

Para a melhoria da síntese articulatória das vogais nasais do português europeu: estudo da duração e de características relacionadas com a fonte glotal

António Teixeira
Francisco Vaz
Lurdes Castro Moutinho
Rosa Lídia Coimbra
Raquel Castro Lisboa
Universidade de Aveiro - Portugal

Abstract

The aim of this research was to continue the work in articulatory synthesis. Thus, a database was collected, containing the nasal sounds of European Portuguese (EP). This corpus includes the different phonetic contexts for the nasal vowels of EP, focusing on the regional variants. The recordings include voice signal, and, simultaneously, information about glottal source, obtained through an Electroglotograph (EGG).

After the signal processing the following parameters were analysed: duration, F0, open quotient, jitter and shimmer. Finally, statistical and inferential analysis was made using the programme SPSS.

I. INTRODUÇÃO

Motivações

Para a continuidade do trabalho em síntese articulatória na Universidade de Aveiro (TEIXEIRA, 2000), tornam-se imprescindíveis dados acerca da produção dos sons do português europeu. Tendo-se a pesquisa, até ao momento, centrado no estudo das vogais nasais, após a recolha de dados usando EMMA (TEIXEIRA & VAZ, 2001), encontra-se em fase de recolha e análise um novo *corpus* acústico (projeto “Fonética Aplicada ao Processamento da Fala: As Nasais do Português”, <http://www.ii.ua.pt/cidlc/gcl/default.asp>).

O sintetizador articulatório necessita, para além da configuração dos articuladores fora do âmbito desta apresentação, de informação acerca da duração e de parâmetros relacionados com a excitação glotal dos quais se destacam a frequência fundamental (F0) e quociente de abertura (OQ). Uma lista completa dos parâmetros utilizados pelo sintetizador relativos à fonte encontra-se na Tabela 1.

Tabela 1 – Parâmetros relativos à fonte glotal utilizados pelo sintetizador articulatório da Universidade de Aveiro. Em negrito apresentam-se os parâmetros estudados no presente trabalho. Refira-se que alguns dos parâmetros não são facilmente obtidos através de análise do sinal de voz e/ou EGG

<i>Parâmetro</i>	<i>Descrição</i>	<i>Mínimo</i>	<i>Máximo</i>	<i>Valor típico</i>	<i>Unidades</i>
ag0	Área glotal mínima	0	10	0	cm ²
agmax	Área glotal máxima	0	10	0.3	cm ²
f0	Freq. fundamental	20	1000	100-200	Hz
openq	Quociente abertura	0	100	60	% of T0
speed	Quociente Velocidade	0	100	2	
jitter	Variação de F0	0	10	2	%
shimmer	Variação de Agmax	0	10	5	%
slope	Declive	0	100	0.03	cm
aspiration	Aspiração	0	5		
lungs	Pressão pulmonar	0	20000	10000	dyne/cm ²
g0	Ganho	0	100		

Geralmente, os *corpora* existentes e as análises publicadas sobre este tema limitam-se a um número reduzido de falantes, regiões do país e contextos fonéticos.

Projeto “Fonética Aplicada ao Processamento da Fala: As Nasais do Português”

Este trabalho é motivado pela existência de uma investigação multidisciplinar, entre lingüistas e engenheiros (Departamento de Línguas e Culturas – CLC e Departamento de Eletrônica e Telecomunicações – IEETA da Universidade de Aveiro), com o objetivo de estudar e aplicar as técnicas da síntese articulatória ao estudo e síntese do PE.

Desde 1995 que se tem desenvolvido pesquisa no domínio da síntese dos sons nasais do português (TEIXEIRA, 2001). As experiências efetuadas mostram que a variação do velo, e mesmo de outros articuladores, influencia a produção e percepção da nasalidade. No entanto, não existe um estudo acústico detalhado, considerando a variação das características ao longo da produção das vogais nasais. Informação acerca do comportamento da fonte glotal durante a produção dessas vogais é também necessária para a continuação deste trabalho, bem como informação acerca da sua variação regional.

Pelo exposto, constata-se que se torna imprescindível a obtenção de dados sobre a forma como esses sons são naturalmente produzidos. Foi já recolhida informação acerca da posição da língua, lábios e véu palatino, durante a produção de palavras e frases contendo nasais (TEIXEIRA & VAZ, 2001a e 2001b), utilizando um sistema de *ElectroMagnetic Midsagittal Articulography* (EMMA). Esta técnica, no entanto, não é viável para um número alargado de falantes nem fornece informação acerca do processo de fonação.

Assim, neste novo projeto – “Fonética Aplicada ao Processamento da Fala: As Nasais do Português” (<http://www.ii.ua.pt/cidlc/gcl/fap/>) – utilizamos o Electroglotógrafo (EGG) para obter informações acerca da excitação glotal ocorrida no momento de produção das nasais do

português; frequência fundamental das diferentes vogais; quociente de abertura e de velocidade. Uma análise baseada unicamente no sinal de fala impossibilita a utilização de métodos que permitam a filtragem do fenômeno de inversão glotal. O EGG, de simples utilização e não interferindo com o sinal de fala, revela-se um instrumento de grande utilidade numa pesquisa deste tipo e indispensável para a análise da variação sincrónica do pitch.

II. NOVO CORPUS

O *corpus* (MOUTINHO et al., 2001) foi organizado de forma a contemplar diferentes contextos fonéticos e a coleta inclui já diversas variantes regionais: Minho, Douro Litoral, Beiras, Alentejo e Algarve. Para além do sinal de voz, foi recolhido sinal de um electroglotógrafo (EGG), para facilitar a obtenção de parâmetros relacionados com a fonte.

