A HARMONIA DOS OBJECTOS DESVOZEADOS

FERNANDO MARTINS
(Faculdade de Letras da Universidade de Lisboa)

No processo de reconhecimento dos segmentos fonéticos, um dos principais problemas diz respeito ao desvozeamento que frequentemente afecta a sequência acústica, sobretudo em situações de fala expontânea. Na presente comunicação propõe-se um método que pretende reconstituir o vozeamento de segmentos que são produzidos como desvozeados.

1. O reconhecimento

O método usado no reconhecimento fonético de fala presente no sistema AuditPT (Martins, 1995) baseia-se nas informações linguísticas incorporadas na base de dados e que permitem a atribuição de unidades linguísticas a sequências acústicas de fala (Cooke e Green, 1988; Green et al, 1987; Haton, 1985; Haton, 1988)). O processo de reconstituição das unidades fonéticas efectua-se pela identificação dos respectivos correlatos acústicos. Os dados analisados neste trabalho foram retirados de um 'corpus' constituído por 4560 palavras.

2. O Vozeamento

O vozeamento é detectado fundamentalmente em dois níveis, através da análise directa do sinal acústico e da análise espectral. Esta necessidade de pelo menos dois níveis justifica-se porque a falha na detecção do vozeamento implica que o sistema considere como não vozeados segmentos que são vozeados. Deste modo, e dado que cada informação extraída do sinal é codificada como uma probabilidade, torna-se necessário aumentar as probabilidades pela extracção da informação mais do que uma vez, sobretudo quando estão em causa informações pertinentes. Se a informação fosse considerada redundante pelo sistema, esta exigência não existiria.

Em português, os pares vozeado / não vozeado são os seguinte [p/b] (pala, bala), [t/d] (tolo, dolo) , [k/g] (calo, galo), [f/v] (faca, vaca), [s/z] (cela, zela) e [ʃ/ʒ] (chá, já). Os restantes segmentos são vozeados, o que implica não ser necessária uma confirmação do vozeamento, dado que não existem os segmentos não vozeados correspondentes.

Os dois níveis referidos anteriormente estão relacionados com uma fase de pré-processamento e uma fase de pós-processamento. No primeiro, é extraída a Frequência Fundamental (Fo) através de um algoritmo de correlação. Apesar de os resultados serem apresentados sob a forma de valores em Hz, apenas são considerados os dados relativos à presença ou não de Fo, directamente relacionada com a presença ou ausência de vozeamento. No segundo nível, a fase do pós-processamento, o sinal acústico corresponde a um espaço de representação espectral, onde é possível obter os dados sobre a distribuição espectral das frequências, depois de aplicada a transformada de Fourier. A presença ou ausência de harmónicas é um factor determinante para a detecção de segmentos vozeados ou não vozeados, respectivamente.

Como se pode verificar na figura 1, estão presentes os 2 níveis referidos: Fo relativo ao nível 1 e as harmónicas relativas ao nível 2. Em ambos os casos, é possível identificar todos os quatro segmentos da palavra 'bobo' como sendo vozeados, quer porque Fo é representado por uma linha contínua, sem interrupções, quer porque as harmónicas não manifestam descontinuidade.

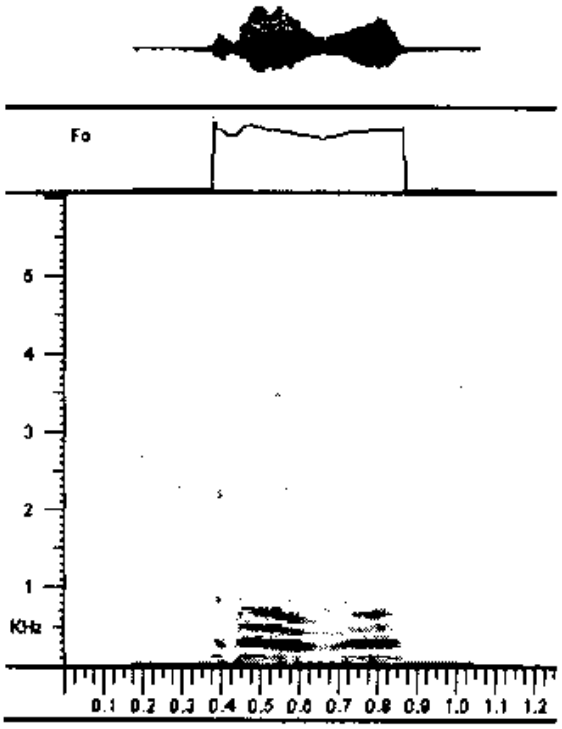

Figura 1 Distribuição harmónica na palavra 'bobo'

Na figura 2, a palavra representada é 'cessar', composta por dois segmentos não vozeados (em ambos os casos [s]).

