ANOTAÇÃO FONÉTICA AUTOMÁTICA DE CORPORA DE FALA TRANSCRITOS ORTOGRAFICAMENTE

Rui Amaral, Pedro Carvalho, Diamantino Caseiro, Isabel Trancoso, Luís Oliveira

IST, Instituto Superior Técnico
INESC- Instituto de Engenharia de Sistemas e Computadores
AV. ALVES REDOL, 9, 1000 - LISBOA
email:{ramaral,pmmc,dcaseiro}@speech.inesc.pt,
{Isabel.Trancoso,Luis.Oliveira}@inesc.pt

Resumo. Descreve-se neste artigo um sistema de anotação fonética automática de *corpora* de fala transcritos ortograficamente. Uma característica deste sistema é a de não estar limitado a uma única transcrição fonética, podendo utilizar um conjunto de transcrições alternativas geradas por regras, num processo de descodificação que irá escolher aquela que mais se adequa ao sinal de fala. Testes efectuados sobre o *corpus* português do EUROM_1 mostraram que a utilização de conhecimento linguístico sobre a forma de regras com vista à obtenção de transcrições alternativas, conduzem a uma maior aproximação à transcrição manual.

1 Introdução

Durante as ultimas décadas, tem-se assistido a uma considerável mobilização de esforços em laboratórios de todo o mundo no sentido de desenvolver produtos de tecnologia da fala. Uma das motivações tem sido a de desenvolver sistemas capazes de estabelecer um interface Homem-Máquina mais natural. A emergência desta tecnologia e a crescente procura de soluções baseadas nestes produtos, em geral suportados por métodos estatísticos, motiva a recolha e anotação de grandes *corpora* de fala. O carácter repetitivo destas tarefas e as dimensões pretendidas para os *corpora* foram factores de "peso" que evidenciaram a necessidade de se recorrer à automatização, pelo menos parcial, destes processos. À anotação manual, tradicionalmente feita por peritos em fonética acústica, estão associados a quatro grandes inconvenientes: é um processo muito moroso, carece de procedimentos standard amplamente aceites e, finalmente, está sujeita a erros humanos e à falta de coerência na aplicação dos critérios entre anotadores. A consciencialização destes problemas sublinhou a necessidade de recorrer a sistemas que executassem de uma forma automática o processo de anotação. A automatização deste procedimento torna possível a anotação de grandes quantidades de material de fala através da aplicação de um conjunto fixo de critérios objectivos, sem intervenção humana. Claro está, que os resultados da anotação automática não têm a precisão de uma anotação manual; todavia, uma vez que os erros são de natureza sistemática e os critérios são explícitos, poderá ser possível a sua correcção.

No estado actual da tecnologia, a arquitectura dos sistemas de anotação automática que reúne maior consenso é aquela que se baseia nos modelos de Markov não observáveis (HMM - *Hidden Markov Models*) e em técnicas que visam melhorar o seu desempenho. A vantagem desta abordagem explica-se pelo facto de permitir resolver simultaneamente os processos de segmentação e etiquetagem.

O sistema de anotação descrito ao longo deste artigo é fiel a esta vertente e resultou de um esforço consertado entre engenheiros e linguistas. Só assim foi possível contar com um pequeno *corpus* de fala anotado foneticamente, muito útil ao processo de criação e treino dos modelos HMM e imprescindível ao processo de avaliação do desempenho do anotador automático. Este trabalho pretende ser uma contribuição para a automatização do processo de anotação fonética de *corpora* de fala, transcritos ortograficamente, para a língua Portuguesa .

Este artigo encontra-se estruturado da seguinte forma: na secção 2 descreve-se a arquitectura do sistema de anotação automática. O *corpora* de treino e teste utilizado no desenvolvimento e na avaliação dos resultados do anotador é descrito na secção 3, sendo o processo de avaliação descrito na secção 4. Após uma breve discussão dos resultados obtidos, na secção 5 apresentam-se algumas conclusões e discutem-se procedimentos que a serem implementados iram permitir melhoramentos no desempenho do sistema.

