Síntese de formantes para a emissão de vogais do português brasileiro

MODALIDADE: COMUNICAÇÃO

SUBÁREA ou SIMPÓSIO: **SA-1**. Composição e Sonologia

Anselmo Guerra
anselmo@ufg.br

Resumo. Trata-se de um estudo sobre o método de síntese de formantes, com o objetivo de criar a simulação computacional da emissão vocal, focado nas características dos sons das vogais do português falado no Brasil. Nosso referencial teórico parte dos conceitos fundamentais de acústica referentes à produção de formantes da voz: harmônicos, ressonância, formantes, passando para a implementação computacional em ambiente de programação PureData. Apresentamos como resultado uma ferramenta que pode ser utilizada em vários contextos: estudos de fonologia, composição musical, educação musical, musicologia.

Palavras-chave. Método de síntese de formantes. Fonologia. Características dos sons das vogais. Sonologia.

Title: **Synthesis of formants for voice simulation**

Abstract. This is a study on the method of synthesis of formants, intending to create a computer simulation of vocal emission, focused on the characteristics of vowel sounds in Portuguese spoken in Brazil. Our theoretical framework starts from the fundamental concepts of acoustics referring to the production of voice formants: harmonics, resonance, formants, passing to the computational implementation in a PureData programming environment. As a result, we present a tool that can be used in various contexts: phonology studies, musical composition, music education, musicology.

Keywords. Formant synthesis method. Phonology. Characteristics of vowel sounds. Sonology.

1. Introdução

A síntese de formantes pode ser definida como uma particularidade do método de síntese subtrativa (ROADS 1996). Em linhas gerais, os sons produzidos por síntese subtrativa são derivados de geradores de sons complexos, que passam por processamentos espectrais, sobretudo no uso de filtros e geradores de envoltória. Nosso objetivo é aplicar o método na simulação da emissão vocal, em particular os sons das vogais do português falado no Brasil.

Partimos de fontes bibliográficas da área de acústica musical, especificamente das características dos sons periódicos harmônicos, do fenômeno da ressonância. A partir desse ponto entramos em aspectos da fisiologia da voz, a produção do som pelas pregas vocais e a relação entre o controle de formantes e a movimentação do trato vocal. A partir do entendimento dessa dinâmica propomos uma implementação computacional capaz de simular a produção vocal na emissão de vogais cantadas, a partir de dados obtidos em estudos de fonologia. O ambiente de programação utilizado foi o PureData, com possibilidade de

alteração de parâmetros dos formantes, ou mesmo a inclusão de dados, permitindo a experimentação de novos sons.

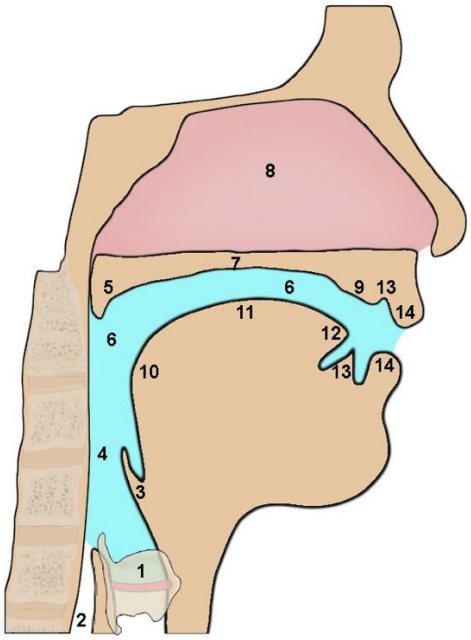
2. Fisiologia da voz e os parâmetros acústicos envolvidos

A produção sonora vocal tem sua origem nas pregas vocais (figura 1). O ar que se movimenta nos pulmões encontra na laringe um obstáculo que orienta de forma controlada a passagem desse fluxo, de modo a produzir vibrações. Quando essas vibrações são periódicas obtemos alturas de som definidas. Durante uma expiração, o ar sai dos pulmões e passa pela laringe. Se as pregas vocais forem tensionadas, elas irão vibrar conforme uma onda complexa. Podemos entender essa onda complexa como uma fundamental acompanhada de harmônicos. A fundamental normalmente é a vibração que dá origem ao nome das notas musicais. Os harmônicos são vibrações associadas a essa fundamental numa razão de números inteiros em relação a esta. Assim, controlando a tensão da pregas vocais, obtemos a gama de sons graves aos agudos, numa tessitura que conseguimos através da variação de tensão que induzimos a elas.

