

O Efeito do Telefone Celular no Sinal da Fala: uma análise fonético-acústica com implicações para a verificação de locutor em português brasileiro

Renata Regina Passetti¹, Plínio Almeida Barbosa¹

¹ Universidade Estadual de Campinas, Brasil

re.passetti@gmail.com, pabarbosa.unicampbr@gmail.com

Resumo

O presente estudo tem como principal objetivo avaliar os efeitos que a transmissão telefônica exerce sobre as vogais orais do português brasileiro. As gravações nas condições diretas e via celular foram realizadas de forma simultânea, permitindo que apenas a interferência da transmissão telefônica no sinal da fala fosse avaliada. Os resultados revelam que a transmissão telefônica agiu de forma a aumentar os valores de frequências do primeiro formante em, aproximadamente, 14% e a diminuir as frequências do terceiro formante em, aproximadamente, 12%. Em relação às frequências do segundo formante, os resultados da análise de dispersão mostraram que a transmissão telefônica agiu de forma a aumentar artificialmente as frequências de vogais com baixos valores de F2 e a diminuir as frequências de vogais com altos valores de F2. O efeito telefônico nas frequências formânticas provocou alterações nas disposições das vogais no espaço acústico dos locutores que podem ter consequências do ponto de vista perceptual.

Palavras-chave: fonética acústica, fonética forense, verificação de locutor, canal telefônico, distorção de sinal.

1. Introdução

A tarefa de verificação de locutor pode ser afetada por aspectos relacionados à língua, como aqueles envolvendo dialetos, sotaques e disfarces de fala, e também por aspectos exteriores a ela, como é o caso dos efeitos causados ao sinal da fala por aparelhos telefônicos e outros aparelhos eletrônicos. Com a crescente evolução tecnológica, teve-se também um aumento no número de investigações de crimes com base em amostras de fala provenientes de telefones fixos e celulares. Em crimes envolvendo o uso destes aparelhos de comunicação, é preciso identificar, além das características linguísticas dos sujeitos investigados, também o grau de modificação causado pela transmissão telefônica ao sinal da fala.

A distorção envolvida na transmissão telefônica pode ser manifestada como resultado do filtro passa-faixas que passa a energia em uma determinada frequência de banda que varia entre 300 Hz a 3500 Hz, aproximadamente. Conforme explica [1], toda energia acima e abaixo dessas frequências de corte serão atenuadas e as informações próximas a esses limiares podem ser severamente comprometidas. Outro efeito da transmissão telefônica está relacionado com a “distorção espectral” (*spectral distortion*), que é capaz de aumentar as frequências que se encontram ligeiramente acima do filtro passa-baixa e diminuir as frequências que se encontram ligeiramente abaixo do filtro passa-alta.

Estudos [2; 3] sobre os efeitos de transmissões telefônicas no sinal da fala mostraram que as principais alterações provocadas pela transmissão telefônica estão relacionadas aos valores de frequências do primeiro e terceiro formantes, com maiores alterações nas frequências formânticas de amostras de fala obtidas de linhas telefônicas móveis [3]. Com base nos estudos desenvolvidos acerca dos efeitos de transmissões telefônicas no sinal da fala, esta pesquisa tem como principal objetivo determinar o grau de modificação fonético-acústica causada pelo filtro de banda do canal telefônico às vogais orais do português brasileiro, a fim de compreender quais parâmetros acústicos da fala são afetados por esse tipo de transmissão.

2. Metodologia

2.1. O corpus “Sem Fio”

O *corpus* de fala espontânea elaborado para esta pesquisa era composto por gravações de 10 locutores do sexo masculino, pertencentes à faixa etária de 20 a 25 anos de idade. Todos eram estudantes de graduação da Universidade Estadual de Campinas, exibiam sotaque do sudeste brasileiro (cidades do interior do estado de São Paulo) e não apresentavam problemas auditivos ou fonoarticulatórios. A opção por indivíduos do sexo masculino teve o intuito de se aproximar de situações forenses reais, pois, de acordo com dados do Ministério da Justiça do Brasil [4], crimes cometidos por homens são mais frequentes no cenário forense real. Aproximadamente 93,6% da população carcerária do Brasil em 2012 era composta por indivíduos do sexo masculino.