Contextos fonéticos

Na constituição do novo *corpus*, teve-se a preocupação de incluir todas as vogais nasais do português, inseridas nos diversos contextos fonéticos em que podem ocorrer:

- consoante oclusiva surda – vogal nasal – consoante oclusiva surda;
- consoante oclusiva sonora – vogal nasal – consoante oclusiva sonora;
- consoante oclusiva surda – vogal nasal – consoante oclusiva sonora;
- consoante oclusiva sonora – vogal nasal – consoante oclusiva surda;
- consoante fricativa – vogal nasal – consoante fricativa;
- outra consoante – vogal nasal – consoante fricativa;
- consoante fricativa – vogal nasal – outra consoante;
- vogal nasal seguida ou precedida de consoante líquida;
- vogal nasal no início e no fim de palavra [*corpus* lido];
- seqüências (incluindo vogal nasal) não existentes dentro de palavra;
- vogal oral – vogal nasal.

Foram também contemplados, no *corpus*, contextos para as vogais orais, com o objetivo de confrontar, nos mesmos falantes e nos mesmos contextos, as características das vogais nasais estudadas com as orais correspondentes.

Este *corpus* consideramos vir colmatar uma lacuna no conhecimento dos sons nasais do PE, pelo fato de contemplar variantes regionais e fornecer informações, pela utilização do EGG, acerca da fonte glotal, até ao momento, não considerada.

Cobertura Regional

A análise que aqui será apresentada fundamentar-se-á nos resultados referentes a seis informantes do sexo masculino provenientes de três das regiões já disponíveis: Beira Litoral, Alentejo e Algarve.

A escolha recaiu sobre informantes com idades superiores a 35 anos, naturais e residentes nas localidades escolhidas, não tendo uma escolaridade superior ao atual ensino obrigatório ou seu correspondente. Esta preocupação prende-se com o fato de se pretender obter uma amostragem homogênea, de modo a apenas fazer variar o parâmetro geográfico. Note-se, ainda, que este perfil de informante deverá garantir-nos uma maior representatividade da respectiva variante geográfica.



Figura 1 - Mapa de Portugal com indicação das zonas onde foi efetuada a coleta do material analisado neste artigo.

Recolha



Figura 2 - Fotografia de uma sessão de gravação. São visíveis o sistema de aquisição e o computador portátil (*notebook*)

As gravações foram sempre realizadas no terreno, utilizando o seguinte equipamento: um CSL 4400 e EGG 6103, ambos da Kay Elemetrics Corp., e um computador portátil, tendo o sinal sido captado através de um microfone Shure SM48 diretamente para o disco duro. Durante a coleta, foram utilizados, sempre que possível, estímulos visuais, através de imagens que conduzissem o informante a produzir as palavras pretendidas, evitando, assim, a sua leitura. A cada informante foram solicitadas duas repetições do *corpus*.

Software utilizado

Para este projeto escolhemos o sistema SFS desenvolvido Mark Huckvale do University College London que pode ser obtido a partir de <http://www.phon.ucl.ac.uk/resource/sfs>.

Tanto para a gravação, como para a segmentação, foi utilizado o programa SFSWin, o qual possibilita a gravação simultânea dos dois canais de entrada que se pretendem utilizar: o microfone e o electroglotógrafo. O SFSWin permite visualizar todas as formas de onda, em simultâneo ou individualmente, (oscilograma, espectrograma e

o sinal acústico que representa os movimentos das cordas vocais na produção do som). Para além disso, permite também operar sobre determinados segmentos selecionados, ouvir, apagar, anotar, etc. O sistema SFS permite a conversão dos arquivos no formato SFS em arquivos WAV, de forma a poderem ser processados por outros programas, nomeadamente pelo MatLab. O programa possui uma ferramenta que permite a visualização das formas de onda, em tempo real, do sinal acústico. Assim, é possível monitorizar a forma de onda e a forma espectral dos sons que estão sendo gravados.

Segmentação e Etiquetagem

Após gravação, procedeu-se a uma primeira audição, com vista a selecionar as palavras que constituirão os arquivos a analisar. Para cada um dos segmentos correspondentes às palavras do *corpus* é criada uma anotação no seu início e fim. Com a ajuda do comando **wordchop** do SFS são depois criados arquivos individuais. O nome desse arquivo contém informação acerca do informante, tabela e item do *corpus* a que se refere, e qual a repetição. Por exemplo, o arquivo **clrT01I01r1.sfs** contém os dados referentes à primeira repetição do primeiro item da tabela 1 (vogais nasais entre oclusivas surdas) preferido pelo informante CLR.

A segmentação consiste na identificação dos segmentos consonantais que precedem e seguem a vogal nasal, ficando esta, conseqüentemente, delimitada. Esta anotação é realizada utilizando o sistema SAMPA para o português. Apresenta-se, na Figura 3, um exemplo de um arquivo anotado.

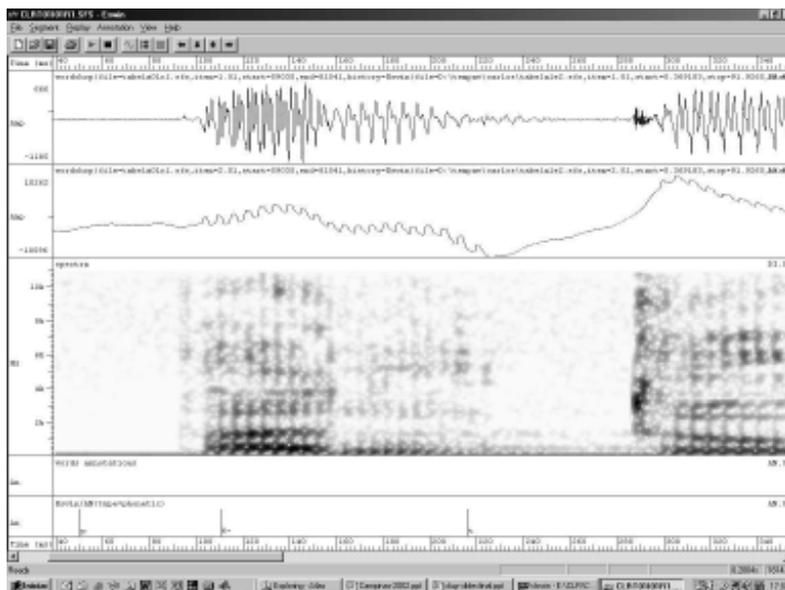


Figura 3 - Exemplo de uma anotação do *corpus*. De cima para baixo, podem ser vistos o sinal de voz, o sinal do EGG, anotação ao nível das palavras e anotação fonética. Nesta última podem ser vistos os pontos de início da vogal nasal e da consoante que a segue (lt) assim como a indicação de qual o segmento que antecede a vogal nasal (neste caso o [p]).

