

MODELOS MORFOLÓGICOS TRIDIMENSIONAIS POR IRM DO TRACTO VOCAL PARA AS PRINCIPAIS VOGAIS DO PORTUGUÊS EUROPEU

Sandra M. Rua^{1*} e Diamantino R. Freitas²

1: Área Científica da Radiologia
Escola Superior de Tecnologia da Saúde do Porto
Instituto Politécnico do Porto
Praça Coronel Pacheco n.º15, 4050-453 Porto
e-mail: smr@estsp.ipp.pt web: <http://www.estsp.pt>

2: Departamento de Engenharia Electrotécnica e de Computadores
Faculdade de Engenharia
Universidade do Porto
Rua Dr. Roberto Frias, s/n 4200-465 Porto
e-mail: dfreitas@fe.up.pt web: <http://www.fe.up.pt>

Palavras-chave: Imagem por Ressonância Magnética, Tracto Vocal, Articuladores, Vogais, Modelos Morfológicos

Resumo. *O entendimento da produção da fala tem sido amplamente procurado, recorrendo à imagem por ressonância magnética (IRM), mas não é totalmente conhecido, particularmente no que diz respeito aos sons do Português Europeu (PE). O principal objectivo deste estudo foi a caracterização das vogais do PE. Com base na IRM recolheram-se conjuntos de imagens bidimensionais, em cinco posições articatórias distintas, durante a produção sustentada do som. Após extracção de contornos do tracto vocal procedeu-se à reconstrução tridimensional, constatando-se que a IRM fornece informação morfológica útil e com considerável precisão acerca da posição e forma dos diferentes articuladores da fala.*

1. INTRODUÇÃO

A fala é uma “ferramenta” singular de comunicação, que envolve um conjunto de órgãos que trabalham sincronamente na produção dos sons – o tracto vocal [1]. Os órgãos vocais podem ser divididos, de acordo com a função que desempenham na produção da fala, em três partes: no sistema sub-laríngeo (pulmões, brônquios e traqueia); no sistema laríngeo (cordas vocais e glote); e no sistema supra-laríngeo (faringe, cavidades bucal e nasal).

Assim, a produção da fala pode ser caracterizada em três etapas distintas (de acordo com estas três partes anatómicas): na respiração (fonte de energia), na fonação (produção da voz, fonte de som) e na articulação (formação do som) [2]. As estruturas anatómicas utilizadas na articulação dos sons, designadas de articuladores, modificam a sua posição e incluem: os lábios (inferior e superior), os dentes incisivos superiores, a região alveolar, o palato duro, o véu do palato (palato mole) e a língua (Figura 1). A língua é um órgão muscular, volumoso, que desempenha um papel importante na mastigação, na gustação e deglutição, e que na fala, é sem dúvida fulcral, pois é o articulador com maior mobilidade e mais flexível. Estruturalmente, apresenta um ápex, um corpo, um dorso e uma raiz [3].

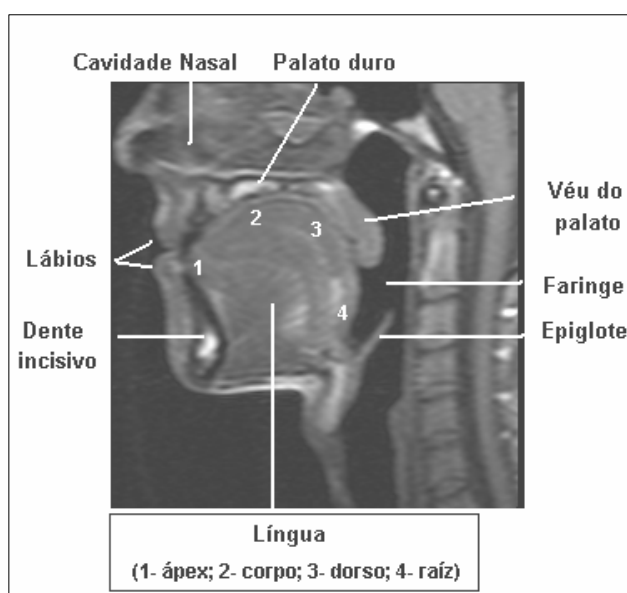


Figura 1. Articuladores envolvidos na produção da fala.