2.2. Procedimentos Experimentais

O procedimento para obtenção de amostras de fala espontânea baseou-se na descrição espontânea de rotas (*spontaneous route descriptions*) [5], na qual os locutores deveriam descrever o trajeto da Praça Arcádia, no Instituto de Estudos da Linguagem da UNICAMP, até o supermercado Pão de Açúcar, no centro de Barão Geraldo, Campinas, com base em imagens do trajeto e instruções sobre como chegar ao destino final, apresentadas a eles através de *slides*.

De forma a garantir robustez ao cálculo dos parâmetros fonético-acústicos e à análise dos resultados, foram realizadas duas gravações simultâneas: (1) na condição face a face, diretamente pelo posicionamento de um microfone em frente aos sujeitos e (2) via telefone celular, conforme apresentado na Figura 1. Esse procedimento permitiu que apenas o efeito causado pela transmissão telefônica ao sinal da fala fosse analisado, uma vez que os pares de amostras de fala de cada sujeito possuíam exatamente o mesmo conteúdo.

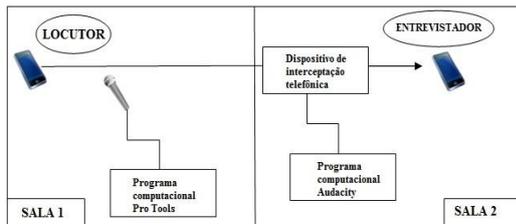


Figura 1: Representação dos ambientes de gravação simultâneos.

Posteriormente à segmentação das vogais orais no programa PRAAT [6], deu-se início à extração automática das frequências dos três primeiros formantes por meio do script ForensicDataTracking [7]. Cada gravação continha, aproximadamente, 700 vogais orais e a duração de, aproximadamente, 5 minutos. A extração automática dos parâmetros foi seguida de verificação e correção manual, quando necessárias.

2.3. Análises estatísticas

A análise do efeito da transmissão telefônica em amostras de fala do português brasileiro foi realizada em duas etapas. Primeiramente, analisou-se o efeito da transmissão telefônica sobre as vogais orais do português brasileiro, seguida da análise do efeito da transmissão telefônica no espaço acústico dos locutores.

As análises estatísticas tinham como objetivo comprovar a significância da diferença entre as médias dos parâmetros em ambos os contextos de gravação. A análise do efeito da transmissão telefônica sobre as vogais e das modificações intralocutor foi testada por meio de testes t de variáveis dependentes, assumindo um nível de significância de 5%, que tinham como finalidade testar a hipótese nula de que não há diferenças significativas entre valores de parâmetros acústicos entre a gravação telefônica e a direta. Para elaboração do espaço acústico dos locutores, foi utilizado o pacote *vowels* do software R [8], que permite a manipulação, normalização e plotagem de dados acústicos referentes aos formantes das vogais. Para normalização dos valores de frequências dos formantes do grupo de locutores utilizou-se o procedimento de Lobanov. Conforme explicam [9], o método utilizado para essa normalização calcula os *z-scores* das frequências formânticas de cada vogal dos falantes utilizando a média e o desvio-padrão das frequências de cada formante.

3. Resultados

3.1. Efeitos da transmissão telefônica sobre as vogais orais do Português Brasileiro

A análise do efeito telefônico sobre as vogais considerou os valores médios das frequências dos três primeiros formantes nas sete vogais orais do português brasileiro, em posição tônica ([a], [e], [ɛ], [i], [o], [ɔ], [u]) e átona ([ɐ], [ɪ], [ʊ]), para todos os locutores.