III. METODOLOGIA DE ANÁLISE

- Procura de uma forma automática, em todos os arquivos, dos segmentos anotados como vogal (oral e nasal) obtenção, para cada um, de: duração, F0, quociente de abertura, *jitter* e *shimmer*. Para cada segmento, informação acerca das medidas obtidas e dos vários fatores (falante, região, sexo do falante, tabela, etc.) é guardada para análise estatística;
- execução de análises estatísticas usando os programas SPSS (Bryman, Cramer 2001) e MacAnova;
- interpretação dos resultados em função da variação regional, das diferentes vogais nos vários contextos fonéticos.

Apresenta-se a seguir alguma informação sobre a forma de obtenção dos parâmetros analisados.

Duração

Como no processo de anotação indica-se onde começa a vogal e onde começa o segmento que a segue. O processo de obtenção da duração é muito simples, envolvendo apenas a utilização do comando **lengthn** do SFS. Como os nossos estudos anteriores apontam para que as vogais nasais sejam constituídas em geral por três fases, constituindo o murmúrio nasal a fase final, não medimos este segmento em separado considerando-o parte integrante da vogal nasal.

Freqüência fundamental (F0)

Tendo sido gravado o sinal EGG, a obtenção de freqüência fundamental foi efetuada de duas formas: uma baseada no EGG, outra usando apenas o sinal de voz .

No primeiro caso, o processo consiste em obter informação acerca do início de cada período do sinal EGG, usando o comando **vtx**, obtendo depois o valor de F0 com base nesta informação usando o comando **fx**. O comando **vtx** estima a posição dos períodos de excitação glotal a partir do sinal do EGG, através do método utilizado no Voiscope, que consiste em: (i) filtragem, (ii) controle de ganho automático e (iii) detecção de picos. Este processo é exemplificado na Figura 4, onde se pode ver o sinal EGG, informação acerca do início e fim de cada período e F0.

No segundo caso, a análise é efetuada usando o comando SFS **fxanal**. Este comando estima F0 do sinal de voz através do método da autocorrelação.

Para possibilitar o estudo de uma possível variação deste parâmetro ao longo da realização da vogal, o valor de F0 foi calculado em três pontos: próximo do início (concretamente a 10% da duração da vogal), no meio e próximo do final (a 90% da duração).

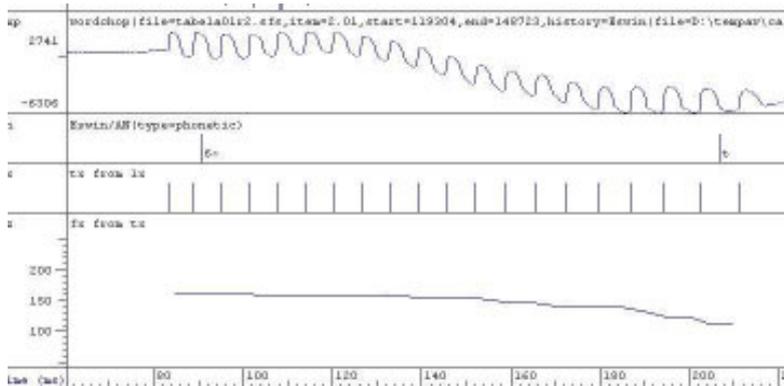


Figura 4 - Exemplo de cálculo de F0 com base no sinal EGG. Com base no sinal EGG, apresentado no topo, é primeiro calculado o início de cada período glotal (chamado de TX no sistema SFS) e depois calculada a frequência fundamental (FX no sistema SFS).

Quociente de abertura

Diversos eventos do processo de fonação estão relacionados com eventos (segmentos ou picos) correspondentes ao sinal EGG ou a sua derivada (DEGG). Por exemplo, o instante de abertura das cordas vocais e o instante do pico positivo no sinal DEGG são coincidentes, assim como o instante de fecho das cordas vocais e o instante do pico negativo do DEGG (CHILDERS, 2000, p. 87). O quociente de abertura (em inglês *open quotient* – OQ) medido da área glotal e o OQ medido a partir do sinal EGG são quase o mesmo, sendo: $OQ = (\text{duração de abertura das cordas vocais}) / (\text{duração do período glotal})$.

O quociente de abertura foi calculado com base na informação acerca da abertura e fecho das cordas vocais gerada pelo comando SFS **txgen**. O processamento efetuado é o seguinte:

- 1) Filtragem passa-baixo do sinal EGG;
- 2) Diferenciação do sinal EGG (obtendo DEGG);
- 3) Procura do máximo e mínimo do sinal DEGG que se encontra acima de um determinado limiar. A posição destes é utilizada como posição de fecho e início de abertura.

O resultado desse processamento é exemplificado na Figura 11. Com base nesta informação foi desenvolvido um programa em SML (*Speech Measurement Language*) para calcular o OQ. Infelizmente, a detecção efetuada pelo **txgen** apresenta muitas deficiências tornando muito difícil a obtenção de uma boa estimativa de OQ em arquivos com o sinal EGG de menor qualidade.

