

Periodic Cycles of Hesitation Phenomena in Spontaneous Speech

Sandra Merlo, Plínio A. Barbosa

Speech Prosody Studies Group, Department of Linguistics, State University of Campinas, Brazil

sgmerlo@gmail.com, pabarbosa.unicampbr@gmail.com

Abstract

To verify whether hesitation phenomena are distributed periodically in spontaneous speech, twenty speech samples produced by five male adults were analyzed. Spectral analysis allowed for three main findings. First, hesitations present stationary behavior, which implies they did not accumulate in the beginning, in the middle, or in the end of speech samples. Second, periodic cycles of hesitation phenomena were detected in all speech samples (mean cycle duration around 13 seconds). This implies that regions with more hesitations tended to regularly alternate with regions with fewer hesitations. Third, periodic cycles accounted for about 30% of variance in data.

Index Terms: hesitation phenomena, time series, periodic cycles

1. Introduction

Hesitation phenomena are signals that reveal the speaker is having difficulties to process information (“what to say”) and/or to verbalize information (“how to say”) during spontaneous speech [1, 2]. However, speakers do not hesitate all the time. They seem to talk more fluently in some moments and less fluently in others. This raises the question whether hesitation phenomena may be periodically distributed in spontaneous speech. In other words, regions with more hesitations would regularly alternate with regions with fewer hesitations.

Several studies tried to determine if hesitation phenomena occur at regular intervals of time in spontaneous speech [3-11]. There are two main problems with previous research.

First, hesitation phenomena were usually not clearly identified. In most studies, hesitation phenomena were mixed or confounded with silence/pause events. However, not every silence may be considered a pause, not every pause may be considered a hesitation pause, and not every phonation may be considered fluent.

Second, statistical methods were usually not applied to detect periodic cycles in data. This is understandable because spectral analysis became really accessible only in the 1990s. N^2 numerical operations are necessary to calculate the spectrum of a time series and it is hard to manually do it. But high speed computers simplified the work with spectral analysis [12]. It is necessary to apply spectral analysis when the aim is to find out periodic cycles in data, because the “detection” of cycles without statistical analysis may simply denote overestimation of periodic structure by human subjects [6, 9].

The purpose of this research is to analyze hesitation phenomena in spontaneous speech using time series analysis. Specifically, an analysis in the frequency domain was done to find out whether hesitation phenomena are periodically distributed in speech samples.

2. Method

2.1. Subjects

Subjects were five male adults 20 to 34 years old. All subjects were native speakers of Brazilian Portuguese (henceforth BP). All of them were considered to be in good health according to an extensive assessment that included language and speech, hearing, cognitive, and psychiatric tests [13-19]. All subjects reported they were neither anxious nor affected by alcohol and were sufficiently aware during data collection. Written informed consent was obtained from all subjects.

2.2. Recording procedures

Speech recordings were done in a sound-proof room, using 22,050 Hz as sampling rate and 16 bits.

Subjects were first presented a picture of a bedroom; this picture was intended to elicit a descriptive text. Then subjects were asked to describe their own bedrooms. After that, subjects were presented a comic strip; this picture was intended to elicit a narrative text. Finally, subjects were asked to talk about a funny experience they have lived.

Therefore, each subject produced four speech samples (20 all in all). The samples were subsequently compared regarding text type (description vs. narrative) and picture support (present vs. absent).

2.3. Acoustical analysis procedures

Speech samples were segmented according to hesitation phenomena. The following signals are considered as hesitation phenomena in BP [2]:

- Non-fluent silent pauses: unfilled pauses at minor syntactic boundaries.
- Filled pauses: phones that do not belong to the syntactic structure of the utterance.
- Phone lengthenings: word-initial or word-final prolongations that do not signal stress, emphasis, or prosodic boundaries.
- Phrase, word, or part-word repetitions: successive occurrences of the same speech segment that do not signal emphasis, synthesis, or any other semantic strategy.
- Unretraced or retraced false starts: syntactic constructions started with some problem that may be stopped or corrected, respectively.