Na produção das vogais e das semivogais, os articuladores mantêm-se, suficientemente, afastados de modo que o fluxo de ar passe livremente e quase sem obstáculos. A principal diferença entre os diversos sons vocálicos (vogais) decorre da posição em que se encontram os lábios e a língua [1].

Por isso, a análise da articulação dos sons apoia-se basicamente na natureza cinestésica sobre as estruturas articulatórias que intervêm na sua produção, e que pode ser controlada pelo próprio falante. Esta análise está na origem da classificação tradicional dos sons, e ainda hoje é utilizada para os referir globalmente e integrar em grandes classes [4].

Neste sentido, a busca pelo entendimento da produção da fala tem sido amplamente estudada, com recurso a diferentes técnicas, mas ainda não totalmente conhecido, principalmente no que diz respeito ao Português Europeu (PE).

Dada a complexidade anatómica e fisiológica do tracto vocal, várias técnicas de análise (morfológicas e/ou funcionais) têm sido desenvolvidas, na tentativa de perceber todos os mecanismos da fala [5-7]. Apesar da similaridade dos aspectos anatómicos e fisiológicos entre os humanos, a fala é única e individual, e acima de tudo, particularmente complexa, pois o próprio indivíduo é incapaz de produzir um som várias vezes da mesma forma, dependendo de factores como a entoação, o timbre, e até mesmo culturais (dialectos da língua).

No mecanismo da produção da fala, o estudo dos articuladores constitui uma tarefa complicada, pois estes possuem um elevado grau de liberdade durante o processo, originando um problema no seu controlo e observação.

A imagem por ressonância magnética (IRM) pelo elevado detalhe anatómico que fornece, destaca-se de outras técnicas, possibilitando o estudo morfológico e dinâmico, bidimensional e tridimensional do tracto vocal, com utilidade quer para a Medicina, quer para a síntese da fala [8-14].

Assim, o principal objectivo deste estudo foi a caracterização morfológica das principais vogais do Português Europeu, culminado na construção de modelos tridimensionais representativos de cada som.

2. MATERIAL E MÉTODOS

Após revisão da literatura pode-se constatar que não existe uniformidade entre os diferentes estudos, quer no que diz respeito à técnica de IRM, quer às técnicas de processamento de imagem. No entanto, quanto à técnica de IRM é inequívoca a necessidade de utilização de sequências cada vez mais rápidas, de forma a possibilitar não só o estudo estático (aquisição de imagens durante a sustentação de um som) como o estudo dinâmico (aquisição de imagens durante a fala) minimizando assim o esforço dos sujeitos. Para tal, é crucial a utilização de equipamentos com um campo magnético de, pelo menos, 1,5T ou acima, mas cuja acesibilidade é limitada face ao número de equipamentos existentes. Quanto às técnicas de processamento de imagem, a sua utilização tem vindo a ser no sentido de tornar o processo de segmentação o mais automático possível (contribuindo para um menor tempo e esforço dos investigadores).

As condicionantes da investigação da fala são várias, quer por limitações do equipamento (técnica usada), ou da técnica de processamento, como também pelo facto de existir variabilidade na produção individual da fala e variabilidade inter-oradores; estes factores condicionaram, a maioria dos estudos revistos, ao recurso de amostras muito reduzidas.

2.1. Sujeitos e corpus da fala

A amostra deste estudo foi constituída por dois sujeitos jovens adultos (do sexo masculino e do sexo feminino), voluntários, saudáveis e sem história prévia de patologia orofacial. O corpus da fala consistiu num conjunto de cinco sons vocálicos do PE sustentados durante a aquisição das imagens.