Jitter

O *Jitter* é uma medida da perturbação do período de oscilação das cordas vocais. Existem várias formas propostas para calculá-lo. No sistema SFS o comando **txstat** permite obter informação do *jitter*. Devido a falta de documentação deste comando, não está disponível a informação acerca do método utilizado. Optamos por desenvolver um novo programa de cálculo do *jitter* usando SML optando pelo cálculo do *Jitter Factor* (LALWANI, 1991). O *Jitter Factor* (JF) é definido como a média do valor absoluto das perturbações na frequência fundamental, normalizada pela média do valor da frequência fundamental, ou seja:

$$JF = \frac{\frac{1}{N-1} \sum_{i=2}^N |F_{0_i} - F_{0_{i-1}}|}{\frac{1}{N} \sum_{i=1}^N F_{0_i}} = \text{Média do módulo das variações} / \text{Média de F0}$$

O valor do JF é calculado usando uma janela (de comprimento N). Para permitir estudar possíveis variações ao longo da vogal, o JF foi calculado: (i) considerando toda a duração da vogal, (ii) uma janela de comprimento fixo (N=5) no início, (iii) no meio e (iv) no fim. Como foi notado, em muitas realizações das vogais, um decréscimo do valor de F0 ao longo da vogal, o valor do *jitter* devido à utilização de um menor valor de F0 média na normalização terá tendência para atingir valores mais elevados. Por este motivo, para além de guardar o JF calculado das 4 formas anteriormente referidas, foi guardado para análise o valor do numerador da fórmula de cálculo do JF.

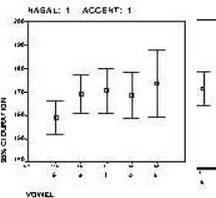
IV. RESULTADOS

Duração

A primeira análise com interesse para aplicação em síntese consiste em obter os valores médios das durações. Na tabela 2, apresentam-se os valores médios, utilizando 6 falantes de 3 regiões, para as 8 vogais orais e as 5 vogais nasais. Os resultados foram decompostos tendo em conta se a vogal se encontrava acentuada ou não.

Tabela 2 – Durações médias das vogais analisadas.

	oral		nasal	
	não acentuada	acentuada	não acentuada	acentuada
a	,	109,38	,	,
Λ	65,81	121,98	133,94	159,14
e	,	112,03	147,49	169,12
ε	,	105,97	,	,
i	,	104,78	147,35	170,76
o	,	116,2		168,55
ɔ	83,28	111,0		,
u	66,56	126,1		173,66



É notória a contribuição da nasa para a duração. As vogais orais não acentuadas apre as vogais nasais em posição acentuada, as maiores durações. uração média;

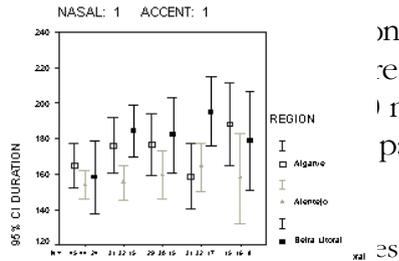
Para além dos valores médios, interessa também a dispersão em torno desses mesmos valores. Na Figura 5 apresenta-se o intervalo de confiança a 95% da média.

Figura 5 – Intervalo de confiança a 95% e valor médio da duração das vogais orais e nasais. Os dois gráficos da esquerda referem-se a vogais orais (nasal=0), o primeiro e terceiro gráfico referem-se ao caso não acentuado (accent=0).

Será que este comportamento se verifica para as três regiões? Para tentar obter uma resposta a esta questão foi efetuada a análise cujos resultados se apresentam na Figura 6. Como estamos mais interessados nas nasais, apenas se apresentam os gráficos referentes a estas vogais.

Figura 6 – Intervalo de confiança das durações para as vogais nasais em posição não acentuada (esquerda) e acentuada (direita) apresentando separadamente os valores obtidos para cada região.

Apesar das médias serem algo sobrepostas. No caso não acentuado de 100 e 180 ms, no caso acentuado de 100 e 180 ms. Na Tabela 3 apresentam-se os valores obtidos para as três regiões analisadas.



Intervalo de confiança de cerca de 100 ms para as vogais nasais.

Tabela 3 – Médias de duração

		VOWEL		Litoral
		nasal	accent	
A	não acentuada	146,03	125,49	132,68
	acentuada	164,80	153,80	158,33
e	não acentuada	161,65	130,69	151,35
	acentuada	176,17	155,14	184,36
i	não acentuada	153,22	135,37	148,80
	acentuada	176,59	159,43	182,15
o	não acentuada	141,86	121,25	147,08
	acentuada	158,75	163,93	195,10
u	não acentuada	151,03	132,92	135,87
	acentuada	187,91	157,85	178,57

Para possibilitar a comparação com o português do Brasil, apresenta-se, na Tabela 4, um resumo das durações médias do presente estudo e do estudo realizado por Sousa (1994). Comparando os valores apresentados nas duas últimas colunas, constatamos não serem detectáveis diferenças significativas entre os valores obtidos para estas duas variedades do português.

Tabela 4 – Comparação dos valores médios obtidos com os de Sousa (1994) para o português do Brasil. As duas primeiras colunas de valores apresentam os resultados de Sousa; a última, os resultados obtidos no presente estudo.

Vogal	de Sousa (1994)		Nosso resultado Para nasal
	Oral	Nasal	
[a] x [ã]	123,03	163,54	159,14
[e] x [ẽ]	109,77	170,05	169,12
[e] x [ē]	98,17		
[i] x [ĩ]	86,85	155,6	170,76
[o] x [õ]	123,2	161,58	168,55
[o] x [ō]	106,11		
[u] x [ũ]	88,54	144,38	173,66

Embora a influência de outros fatores na duração seja também passível de ser investigada com o material disponível, a ainda não disponibilidade para análise da totalidade do *corpus* levou-nos a adiar essas análises para um futuro próximo.