2 Arquitectura do Sistema

A arquitectura do sistema automático de anotação descrito neste artigo encontra-se ilustrada na figura seguinte:

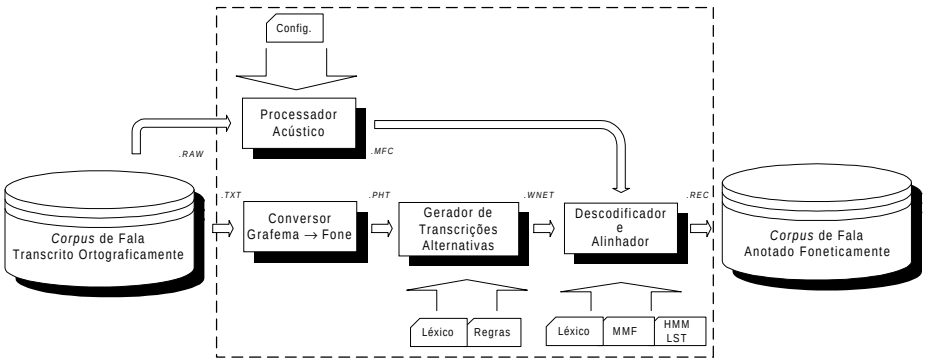

Fig. 1. Arquitectura do sistema de anotação

Este sistema é constituído por quatro módulos, que passaremos a descrever:

2.1 Módulo 1: Conversor de grafema para fone (G2P)

A normalização do texto bem como a conversão grafema-fone é feita usando os módulos correspondentes do sistema de síntese de fala para português, DIXI [6], [7].

Dos vários procedimentos de análise de texto que este sistema possui utilizaram-se apenas os seguintes:

- Normalização de texto: tendo em consideração símbolos especiais, numerais, abreviaturas e acrónimos.
- Processamento das vogais acentuadas e cedilhas: utilização de um formato interno normalizado para representar tanto as vogais acentuadas como as letras cedilhadas.
- Transcrição fonética: transcrição fonética larga feita ao nível da palavra com a ajuda de um pequeno dicionário interno e de um conjunto de 200 regras.

Designou-se por transcrição canónica a transcrição fonética larga produzida por este módulo.

2.2 Módulo 2: Gerador de transcrições alternativas (GTA)

A transcrição obtida no módulo anterior corresponde a uma transcrição fonética larga e pode, portanto, não corresponder exactamente ao que foi dito pelo falante. Por exemplo a frase "vou para a praia" é transcrita pelo módulo 1 em [v"o p6r6 6 pr"aj6] utilizando o alfabeto fonético SAMPA[1] . No entanto a sua realização mais frequente é [v"o pr"a pr"ai6][2] ilustrando um fenómeno de junção entre as palavras (*sandhi*) "para a" e uma redução vocálica no interior da palavra "para". A principal motivação do presente módulo é a de gerar estas transcrições alternativas. A sequência de fones de entrada é transformada num reticulado de fones em que cada caminho, desde o fone inicial, até ao fone final, corresponde a uma transcrição fonética alternativa. A escolha da transcrição mais adequada é feita posteriormente por um descodificador de fones.

As regras fonéticas são especificadas usando uma gramática de estado finito. Cada regra é representada por uma expressão regular aumentada com o operador →, *transdução simples*, tal que *(a→b)* significa que o símbolo terminal *a* é transformado no símbolo terminal *b*.

Considere-se por exemplo a regra que produz a transcrição alternativa para a palavra "para" no exemplo anterior:

```
DEF_RULE pra, ( p ("6 → NULL) r 6 )
```

ou para implementar a transcrição alternativa para a *sandhi* referida no mesmo exemplo ("sil" foi utilizado para identificar um silêncio ou separador de palavras):

```
DEF_RULE 6a, ( (6 → NULL) (sil → NULL) (6 → a) )
```

A linguagem de especificação de regras permite ainda a definição de símbolos não-terminais, por exemplo $glide definido como (w | j). A sintaxe da sua definição é semelhante à das regras BNF, aumentada com expressões regulares (incluindo os símbolos: ‘A B’ sequência, ‘A | B’ alternativa, ‘[A]’ opção, ‘{A}’ zero ou mais repetições, ‘<A>’ uma ou mais repetições, NULL, epsilon) e limitada de forma a que um não-terminal só possa ser usado depois de definido. Isto limita o seu poder expressivo ao das gramáticas regulares, condição necessária para permitir usar as propriedades de fecho, e respectiva álgebra de autómatos finitos [5].

