

Estudos sobre Pausas na Leitura de Textos por um Locutor do Português Brasileiro Visando à Síntese de Fala

Izabel C. Seara, Augusto H. Hentz, Rui Seara Jr., Fernando S. Pacheco,
Sandra G. Kafka, Rui Seara e Simone Klein

Resumo—Neste artigo, são examinados parâmetros referentes a pausas de silêncio, tais como sua relação com a pontuação (vírgula, ponto e vírgula, dois pontos), posição na sentença e duração, inseridas por um locutor nativo do português brasileiro na leitura de um *corpus* que serve de base para um sistema de síntese de fala (TTS). Através da análise de pausas desse *corpus* de aproximadamente 17 horas, verificou-se que um grande número de pausas não está associado à pontuação (61,36%), diferentemente do que as gramáticas normativas preconizam. Os dados reforçam também a presença de estruturas tópico/comentário. Os resultados aqui apresentados estão sendo considerados para a construção de um modelo de pausas a ser integrado a um sistema de síntese de fala.

Palavras-Chave—Marcas de pontuação, pausas na fala lida, português brasileiro, síntese de fala.

Abstract—In this paper we investigate silent pauses in read speech, considering parameters such as duration and position within sentences. We have also assessed the relationship between pauses and punctuation marks (comma, semicolon and colon). Thus, through the analysis of a corpus of approximately 17 hours of recording, carried out by a female professional speaker (native) of the Brazilian Portuguese language, we observe a large proportion of pauses without punctuation (61.36%). This fact has not been considered in grammar books. Besides, our data reinforce the presence of topic/comment structures in reading. The results here presented are being considered to generate a pause model to be integrated in a speech synthesis system.

Keywords—Punctuation marks, pauses in read speech, Brazilian Portuguese, speech synthesis.

I. INTRODUÇÃO

Nas gramáticas normativas do português do Brasil (PB), encontra-se a afirmação de que onde há pausa há pontuação e

onde não há pausa não há pontuação. Se isso fosse verdadeiro, um modelo de pausas para a síntese de fala estaria totalmente atrelado ao parâmetro pontuação e as pausas seriam facilmente modeladas. Estudos recentes sobre pausa [1]-[3] têm mostrado que existe uma idéia equivocada sobre a relação pausa/pontuação. Dizem também que a duração das pausas depende da língua natural. Dessa forma, justifica-se o presente artigo que analisa um *corpus* de fala do português brasileiro, resultante da leitura em voz alta, e investiga a relação entre pausa e pontuação. Esses dados referem-se a um único estilo de fala: a leitura cuidada de textos dissertativos. Alguns autores [4], [5] sugerem ainda que muitas situações que parecem inadequadas na leitura são reflexos de estruturas de uso corrente na língua falada, como as estruturas tópico/comentário. Neste estudo, será verificada a existência dessas estruturas.

Este artigo propõe-se assim a avaliar as situações em que um locutor do português brasileiro insere pausas quando realiza a leitura de textos. Focaliza-se o estudo nesse único locutor, uma vez que ele foi o leitor do *corpus* que serve de base para um sistema de síntese de fala (TTS). Serão observados os parâmetros: relação pausa/pontuação, posição da pausa na sentença e duração. Para abordar esses tópicos, na Seção II, apresentamos os dados avaliados e a metodologia de análise. Na Seção III, mostramos o que as gramáticas normativas têm a dizer sobre a relação pausa e pontuação, contrapondo as colocações dos gramáticos aos dados analisados no presente estudo. A Seção IV apresenta os resultados concernentes ao número e duração das pausas, associadas ou não à pontuação. Finalmente, na Seção V, temos as conclusões e encaminhamentos acerca deste tema.

II. CORPUS E METODOLOGIA

Nesta pesquisa, avaliamos um *corpus* composto de aproximadamente 17 horas de fala de uma locutora profissional, falante nativa do português brasileiro. A gravação de fala corresponde à leitura com velocidade de fala normal de textos dissertativos, tais como notícias de jornais e revistas, resenhas, textos didáticos, e foi realizada em estúdio com os cuidados necessários para se assegurar um sinal de boa qualidade. O *corpus* gravado foi previamente transcrito e

I. C. Seara, A. H. Hentz, R. Seara Jr., F. S. Pacheco, S. G. Kafka, R. Seara e S. Klein, LINSE – Laboratório de Circuitos e Processamento de Sinais, Departamento de Engenharia Elétrica, Universidade Federal de Santa Catarina, Florianópolis, SC, E-mails: {izabels, ahentz, ruijr, fernando, kafka, seara, klein}@linse.ufsc.br.

Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Financiadora de Estudos e Projetos (FINEP) e Dígito Tecnologia Ltda.

etiquetado segundo sua classificação morfossintática (a partir de um *parser* elaborado para um sistema de síntese de fala [6]). Além disso, foram também identificados o tipo de pontuação e a duração das pausas. Essa etiquetagem foi corrigida manualmente por um *expert* linguísta.

As pausas preenchidas por hesitações ou pelo alongamento de segmentos fonéticos não foram consideradas, sendo analisadas somente aquelas referentes a intervalos de silêncio. Para a etiquetagem das pausas acústicas no *corpus*, fixamos dois tipos de pausas: aquelas que correspondem a intervalos de silêncio superiores a 300 ms, definidos como correlatos acústicos das pausas longas, e as que compreendem intervalos de silêncio entre 90 e 299 ms, definidos como correlatos acústicos das pausas curtas. A partir de dados sobre a duração do intervalo de silêncio relativo ao período de oclusão das plosivas no português brasileiro, encontrados em [7], calcula-se uma média dessas durações que resultou em 45 ms (intervalos de silêncio com maior duração correspondem aos das bilabiais). Assim, a escolha do limite inferior de 90 ms¹, definido para as pausas curtas, pode assegurar que, mesmo se o intervalo de silêncio das plosivas estiver incluído nas pausas detectadas, ainda teremos um silêncio referente apenas à pausa.

Os dados foram coletados através do processamento automático de um detector de pausas que, além de calcular os intervalos de silêncio, recupera no *corpus* todos os contextos em que foram inseridas pausas e/ou pontuação, inclusive a classe morfossintática das palavras próximas às pausas.

III. RELAÇÃO DAS PAUSAS COM A PONTUAÇÃO

Professores da disciplina de Língua Portuguesa geralmente se queixam de um grande equívoco internalizado por seus alunos referente à pontuação. É a idéia de que, para se colocar vírgula em um texto, é preciso observar onde se fazem pausas. Nesses locais de pausa, são colocadas as vírgulas. A partir daí, percebe-se a dificuldade de ensinar aos alunos que a pontuação baseia-se na estrutura sintática da oração e não nas pausas feitas.

Em um estudo sobre discrepâncias entre pausas e pontuação [9], é salientado que as justificativas dadas por alunos para inserção de pontuação em seus textos mostram que a consciência metalingüística dos informantes limitava-se à regra geral de isomorfia entre pausas e pontuação colocada pela gramática normativa, ou seja, onde se faz pausa deve ser inserida pontuação.

Assim, se observarmos as gramáticas normativas, veremos que quase todas fazem menção a pausas quando falam de sinais de pontuação. Vejamos: em [10] é colocado que “Alguns sinais de pontuação servem, fundamentalmente, para marcar pausa (o ponto, a vírgula, o ponto e vírgula).” Mais adiante, esse autor explica que “vírgula é o sinal que indica uma pausa de curta duração, sem marcar o fim do enunciado” [10, p. 334]. E cita ainda que “O ponto e vírgula marca uma

pausa mais longa do que a vírgula, porém menor que a do ponto” [10, p. 339].

Em [11], sinais de pontuação são conceituados como o conjunto de sinais gráficos indicativos na escrita das pausas da linguagem oral. Esses autores, como muitos outros [12]-[14], salientam que não se emprega vírgula entre o sujeito e o verbo da oração ou entre o verbo e seus complementos. Nesses casos, podemos dizer que não teríamos pausas na linguagem oral, já que, segundo as gramáticas, não se colocariam os sinais de pontuação indicativos dessas pausas. No entanto, no *corpus* em análise, observamos pausas nessas seqüências, conforme pode ser observado nos exemplos a seguir:

Exemplo 1:

Os pesquisadores afirmam [pausa curta] que os resultados são a primeira evidência de que os transgênicos [pausa longa] podem gerar conseqüências catastróficas [pausa curta] às suas próprias espécies.

Exemplo 2:

... assim como um arco-íris [pausa curta] separa a luz solar em várias cores.