Os testes estatísticos (análise de variância de medidas repetidas – *repeated measures ANOVA*) confirmam como significativa a influência da vogal [F(4,20)=22.033, p=0.000], nasalidade [F(1,3)=11.971, p=0.041], acento [F(1,3) =10.786, p=0.046 (muito próximo de 0.05)] e posição [F(2,6)=48,702, p=0.000] na duração. A influência da região [F(2,3)=.355, p=0.727 (ns)] não se revelou significativa, assim como a influência da repetição [F(1,3)=8.604, p=0.061 (ns)]. Relativamente a este último resultado, pode considerar-se como um elemento de validação das coletas. Diversas interações entre os fatores revelaram-se também significativas – como por exemplo vogal x nasalidade. Não analisaremos aqui esses resultados.

Frequência Fundamental (F0)

Durante o processo de cálculo, foram monitorizados os valores obtidos pelos dois métodos (usando EGG ou o sinal de voz). Para alguns informantes, o processo baseado no EGG não foi capaz de determinar corretamente o início de cada período, falhando um ou mais inícios ou detectando inícios onde estes não existem. Em consequência dessa deficiente detecção, os valores de F0 apresentam valores ora muito altos ora muito baixos. Para analisar esses resultados foram eliminadas, usando as facilidades do comando **fx**, todas medições de F0 inferiores a 100 e superiores a 200 Hz, visto tratar-se de informantes masculinos e os valores obtidos pela análise do sinal de fala não excederem estes limites.

Começamos por investigar se os dois métodos fornecem valores semelhantes. Na Tabela 5 apresentam-se as correlações entre as medidas efetuadas pelos dois métodos. F0 medido no início e no meio apresentam um valor de correlação próximo de cerca de 0.9 para os dois métodos. Os valores mais baixos de correlação aparecem entre F0 iniciais/médios e médios/finais. Devido aos valores elevados da correlação entre os dois métodos e às falhas do método baseado no EGG, grande parte das análises foram efetuadas utilizando os valores obtidos com base no sinal de voz.

Tabela 5 – Correlações entre as 3 medidas pelos 2 métodos

		FOSTART	FOMED	FOEND	FOSTART2	FOMED2	FOEND2
FOSTART	Pearson Correlation	1	,914**	,660**	,862**	,741**	,672**
	Sig. (2-tailed)	,	,000	,000	,000	,000	,000
	N	901	736	837	901	900	901
FOMED	Pearson Correlation	,914**	1	,617**	,774**	,708**	,586**
	Sig. (2-tailed)	,000	,	,000	,000	,000	,000
	N	736	749	695	749	749	749
FOEND	Pearson Correlation	,660**	,617**	1	,596**	,490**	,837**
	Sig. (2-tailed)	,000	,000	,	,000	,000	,000
	N	837	695	909	909	908	909
FOSTART2	Pearson Correlation	,862**	,774**	,596**	1	,868**	,728**
	Sig. (2-tailed)	,000	,000	,	,000	,000	,000
	N	901	749	909	909	1013	1015
FOSTART2	Pearson Correlation	,741**	,708**	,490**	,868**	1	,607**
	Sig. (2-tailed)	,000	,000	,000	,000	,	,000
	N	900	749	908	1013	1013	1013
FOEND2	Pearson Correlation	,672**	,586**	,837**	,728**	,607**	1
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,
	N	901	749	909	1015	1013	1015

** Correlation is significant at the 0.01 level (2-tailed).

Na Figura 7 apresenta-se o estudo da variação de F0 ao longo da vogal em função da vogal, nasalidade e acento. Nota-se um valor de F0 inferior no final da realização das vogais, especialmente notório nas nasais acentuadas.

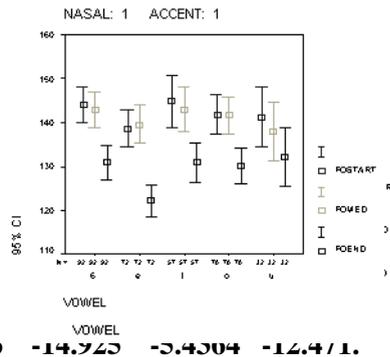
Figura 7 – Intervalo de confiança para nasalidade (nasais em baixo), do acent

Calculando a diferença em vogal, obtém-se:

-15.941 -15.565 -17.272 -16.2

Efetuando o mesmo cálculo

-23.418 -11.851 -18.376



, da reita)

cada

28

1-se:

Ou seja, para todas as vogais e para todos os informantes o valor médio final é cerca de 10 a 20 Hz inferior ao valor médio inicial. O valor médio para vogais orais é -12.854 e -15.02 Hz para as nasais.

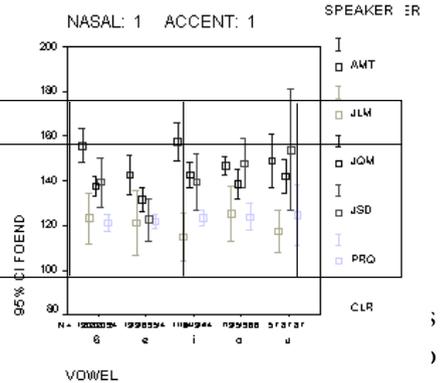
Na Figura 8, apresenta-se o estudo da variação de F0 para os vários informantes. Para a mesma vogal os vários informantes apresentam valores médios e intervalos de confiança díspares. Não consideramos que se possa apresentar um valor médio por vogal. Os fatores intrínsecos ao falante parecem sobrepor-se à vogal. Na Tabela 6

apresentam-se os valores médios no início, meio e fim para cada um dos seis informantes.