---

[1] A transcrição equivalente usando os símbolos do alfabeto fonético internacional (IPA) seria [vˈo pɐɾɐ ɐ pɾˈajɐ].

[2] Ou [vˈo pɾˈa pɾˈajɐ] no IPA

A ordem de aplicação é especificada pela sequência da definição das regras, ou seja se a regra R2 é definida depois de R1, então será aplicada ao resultado de R1.

A aplicação das regras à transcrição fonética larga consiste na transformação desta num autómato, que posteriormente será composto com os autómatos que implementam cada uma das regras, sucessivamente. Utilizando a propriedade associativa da composição de autómatos todas as regras poderiam ser combinadas num único autómato. No entanto, o autómato resultante poderia ter dimensões incomportáveis, pelo que se optou pela composição, em sequência, do autómato de cada uma das regras começando com a frase inicial. As restrições que esta impõe são, em geral, suficientes para evitar uma explosão na dimensão dos transdutores obtidos.

Finalmente é feita a projecção do transdutor resultante de todas as composições, obtendo-se um aceitador (*acceptor*) que é convertido para o para um reticulado com etiquetas nos nós no formato adequado para a utilização no descodificador (ferramentas HTK *Hidden Markov Model Toolkit* da *Entropic Research Labs*).

A aplicação dos dois exemplos anteriores (regras *pra* e *6a*) à sequência "para a praia" produziria um reticulado equivalente a:

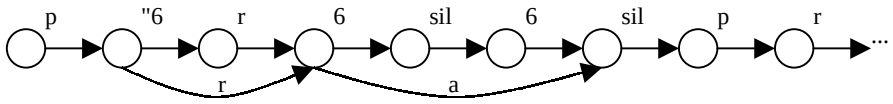

Fig. 2. – Ilustração parcial do reticulado para o exemplo "para a praia" após aplicação das regras *pra* e *6a* onde é possível obter quatro transcrições alternativas.

As regras que foram utilizadas implementam alguns fenómenos mais comuns de *sandhi*, redução vocálica e pronúncia alternativa. Algumas das regras usadas, mais especificamente as de *sandhi*, foram retiradas do sistema DIXI. As restantes foram extraídas de forma semi-automática a partir dos resultados da comparação das transcrições fonética estreita (referência) e larga (hipótese).

A título de exemplo apresentam-se de seguida alguns dos casos tratados.

Tipo	Texto	T.F. Larga	T.F. alternativa
sandhi com alteração de qualidade vocálica	de uma	[d@ um6]	[djum6]
	mesmo assim	[m"eZmu 6s"i~]	[m"eZmw6s"I~]
sandhi e consequente de redução vocálica	de uma	[d@ um6]	[dum6]
	mesmo assim	[m"eZmu 6s"i~]	[m"eZm6s"i~]
redução vocálica	semana	[s@m"6n6]	[sm"6n6]
	oito	["ojtu]	["ojt]
pronuncias alternativas	restaurante	[R@Stawr"6~t]	[R@StOr"6~t]
	viagens	[vj"aZ6~j~S]	[vj"aZe~S]

Tabela 1. Exemplos de alguns dos casos processados pelo sistema de anotação.