2.2. Equipamento e procedimentos gerais

Para a recolha das imagens do tracto vocal foi utilizado um equipamento de ressonância magnética Siemens Magnetom Symphony ® com um campo magnético de 1,5T e, foi usada uma antena Phased Array de cabeça para recepção do sinal.

Os sujeitos foram posicionados em decúbito dorsal, com a cabeça devidamente imobilizada, e foram colocados auscultadores, de forma a facilitar também a intercomunicação com os mesmos, devido ao elevado ruído provocado pelo equipamento; este ruído inviabilizou a gravação do sinal acústico da fala.

2.3. Estudo estático

Neste estudo foram obtidos dois conjuntos de imagens que representam o volume em estudo – denominados de *stacks*, primeiro no plano sagital (três cortes contíguos com 5mm de espessura) e depois no plano coronal (quatro cortes espaçados 10mm entre si e com 6mm de espessura), durante 9s e 9,9s para cada som, respectivamente.

Antes do início de cada aquisição, o sujeito era informado acerca do som pretendido através do intercomunicador, tendo-se obtido imagens sagitais e coronais ponderadas em T1 (sequência Turbo Spin-eco) e com um *field of view* de 150mm.

De seguida, apresentam-se algumas das imagens obtidas neste estudo estático, nos dois planos de imagem, realçando-se também qual a sua relevância no estudo da produção da fala.

O plano médiosagital corresponde ao plano central (dividindo a cabeça em duas partes, a direita e esquerda) e, é sem dúvida, o plano mais informativo pois permite demonstrar toda a anatomia do tracto vocal, bem como a maioria dos articuladores. As imagens obtidas neste plano (Figura 2) demonstram claramente as diferenças na posição e forma dos lábios e da língua, para cada vogal.

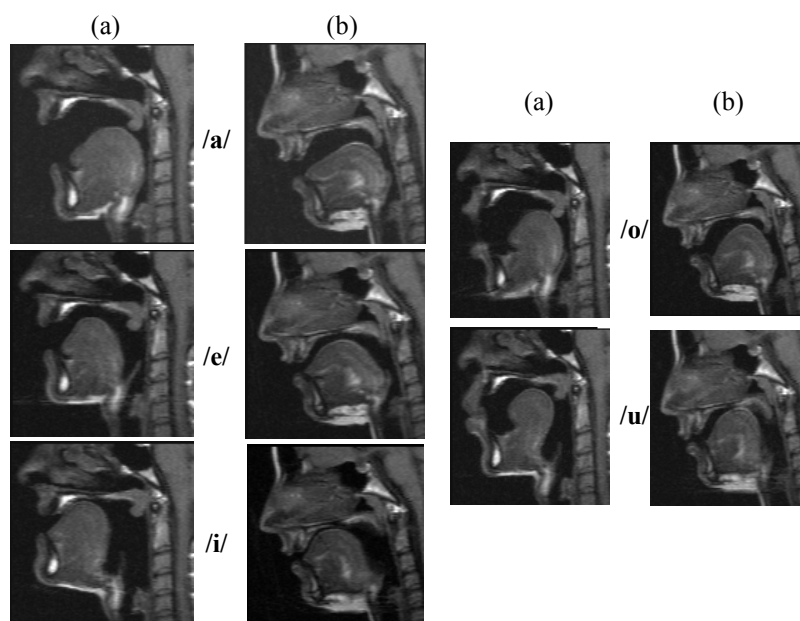


Figura 2. Imagens médiosagittais das vogais sustentadas por um sujeito do sexo masculino (a) e um do sexo feminino (b).

Por outro lado, estas imagens comparativas ilustram bem as diferenças anatómicas e posturais entre os sujeitos dos dois sexos, nomeadamente quanto ao comprimento (extensão) do tracto vocal. Em termos morfológicos nos dois sujeitos, e para as vogais sustentadas, os movimentos dos articuladores são similares, nomeadamente da língua.