Figura 8 – F0 para as vogais nasais (acentuadas ou não) em função da vogal e do falante

Tabela 6 - Valores médios de F0 medido cada um dos sei:

	1 - AMT	2 - JLM	3
F0 inicial	172.81	129.67	161.14
F0 medial	171.91	132.6	161.14
F0 final	149.39	117.82	161.14



Foram efetuadas análises de v no início e fim. Não se justifica a ar pela elevada correlação com o valo

Para o início, o efeito do falante em F0 é estatisticamente significativo, assim como o efeito da posição. Os outros fatores não se revelaram como significativamente afetando o valor de F0 na fase inicial da vogal. Nesta análise, utilizou-se a medida baseada no sinal de fala, para evitar ter de lidar com os casos desconhecidos criados pelas falhas do outro método de obtenção de F0. Detalhes da análise encontram-se na Tabela 7.

Tabela 7 – Resultados da análise de variância (efetuada em Macanova) para os valores de F0 medidos no início

	DF	SS	MS	F	P-value
CONSTANT	2	2.2183e+07	2.2183e+07	28804.00822	0
speaker	5	2.9178e+05	58356	75.77499	0
vowel	7	10657	1522.5	1.97695	0.086576
ERROR1	35	26954	770.12	2.31889	0.17631
nasal	1	8.8699	8.8699	0.02671	0.87658
ERROR2	5	1660.5	332.11	0.19370	0.95207
accent	1	91.851	91.851	0.05357	0.82614
ERROR3	5	8572.9	1714.6	4.86012	0.016292
position	2	4627	2313.5	6.55785	0.015152
ERROR4	10	3527.9	352.79	0.95319	0.48318
ERROR5	943	3.4901e+05	370.11		

Para a medida no final da vogal, além do falante revela-se como também significativa a influência da vogal. A posição onde ocorre a vogal deixou de ser um fator estatisticamente significativo. Detalhes da análise encontram-se na Tabela 8.

Tabela 8 – Resultados da ANOVA realizada para F0 medido no final da vogal.

	DF	SS	MS	F	P-value
CONSTANT	2	1.8089e+07	1.8089e+07	34471.12478	0
speaker	5	1.4281e+05	28563	54.43007	0
vowel	7	15240	2177.2	4.14887	0.020399
ERROR1	35	18367	524.76	0.62862	0.81191
nasal	1	999.09	999.09	1.19614	0.32396
ERROR2	5	4176.3	835.26	0.69276	0.65154
accent	1	549.85	549.85	0.45604	0.52945
ERROR3	5	6028.5	1205.7	1.39928	0.30386
position	2	3078	1539	1.78607	0.21715
ERROR4	10	8616.6	861.66	2.55187	0.004829
ERROR5	943	3.1841e+05	337.66		

Quociente de abertura (OQ)

Na Figura 9 apresenta-se o diagrama de extremos e quartis (*boxplot*) do quociente de abertura em função da vogal e da nasalidade. A figura não revela diferenças de valores médios nem de dispersão para as várias vogais, tanto para as orais como para as nasais.

Figura 9 – Diagrama de extremos e quartis (*boxplot*) dos valores do quociente de abertura em função da vogal e nasalidade.

Na Figura 10 apresenta-se o diagrama de extremos e quartis (*boxplot*) do quociente de abertura para os diferentes informantes. Nota-se uma variação em função do falante, não sendo a região um fator significativo. Falantes da mesma região exibem comportamentos bastante diferentes entre si.

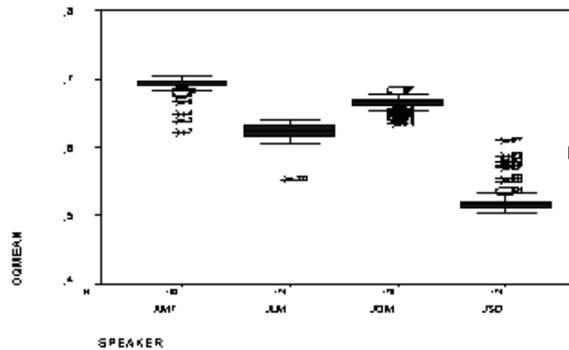


Figura 10 – *Boxplots* para cada falante. Os dois da esquerda do Algarve, os do meio do Alentejo e os da direita da Beira Litoral.

Analisando com mais detalhe, se olharmos para os valores de cada informante para cada uma das vogais, inscritos na Tabela 9, constatamos que todos os valores para o mesmo falante são muito semelhantes, tanto para as orais como para as nasais.

Tabela 9 – Valores médios de OQ por vogal e falante. Na primeira coluna apresenta-se a sigla pela qual é identificado o falante.

oral

	a	Λ	e	ε	i	o	ɔ	u
AMT	,69	,69	,69	,69	,69	,69	,69	,69
JLM	,62	,62	,61	,62	,62	,61	,62	,62
JQM	,67	,66	,67	,67	,67	,67	,67	,67
JSD	,51	,51	,52	,51	,52	,52	,52	,52
PRQ	,62	,62	,58	,58	,58	,58	,58	,58
CLR	,50	,50	,50	,51	,51	,50	,51	,51

nasal

	Λ	e	i	o	u
AMT	,69	,69	,69	,69	,69
JLM	,62	,61	,62	,61	,62
JQM	,66	,67	,67	,67	,67
JSD	,51	,52	,51	,52	,52
PRQ	,62	,58	,58	,58	,58
CLR	,50	,50	,51	,51	,51

Não é utilizável um valor médio global, nem mesmo para cada vogal, visto que cada informante apresenta um valor médio “próprio”. Os histogramas de OQ para cada vogal, em função da região e em função da nasalidade, apresentam vários máximos não sendo representados corretamente por uma distribuição unimodal.

Foi também investigado o efeito do acento, não se tendo constatada uma influência significativa desse fator no quociente de abertura.

Os valores obtidos para quatro dos informante estão de acordo com valores médios publicados para a voz modal (CHILDERS, 2000, p. 313). Os informantes AMT e JQM apresentam valores que já se podem considerar característicos de outros tipos de voz.