f = frequência fundamental em Hz
2f = segundo harmônico
3f = 3º. harmônico
4f = 4º. harmônico
5f = 5º. harmônico
...
nf = enésimo harmônico

Figura 1: imagem das pregas vocais e frequências produzidas. Fonte: www.voicescienceworks.org

Nesse estágio inicial, ainda na laringe, o som é, portanto, rico em harmônicos. Cada harmônico possui uma amplitude e normalmente a fundamental é o harmônico mais forte, e por isso, o que é usado para dar o nome da nota. Assim, se variarmos as amplitudes de uns harmônicos em relação aos outros, obtemos diferentes qualidades tímbricas que caracterizam cada vogal, por exemplo, mesmo mantendo a mesma fundamental. Os formantes são o resultado dessa filtragem, que são produzidos no trato vocal (figura 2).



1. *Laringe (pregas vocais)*
2. *Esôfago*
3. *Epiglote*
4. *Faringe*
5. *Véu palatino*
6. *Cavidade oral*
7. *Palato*
8. *Cavidade nasal*
9. *Crista alveolar*
10. *Raiz da língua*
11. *Corpo da língua*
12. *Ponta da língua*
13. *Dentes*
14. *Lábios*

Figura 2: fisiologia do trato vocal. Fonte: www.voicescienceworks.org, editado pelo autor

Os espaços criados no trato vocal passam então a atuar como ressonadores, comportando-se como filtros, aumentando ou atenuando faixas de frequências, portanto, alterando os harmônicos produzidos na laringe. A envoltória do espectro resultante é o que chamamos de formantes.

Formantes, portanto, são faixas de frequências que estão destacadas na série harmônica, de acordo com a posição de partes específicas do trato vocal. Para efeito didático, ilustraremos os dois principais formantes, resultantes da ressonância produzida pelo espaço antes da língua (primeiro formante) e a ressonância produzida acima da língua até a ponta dos lábios (segundo formante).

A ressonância é apresentada em (ROSSING *et al.*, 2002, p.26) como um dos modos sistemas vibrantes simples, através do exemplo clássico dos Ressonadores de Helmholtz (1821-1894) que em sua época usava recipientes em tamanhos e formatos dimensionados para vibrar em presença de frequências específicas, servindo, então, como analisador de espectro. Mais adiante, os autores (ROSSING *et al.*, 2002, p. 67) apresentam mais exemplos de aplicação de ressonadores, sobretudo na aplicação do fenômeno nas caixas de ressonância dos instrumentos acústicos, como vemos por exemplo nos instrumentos de cordas, no violão, no piano, que funcionam como amplificadores naturais do som produzido pelas cordas.

A produção de som pelo trato vocal tem a qualidade única de poder alterar os formantes no som, movimentando seus espaços internos, o que não ocorre com os instrumentos musicais. Na figura 3, ilustramos a posição dos formantes.

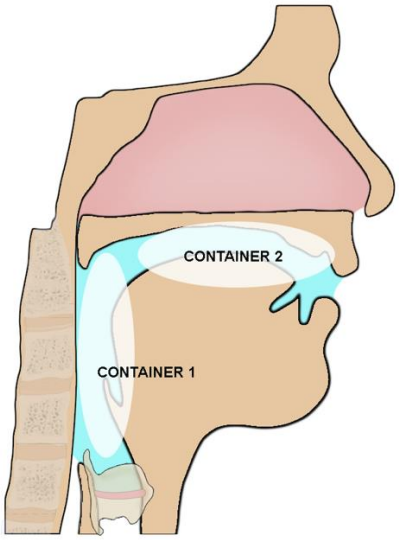


Figura 3: fisiologia do trato vocal. Fonte: www.voicescienceworks.org, editado pelo autor

Conforme alteramos nossas cavidades internas, movimentamos os formantes:

- Se a mandíbula descer, o container 1 (atrás da língua) fica menor e o container 2 (na frente da língua) fica maior. O tom do F1 fica mais agudo e tom do F2 fica mais grave.
- Se os lábios se aproximam, os dois containers de ar ficam maiores. Os tons de ambos os containers de ar (F1 e F2) diminuem.
- Se a língua avançar, o container 1 aumenta e o container 2 diminui. A altura do F1 fica mais grave e altura do F2 fica mais aguda.
- Se a língua retrair, o container 1 ficará menor e o container 2 ficará maior. A altura do F1 fica mais aguda e a altura do F2 fica mais grave.
- Se a faringe se estreitar, o container 1 ficará menor. A altura do F1 fica mais aguda.
- Se a faringe permanecer neutra ou "aberta", o container 1 permanecerá grande. A altura do formante 1 é mais grave.
- Se a laringe alargar, os containers 1 e 2 ficam menores. A alturas de F1 e F2 ficam mais agudas.
- Se a laringe diminuir, os containers 1 e 2 aumentam. A altura do F1 e F2 ficam mais graves.

Resumindo, os ressonadores são recipientes de ar. Eles não iniciam o som. No entanto, se entrarem em contato com uma onda sonora semelhante à que desejam vibrar, eles amplificam essa vibração. Assim, os containers menores ressoam frequências mais agudas. Containers maiores ressoam frequências mais graves. Esse é o mecanismo que nos permite articular diferentes vogais.

3. Caracterização dos sons das vogais do português falado no Brasil

Os sons das vogais são caracterizados pelas regiões de frequências de ressonância através de dados obtidos na análise da média dos formantes. Tomamos como base a pesquisa desenvolvida na INATEL (SIQUEIRA et all, 2021), que criou um banco de dados com a gravação de voz de cinco amostras de cada vogal de 10 vozes masculinas e 10 vozes femininas para obter as regiões dos formantes.

> As vogais possuem som contínuo e o trato vocal supraglótico sem bloqueio na passagem de ar. As características do som de cada segmento vocálico dependem da formação das cavidades supraglóticas que geram as frequências no trato vocal que são denominadas formantes. A frequência da primeira formante (F1) e da segunda formante (F2) são essenciais para determinar a característica de uma vogal. Na produção das vogais há movimentos nos articuladores e os estudos de Lindblom e Sudbergn mostram que a formante F1 está interligada com a mandíbula e a formante 2 está relacionada com a língua. Entretanto, a faringe influencia em todas a formantes. (SIQUEIRA *et al.*, 2021, p.2)

Essa pesquisa nos trouxe elementos relevantes, demonstrando a relação das vogais com a articulação interna que controla os formantes especificamente estudados em falantes do português falado no Brasil. Entendemos assim como podem ser obtidos resultados diferentes de acordo com os diferentes idiomas e suas variações culturais. Mesmo considerando o português falado no Brasil, podem-se obter variações de resultados pelas pronúncias regionais, pela anatomia das vozes masculinas, femininas e infantis, como podemos observar em pesquisas, por exemplo (BROD e SEARA, 2014). Assim, vamos supor que a amostragem utilizada contempla a diversidade vocal brasileira.

Na mesma linha de pesquisa, encontramos também a pesquisa desenvolvida por (SILVA et all, 2018). Aponta que os sons da fala constituem o primeiro aspecto que chama atenção ao depararmos com uma língua qualquer ou uma variação regional de nossa própria.

Explica didaticamente que os estudos da fala se dividem em fonética e fonologia. Fonética é o estudo dos sons como entidades físico-articulatórias do aparelho fonador. Seu objetivo é descrever os sons da linguagem e analisar suas particularidades articulatórias, acústicas e perceptivas, ou seja, identificar o mecanismo envolvido na produção e recepção dos sons da fala. A Fonologia é o estudo das diferenças fônicas que geram significados e suas relações entre os elementos e condições de diferenciação que combinados podem formar morfemas, palavras e frases (SILVA et all, 2018, p. 28).