2.3 Módulo 3: Processador Acústico do Sinal

O processamento acústico do sinal tem como objectivo a extracção de características adequadas ao processamento final que se tem em vista (reconhecimento, síntese ou codificação). Neste sistema, como exemplo, é feita uma parametrização do sinal do tipo MFCC (Mel - *Frequency Cepstral Coefficients*) embora outras pudessem ser facilmente integradas. Esta parametrização permite alcançar uma redução da dimensão espectral do sinal de uma forma consistente com as faculdades auditivas humanas (decréscimo da resolução espectral auditiva com a frequência, acima dos 800 Hz). Os ficheiros contendo as locuções de fala devem estar em formato RAW, tipicamente amostrados a 16 kHz com 16 bits por amostra. Os parâmetros necessários à realização da codificação desejada são fornecidos ao sistema através de um ficheiro de configuração, sendo tipicamente os seguintes: 12 MFCC calculados cada 5 ms a partir de um banco de 26 filtros, utilizando janelas de Hamming de 25ms. Vectores de 39 componentes são construídos com os 12MFCC, a energia do sinal e ainda as respectivas diferenças de primeira e segunda ordem.

2.4 Módulo 4: Descodificador fonético

A descodificação fonética é feita usando o algoritmo de Viterbi podendo este processo ser descrito da seguinte forma: numa primeira fase a transcrição ortográfica é convertida numa transcrição fonética larga pelo módulo 1. A transcrição resultante desta conversão grafema para fone é posteriormente transformada num autómato e composta com os autómatos responsáveis por gerar transcrições fonéticas alternativas. O autómato final é convertido num reticulado de fones que posteriormente será utilizado pelo descodificador. Esta informação constitui uma das três restrições principais a serem usadas no processo de descodificação. A segunda restrição provém da locução de fala sob a forma de uma sequência de vectores de parâmetros, conforme descrito no ponto anterior. A terceira e última restrição, está presente na informação dos modelos HMM que modelam o universo das unidades a descodificar. Reunidas estas fontes de informação, desencadeia-se o processo de descodificação com vista a encontrar entre um conjunto de transcrições alternativas, aquela que mais se "adequa" à respectiva locução de fala.

3 *Corpora* de Treino e Teste

O *corpus* de fala em português utilizado no desenvolvimento e na avaliação do anotador, mais especificamente no treino dos modelos HMM e na posterior avaliação do anotador automático, foi recolhido no âmbito do projecto comunitário ESPRIT 6819 SAM-A (Speech Assessment in Multilingual Applications) [3]. Este projecto, motivado pela construção de recursos comuns para as línguas europeias, teve como um dos principais resultados a recolha de uma base de dados multilíngue EUROM_1 para o Português Europeu, o Espanhol e o Grego, à semelhança do que tinha sido feito em projecto anterior para as outras línguas como o Dinamarquês, Holandês, Alemão, Francês, Italiano e Norueguês. A recolha e processamento da parte portuguesa deste *corpus* multilíngue foi feita pelo INESC com a cooperação de linguistas do Centro de

Linguística da Universidade de Lisboa (CLUL). Foi utilizada uma câmara anecóica e uma frequência de amostragem de 20 kHz, com 16 bits por amostra, tendo sido efectuada a conversão para 16 kHz. Parte do *corpus* foi transcrito manualmente, por um perito em fonética. Este subconjunto constitui a totalidade do material de fala usado neste trabalho. Contém fala lida de 5 falantes do sexo masculino e 5 do sexo feminino (denominadas "few speakers"). Os textos que constituem as mensagens de ponto lidas deste *corpus* encontram-se estruturados da seguinte forma:

- Passagens: parágrafos compostos por 5 frases relacionadas entre si. Existem 40 passagens diferentes, agrupadas em blocos de 10 cada. O conteúdo das secções O0-O9 e P0-P9 foi traduzido directamente a partir das correspondente versão inglesa do EUROM 1. Nas secções Q0-Q9 e R0-R9 foram adaptadas material recolhido de livros e jornais escritos na língua portuguesa.
- Frases de preenchimento: frases construídas por forma a compensar o desequilíbrio fonético existente nas passagens introduzindo difones inexistentes no conjunto das passagens. São 50 frases agrupadas em 10 blocos de 5 frases cada (F0 – F9).

