read and Spontaneous Speech in Rusian, Finnish and Dutch," in *Proceedings of the ICPhS'03*, forthcoming.

[5] Volskaya N.B., Svetozarova N.D., Skrelin P.A. "Modeling of prosody for the Russian text," *Bulletin of Phonetic Fund of the Russian Language*, N 6, pp. 85-100, 1997 (in Russian).

## Appendix

**Example of calculation of one boundary pause duration for Speaker 2**

(on the left – real duration of sounds and number of syllables in the intonation unit; on the right – the speaker–dependent data necessary for calculation and calculation result)

| **syllables** | **Snds** | **T** *ms* | Average sound duration | **74,8 ms** |
|---|---|---|---|---|
| 1 | m | 88 | Current | |
| | y | 67 | sound duration | **68,1 ms** |
| 2 | p | 83 | Tempo coef- | |
| | r' | 30 | ficient $K_t$ | **0,91** |
| | i | 49 | Average sylla- | |
| 3 | v | 36 | ble duration | **173,4 ms** |
| | l' | 50 | Average number | |
| | i | 55 | of syllables in the | |
| 4 | k | 86 | intonation unit | **10** |
| | l' | 50 | Number of sylla- | |
| | i | 73 | bles in the current | |
| 5 | v | 45 | intonation unit | **11** |
| | n' | 59 | | |
| | i | 37 | Syllable factor | **1,1** |
| 6 | m | 59 | | |
| | a | 127 | Current value of N | **3** |
| 7 | n' | 48 | | |
| | e | 61 | Calculated | |
| 8 | p | 86 | duration of pause: | |
| | a | 35 | | |
| 9 | s | 85 | 3 ·173,4·0,91·1,1 = | **519,5 ms** |
| | a | 107 | | |
| 10 | zh | 100 | | |
| | y | 132 | Difference between | |
| 11 | r | 36 | **real (521 ms)** and | **1,5 ms** |
| | a | 58 | **calculated (519,5** | |
| | f | 97 | **ms)** pause duration | **(0,29 %)** |
| **pause** | | **521** | | |