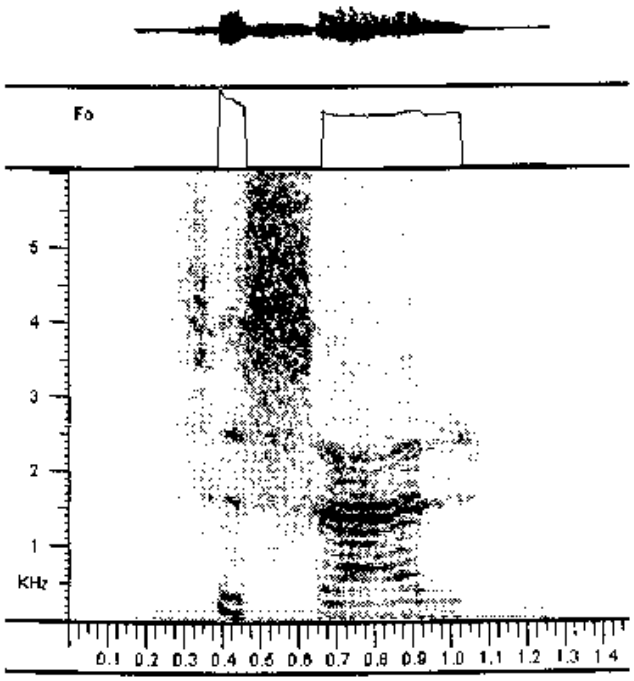

Figura 2 Distribuição harmónica na palavra 'cessar'

Como se pode comprovar, o sistema não detecta a presença de vozeamento nos dois casos, pela inexistência quer de Fo quer de harmónicas.

3. O Desvozeamento

Vejamos agora alguns casos em que durante a produção das sequências fonéticas, ocorreram desvozeamentos de segmentos. É o caso apresentado na figura 3, durante a produção da palavra 'dedo'. Os quatro segmentos desta palavra são (ou devem ser) vozeados. No entanto, o segmento intervocálico é detectado como não vozeado, quer no nível de Fo, quer no nível da análise harmónica.

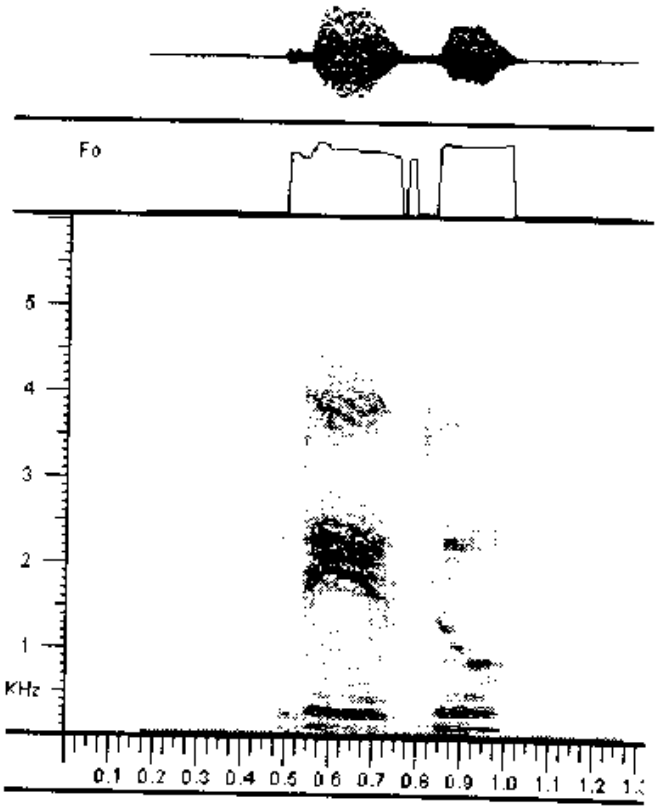

Figura 3 Distribuição harmónica na palavra 'dedo'

Foram analisados 220 casos de desvozeamento presentes no 'corpus', o que na prática corresponde a 4% de palavras não reconhecidas (de um total de 4560) por este motivo. Todos os casos foram sujeitos a um teste de percepção com dez informantes. Em 100% dos casos, foram reconhecidos os segmentos como sendo vozeados. Muitos destes casos podem ser explicados pelo facto de não existirem as palavras em que fosse possível constituir um par mínimo pela substituição do segmento vozeado. Por exemplo não existe 'deto', por oposição a 'dedo'. No entanto, muitas das palavras podem ser colocadas em oposição, como por exemplo 'casa' e 'caça'. Como se disse, mesmo nestes casos, os segmentos foram reconhecidos como vozeados.