Uma distinção entre sinais pausais e sinais melódicos é apresentada em [15]. Para esses autores, os sinais de pontuação compreendem um grupo que corresponde aos sinais que marcam pausas, como a vírgula, o ponto e o ponto e vírgula e o grupo de sinais cuja função é marcar a entoação, como os dois pontos e os demais sinais de pontuação. Dizem esses autores, no entanto, que essa distinção “didaticamente cômoda” não é rigorosa, uma vez que os sinais de pontuação podem indicar ao mesmo tempo pausa e melodia [15, p.625]. E continuam relacionando vírgula à pausa de curta duração, ponto à de longa duração e ponto e vírgula como algo intermediário entre a vírgula e o ponto.

As pausas também são mencionadas em [16, p.181], que conceitua a pontuação como um “sistema de sinais com que se representam os fonemas supra-segmentais, principalmente as pausas.”

Gramáticos, como [13], [14], [17], salientam todavia que se deve esclarecer que nem toda pausa da língua oral corresponde a uma vírgula na língua escrita.

Em [14], é afirmado ainda que podemos seguramente dizer que: “ONDE NÃO HÁ PAUSA NÃO HÁ VÍRGULA” (grifo do autor) [14, p.571]. Porém, como nossos dados mostram situações em que não havia pausas, mas havia vírgulas (3%), seria melhor dizer que, na maior parte das vezes, onde não há pausa não há vírgula. Em um estudo de pausas na leitura em língua francesa [5], que utiliza dados referentes a um locutor, foram encontrados 4,6% de pontuações que não correspondiam a pausas. Em outra pesquisa sobre a leitura [3], em que foram avaliadas cinco horas de fala lida por 50 locutores de cinco línguas diferentes (francês, inglês, espanhol, italiano e alemão), os resultados são bastante divergentes dos aqui apresentados. Segundo esse estudo, para

¹ Autores, como [8], estabeleceram, como limite inferior para a detecção de pausas, um intervalo de silêncio de 100 ms.

o francês, 33% das ocorrências de pontuação não correspondiam a pausas².

Em nossos dados, referentes à pausa não associada à pontuação, é mostrada a separação por pausa do sujeito e seu verbo e do verbo e seu complemento. Vejamos alguns exemplos extraídos do *corpus* de nossa locutora:

Exemplo 3:

... a preocupação dos médicos e nutricionistas [pausa longa] é descobrir o princípio ativo desse legume ...

Exemplo 4:

... um sensor de radiação infravermelho [pausa longa] usa o calor do alvo para criar uma imagem em um monitor.

Exemplo 5:

... o número do telefone é [pausa longa] meia dois cinco zero...

Exemplo 6:

... meu médico jamais chegaria a descobrir [pausa longa] que a nicotina [pausa curta] me contaminava o sangue.

Novamente, apresentamos divergências nestes resultados se comparados aos mostrados em [3] e [5]. Os primeiros apresentam uma média de 15,4% de ausência de pontuação relacionada às pausas nas línguas estudadas por eles [inglês, francês, alemão, italiano e espanhol (ver Tabela I)], sendo o italiano a que apresenta maior frequência de pausas sem pontuação (33%). Para o francês, eles apontam 11,9%, enquanto em [5], para essa mesma língua, é indicado que houve ausência de pontuação em 36% das pausas avaliadas. Nossos resultados mostram que, do total de 9985 pausas internas a sentenças, 6127 compreendiam intervalos de silêncio não relacionados a sinais de pontuação, correspondendo a 61,36% dos dados analisados. Por outro lado, mesmo assumindo um limite inferior maior do que o estabelecido anteriormente, por exemplo, de 200 ms (desconsiderando dessa forma grande parte de nossas pausas curtas), ainda obteremos 41,13% de pausas que não correspondem a sinais de pontuação. Nesse caso, o número total de pausas (associadas ou não à pontuação) será de 5831. Assim, pode-se inferir que os dados apresentados pelo locutor em estudo parecem indicar que o português brasileiro é a língua, dentre alemão, inglês, italiano, espanhol e francês, que tem a mais alta ocorrência de pausas sem pontuação, conforme pode ser visto na Tabela I.