O plano coronal (ou plano frontal) permite obter informação específica quanto às dimensões laterais do tracto vocal, e é particularmente útil no estudo dos lábios e da língua (relação com o palato). As imagens obtidas neste plano permitem observar diferenças substanciais na forma dos lábios (arredondada, estreita, aberta), assim como na relação entre a língua e o palato, e posição da língua.

Por exemplo, comparando a vogal /i/ com a vogal /o/, na figura 3, pode-se observar que: os lábios estão arredondados na vogal /o/ enquanto que na vogal /i/ a abertura da boca é estreita; a língua está em posição mais recuada na vogal /o/; e na vogal /i/ a língua está mais próxima do palato.

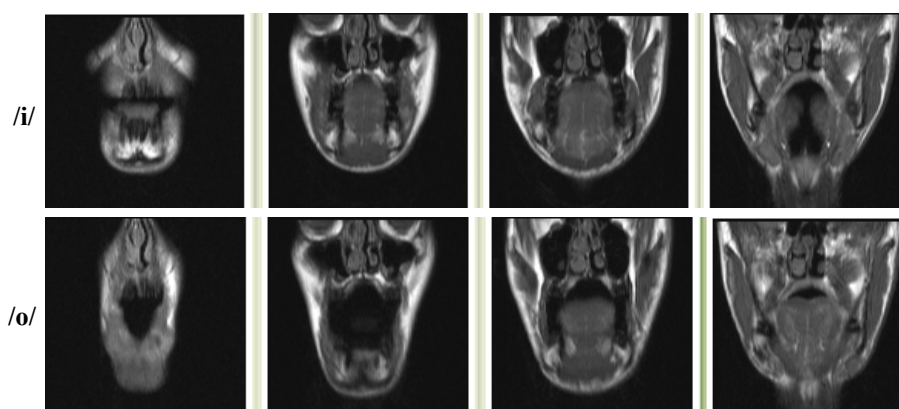


Figura 3. Imagens coronais das vogais /i/ e /o/ do Português Europeu.

2.4. Interpretação e processamento das imagens

As imagens médicas das diferentes modalidades em Radiologia são apresentadas no formato DICOM, que é o *standard* global de comunicação mais usado actualmente. Este *standard* de transmissão de imagens digitais possibilita a recuperação de imagens e de toda a informação associada (metadados), assim como a partilha de informação entre equipamentos.

A análise da geometria do tracto vocal requer o recurso a técnicas de processamento de imagem ajustadas à finalidade de cada estudo, de forma a permitir a delimitação das diferentes cavidades deste tubo de passagem do ar dos restantes tecidos que o circundam. Para tal, é necessário realizar duas tarefas fundamentais: (1) a localização e segmentação da região de interesse (ROI), e (2) a reconstrução tridimensional desse objecto; a primeira etapa é determinante para o sucesso da segunda.

Apesar do menor ou maior automatismo na realização das tarefas de segmentação, este processo continua a ser complexo, dados os vários problemas que ainda se verificam, relacionados por um lado com a técnica de IRM, e por outro com as características anatómicas do tracto vocal:

- Ainda se desconhece qual a acção da epiglote e dos seios piriformes na produção da fala, existindo, por isso, dúvidas quanto à importância da sua delimitação e inclusão na segmentação do tracto vocal;
- A não identificação dos dentes em IRM;
- O término do tracto vocal nos lábios ainda não está bem definido;
- Existem estruturas, que pela similaridade de intensidade de sinal em IRM com o tracto vocal (como os discos intervertebrais da coluna, a cavidade nasal, a porção óssea cortical do palato, e o ar circundante na face e pescoço) dificultam a correcta delimitação no caso de utilização de técnicas semi-automáticas ou automáticas, devido à *fuga* desses contornos;
- O processo de localização do tracto vocal é também complexo e exigente (pela necessidade de conhecimentos anatómicos) consoante o plano em estudo,

principalmente nos cortes axiais e coronais, pela dificuldade na identificação dos dentes e alvéolos e, na interpretação das próprias imagens.