3.1.1. Frequências de F1

Os resultados obtidos pela análise das frequências do primeiro formante indicam um aumento nos valores desse formante na condição telefônica para todas as vogais analisadas. De acordo com os dados apresentados na Tabela 1, os valores de F1 na gravação via celular são, em média, 14% superiores aos

valores obtidos para o mesmo formante na gravação direta. As vogais fechadas são mais afetadas pela transmissão telefônica, com destaque para a vogal alta anterior [i], mais afetada pela transmissão telefônica, com aumento médio nos valores de F1 de 28%, resultando em uma diferença média de 81 Hz entre as duas condições de análise.

De acordo com a análise estatística, as vogais médias posteriores não apresentaram diferenças significativas para esse parâmetro entre as duas condições de análise. A vogal média-baixa posterior, [ɔ], permaneceu robusta ao efeito telefônico, com aumento médio de apenas 4%.

Vogais	F1 (Hz)		Diferença (%)	Desvio-padrão (Hz)	Teste t
	direta	celular			
[a]	641	693	8,1	18,5	t (17.876) = -3,814, p < 0,05
[e]	524	564	7,6	21,4	t (17.975) = -4,999, p < 0,05
[ɛ]	418	492	17,7	13,9	t (17.988) = -4,146, p < 0,05
[i]	328	422	28,7	20,2	t (17.638) = -3,161, p < 0,05
[e]	519	569	9,6	23	t (17.951) = -8,287, p < 0,05
[i]	341	411	20,5	23,3	t (17.579) = -9,432, p < 0,05
[o]	409	482	17,8	23,8	t (10.885) = -2,127, ns
[ɔ]	590	614	4,1	29,7	t (15.994) = -1,42, ns
[u]	362	438	21,0	23,2	t (16.450) = -4,503, p < 0,05
[ʊ]	382	452	18,3	20,5	t (17.629) = -5,393, p < 0,05
Todas as vogais	446	510	14,1	28,3	

Tabela 1: Frequências médias de F1 nas condições de gravações direta e celular para todas as vogais.

A correlação entre os dados de F1 obtidos na gravação direta e as diferenças médias entre as duas condições de gravação está representada no gráfico de dispersão da Figura 2. A análise do gráfico revela a concentração da maioria dos dados proximamente da linha zero do eixo das ordenadas, indicando que houve pouca disparidade nos valores desse formante entre as condições direta e celular. O gráfico revela ainda que, aproximadamente, 17% da variância dos dados estão associados aos valores de F1 na condição direta (sem interferência da transmissão telefônica).

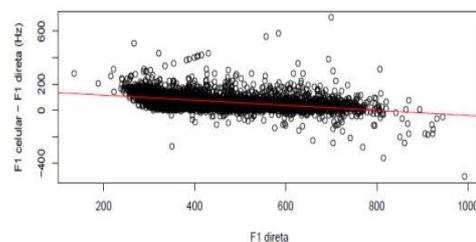


Figura 2: Gráfico de dispersão dos valores de F1. (R^2 ajustado = 0,1677, valor de $p < 0,05$).

3.1.2. Frequências de F2

Os efeitos da transmissão telefônica sobre o segundo formante estão apresentados na Tabela 2. De modo geral, percebe-se que a transmissão telefônica teve menos influência nos valores de F2, visto que a diferença percentual média entre as duas condições foi menor comparada à diferença média estabelecida para F1. Os resultados indicam que, para os valores médios do segundo formante, o efeito causado pela transmissão telefônica consistiu na diminuição das frequências

desse formante em comparação com os valores estabelecidos para a condição direta.

O maior desvio foi atestado para a vogal média-baixa anterior, [ɛ], cujos valores médios de F2 diminuíram, aproximadamente, 13% na condição telefônica. Mais uma vez, a análise estatística revelou que as vogais posteriores permaneceram mais robustas aos efeitos da transmissão telefônica, sem apresentar diferenças estatisticamente significativas entre as duas condições de gravação.