> Tratando-se da língua do português brasileiro, uma vogal sempre compõe uma sílaba e essa vogal em termos fonéticos ocorre em uma intensidade sonora mais elevada que as outras letras que formam a sílaba, tornando-se então o núcleo, que para efeitos práticos, ajuda na sua identificação por meios de análise de frequência. As quantidades de vogais no português brasileiro são refinadas em relação ao modelo internacional em um total de quinze tipos de vogais, dividindo entre dez orais (entre tensas e frouxas) e cinco nasais. (SILVA *et al*., 2018, p. 29)

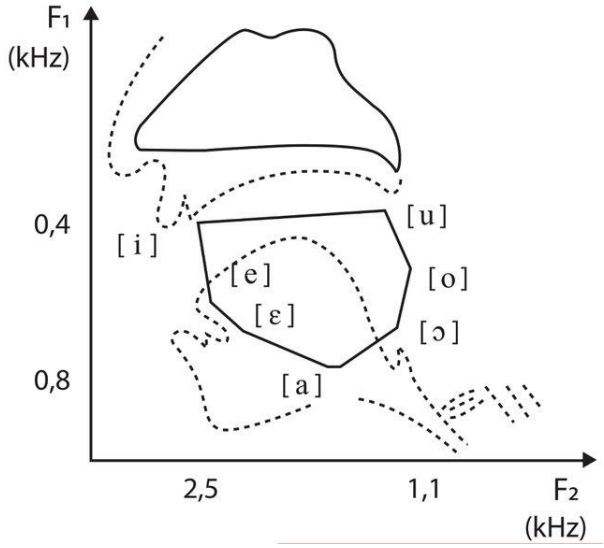


Figura 4: Representação acústico-articulatória das vogais com relação ao quadrilátero vocálico do português brasileiro. Fonte: (SILVA *et al.*, 2018, p. 28)

Empenhados em obter dados das frequências dos harmônicos correspondentes aos primeiros formantes relacionados as vogais, encontramos a pesquisa apresentada em (GONÇALVES et all, 2009), juntamente com a informação relevante das suas intensidades e respectivos desvios padrão de acordo com os biotipos masculino e feminino, considerando as vogais do português falado no Brasil.

VOGAIS	MULHERES			HOMENS		
	F1	F2	F3	F1	F2	F3
/a/	1002,90	1549,95	2959,70	753,87	1278,70	2483,44
/ɛ/	672,45	2242,93	3018,60	588,44	1745,11	2566,00
/e/	437,03	2429,76	3087,09	406,63	1955,60	2540,33
/i/	361,90	2583,89	3378,14	297,80	2150,85	2925,14
/ɔ/	715,34	1073,27	2981,69	580,15	947,25	2525,52
/o/	444,89	914,26	2899,80	411,62	832,84	2376,13
/u/	461,82	763,41	2902,55	345,27	799,51	2351,50

Tabela 1: Médias dos valores das frequências dos harmônicos correspondentes aos três primeiros formantes (F1, F2 e F3) em Hz, para cada vogal. Fonte: (GONÇALVES *et al.*, 2009, p. 682).

Em nossa implementação, iremos elaborar uma sequência de dados unindo as informações da tabela 1 com valores de largura de banda (Q) e amplitudes relativas para obtermos a simulação das vogais. Os símbolos fonéticos da tabela 1 correspondem às vogais especificadas no International Phonetic Alphabet (IPA, 2021), relacionadas, respectivamente, às vogais “a, é, e, i, ó, o, u” do português brasileiro. Assim, doravante usaremos as letras do alfabeto para melhor compreensão do leitor não especializado em fonética.

4. Implementando a síntese de formantes em ambiente PureData

PureData é uma linguagem gráfica de programação para multimídia. Foi desenvolvida por Miller PUCKETTE (2007) com a colaboração de vários pesquisadores responsáveis pela criação de bibliotecas externas que ampliam as funcionalidades do ambiente de programação (www.puredata.info). Utilizamos a versão 0.51-4.

Em nosso levantamento de pesquisas correlatas, encontramos o trabalho desenvolvido por Jim QODE (2021) que realizou uma implementação com base nas vogais da língua turca. Seu sintetizador de formantes vocais apresenta elementos que foram aproveitados em nossa versão em português brasileiro. Acrescentamos a variante feminina nas opções de vogais. Utilizamos a biblioteca de *externals ELSE*, de Alexandre PORRES (2021).