Jitter

Apenas foi possível utilizar o método de cálculo do JF em 4 dos 6 informantes. O problema deveu-se à dificuldade que o programa **hqtx** teve em processar o sinal EGG para dois dos falantes.

Na Figura 13 apresentam-se os resultados das medidas efetuadas no início, meio, fim e na totalidade da vogal, em função da vogal e da nasalidade. Numa primeira análise, os valores dos quatro gráficos não diferem grandemente.

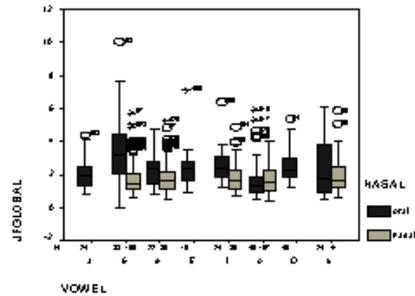


Figura 13 – *Boxplot* do valor do *Jitter* , vogal (início, meio e fim) e usando a totalidade da duração da vogal. São apresentados valores para cada vogal separados pela nasalidade

Os valores médios encontram-se na Tabela 10. Para cada caso apresenta-se o valor de JF no início, meio, fim e na globalidade (de cima para baixo).

Tabela 10 – Valores médios obtidos para o *Jitter Factor*. Na primeira tabela da esquerda por vogal e tendo em conta a nasalidade, a segunda tabela da direita médias globais para os valores medidos nas três posições e na globalidade da vogal. Na última tabela de baixo à esquerda os valores para vogais orais, à direita para vogais nasais.

	a	Λ	e	ε	i	o	ɔ	u
oral	2,16	3,27	2,43	2,82	2,08	1,69	2,42	2,34
	1,53	3,06	1,80	1,81	2,14	1,56	2,26	2,13
	1,96	3,65	2,75	2,36	2,74	1,63	2,20	2,56
	2,07	3,51	2,37	2,47	2,56	1,78	2,59	2,46
nasal	,	1,85	1,88	,	1,91	1,72	,	1,98
	,	1,25	1,51	,	1,28	1,29	,	1,77
	,	2,29	2,17	,	2,47	2,12	,	2,04
	,	1,78	1,85	,	1,83	1,74	,	2,02

Descriptive Statistics

	N	Mean	
	Statistic	Statistic	Std. Error
JFSTART	692	2,0159	,0560
JFMED	692	1,5907	,0526
JFEND	692	2,2996	,0613
JFGLOBAL	692	2,0258	,0433
Valid N (listwise)	692		

Descriptive Statistics

	N	Mean	
	Statistic	Statistic	Std. Error
JFSTART	229	2,3538	,1074
JFMED	229	2,0568	,1062
JFEND	229	2,4284	,1043
JFGLOBAL	229	2,4587	,0972
Valid N (listwise)	229		

a. NASAL = oral

Descriptive Statistics

	N	Mean	
	Statistic	Statistic	Std. Error
JFSTART	463	1,8487	,0560
JFMED	463	1,3602	,0526
JFEND	463	2,2359	,0613
JFGLOBAL	463	1,8116	,0433
Valid N (listwise)	463		

a. NASAL = nasal

Nota-se um valor inferior, relativamente ao início e fim de JF, para as vogais nasais medido a meio da vogal. O valor final é o mais elevado. No entanto, estando o valor de JF dependente do valor médio de F0, é preciso determinar se essa diminuição se deve a um aumento desse valor. Para isso analisamos o valor do numerador da expressão de cálculo de JF, apresentado na Figura 14. Continua a aparecer um valor mais baixo da medida efetuada a meio, pelo menos para 4 das vogais nasais.

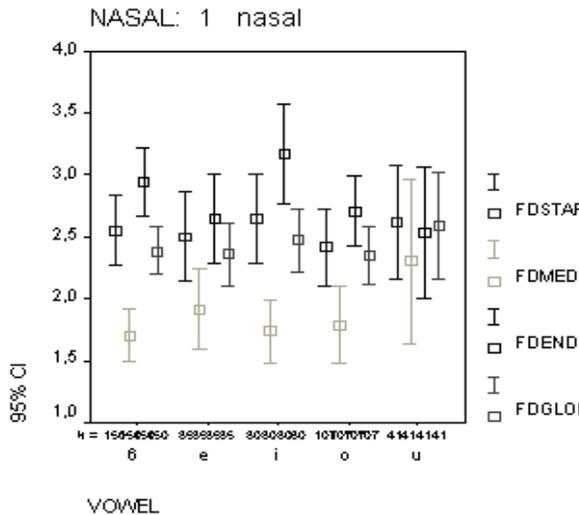


Figura 14 – Numerador do cálculo

Vendo os gráficos e estatísticas associadas com o contexto nasal, o efeito aparece nos dois casos. Este resultado deve ser objeto de mais análises.

A análise de variância revela como não significativa a influência da vogal, nasalidade, acento, posição, repetição e região no valor de JF.

V. DISCUSSÃO

Quanto à variação regional, a amostra apresenta-se ainda demasiado pequena. Por agora, o *corpus* tem apenas como vantagem não ser exclusivamente constituído por falantes de uma zona do país, ser recolhido no terreno, integrando por isso mais variação do que o habitual.

Apesar do *corpus* incluir todos os contextos em que ocorrem vogais nasais em português, neste estudo apenas foram analisados os contextos envolvendo oclusivas, essencialmente por serem os mais representativos e estarem já etiquetados. Os resultados apresentados devem ser lidos tendo em conta a limitação a esses contextos.

A utilização do EGG possibilitou a obtenção de parâmetros como o quociente de abertura e o *jitter*. Na fase de etiquetagem o sinal EGG foi também de grande utilidade, permitindo mais precisão na identificação sobretudo da fase inicial da vogal.