Do total de pausas analisadas, 3858 delas estavam associadas a sinais de pontuação e destas, 3633

² Os resultados apresentados em [3] são prejudicados pela falta de dados referentes aos limites inferiores das classes de pausas fracas (correspondentes às nossas pausas curtas) analisadas, já que, para a verificação da interferência ou não do intervalo de silêncio de consoantes plosivas, seria necessário uma referência que estabelecesse qual a média ou o intervalo dos valores concernentes ao silêncio da oclusão dessas consoantes. Com isso, seria possível dizer que um silêncio em contexto de plosiva referente a uma pontuação era de fato um intervalo de silêncio somente da plosiva e não havia uma pausa associada a ele.

acompanhavam a vírgula; 52, o ponto e vírgula; 173, os dois pontos.

TABELA I
FREQUÊNCIA DE OCORRÊNCIA DE PAUSA SEM PONTUAÇÃO
EM SEIS DIFERENTES LÍNGUAS³

Língua	Pausas com ausência de pontuação
Inglês	6,4%
Francês	11,9%
Alemão	11,0%
Italiano	33,7%
Espanhol	14,0%
Português brasileiro	61,36% (41,13%)

Para se ter uma noção da dinâmica de inserção de pausas, principalmente nos casos em que não se tem associação a sinais de pontuação, faremos a análise das seqüências sintáticas anteriormente citadas (sujeito/predicado e verbo/complemento), a fim de estabelecer um modelo de pausas com uma distribuição mais coerente com aquela apresentada pela locutora estudada. Essa discussão é feita a seguir.

IV. POSIÇÃO DAS PAUSAS NAS SENTENÇAS LIDAS

Para que pudéssemos avaliar as pausas realizadas dentro de estruturas que sintaticamente as censurariam, já que não há inserção de pontuação, observamos nos dados, analisados morfossintaticamente, as estruturas sujeito/predicado (Grupo 1) e verbo/complemento (Grupo 2).

Conforme as Tabelas II e III, podemos verificar que há uma tendência maior de inserção de pausas entre sujeito e predicado do que entre verbo e complemento.

TABELA II
NÚMERO DE CASOS E FREQUÊNCIA DE OCORRÊNCIA DAS SEQÜÊNCIAS
RELACIONADAS A SUJEITO/VERBO E SUA OCORRÊNCIA COM E SEM PAUSAS

Grupo 1	Sem pausas		Pausas longas (maior que 300 ms)		Pausas curtas (de 90 a 299 ms)	
	Número	%	Número	%	Número	%
Sujeito e verbo	1658/3196	51,88	459/3196	14,40	1079/3196	33,80
			459/1538	29,84	1079/1538	70,16
			1538 (48,12%)			

As pausas não relacionadas à pontuação são associadas a fronteiras de sintagmas⁴, quer sejam oracionais, preposicionais ou adverbiais (29,23% corresponderam a conjunções; 17,48% a preposições; 4,13% a advérbios). Nesses casos, 81,22% dos intervalos de silêncio corresponderam aos correlatos acústicos de pausas curtas.

Analisando-se os dados que apresentam pausa entre sujeito e predicado, vê-se que parecem remeter à estrutura tópico/comentário. No entanto, é ressaltado em [1] que, caso se quisesse remarcar tópico e comentário, a vírgula deveria

³ Os dados do inglês, francês, alemão, italiano e espanhol são obtidos de [3].

⁴ Grupo de elementos lingüísticos (no nível sintático) composto por um núcleo (por exemplo, um verbo, um nome, um adjetivo, etc.) e por outros termos que se unem ao núcleo, formando uma locução que entrará na formação da oração.

aparecer na escrita. Sendo a estrutura tópico/comentário uma característica da oralidade e como, nos textos lidos, a vírgula não estava presente, esperava-se que na leitura em voz alta não ocorresse a pausa. O autor diz ainda que, ao se transcrever enunciados com a estrutura tópico/comentário, essa estrutura é transformada em sujeito e predicado. Da mesma forma, o leitor, quando lê um texto com a estrutura de sujeito e predicado, lê como se fosse tópico/comentário, através da inserção de pausa.