Os dados ficam armazenados em arquivo texto juntamente com o patch no mesmo diretório. Os arquivos texto são estruturados na forma:

Criamos um subdiretório *data*/Formant$1, onde $1 é a variável que representa números de 1 a 14, na sequência em que apresentamos na Tabela 2. Assim, o parâmetros abaixo podem ser alterados e atualizados facilmente:

<Frequência central (Hz)> <largura da banda (Q)> <amplitude relativa (db)>

Complementando a tabela anterior, obtemos a Tabela 2:

homens		**F1**			**F2**			**F3**	
VOGAIS	**Fc(Hz)**	**Q(Hz)**	**A(db)**	**Fc(Hz)**	**Q(Hz)**	**A(db)**	**Fc(Hz)**	**Q(Hz)**	**A(db)**
a	753,87	42	6	1278,79	72	6	2483,44	140	-4
é	588,44	33	6	1754,11	98	5	2566,00	144	0
e	406,63	22	5	1955,6	109	2	2540,33	142	0
i	297,80	13	9	2150, 85	120	2	2925,14	164	0
ó	580,15	32	7	947,25	53	3	2525,52	141	-4
o	411,62	23.	8	832,84	47	3	2376,13	133	-4
u	345,27	30	4	799, 51	80	4	2351,50	131	-2

mulheres		**F1**			**F2**			**F3**	
VOGAIS	**Fc(Hz)**	**Q(Hz)**	**A(db)**	**Fc(Hz)**	**bandw(Hz.**	**A(db)**	**Fc(Hz)** **A(db)**		**Q(Hz)**
a	1002,90	57	6	1549,95	87	6	2959,70	166	-4
é	672,45	37	6	2242,93	125	5	3018,60	169	0
e	437,03	25	5	2429,76	136	2	3087,09	173	0
i	361,90	20	9	2583,89	145	2	3378,14	189	0
ó	715,34	40	7	1073,27	60	3	2981,69	167	-4
o	444,89	25	8	914,26	51	3	2899,80	163	-4
u	461,82	26	4	763,41	43	4	2902,55	162	-2

Tabela 2: Frequência central, largura de banda (Q) e amplitude relativos aos três formantes principais relacionados com as vogais. Fonte: do autor.

Experimentalmente, estabelecemos uma largura de banda (Q) aproximadamente um semitom cromático acima e um abaixo das frequências centrais. As amplitudes são arbitrárias. O patch principal recebe os seguintes *inputs*:

- frequência fundamental (*fundamental*)
- largura de pulso (*largura_de_pulso*)
- vogal (entrada numérica de acordo com tabela anexa)

fundamental e *largura_de_pulso* alimentam o gerador de pulso [pulse~] que então é conectado com um filtro passa baixa [lowpass~]. Seus argumentos são; 1º *frequência de corte* em Hz; e 2º *ressonância* (de 0 a 1).

A função [forfil3] recebe o sinal filtrado na primeira entrada e os dados do formante da vogal escolhida. O sinal processado é enviado a um filtro passa alta, com

frequência de corte em 100 Hz. O resultado final é enviado ao conversor DA do hardware utilizado por meio da função [output~].