A análise das imagens consistiu em duas etapas distintas, primeiro, na segmentação das imagens usando o *plugin* 3D editing do programa de processamento de imagem Image Jⁱ, e posteriormente, na sua reconstrução tridimensional. A representação gráfica e combinação dos modelos tridimensionais de cada *stack* foi possível graças a um software para criação de gráficos e animações com aplicações interactivas 3D, denominado Blenderⁱⁱ.

Neste estudo em particular, optou-se pelo método de *threshold*, por vários motivos: (1) pelo automatismo (delimitação é quase automática, dado o limiar da ROI); (2) pela sua flexibilidade e adaptabilidade (contorno facilmente ajustável, e perfeitamente adaptável à morfologia do tracto vocal); e (3) pela menor dependência do investigador, comparativamente aos métodos manuais.

Assim, a segmentação das imagens do presente estudo (Figura 4), foi efectuada de acordo com os seguintes procedimentos:

- (1) Conversão dos cortes de cada plano em *stacks*, para cada som;
- (2) Identificação da região a segmentar e adição manual de objectos opacos para permitir o fecho da mesma, em cada um dos cortes;
- (3) Adição da imagem dos dentes apenas nas imagens das diferentes *stacks* sagitais, obtida pela extracção dos contornos dos mesmos de um estudo anatómico feito inicialmente (em posição de repouso);
- (4) Delimitação dos contornos (fronteira ar-tecidos moles), de cada imagem bidimensional, para cada um dos sons, usando a técnica semi-automática de *threshold*, através da ferramenta SegmentingAssistant.

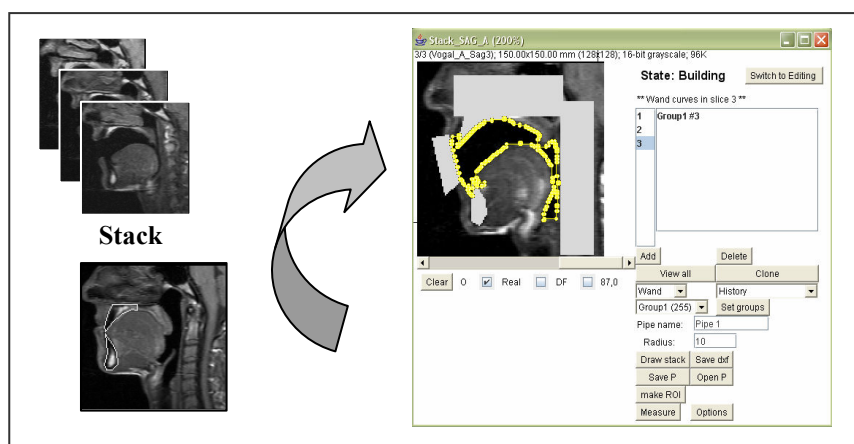


Figura 4. Passos efectuados para a segmentação das imagens.

ⁱ Desenvolvido no National Institutes of Health. Disponível em: <http://rsb.info.nih.gov/ij>.

ⁱⁱ Disponível em: <http://www.blender.org/cms/Home.2.0.htm>.

O conjunto de contornos bidimensionais – *outlines*, obtidos pela delimitação em cada imagem, para cada som, definidos com base no limiar do histograma na região em estudo, permitiu a construção dos modelos tridimensionais (através da união das *outlines*).

Graças ao software utilizado na segmentação das imagens foi possível aceder à informação DICOM que contém todos os metadados referentes à gestão da imagem, impressão, transferência e arquivo (como os parâmetros de aquisição, características de imagem, e outros objectos relativos à orientação do paciente) e, também informação das *stacks*. Após esta interpretação foi possível o alinhamento dos modelos tridimensionais num sistema de coordenadas comum – aqui denominado registo – usando um processo, o mais automático possível, e cuja localização espacial correspondesse de facto à sua posição de origem no sujeito.