Como se afirmou anteriormente, o sistema de reconhecimento extrai informações do sinal e atribui probabilidades a essas informações. No caso do vozeamento, se existir uma probabilidade de 90% de um segmento ser vozeado devido a Fo extraído no nível 1 e uma probabilidade de 92% de ser vozeado devido à presença de harmónicas, então o segmento é considerado vozeado. Mas pode acontecer, como o caso apresentado na figura 4, que as probabilidades não sejam semelhantes. Na palavra 'jogo', o /g/ intervocálico é detectado como não vozeado por Fo e como vozeado pela presença de harmónicas, ainda que estas se apresentem bastante enfraquecidas. Nestes casos, a probabilidade referente às harmónicas sobrepõe-se à probabilidade referente a Fo.

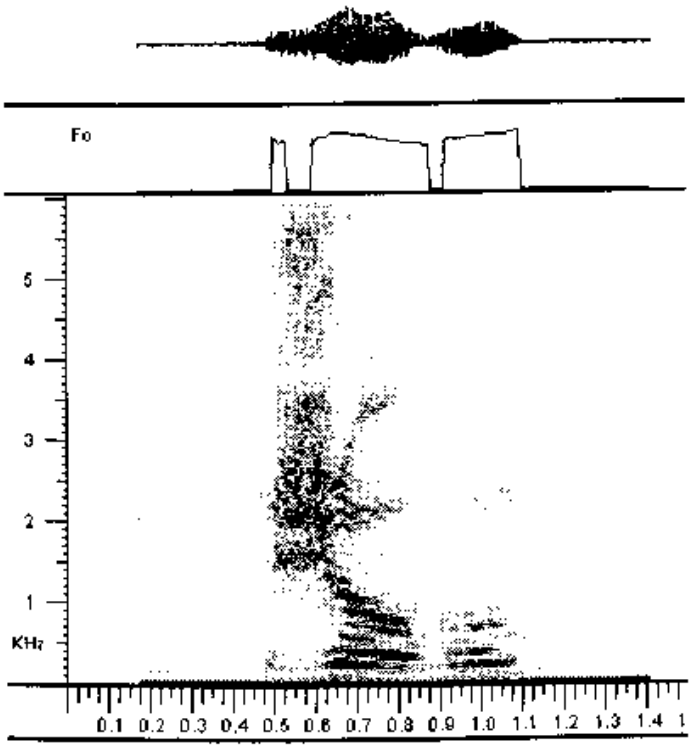

Figura 4 Distribuição harmónica na palavra 'jogo'

A figura 5 confirma esta mesma situação: na palavra 'zeloso', o /z/ intervocálico apenas é detectado como vozeado pela análise harmónica.

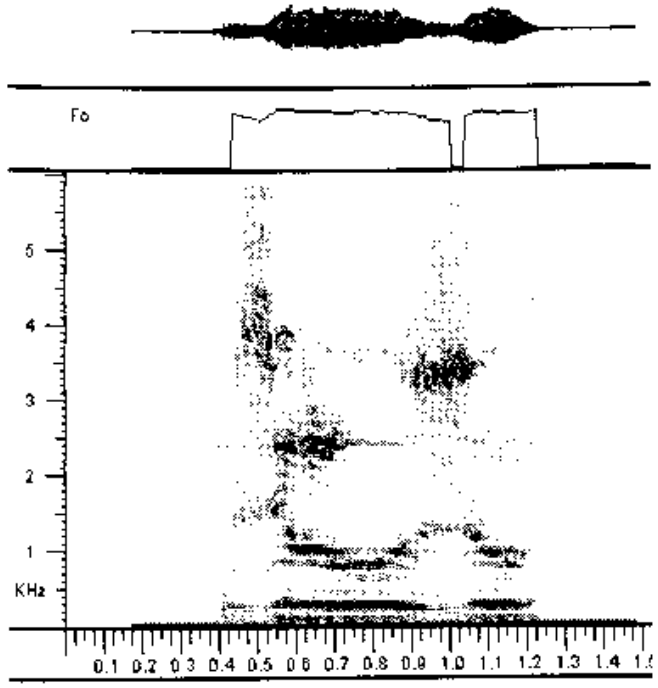

Figura 5 Distribuição harmónica na palavra 'zeloso'

Na grande maioria dos casos de desvozeamento (75%), no entanto, em ambos os níveis de verifica a probabilidade de os segmentos serem não vozeados. O teste de percepção referido anteriormente, por oposição, demonstra que os segmentos são vozeados. Este facto demonstra que as características inerentes ao vozeamento devem ser encontradas a partir de outras informações acústicas, nomeadamente pela análise das fronteiras entre segmentos. Esta necessidade justifica-se porque também os correlatos acústicos do ponto de articulação podem ser detectados no início ou final dos segmentos vizinhos.