Vogais	F2 (Hz)		Diferença (%)	Desvio-padrão (Hz)	Teste t
	direta	celular			
[a]	1495	1366	-8,6	29,4	t (17.458) = 2.259, p < 0,05
[e]	1462	1396	-4,5	34,9	t (17.239) = 2.961, p < 0,05
[ɛ]	1654	1604	-3,0	78,8	t (17.746) = 2.469, p < 0,05
[ɛ]	1712	1494	-12,8	93,3	t (18) = 3.062, p < 0,05
[i]	1950	1855	-4,8	151,8	t (16.702) = 3.388, p < 0,05
[i]	1922	1827	-4,9	109,8	t (17.997) = 3.500, p < 0,05
[o]	1042	944	-9,4	310,2	t (11.605) = 1.448, ns
[ɔ]	1100	1072	-2,5	172,5	t (12.937) = 1.070, ns
[u]	1027	991	-3,5	164,8	t (12.879) = 1.160, ns
[v]	1299	1164	-10,4	132,4	t (16.854) = 1.055, ns
Todas as vogais	1434	1299	-9,5	148,7	ns

Tabela 2: Frequências médias de F2 nas condições de gravações direta e celular para todas as vogais.

O gráfico de dispersão representado na Figura 3 apresenta uma melhor compreensão das modificações ocorridas nos dados de F2. A dispersão de dados em F2 referente aos dados de análise dessa pesquisa segue, em parte, a mesma explicação encontrada para os dados de [3]. Aparentemente, o filtro passa-faixas do telefone celular possui um efeito de deslocamento ascendente para vogais com baixos valores de F2, resultando em um aumento de valores, e o efeito contrário, ou seja, um deslocamento descendente para vogais com altas frequências de F2. De acordo com o gráfico, 24,5% da variância dos dados estão associados aos valores de F2 na condição direta. O gráfico de dispersão de F2 revela, ainda, a complexidade do efeito telefônico sobre as frequências desse formante. Observa-se a elevada concentração de dados na parte inferior direita do gráfico, marcada pela cruz pontilhada em azul, em torno dos 2 kHz (faixa próxima da distribuição formântica de F2 em vogais altas anteriores, como [i] e [i]). Essa distribuição revela um comportamento particular cujo espalhamento dos dados apresenta-se de forma concorrente à distribuição de base que contém a maioria dos dados.

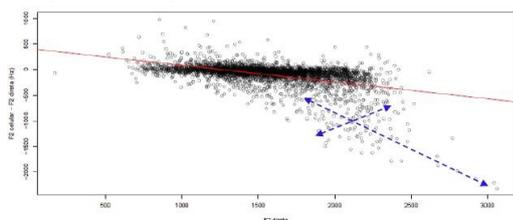


Figura 3: Gráfico de dispersão dos valores de F2. (R^2 ajustado = 0,2457, valor de $p < 0,05$).

3.1.3. Frequências de F3

Os resultados obtidos para a análise do terceiro formante seguem o mesmo padrão estabelecido para as frequências de F2. De acordo com os dados apresentados na Tabela 3, todas as vogais estiveram sujeitas a uma diminuição em seus valores na condição telefônica. De modo geral, a transmissão telefônica alterou as frequências de F3 em 11,5%. A vogal média-baixa anterior, [ɛ], apresentou o maior desvio em F3 e, conseqüentemente, foi a mais afetada pelas características da transmissão telefônica, com diminuição em suas frequências de, aproximadamente, 20%. Mais uma vez não foram atestadas diferenças significativas para todas as vogais posteriores estudadas, sendo a vogal alta posterior a menos afetada dentro o grupo pelo filtro telefônico de passa-faixa.

Vogais	F3 (Hz)		Diferença (%)	Desvio-padrão (Hz)	Teste t
	direta	celular			
[a]	2492	2347	-5,8	189	t (17.967) = 4.670, p < 0,05
[e]	2582	2189	-15,2	150,7	t (15.618) = 6.846, p < 0,05
[ɛ]	2610	2231	-14,5	227,5	t (16.531) = 5.237, p < 0,05
[ɛ]	2613	2108	-19,4	141,2	t (16.704) = 9.463, p < 0,05
[i]	2864	2327	-18,7	247,9	t (17.573) = 3.780, p < 0,05
[i]	2587	2364	-8,6	281,9	t (15.551) = 3.892, p < 0,05
[o]	2259	1965	-13,0	174	t (11.023) = 1.434, ns
[ɔ]	2655	2417	-9,0	164,5	t (16.24) = 1.630, ns
[u]	2643	2441	-7,7	155,3	t (12.726) = 0.761, ns
[v]	2710	2454	-9,5	208,4	t (14.653) = 2.169, p = 0,05
Todas as vogais	2641	2339	-11,5	215,4	ns

Tabela 3: Frequências médias de F3 nas condições de gravação direta e celular para todas as vogais.