As imagens que se seguem (Figura 5) representam duas perspectivas diferentes do modelo tridimensional obtido para a vogal /u/, após importação das *outlines* para o espaço gráfico do software Blender. A *skin* (objecto 3D) representada a azul (5a) foi gerada pela união das três *outlines* sagitais extraídas das imagens bidimensionais nesse plano; a *skin* representada em tons de vermelho (5b) resultou da união das quatro *outlines* extraídas da *stack* coronal.

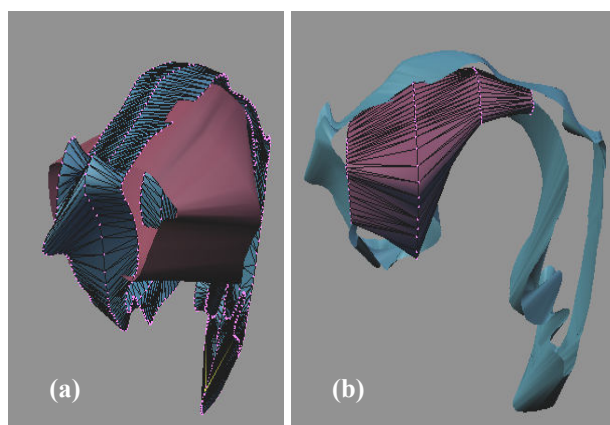


Figura 5. Modelo tridimensional da vogal /u/.

3. RESULTADOS

Os modelos tridimensionais que a seguir se apresentam (Figura 6) demonstram a posição e forma dos diferentes articuladores envolvidos na produção das diferentes vogais, em diferentes perspectivas. Nestes modelos é possível constatar a importância quer da informação sagital, quer da informação coronal para uma melhor compreensão da morfologia do tracto vocal durante produção da fala. As modificações posicionais dos diferentes articuladores são particularmente bem demonstradas no plano sagital, enquanto que a informação coronal possibilita uma melhor observação das dimensões laterais do tracto vocal, e em particular da cavidade oral.