Para confirmar esta posição, efectuou-se outro teste de percepção tendo como base palavras sintetizadas em que são eliminadas totalmente as sequências correspondentes a segmentos vozeados. Compare-se a figura 6 com a figura 1: a palavra é exactamente a mesma, a diferença reside na eliminação do /b/ intervocálico. Efectuado o teste de percepção com fez informantes, o resultado de 100% de reconhecimento do segmento eliminado demonstra que a identificação do mesmo é possível a partir das propriedades presentes nos segmentos vizinhos. Esta conclusão tem outras consequências sobre a identificação dos segmentos, nomeadamente sobre os correlatos acústicos do modo e ponto de articulação, mas vamos tratar apenas da questão relativa ao vozeamento.

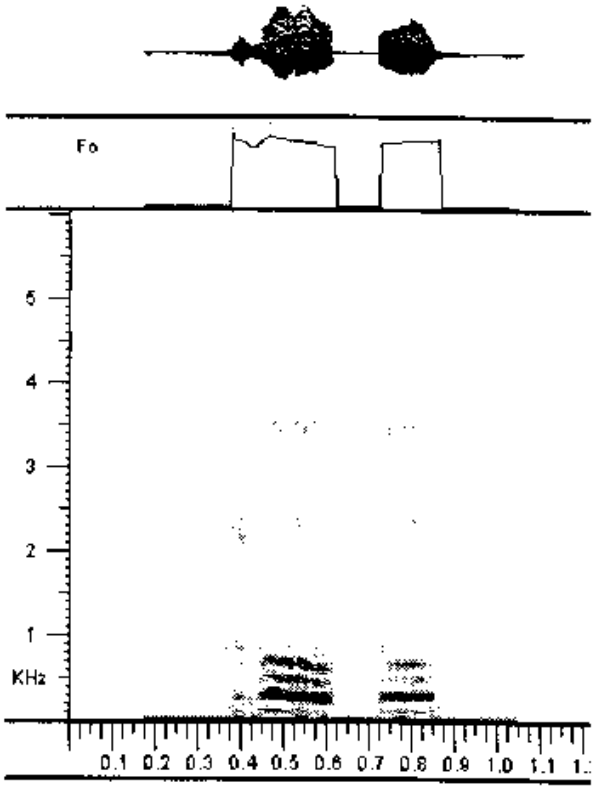

Figura 6 Distribuição harmónica na palavra sintetizada 'bobo'

4. Os objectos harmónicos

As harmónicas acompanham a subida ou descida de Fo, na medida em que são múltiplas desta. Por efeitos de coarticulação e à semelhança do que acontece com os correlatos acústicos do ponto e modo de articulação, a influência está sobretudo presente no início ou final das transições entre segmentos.

No conjunto das 4560 palavras, foram analisadas as harmónicas de 2554 segmentos que fazem parte do par vozeado / não vozeado (por exemplo, p/b, f/v, etc). A análise centrou-se na transição entre segmentos, nomeadamente entre os segmentos alvo e os segmentos vizinhos no contexto.