O gráfico de dispersão apresentado na Figura 4 mostra que a correlação entre as frequências de F3 nas duas condições de análise foi maior comparada à estabelecida para os dois primeiros formantes. Aproximadamente, 33% da variância dos dados em F3 está relacionada aos valores desse formante na condição direta. Pela análise dos dados observa-se, ainda, que vogais cujos valores de F3 estão próximos a 3 kHz sofreram maiores variações entre as duas condições de análise, como atestado na Tabela 3 para as vogais alta e média-alta anteriores.

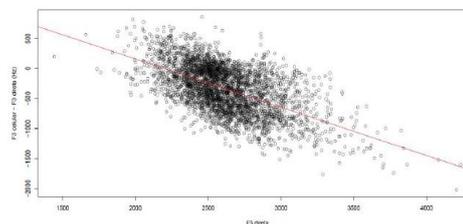


Figura 4: Gráfico de dispersão dos valores de F3. (R^2 ajustado = 0,3285, valor de $p < 0,05$).

3.2. Efeitos da transmissão telefônica no espaço acústico dos locutores

A análise das modificações causadas no espaço acústico dos locutores lançou luz sobre o efeito da transmissão telefônica nas vogais analisadas em função de características individuais,

que foram normalizadas para efeito de comparação. A Figura 5 apresenta o espaço acústico normalizado dos sujeitos nas condições de gravação direta e via celular. De modo a facilitar a visualização do efeito da transmissão telefônica nas vogais orais do português brasileiro, calcularam-se os valores médios das frequências do primeiro e segundo formantes de cada vogal analisada.

Pela análise da figura, observa-se um abaixamento global do espaço acústico na condição “celular”, motivado pelo aumento nos valores de frequências do primeiro formante, acompanhado de um efeito de posteriorização na disposição das vogais, em função da diminuição dos valores de F2. Observa-se, ainda, que as vogais altas foram mais afetadas pela transmissão telefônica, possivelmente em razão da proximidade aos limiares de frequência do filtro de passa-baixa do telefone celular, em torno dos 300 Hz.

A análise do espaço acústico também revelou que a vogal aberta [a] permaneceu robusta ao efeito da transmissão telefônica para maioria dos sujeitos. A variação em menor grau para essa vogal pode estar associada aos seus altos valores de F1 e valores intermediários de F2, portanto, distantes dos limiares de frequências dos filtros passa-faixa, o que a torna uma boa candidata para análises forenses.

A reconfiguração do espaço acústico dos locutores como efeito da transmissão telefônica também possui consequências perceptuais. Com as modificações nas frequências do primeiro e segundo formante na condição telefônica, muitas vogais moveram-se espacialmente para posições ocupadas por vogais com qualidades vocais distintas. Pela observação do espaço acústico normalizado, nota-se que vogais originalmente produzidas como altas passaram a ocupar posições de vogais médias. Em relação às alterações do segundo formante, na condição telefônica a diminuição das frequências de F2 para vogais anteriores provocou o deslocamento destas para posições centrais do espaço acústico, o que auditivamente pode resultar em percebê-las como mais abertas.

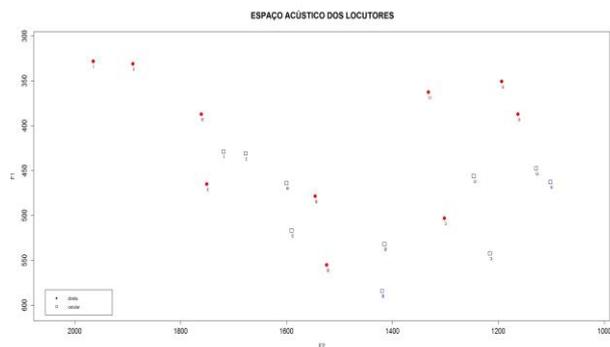


Figura 5: Espaço acústico normalizado $F1 \times F2$, em Hertz, das vogais orais dos locutores nas condições “direta” e “via celular”.