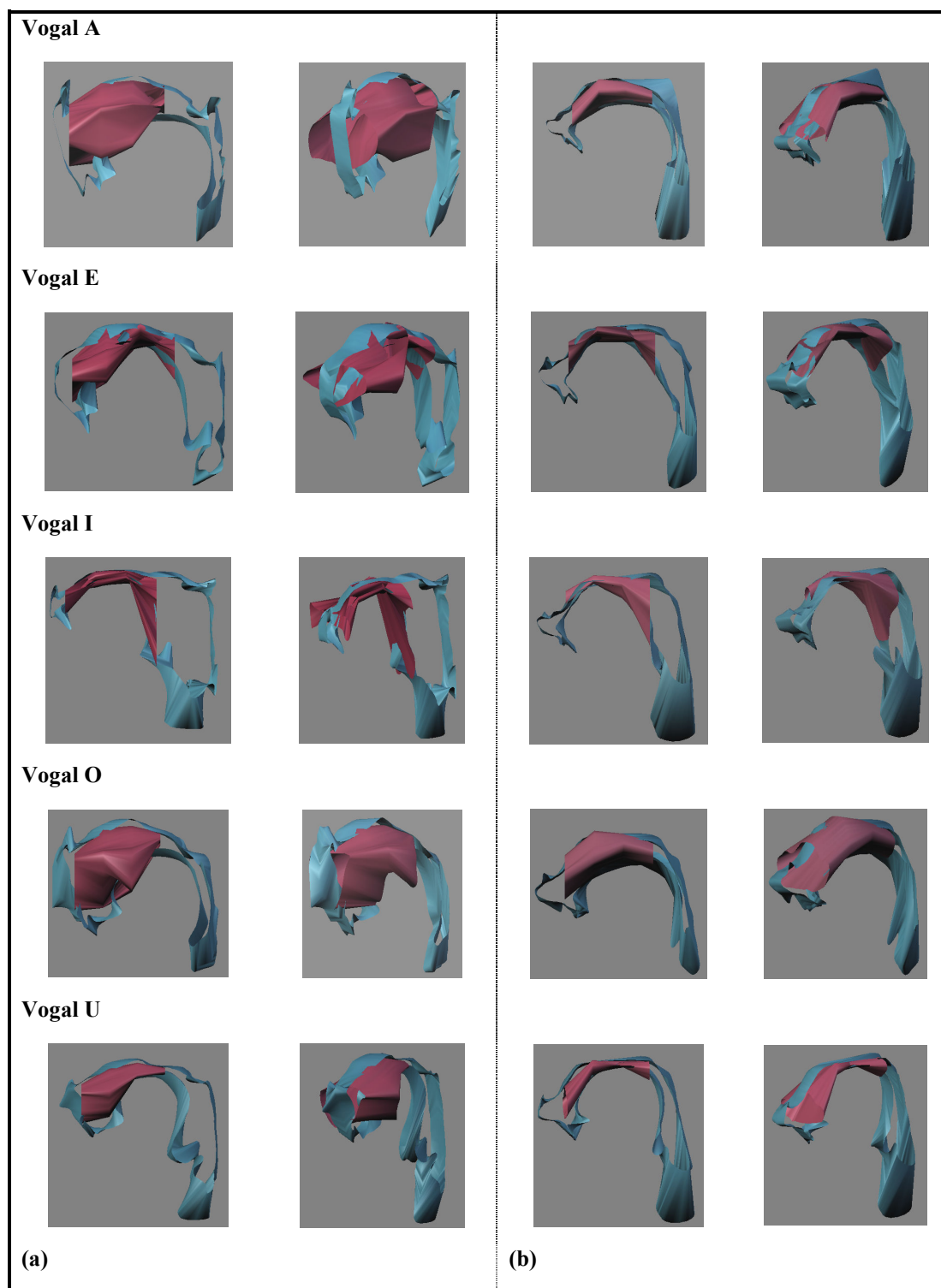


Figura 6. Modelos tridimensionais das vogais do PE produzidas por um sujeito do sexo masculino (a) e do sexo feminino (b).

4. CONCLUSÃO E PERSPECTIVAS FUTURAS

O estudo da produção da fala, pela complexidade anatómica e funcional do tracto vocal, tem despertado um enorme interesse multidisciplinar. A utilização e evolução da IRM tem contribuído fortemente para o seu grande desenvolvimento. Apesar das vantagens inequívocas da IRM comparativamente a outras técnicas, persistem algumas limitações fortes face às necessidades já conhecidas, relacionadas em parte com a pouca acessibilidade destes equipamentos, e com as limitações ainda existentes de rapidez dos mesmos e com a inexistência de protocolos específicos para as recolhas.

No entanto, verificaram-se algumas dificuldades, que influenciaram de forma não desprezável os resultados obtidos, nomeadamente:

- (1) Dificuldades na delimitação dos contornos, em parte devido à não identificação dos dentes em IRM, informação esta que é particularmente importante nos cortes coronais;
- (2) Limitações no número de imagens obtidas, condicionado quer pelo tempo de aquisição, quer pelo compromisso entre tempo e resolução do equipamento, e também quer no sentido de minimizar o esforço e variabilidade da articulação dos sujeitos durante a sustentação dos sons;
- (3) Dificuldades no registo geométrico das *stacks*, e cuja interpretação segundo a norma DICOM ainda não foi totalmente concretizada;
- (4) A impossibilidade de registo acústico da fala, pelo elevado ruído provocado pelo equipamento durante as aquisições.

Por aperfeiçoar está ainda o moroso e difícil processo de segmentação das imagens, apesar do método de segmentação *threshold* ser quase automático. Os modelos tridimensionais criados ainda não estão totalmente fechados lateralmente por inexistência dessa possibilidade nas aplicações de processamento conhecidas, pelo que, em investigações futuras será necessário completar estes modelos híbridos, de forma a permitir a construção de um modelo anatómico 3D digital do tracto vocal que permita, entre muitas outras coisas extrair informação completa da superfície delimitadora e da sua dinâmica. Com a finalização destes modelos tridimensionais, será possível também o cálculo de funções de área e dos parâmetros articulatorios para síntese da fala.