O sistema SFS revelou-se uma ferramenta poderosa, em especial a linguagem SML – *speech measurement language* –, tendo facilitado muito a separação do *corpus* em arquivos individuais e a análise de forma automática de todo o *corpus* já anotado. Foi também fácil criar os arquivos no formato utilizável pelo programa de análise estatística. “Facilmente” se podem criar arquivos para treino de modelos CART para utilização num sistema de síntese. Será interessante analisar as árvores de decisão criadas.

Os resultados apresentados são essencialmente exemplificativos do que se pode obter quando o *corpus* estiver completo. Tencionamos no final do projeto, refazer as análises agora apresentadas utilizando todas as regiões e informantes recolhidos.

Foi feito um esforço de comparação dos resultados obtidos com resultados publicados (ANDRADE, 1987; DELGADO MARTINS, 2002). No entanto, o tipo de parâmetro por nós analisados e a metodologia diferente – tendo em conta os nossos objetivos – adotada para outros não nos permitiu grandes comparações. Por exemplo, muitas das análises de duração das vogais nasais não incluem o segmento consonântico nasal final; muitos dos resultados referem-se apenas a um informante e a voz de “laboratório” não espontânea.

VI. CONCLUSÕES E TRABALHO FUTURO

Motivados pela necessidade de obtenção de parâmetros relativos à fonte glotal para continuação do trabalho que se tem desenvolvido na área de síntese articulatória na Universidade de Aveiro, estamos procedendo a coleta de um *corpus* contemplando as variações contextuais e regionais das vogais nasais. Neste artigo, apresentamos a metodologia de análise e os resultados obtidos relativamente aos seguintes parâmetros: duração, frequência fundamental, quociente de abertura e *jitter*.

Dos resultados obtidos salienta-se:

1. Duração – Os testes estatísticos confirmam como significativa a influência da vogal, nasalidade, acento e posição na duração. A influência da região não se revelou significativa, assim como a influência da repetição. São apresentados: valores médios, utilizando 6 falantes de 3 regiões, para as 8 vogais orais e as 5 vogais nasais; médias para cada vogal para as três regiões analisadas; e a comparação com os valores medidos para o português do Brasil de Sousa (1994);
2. Frequência fundamental – Verifica-se uma grande correlação entre os valores obtidos com base no EGG ou apenas no sinal de voz; um valor de F0 inferior no final da realização das vogais é especialmente notório nas nasais acentuadas; que para a mesma vogal os vários informantes apresentam valores médios e intervalos de confiança díspares, os fatores intrínsecos ao falante parecem sobrepor-se à vogal;
3. Quociente de abertura – Revelou-se uma característica idioletal, cada informante com um valor muito estável com a variação da vogal, nasalidade etc. Falantes da mesma região exibem comportamentos bastante diferentes entre si;
4. *Jitter* - Numa primeira aproximação podem usar-se valores de JF da ordem dos 2%. Caso se esteja interessado em incluir na modelação mais detalhe – cujos efeitos em termos de qualidade deverão ser estudados por testes perceptuais – poderá utilizar-se um valor um pouco inferior a 2% na zona média e um valor acima de 2% na parte final.

Trabalho futuro

1. Analisar todo o *corpus* assim que estiver disponível, estendendo esta análise para todas as regiões do *corpus* e ao sexo feminino;
2. Melhorar a técnica de obtenção do quociente de abertura;
3. Adicionar a análise da duração da fase final de realização das vogais nasais (o habitualmente designado de murmúrio nasal);
4. Medir outros parâmetros da fonte como quociente de velocidade (*Speed Quotient*);
5. Finalmente, utilizar os dados no sintetizador articulatório da Universidade de Aveiro.

Agradecimentos

A todos os que contribuíram para o *corpus*, FCT pelo financiamento (Projeto PCTI 1999 PLP 36427).

REFERÊNCIAS

ANDRADE, Amália. *Um Estudo Experimental das Vogais Anteriores e Recuadas em Português*. 1987. 355 p. Dissertação (Mestrado em Linguística Portuguesa para acesso à categoria de Investigador Auxiliar) - Centro de Linguística da Universidade de Lisboa, Lisboa.

BRYMAN, Alan; CRAMER, Duncan. *Quantitative Data Analysis with SPSS Release 10 for Windows – A Guide for Social Scientists*. Hove: Routledge, 2001. 295 p.

CHILDERS, Donald. *Speech Processing and Synthesis Toolboxes*. New York: John Wiley & Sons, Inc, 2000. 483 p.

de SOUSA, Elizabeth M. G. *Para a Caracterização Fonético-Acústica da Nasalidade no Português do Brasil*. 1984. 170p. Dissertação (Mestrado) - Universidade Estadual de Campinas, Campinas.

DELGADO-MARTINS, Maria Raquel. *Fonética do Português. Trinta anos de investigação*. Lisboa: Edit. Caminho, 2002.

LALWANI, Ajit. *Flexible Formant Synthesizer: A Tool for Improving Speech Production Quality*. 1991. 318 p. Dissertação (Doctor of Philosophy) - University of Florida, Gainesville.

MOUTINHO, Lurdes Castro et al. Contributo para o estudo da variação contextual e regional das vogais nasais do Português Europeu. In: DUARTE, Isabel Margarida; BARBOSA, Joaquim; MATOS, Sérgio; HÜSGEN, Thomas (Org.). *Encontro Comemorativo dos 25 anos do Centro de Linguística da Universidade do Porto*. Porto: Universidade do Porto, 2001 (Julho 2002). 2 v. v. 2, p. 5-17.

TEIXEIRA, António; VAZ, Francisco. European Portuguese nasal vowels: An EMMA Study. In: *Proceedings Eurospeech*. Aalborg, Dinamarca, 2001.

TEIXEIRA, António. *Síntese articulatória das vogais nasais do Português Europeu*. julho 2000. 235 p. Tese (Doutoramento em Engenharia Electrotécnica) - Universidade de Aveiro, Departamento de Electrónica e Telecomunicações.