Intervals with hesitation phenomena received the numerical label “0” and intervals without hesitations the label “1”. Therefore, we analyzed speech as a binary process as regards hesitation phenomena (“yes” or “no”).

The numerical layer was sampled at 200 milliseconds to time series generation (Table 1). This sampling rate was chosen because 200 ms is an estimation of syllable duration mean in BP [20].

Table 1. An excerpt of the picture description sample produced by Subject 1

Utterances	Classification	Duration (seconds)	Numerical layer (time series)
“ (...) entre o tapete	Non-hesitation phenomena	0.928	1
			1
			1
			1
			0.64
e:	Hesitation phenomena (lengthening)	0.237	0.18
e a	Non-hesitation phenomena	0.257	1
e a	Hesitation phenomena (word repetitions)	0.340	0.11
			0.19
mesa do computador (...)”	Non-hesitation phenomena	1.036	1
			1
			1
			1

2.4. Statistical analysis

The first step in statistical analysis was to plot time series data against time. Time plots were used to observe oscillatory behavior and possible trends in the data [12, 21]. Figure 1 shows the time plot of the personal description produced by Subject 5. This example was chosen because it has the highest explained variance.

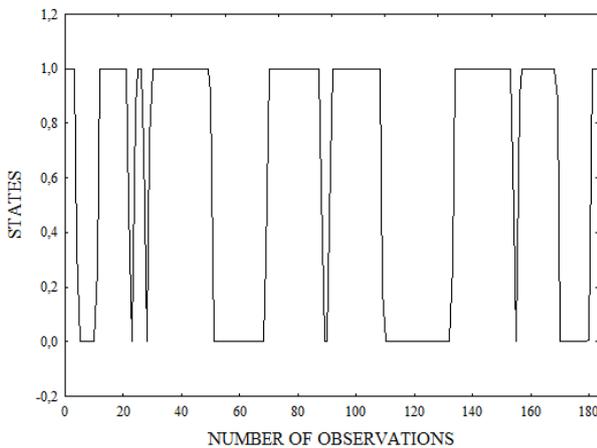


Figure 1: Time time plot of the personal description sample produced by Subject 5. The states are in y-axis (1 = non-hesitation phenomena and 0 = hesitation phenomena). The number of observations is in the x-axis (in this case, there are 183 observations)

Each time series has N observations, which depends on the duration of the speech sample and the sampling rate [12, 21].

The spectrum consists of N/2 frequencies. The lowest frequency is the fundamental frequency which corresponds to the total duration of each speech sample. The highest

frequency is the Nyquist frequency which is 2.5 Hz (corresponding to a 400 milliseconds period). Spectral analysis estimates the amplitude of each frequency in the spectrum. The graphical representation of the N/2 frequencies against amplitude is called “periodogram” [12, 21].

Fisher test was applied to test the statistical significance of peaks in the periodograms [21].

Sinusoidal models of the significant peaks were then adjusted to time plots [12, 21].

3. Results

All speech samples contained hesitation phenomena. The total duration of the samples ranged from 21 to 222 seconds.

3.1. Stationarity

All time series showed stationary behavior because a peak at 0 Hz in the periodogram was not found when spectral analysis was applied [12]. Therefore, hesitation phenomena did not accumulate in the beginning, in the middle, or in the end of speech samples.

3.2. Periodic cycles

Each periodogram had at least one significant peak according to Fisher test ($p < 0.01$). There were 59 significant peaks in the entire sample. Thus, on the average, three significant cycles of hesitations were found per speech sample. Figure 2 shows the spectral analysis (periodogram) of the personal description produced by Subject 5. The two highest peaks are statistically significant.