Para o futuro perspectiva-se, por um lado, a implementação destes estudos na prática clínica para o diagnóstico e orientação clínica na avaliação de doentes com *deficits* articulatorios, e por outro, a possibilidade de utilização destes dados para o ensino de segundas línguas, neste caso do PE. Esta base de dados poderá também, num futuro próximo, servir como uma ferramenta auxiliar quer para a formação dos Terapeutas da Fala, como para a utilização da mesma por estes profissionais na avaliação de perturbações da fala.

REFERÊNCIAS

- [1] I.H. Faria, E.R. Pedro, I. Duarte, C.A.M. Gouveia. *Introdução à Linguística Geral e Portuguesa*, Lisboa, Caminho, (1996).
- [2] M.R. Martins. *Ouvir falar – Introdução à Fonética do Português*, 3ª Edição, Lisboa, Editorial Caminho, (1988).

- [3] E. Gardner, D.J. Gray, R. O’Rahilly. *Anatomia*, Edição, Rio de Janeiro, Guanabara Koogan, (1988).
- [4] M.H. Mateus, A. Andrade, M.C Viana, A. Villalva. *Fonética, Fonologia e Morfologia do Português*, Lisboa, Universidade Aberta, (1990).
- [5] A.A. Wrench, A.D. McIntosh, W.J. Hardcastle, “Optopalatograph (OPG): a new apparatus for speech production analysis”, *Conference on Spoken Language Processing in Philadelphia*. Vol. **3**, pp. 1589-1592, (1996).
- [6] P. Hoole, N. Nguyen, *Electromagnetic articulography in coarticulation research*. W.H Hardcastle and N. Hewlett eds. *Coarticulation: Theory, Data and Techniques*, 1999, Cambridge University Press., (1999), pp. 260-269.
- [7] T. Kenneth, T. Ingo, E. Hoffman, B. Story, “Three-dimensional vocal tract imaging and formant structure: Varying vocal register, pitch, and loudness”, *The Journal of the Acoustical Society of America*. Vol. **109**, Issue 2, pp. 742-747 (2001).
- [8] B.J. Kröger, R. Winkler, C. Mooshammer, B. Pompino-Marschall, *Estimation of Vocal Tract Area Function from Magnetic Resonance Imaging: Preliminary Results. Proceed. 5th Seminar On Speech Production: Models And Data*, München, 2000, Germany (2000), pp. 333-336.
- [9] P. Badin, L. Pouchoy, G. Bailly, M. Raybaudi, C. Segebarth, J.F. Lebas, M.K. Tiede, E. Vatikiotis-Bateson, Y. Tohkura, *Un modèle articulatoire tridimensionnel du conduit vocal basé sur des données IRM. Actes des 22èmes Journées d'Etude sur la Parole*, Martigny, 1998, Suíça (1998), pp. 283-286.
- [10] O. Engwall, *A revisit to the Application of MRI to the Analysis of Speech Production - Testing our assumptions. Proceed. 6th Int. Seminar on Speech Production*, Sydney (2003).
- [11] M.S. Avila-García, J.N. Carter, R.I. Damper, *Extracting Tongue Shape Dynamics from Magnetic Resonance Image Sequences. Transactions on Engineering, Computing and Technology V2*, (2004), p. 288-291.
- [12] S. Narayanan, K. Nayak, S. Lee, A. Sethy, D. Byrd, “An Approach to Real-time Magnetic Resonance Imaging for Speech Production”. *Journal Acoustical Society of America*. **115**(4), pp. 1771-76 (2004).
- [13] Serrurier, A., Badin, P. A Three-dimensional Linear Articulatory Model of Velum based on MRI data. *InterSpeech*, Lisbon. 2005, pp. 2161-2163.
- [14] BADIN, P., SERRURIER, A. Three-dimensional Modeling of Speech Organs: Articulatory Data and Models, IEICE Technical Committee on Speech, Kanazawa, Japan. 2006, pp (In print).