TABELA III
NÚMERO DE CASOS E FREQUÊNCIA DE OCORRÊNCIA DAS
SEQÜÊNCIAS RELACIONADAS A VERBO/COMPLEMENTOS
E SUA OCORRÊNCIA COM E SEM PAUSAS

Grupo 2	Sem pausas		Pausas longas		Pausas curtas	
	Número	%	Número	%	Número	%
Verbo e complemento	1489/1515	98,30	11/1515	0,70	15/1515	1
			26 (1,70%)			

Na produção de estruturas do tipo tópico/comentário, é afirmado também que o falante divide o enunciado em dois grupos tonais, podendo haver uma pausa entre um grupo tonal e outro, mas ela não é necessária [1]. Em [18], é ratificado tal fato colocando que, nesse caso, há a presença de pausa e/ou de curva entonacional ascendente na fronteira entre sujeito e verbo. Nossos dados mostram que a locutora produz os enunciados reforçando a estrutura tópico/comentário, pois, na fronteira entre sujeito e verbo, além da presença de pausa, são muitas vezes observados dois grupos tonais aparentes com curva entonacional ascendente (Fig. 1).

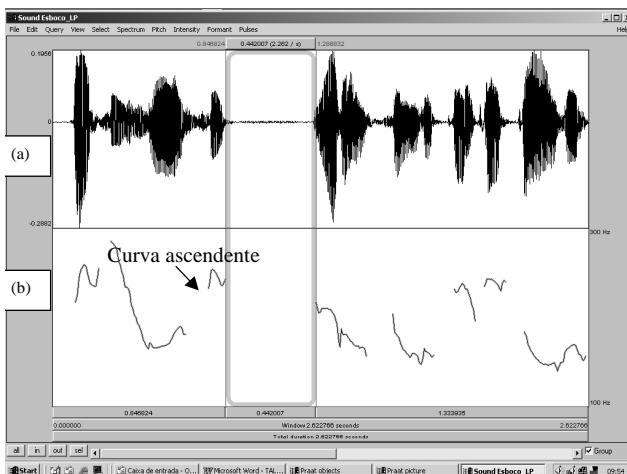


Fig. 1. Exemplo de pausa entre sujeito e predicado: “Esse esboço [pausa: 442 ms] irá sendo completado...”. (a) Forma de onda. (b) *Pitch*.

Um estudo bastante completo sobre o tópico no português brasileiro [2] constata que, no Brasil, a língua coloquial espontânea apresenta uma enorme quantidade de construções topicalizadas com uma grande variedade de tipos (língua com proeminência de tópico). Porém, como há uma censura gramatical muito grande às sentenças com tópico, elas não aparecem na língua escrita na abundância em que aparecem na língua falada.

Os resultados apurados mostram que, apesar da censura gramatical sobre a estrutura de tópico, ela aparece na leitura com bastante evidência, pois nossa locutora transformou 48,12% das estruturas sujeito/predicado da escrita em estruturas tópico/comentário, próprias da fala espontânea. Por outro lado, a forte tendência a pausas curtas (70,2%) nesses dados (ver Tabela II) pode mostrar que, de alguma maneira, na leitura, a censura gramatical à estrutura de tópico freia a ocorrência de pausas longas.

Assim, a partir das constatações anteriormente apontadas, pode-se ratificar a presença da estrutura tópico/comentário na leitura de textos, corroborando com fatos já verificados por outros autores [1], [2], [18].

Vários estudos [8], [19], [20], que investigam a correlação entre pausas de silêncio e limites discursivos, também têm destacado a importância das pausas para, por exemplo, topicalização de estruturas ou introdução de um novo tópico e remarcam que elas organizam a mensagem. No entanto, também há um consenso de que a pausa muitas vezes aparece em combinação com outros fatos prosódicos, como, por exemplo, a alteração de *pitch*, conforme verificado na Fig. 1.

Os casos relacionados à estrutura tópico/comentário, aqueles em que houve a inserção de pausas entre o sujeito e o predicado (estruturas da escrita), não apresentaram violação de continuidade, isto é, produziram constituintes inteiros sem interrupções, conforme definido em [21], pois essas pausas são geralmente colocadas antes de fronteiras sintáticas. Essa autora salienta que, na fala lida, pouquíssimos são os dados em que se percebem esses limites em posições não motivadas sintaticamente. Encontramos, em nossos dados, raríssimos casos de violação de continuidade (menos do que 1%), conforme o Exemplo 7. Em [9], são constatados casos desse tipo e, para essa autora, são classificados como pausas internas aos constituintes sintáticos.

Exemplo 7:

... entre outros [pausa curta] agentes de doenças.

Neste exemplo, o pronome adjetivo *outros*, que antecede o substantivo *agentes*, formando com este um constituinte sintático do qual *agentes* é o seu núcleo, é interrompido por uma pausa curta.