4. Conclusões

As análises dos efeitos da transmissão telefônica sobre o sinal da fala conduzidas neste estudo apresentam importantes contribuições às pesquisas em Fonética Forense, área da ciência da fala que se encontra em desenvolvimento no Brasil. Os experimentos realizados permitiram avaliar o grau de modificação de parâmetros fonético-acústicos em função da transmissão telefônica nas vogais orais do português brasileiro e, com isso, classificar vogais mais robustas a essas variações e, portanto, boas candidatas para as análises forenses reais de

acordo com os parâmetros analisados. Apesar das frequências dos três primeiros formantes terem se mostrado parâmetros acústicos suscetíveis a alterações em suas medições devidos aos efeitos telefônicos, não devem ser banidos das análises forenses com amostras de fala provenientes de aparelhos celulares, mas sim usados com cautela e sob os cuidados de profissionais aptos com formação em Fonética Acústica. Os resultados obtidos por meio da análise do espaço acústico normalizado podem ter consequências perceptuais, uma vez que vogais com baixos valores de F1 (vogais altas, por exemplo), cujas frequências foram aumentadas artificialmente pelo filtro telefônico, passaram a se situar proximalmente de vogais com valores mais altos de F1 (vogais médias). O mesmo foi atestado para as frequências de F2, cujas alterações nos valores originais deste formante para o conjunto de vogais fizeram com que se deslocassem para posições posteriores do espaço acústico. Essas alterações podem resultar em perceber essas vogais como mais abertas e/ou mais posteriores em ligações telefônicas.

5. Agradecimentos

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo financiamento desta pesquisa de Mestrado (processo nº 2013/12516-1). As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade dos autores e não refletem necessariamente as visões da FAPESP.

6. Referências

- [1] ROSE, P., “Effect of telephone transmission”. In: SELBY H. & FRECKELTON I. (eds.), *Expert Evidence*. Sydney: Thomson Lawbook Company, 2003.
- [2] KÜNZEL, H. J., “Beware of the ‘telephone effect’: the influence of the telephone transmission on the measurement of formant frequencies”. In: *Forensic Linguistics 8(1)*, pp. 80-99, 2001.
- [3] BYRNE, C.; FOULKES, P. “The ‘Mobile Phone Effect’ on Vowel Formants”. *The International Journal of Speech, Language and the Law – Equinox Publishing*, v. 17, n.1, pp. 83-102, 2004.
- [4] MINISTÉRIO DA JUSTIÇA. Departamento Penitenciário Nacional. “Formulário Categoria e indicadores preenchidos: todas UF’s. Categoria: Quantidade de Presos/Internados”. Disponível em: <<http://portal.mj.gov.br/data/Pages/MJC4D50EDBPTBRNN.htm>>. Acesso em: 8 set. 2014.
- [5] WILLIAMS, S.; WATSON, C. I. “A profile of the discourse and intonational structures of route descriptions”. In: *Proceedings of Eurospeech 1999*. Budapest, Hungary, 5-9 September, 1999.
- [6] BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer (Versão 5.3.34)*. Disponível em: <<http://www.praat.org>>. Acesso em 2013.
- [7] BARBOSA, P. A. *ForensicDataTracking*. 2013.
- [8] R DEVELOPMENT CORE TEAM, version 3.0.1. “R: A language and environment for statistical computing”. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0. Disponível em: <<http://www.Rproject.org>>. Acessado em 2013.
- [9] BARBOSA, P.; MADUREIRA, S. “A normalização de dados de frequência de formantes pelo método Lobanov”. In: *Manual de Fonética Acústica Experimental: aplicações a dados do português*. São Paulo: Cortez, pp. 302-303, 2015.