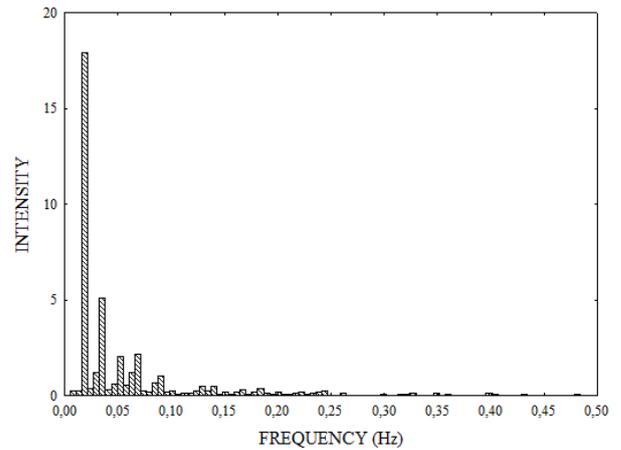


Figure 2: Spectral analysis of the personal description sample produced by Subject 5. The highest two peaks are statistically significant ($p < 0.05$) and correspond to cycles of 12.13 and 6.07 seconds, respectively

Figure 3 shows the sinusoidal model adjusted to the time plot of the personal description sample produced by Subject 5.

The median duration of cycles was 8.84 seconds, the mean was 12.82 seconds, and the standard-deviation was 12.13 seconds. The histogram of the results can be seen in Figure 4.

Periodic cycles explained an average of 29% (ranging from 11 to 58%) of the variance of hesitation phenomena data in frequency domain. The two cycles in the personal description sample produced by Subject 5 (Figures 2 and 3) explained 58% of the variance in that speech sample.

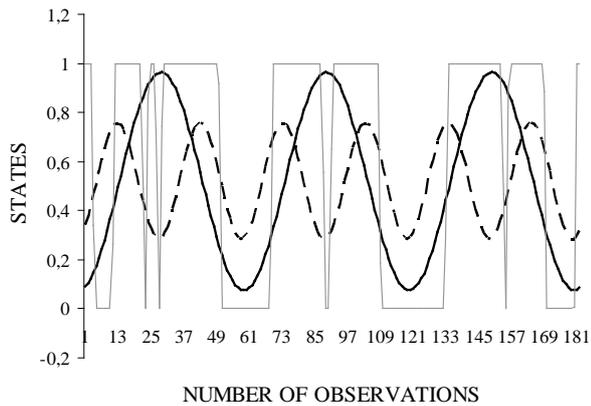


Figure 3: *Sinusoidal model adjusted to the time plot of the personal description sample produced by Subject 5. The wave with solid line corresponds to the cycle of 12.13 seconds and the wave with dashed line corresponds to the cycle of 6.07 seconds*

There was a significant and negative correlation of 60% ($p < 0.05$) between explained variance and duration of speech samples, that is, the longer the spoken text, the lesser the variance explained.

There were no significant differences for cycles' duration regarding text types ($p > 0.10$) or picture support ($p > 0.10$).

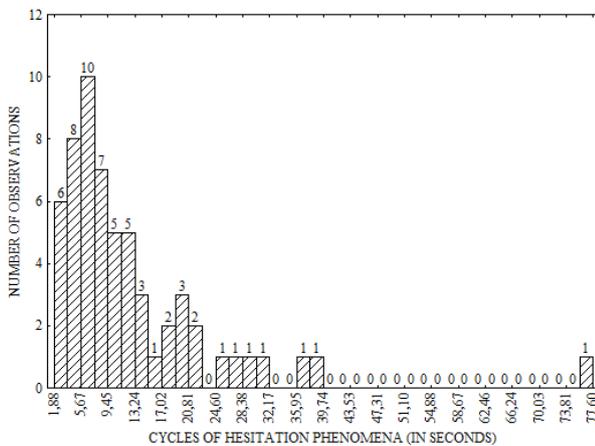


Figure 4: *Histogram of periodic cycles of hesitation phenomena found in data. 59 cycles were found in the data, a mean of 3 cycles per sample*

4. Discussion

4.1. Stationarity

It seems to be a common belief that hesitation phenomena accumulate in the beginning of speech samples [22]. This would be due to spoken language planning. The present study did not find such an effect. Hesitation phenomena were equally distributed throughout examined speech samples. Language planning during spontaneous speech is done little by little. The speaker did not previously plan the entire text [23]. This fact could explain why hesitations are spread over textual duration and do not accumulate in the beginning of texts. However, the speech samples analyzed in this study were relatively short in duration as compared to daily interactions.