V. DURAÇÃO DAS PAUSAS

De forma geral, os resultados mostraram que pausas não relacionadas no texto a caracteres de pontuação apresentaram-se em sua maioria como curtas (ver Tabela IV). Na Fig. 2, pode-se observar que as pausas não associadas à pontuação concentram suas durações entre 90 e 270 ms.

Outro ponto discrepante de nossos dados, comparados aos apresentados em [3], relaciona-se à duração de pausas inferiores a 200 ms e correspondentes aos intervalos de silêncio das pausas curtas (de 90 a 299 ms). Para esses autores, apenas 8,2% do total de pausas analisadas têm duração inferior a 200 ms. Já, em nosso *corpus*, temos 78,42% de pausas curtas.

TABELA IV
NÚMERO DE CASOS E FREQUÊNCIA DE OCORRÊNCIA DAS PAUSAS NÃO ASSOCIADAS À PONTUAÇÃO

Tipo de pausas	Pausas Curtas		Pausas Longas		Total	
	Número	Freq.	Número	Freq.		
Sem pontuação	5367	87,61%	760	12,39%	6127	61,36%
Com pontuação	2465	63,89%	1393	31,11%	3858	38,64%
Total	7832		2153		9985	

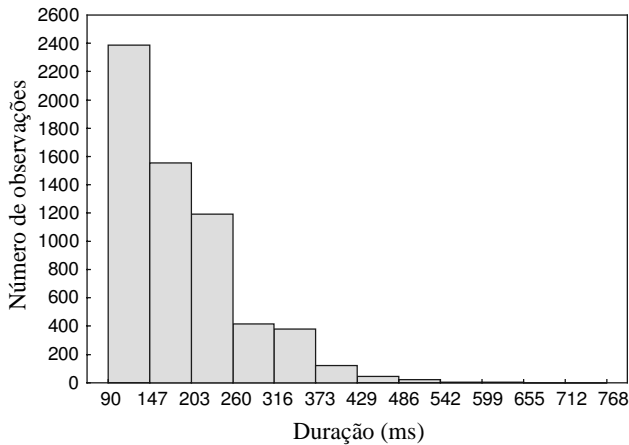


Fig. 2. Histograma de duração de pausas não relacionadas à pontuação.

As pausas relacionadas a caracteres de pontuação, cujos resultados são mostrados na Tabela V, apresentam uma média de duração de 282 ms, reforçando a tendência a pausas mais curtas do que as apresentadas em estudos de outras línguas naturais. Dados que apresentaram caracteres de pontuação aos quais não correspondiam uma pausa foram muito poucos (menos de 3%).

TABELA V
NÚMERO DE CASOS E FREQUÊNCIA DE OCORRÊNCIA DAS PAUSAS ASSOCIADAS À PONTUAÇÃO

Tipo de pontuação	Pausas Curtas		Pausas Longas		Total
	Número	Frequência ocorrência	Número	Frequência ocorrência	
Vírgula	2439	67,14%	1194	32,86%	3633
Ponto e vírgula	8	15,39%	44	84,61%	52
Dois pontos	18	9,25%	155	90,75%	173
Total	2465/3858		1393/3858		3858

As Fig. 3, 4 e 5 apresentam histogramas de duração das pausas correspondentes à vírgula, ponto e vírgula e dois pontos, respectivamente. Esses gráficos mostram que a vírgula está relacionada a uma pausa mais curta do que a dos outros sinais de pontuação aqui estudados; os dois pontos relacionam-se a uma pausa mais longa; o ponto e vírgula a uma pausa de duração intermediária. Esses resultados (ver Tabela VI) são consistentes com as colocações das gramáticas

normativas sobre a duração das pausas referentes aos diferentes sinais de pontuação. Contudo, os dados não formam *clusters* distintos, havendo muitas sobreposições entre essas três classes de pontuação. A média geral da duração das pausas é de 315 ms, se consideradas as não relacionadas à pontuação.