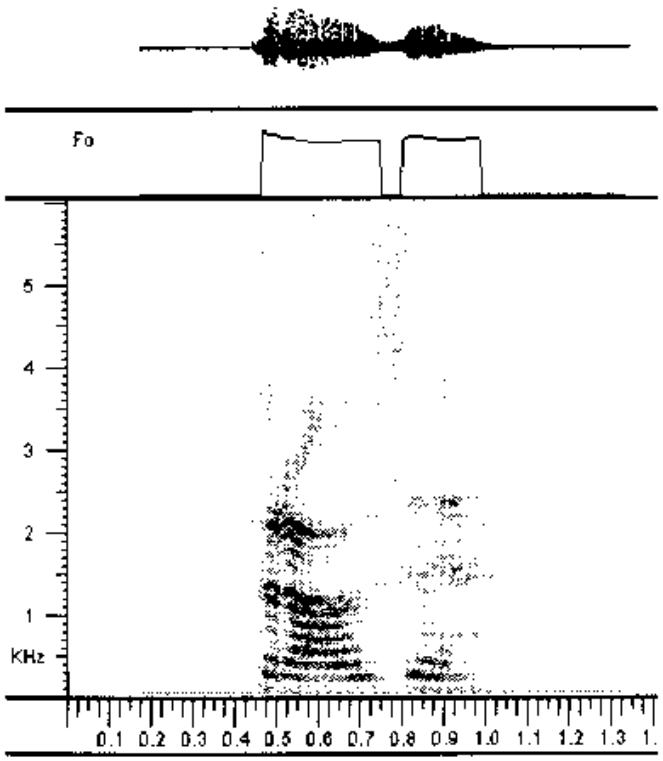

Figura 7 Distribuição harmónica na palavra 'prosa'

As transições das harmónicas foram etiquetadas como A (ascendentes), D (descendentes) e C (contínuas) em função da respectiva trajectória nos primeiros 30 µs no caso do segmento seguinte e nos últimos 30 µs no caso do segmento antecedente. As leituras foram efectuadas nas harmónicas mais baixas. Por exemplo, na figura 7, no segmento que segue o /p/ inicial a transição D, enquanto que relativamente ao /z/, a transição do segmento antecedente é C e do segmento seguinte também.

No total dos casos analisados, os resultados foram os seguintes:

Segmento	Antes		Depois	
/p/	v	D=80	A=5	C=10
/b/	v	D=2	A=89	C=9
/t/	v	D=83	A=4	C=13
/d/	v	D=4	A=88	C=8
/k/	v	D=91	A=3	C=6
/g/	v	D=3	A=92	C=5
/f/	v	D=80	A=4	C=16
/v/	v	D=5	A=79	C=16
/s/	v	D=83	A=3	C=14
/z/	v	D=6	A=82	C=12
/ʃ/	v	D=87	A=3	C=10
/ʒ/	v	D=4	A=91	C=5

No quadro dos resultados *v* representa valores variáveis que se distribuem por A D C sem regularidade. Por essa razão, os respectivos valores não foram apresentados, dado que não são pertinentes. Os valores numéricos representam as percentagens de ocorrência.

Os resultados apontam claramente no sentido de se considerar como pertinente na análise das trajectórias o contexto seguinte e não o antecedente. Assim, se considerarmos o contexto seguinte, podemos afirmar que quando o segmento em análise é fonologicamente não vozeado a trajectória para o segmento seguinte é claramente D (84%): quando é fonologicamente vozeado, a trajectória para o segmento seguinte é A (87%).

5. Conclusão

No conjunto das 220 palavras não reconhecidas (4%) pelo sistema AuditPT, a introdução de informações relativas à trajectória das harmónicas teve como consequência um aumento na taxa de sucesso do sistema. Dos segmentos não identificados inicialmente, devido ao desvozeamento, 86% foram reconhecidos correctamente depois de introduzidas as regras correspondentes às trajectórias analisadas. Demonstrou-se, assim, a importância de estender a análise de um segmento aos segmentos vizinhos e a importância que as harmónicas desempenham na determinação do vozeamento ou não vozeamento, particularmente através das transições

Bibliografia

COOKE, M.P. e GREEN, P.D. (1988), On Finding Objects in Spectrograms: A Multiscale Relaxation Labelling Approach. In: Niemann, H.; Lang, M. e Sagerer, G. (ed). *Recent*

Advances in Speech Understanding and Dialog Systems. NATO ASI Series F. Vol. 46, Berlin: 129-133.

GREEN, P.D.; COOKE, M.P.; LAFFERTY, H.H. e SIMONS, A.J.H. (1987), A Speech Recognition Strategy Based on Making Acoustic Evidence and Phonetic Knowledge Explicit. *Eurospeech*, Edinburgo: 373-375.

HATON, J.P. (1985), Knowledge-Based and Expert Systems in Automatic Speech Recognition. In: De Mori, R. e Suen C.Y. (ed.). *New Systems and Architectures for Automatic Speech Recognition and Synthesis*. NATO ASI Series F, Vol. 16, Berlin: 249-269.

HATON, J.P. (1988), Knowledge-Based Approaches in Acoustic-Phonetic Decoding of Speech. In: Niemann, H.; Lang, M. e Sagerer, G. (ed). *Recent Advances in Speech Understanding and Dialog Systems*. NATO ASI Series F, Vol. 46, Berlin: 51-69.

MARTINS, Fernando (1995), *Modelo de Reconhecimento de Fala para a Língua Portuguesa - As Invariantes Fonéticas e a Programação por Objectos*, dissertação de doutoramento, Faculdade de Letras da Universidade de Lisboa.