It remains an open question if the same characteristic is to be observed in longer samples.

4.2. Periodic cycles

Cycles of hesitation phenomena were detected in each and every speech sample. However, just about 30% of the variance in the data was explained by periodic cycles. This means that 70% of the variance of the data in frequency domain is not periodic. This finding indicates that hesitations are a stochastic process, that is, hesitation phenomena oscillation pattern can only be partially forecasted [12].

Statistical theory states that each periodic cycle has its own underlying generator mechanism [12, 21]. If we assume that hesitation phenomena are caused by difficulties to process and/or to verbalize information [1, 2], it would be reasonable to assume that the periodic cycles of hesitation phenomena are also generated by the same processes. This would conduct us to the suggestion that if a speech sample has just one periodic cycle, all hesitation phenomena in that sample were caused by one only linguistic cognitive process. Speech samples analyzed had usually three periodic cycles, what suggests that speakers usually had difficulties with three linguistic cognitive processes at the same time during speech production. Moreover, the presence of more than one periodic cycle of hesitation phenomena within the same speech sample would suggest that spoken language is processed in parallel within working memory. Parallel processing is considered a main condition for the production of fluent speech [23].

The experimental manipulations that were done (text types and picture support) did not interfere with hesitation phenomena in terms of periodicity, at least.

Cycles of hesitation phenomena did not differ because of text types (description vs. narrative). Picture descriptions are mainly characterized by constant and simultaneous facts [24]. On the other hand, narratives are mainly characterized by transitory and successive facts that intend to restore the state of equilibrium [24]. Results suggest that the cognitive operations involved in these text types were different but they set the same level of difficulty for the speakers when this was measured as hesitation phenomena in frequency domain.

The same occurred with picture support. When talking about a picture, it is necessary to analyze and interpret visual information [25]. On the other hand, when talking about personal experiences, it is necessary to recall information previously stored in long-term memory. Results again suggested that the different cognitive operations involved in speaking with or without a picture support set the same level of difficulty for the speakers.

However, there was an inverse and moderate correlation between speech sample duration and variance of period cycles of hesitation phenomena. In other words, with shorter speech samples, there was a stronger tendency to regular oscillations in hesitation phenomena occurrence.

The long median and mean of the cycle's periods indicate that hesitation phenomena may be usually related to language processes that take long time to unfold. We reason that macroplanning activities may be such processes [23]. Macroplanning activities involve information selection and ordering, which can be considered as semantic processes. This is in accordance with previous studies, which suggested that hesitation phenomena are related to macroplanning activities [1, 2]. Thus, we intend to do additional experiments to test if perturbations in language planning and in working memory may interfere with cycles of hesitation phenomena.

5. Conclusions

Spectral analysis indicated the presence of periodic cycles of hesitation phenomena in all speech samples that were analyzed. The mean duration of the cycles was around 13 seconds. There was usually three cycles per speech sample. The periodic cycles explained about 30% of the variance in data.

Text type and picture support did not modify the cycles of hesitation phenomena. Thus, these variables did not show up an important underlying mechanism to the generation of periodic cycles.

How much these results may be generalized is still to be further tested.

6. Acknowledgements

We thank “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES” for a research grant. The second author also thanks a grant from CNPq (300371/2008-0).