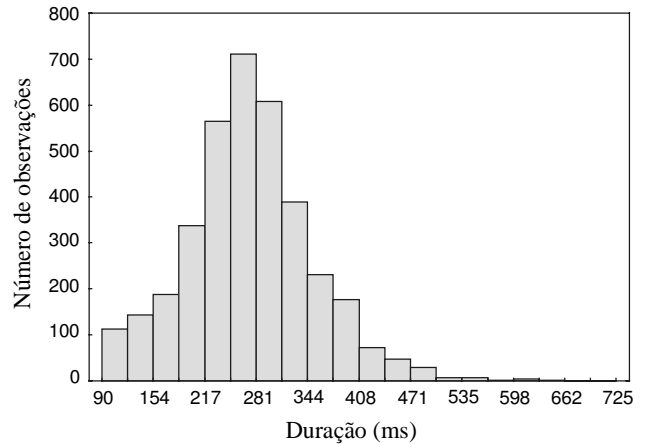


Fig. 3. Histograma de duração de pausas associadas à vírgula.

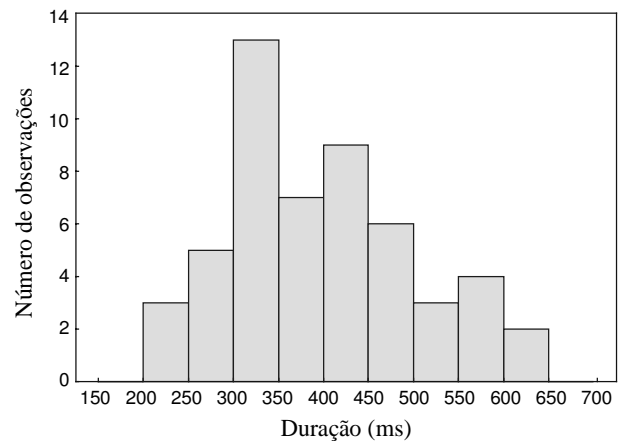


Fig. 4. Histograma de duração de pausas associadas ao ponto e vírgula.

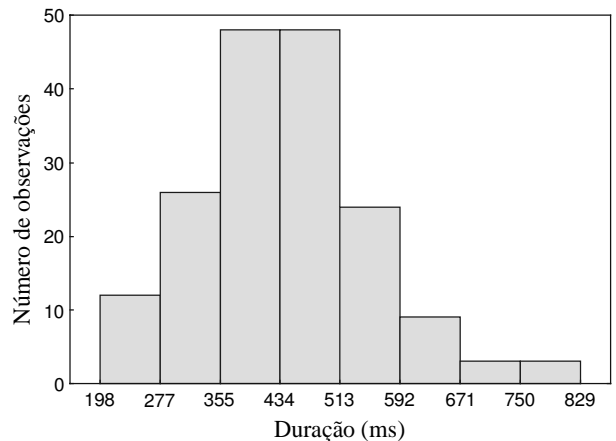


Fig. 5. Histograma de duração de pausas associadas aos dois pontos.

TABELA VI
DURAÇÃO MÉDIA DAS PAUSAS APRESENTADAS NOS
DADOS DA LOCUTORA EM ESTUDO

Tipos de Pausa	Ocorrência		Duração (ms)	
	Número	%	Média	Desv.Pad.
[Pausa longa]	759	7,6	356	58
[Pausa curta]	5368	53,76	163	54
Vírgula [Pausa longa]	1194	11,96	357	54
Vírgula [Pausa curta]	2439	24,43	231	51
Ponto e vírgula [Pausa longa]	44	0,44	426	91
Ponto e vírgula [Pausa curta]	08	0,08	259	23
Dois pontos [Pausa longa]	157	1,57	478	103
Dois pontos [Pausa curta]	16	0,16	252	32
Total	9985	100	315	100

VI. CONCLUSÕES

Neste artigo, discutimos a relação pausa/pontuação na fala lida de um locutor do português brasileiro. Mostramos que as colocações das gramáticas de que **onde há pausa há pontuação** e **onde não há pausa não há pontuação** não são precisas, já que observamos 61,36% de pausas não relacionadas à pontuação e 3% de pontuações não associadas à pausa. A recorrente inserção de pausa entre sujeito e predicado reforça a presença de estruturas tópico/comentário na fala lida. Os resultados referentes à duração das pontuações internas às sentenças são coerentes com aqueles apresentados nas gramáticas normativas. Dessa forma, enfatiza-se a necessidade deste tipo de estudo, visto que se observa uma relação pausa/pontuação bastante divergente daquela apregoada pelas gramáticas normativas e pela intuição dos falantes, já que a maior parte das pausas não apresenta correspondência com os sinais gráficos de pontuação. A partir dos parâmetros observados, um modelo de pausas será gerado para ser incorporado a um sistema de síntese de fala que tem por base a fala da locutora aqui estudada. Para se poder generalizar os dados aqui apresentados como próprios do português brasileiro, novas pesquisas utilizando mais falantes são necessárias.