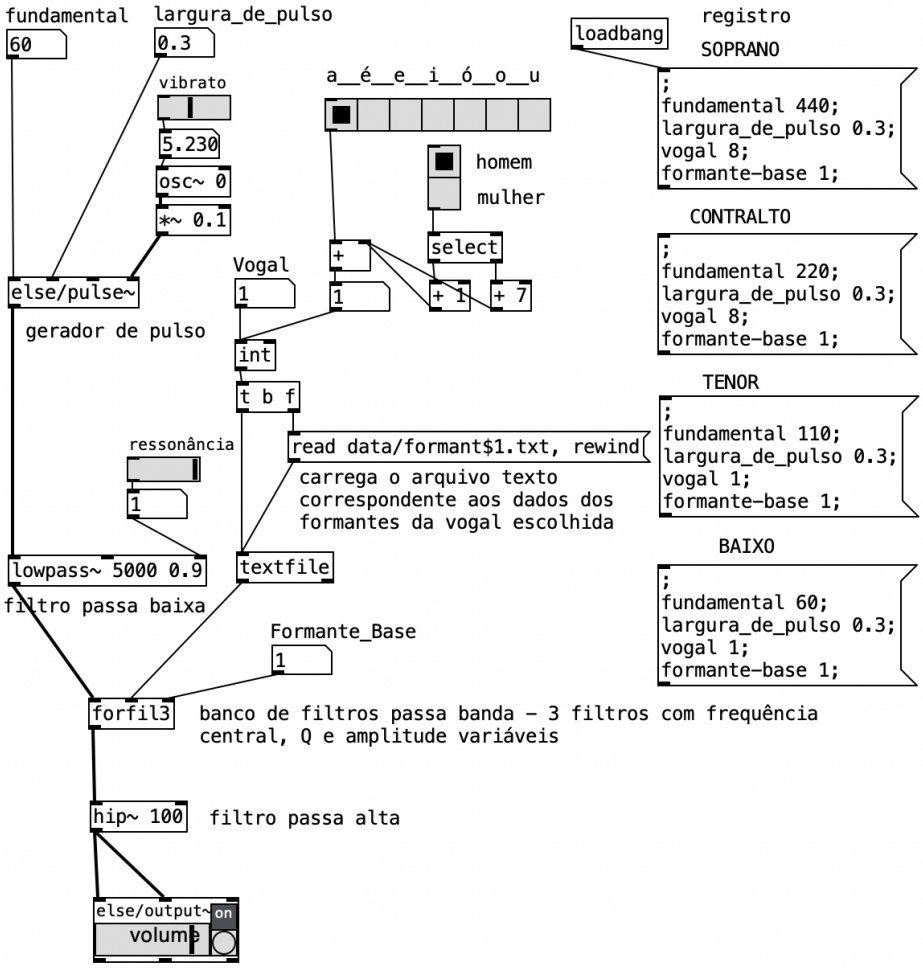


Figura 5: patch principal – síntese de formantes. Fonte: do autor

5. Discussão, resultados e conclusões

Na saída da função [pulse~] obtemos a seguinte forma de onda, comum a todas as vogais, representando o som que é produzido pelas pregas vocais, ainda na laringe:

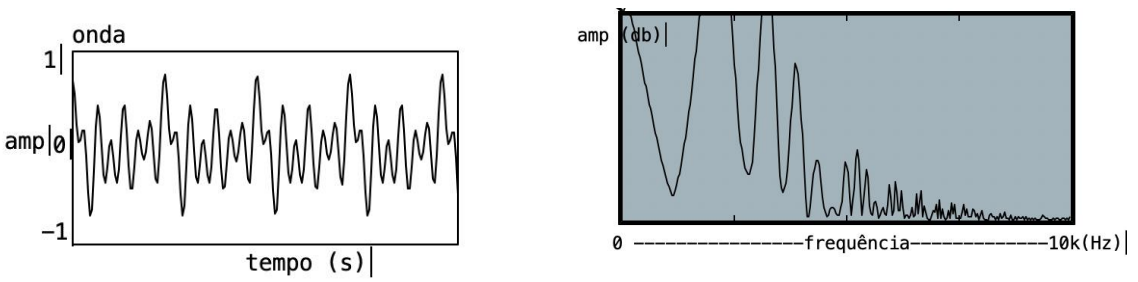


Figura 6: forma de onda do gerador de pulso e seu espectro . Fonte: do autor

Em sequência, apresentamos o espectro de cada vogal na figura abaixo:

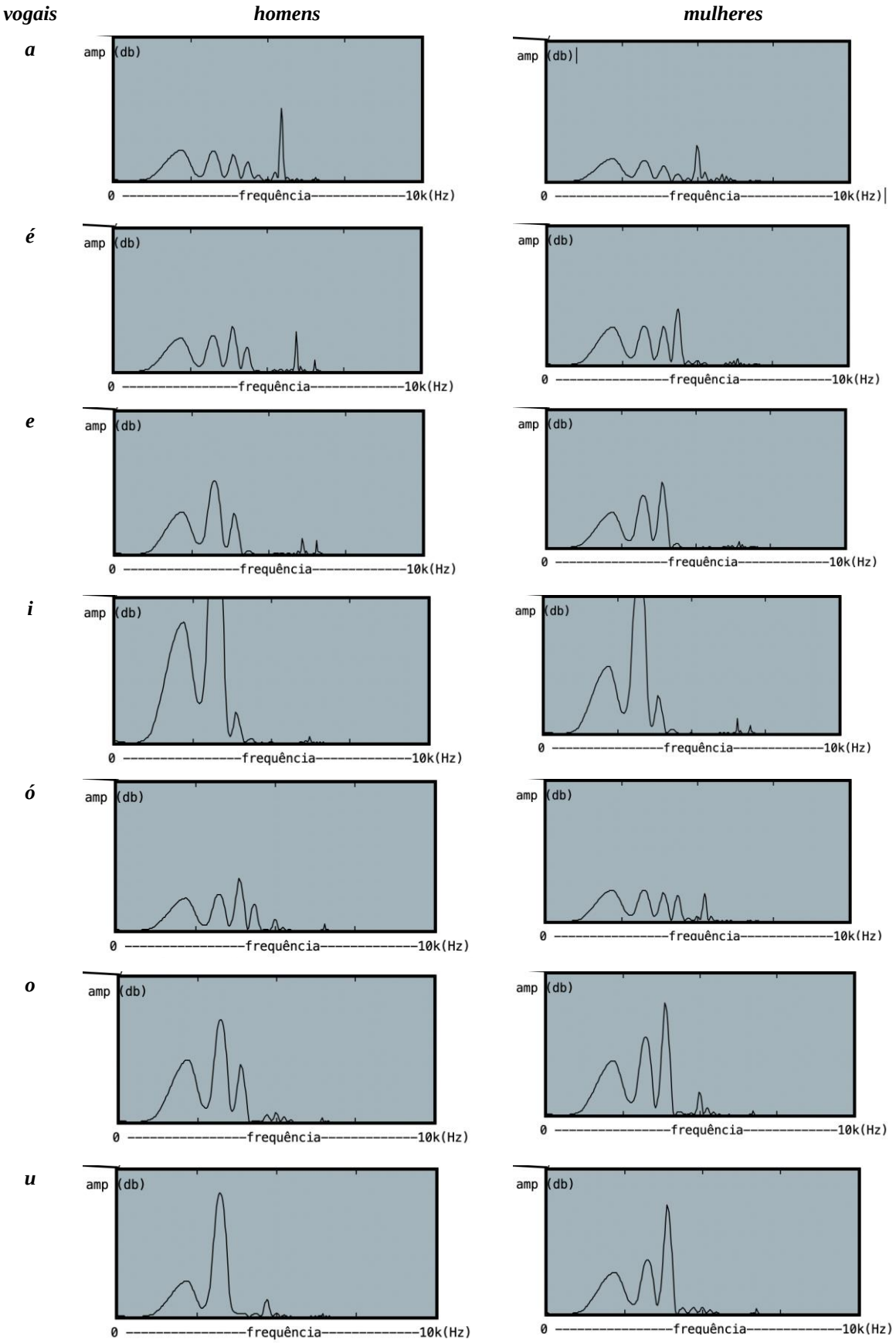


Figura 7: Espectros correspondentes às vogais sintetizadas

Podemos verificar por meio do exame da Figura 7, que os espectros são derivados do espectro da Figura 6, filtrados conforme os formantes apresentados na Tabela 2. Assim, demonstra-se que, mesmo com diferentes frequências centrais, homens e mulheres produzem espectros com a mesma morfologia característica de cada vogal, sendo os formantes da mulher localizados em região mais aguda que os homens.