7. References

- [1] Fox Tree, J. E. and Clark, H. H., “Pronouncing “the” as “thee” to signal problems in speaking”, *Cognition* 62:151-167, 1997.
- [2] Marcuschi, L. A., “Hesitação”, in C. Jubran and I. Koch [Eds], *Gramática do Português Culto Falado no Brasil: Construção do Texto Falado*, Editora da Unicamp, 48-70, 2006.
- [3] Henderson, A., Goldman-Eisler, F., and Skarbek, A., “Temporal patterns of cognitive activity and breath control in speech”, *Lang Speech*, 8:236-242, 1965.
- [4] Henderson, A., Goldman-Eisler, F., and Skarbek, A., “Sequential temporal patterns in spontaneous speech”, *Lang Speech*, 9:207-216, 1966.
- [5] Goldman-Eisler, F., “Sequential temporal patterns and cognitive processes in speech”, *Lang Speech*, 10:122-132, 1967.
- [6] Schwartz, J. and Jaffe, J., “Markovian prediction of sequential temporal patterns in spontaneous speech”, *Lang Speech*, 11:27-30, 1968.
- [7] Butterworth, B., “Hesitation and semantic planning in speech”, *J Psycholinguist Res*, 4:75-87, 1975.
- [8] Warner, R. M., “Periodic rhythms in conversational speech”, *Lang Speech*, 22:381-396, 1979.
- [9] Power, M. J., “Are there cognitive rhythms in speech?”, *Lang Speech*, 26:253-261, 1983.
- [10] Greene, J. O. and Cappella, J. N., “Cognition and talk: the relationship of semantic units to temporal patterns of fluency in spontaneous speech”, *Lang Speech*, 29:141-157, 1986.
- [11] Roberts, B. and Kirsner, K., “Temporal cycles in speech production”, *Lang Cog Process*, 15:129-157, 2000.
- [12] Chatfield, C., *The Analysis of Time Series*, Chapman & Hall, 2003.
- [13] Goodglass, H. and Kaplan, E., *The Assessment of Aphasia and Related Disorders*, 2nd ed., Lea & Febiger, 1983.
- [14] Radanovic, M. and Mansur, L. L., “Performance of a Brazilian population sample in the Boston Diagnostic Aphasia Examination: a pilot study”, *Braz J Med Biol Res*, 35:305-31, 2002.
- [15] Pereira, L. D., “Processamento auditivo”, *Temas Sobre Desenvolvimento*, 2:7-14, 1993.
- [16] Folstein, M. F., Folstein, S. E., and McHugh, P. R. “Mini-mental state: a practical method for grading the cognitive state of patients for the clinician”, *J Psychiatr Res*, 12:189-198, 1975.
- [17] Bertolucci, P. H. F., Brucki, S. M. D., Campacci, S. R., and Juliano, Y., “O Mini-Exame do Estado Mental em uma população geral: Impacto da escolaridade”, *Arq Neuropsiquiatr*, 52:1-7, 1994.
- [18] Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., and Dunbar, G. C., “The miniinternational neuropsychiatric interview (M.I.N.I.): the development and validation of a structured psychiatric interview for DSM-IV and ICD-10”, *J Clin Psychiatry*, 59(Suppl 20):22-33, 1998.
- [19] Amorim, P. MINI: Mini International Neuropsychiatric Interview. Brazilian version 5.0.0, Nucleo de Atenção Psicossocial Novo Mundo, 2001.
- [20] Barbosa, P. A. *Incursoes em Torno do Ritmo da Fala*. Pontes and FAPESP, 2006.
- [21] Morettin, P. A. and Tolo, C. M. C., *Análise de Séries Temporais*, Edgard Blucher, 2004.
- [22] Chafe, W., “Some reasons for hesitating”, in D. Tannen and M. Saviile-Troike [Eds], *Perspectives on Silence*. Ablex, Norwood, 77-89, 1985.
- [23] Levelt, W. J. M., *Speaking: From Intention to Articulation*, MIT Press and Bradford Book, 1989.
- [24] Adam, J.-M. and Revaz, F. A *Análise da Narrativa*, Gradiva, 1997.
- [25] Bitar, M. L., *Produção Oral de Crianças a partir da Leitura de Imagens*, Humanitas, 2002.