REFERÊNCIAS

- [1] E. Campione et J. Véronis, "Étude des relations entre pauses et pontuations pour la synthèse de la parole à partir de texte," in *Proc. Traitement Automatique du Signal*, Nancy, France, Juin 2002, pp. 175-184.
- [2] J. Cao and W. Zhu, "Syntactic and lexical constraint in prosodic segmentation and grouping," in *Proc. Speech Prosody*, Aix-en-Provence, France, Apr. 2002, pp. 203-206.
- [3] G. Vannier et al., "Pauses location and duration calculated with syntactic dependencies and textual considerations for T.T.S. system," in *14th Int. Congress Phonetics Sciences*, San Francisco, USA, Aug. 1999.
- [4] L. C. Cagliari, "Da importância da prosódia na descrição de fatos gramaticais". In R. Ilari, (org.) *Gramática do Português Falado: Níveis de Análise Lingüística*, vol. 2, Campinas, SP: Unicamp, 1992, pp. 39-64.
- [5] E. Pontes, *O Tópico no Português Brasileiro*. São Paulo: Pontes, 1987.
- [6] C. Seara, S. G. Kafka, S. Klein, R. Seara, "Alternância vocálica das formas verbais e nominais do português brasileiro para aplicação em conversão texto-fala," *Revista da Sociedade Brasileira de Telecomunicações*, vol. 17, no. 1, pp. 79-85, Jun. 2002.

- [7] C. Seara, "Análise acústico-perceptual da nasalidade das vogais do português brasileiro," *Tese de Doutorado*, Pós-graduação em Lingüística, Universidade Federal de Santa Catarina, Florianópolis, 2000.
- [8] S. Gustafson-Vapkova e B. Megyesi, "Silence and discourse context in read speech and dialogues in Swedish," in *Proc. Speech Prosody*, Aix-en-Provence, France, Apr. 2002, pp. 363-366.
- [9] B. B. Rodrigues, "Discrepâncias entre a pontuação e as pausas," *Dissertação de Mestrado*, Pós-graduação em Lingüística, Universidade Federal de Santa Catarina, Florianópolis, 1993.
- [10] E. Terra, *Curso Prático de Gramática*, São Paulo: Scipione, 1996.
- [11] R. M. Mesquita e C. R. Martos, *Gramática Pedagógica*, São Paulo: Saraiva, 1991.
- [12] C. E. Faraco e F. M. Moura, *Gramática: Fonética e Fonologia, Morfologia, Sintaxe, Estilística*. São Paulo: Ática, 1991.
- [13] F. P. Savioli, *Gramática em 44 Lições: Com mais de 1700 Exercícios*. São Paulo: Ática, 1992.
- [14] N. M. de Almeida, *Gramática Metódica da Língua Portuguesa*, São Paulo: Saraiva, 1988.
- [15] C. Cunha e L. Cintra, *Nova Gramática do Português Contemporâneo*, Rio de Janeiro: Nova Fronteira, 1985.
- [16] C. P. Luft, *Moderna Gramática Brasileira*. Rio de Janeiro: Globo, 1987.
- [17] H. A. de André, *Gramática Ilustrada*. São Paulo: Moderna, 1974.
- [18] C. Mollica, "Intervalos entre o silêncio e a fala e suas representações na escrita," *Cadernos de Letras*, no. 9, Rio de Janeiro, pp. 143-149, 1993.
- [19] L-C. Yang, "Duration and pauses as cues to discourse boundaries in speech," in *Proc. Speech Prosody*, Nara, Japan, Mar. 2004, pp. 267-270.
- [20] M. E. van Donzel and F. J. Koopmans-van Beinum, "Pausing strategies in discourse in dutch," in *Proc. Fourth Int. Conf. Spoken Language Process.*, Philadelphia, USA, Oct. 1996, pp. 1029-1032.
- [21] E. Strangert, "Speech chunks in conversation: Syntactic and prosodic aspects," in *Proc. Speech Prosody*, Nara, Japan, Mar. 2004, pp. 305-308.