Alguns elementos de interface foram criados para facilitar a entrada de dados e auxiliar na visualização das ações. Foram criadas caixas de mensagem com itens de inicialização reguladas para os registros das vozes de soprano, contralto, tenor e baixo. Assim, os dados adequados são enviados às entradas *fundamental*, *largura de pulso* e *vogal* (para sempre iniciar com a vogal "a"). Outras caixas de mensagem de inicialização podem ser criadas pelo usuário. Os dados dos formantes armazenados em arquivos texto que devem ser dispostos numa subpasta chamada *data*. Assim, os arquivos *formant1.txt* a *formant7.txt* correspondem ao formantes da vogais *a, é, e, i, ó, o, u* em homens, assim como *formant8.txt* a *formant14.txt* corresponde as mesmas às mulheres. O conteúdo do arquivo texto então é lido [read] e armazenado [textfile]. A função [forfil3] é um banco de três filtros passa-banda que utiliza os dados armazenados e [textfile]. Para a seleção das vogais e do gênero usamos seletores, um horizontal para selecionar as vogais e um vertical para selecionar o gênero.

Fazendo a experimentação, pudemos certificar auditivamente o resultado. Mesmo com a simplicidade desse patch os timbres masculinos (tenor e baixo) e os femininos (soprano e contralto) apresentam um certo realismo. Ao experimentarmos combinações não convencionais, por exemplo, homem em registro de soprano, o som resultante se assemelha ao que chamamos de *falsete*. Por sua vez, se pegarmos os formantes femininos e aplicar nos registros graves, soa como uma mulher imitando as vozes graves.

Outro elemento introduzido foi o *vibrato*, aproveitando o *inlet* da função [pulse] relativa ao *input* de modulação. Outros elementos podem ser acrescentados, por exemplo, a entrada de um controlador MIDI externo, permitindo a performance de sequências melódicas, assim como a utilização de geradores de envoltória. Porém, ficam estes como sugestão para próximas versões do *patch*.

Esta implementação mostra-se valiosa para os estudos de sonologia, pela possibilidade de geração, audição e visualização dos espectros produzidos, pela abertura para experimentação e facilmente expansível para novas funcionalidades e aplicações. É uma ferramenta útil para sala de aula, como complemento de aulas de acústica, tecnologia musical,

canto e composição. Vislumbramos possibilidades de uso em pesquisa musicológica, seja na musicologia sistemática ou na etnomusicologia.

A demonstração do funcionamento do patch em audiovisual e a disponibilização do patch são encontrados no endereço https://youtu.be/90h4DotINjo.

Referências

BROD, Lilian e Izabel SEARA. Caracterização acústica de vogais orais na fala infantil: o falar florianopolitano. *Letras de Hoje*. Porto Alegre, v.49, n.1, p. 95-105, 2014.

GONÇALVES, M. Inês, et all. Função de transferência das vogais orais do Português brasileiro: análise acústica comparativa. *BJORL – Brazilian Journal of Otorhinolaryngology*. São Paulo: v. 75, ed. 5, p. 680-684, 2009.

IPA. *The International Phonetic Alphabet*. International Phonetic Association. Disponível em: https://www.internationalphoneticassociation.org/content/full-ipa-chart. Acesso em 17 set 2021.

PORRES, Alexandre. *Live Electronics Tutorials*. Disponível em https://github.com/porres. Acesso em 4 jul 2021.

PUCKETTE, Miller. *The Theory and Technique of Electronic Music*. Hackensack, N.J.: World Scientific Publishing Co., 2007.

QODE, Jim. *Speech Formant Synthesizer*. Forum Pd Patch Repository. Disponível em http://www.pdpatchrepo.info/hurleur/formant.zip. Acesso em 4 jul 2021.

ROADS, C. (editor). *The Computer Music Tutorial*. Mass.: MIT Press, 1996.

ROSSING, Thomas *et al. The Science of Sound*. Edição 3ª. São Francisco: Addison Wesley, 2002, 783 pgs.

SILVA, Adelino *et al*. Identificação de padrões de vogais em registros acústicos: análise por componentes cepstrais e redes neurais. *Pós em Revista do Centro universitário Newton Paiva*. Belo Horizonte: v. 2016/2, n.13, p. 27-35, 2016.

SIQUEIRA *et al*. Características dos sons das vogais do português falado no Brasil. Editora*:* *Instituto Nacional de Telecomunicações – INATEL*. Disponível em: https://www.inatel.br. Acesso em 4 jul 2021.