Falante	**Passagens**			**Frases**
1	O0 - O4	O5 - O9	P0 - P4	F5 - F9
2	O0 - O4	O5 - O9	P0 - 04	F0 - F4
3	P5 - P9	Q0 - Q4	Q5 - Q9	F5 - F9
4	P0 - P4	P5 - P9	Q0 - Q4	F5 - F9
5	O5 - O9	P0 - P4	P5 - P9	F0 - F4
6	P5 - P9	Q0 - Q4	Q5 - Q9	F5 - F9
7	O0 - O4	O5 - O9	P0 - P4	F0 - F4
8	Q0 - Q4	Q5 - Q9	R0 - R4	F0 - F4
9	R5 - R9	O0 - O4	O5 - O9	F5 - F9
10	Q5 - Q9	R0 - R4	R5 - R9	F5 - F9

Treino Teste 1 Teste 2

Tabela 2. Conjuntos de treino e teste

Cada falante deste subconjunto tem apenas 5 passagens e 5 blocos de frases. A divisão do *corpus* em material de treino e teste foi feita tendo em conta as seguintes considerações: pretendendo-se construir modelos independentes do falante e do género, incluiu-se no treino o maior número possível de locutores para cada um dos sexos. No que diz respeito ao teste, utilizou-se todo o material do *corpus* contendo a fala de dois falantes, de ambos os sexos que não estivessem presentes no material de treino. Por conseguinte, criaram-se dois conjuntos de teste sendo o primeiro conjunto caracterizado por conter fala de dois locutores "desconhecidos", mas lendo texto comum a outros locutores presentes no material de treino. O segundo conjunto de teste tem um carácter mais selectivo na medida em que se caracteriza por conter fala dos locutores "desconhecidos", com um texto diferente do lido no material de treino.

4 Experiência de Anotação

A experiência de anotação automática que iremos descrever diz respeito à anotação fonética do *corpus* de teste, descrito na secção anterior, tendo sido utilizada como. referência a anotação manual. Assim, o melhor processo de transcrição automática é aquele que conduz a uma transcrição tão próxima quanto possível da obtida manualmente. O melhor processo de alinhamento é aquele que, conduz a marcas temporais tão próximo quanto possível das obtidas manualmente.

4.1 Experiências de transcrição e alinhamento fonético

Nas experiências de transcrição e alinhamento um conjunto de 60 fones foi modelado por um conjunto de outros tantos modelos HMM com arquitectura do tipo esquerda-direita, 3 estados, 3 misturas e foram treinados de forma diferenciada de acordo com o fim a que se destinam (descodificação ou alinhamento). A descodificação foi feita usando uma gramática do tipo ergódico. Importa referir que devido às reduzidas dimensões dos *corpora* de teste, todas as experiências realizadas com vista ao desenvolvimento do sistema de anotação foram feitas sobre o *corpora* de treino.

Os resultados obtidos nesta experiência encontram-se ilustrados na tabela 7, onde se utiliza como indicadores de alinhamento a percentagem de casos cujo erro absoluto é inferior a 10 ms e erro absoluto médio de forma a incluir 90 % dos casos.

Modelos	**Transcrição**	**Alinhamento**	
	Precisão	**< 10ms**	**Percentil 90%**
HMM (transcrição)	52,8 %	66,87 %	20 ms
HMM (alinhamento)	42,96 %	78,94 %	18 ms

Tabela 3. Resultados das experiências de transcrição e alinhamento fonético.

Os dados na tabela 7 mostram que os melhores resultados para cada uma das tarefas, transcrição e alinhamento, obtêm-se usando os respectivos modelos. Isto é, no alinhamento usando os modelos treinados para o alinhamento e na transcrição modelos treinados para a descodificação ou reconhecimento.

Após ter sido validada a adequação dos modelos a cada uma das tarefas, realizou-se uma experiência de anotação com o objectivo de testar três estratégias possíveis para a utilização dos referidos modelos num sistema de anotação:

- **Estratégia 1**: Descodificador/Alinhador usando modelos HMM treinados para alinhamento.
- **Estratégia 2**: Descodificador/Alinhador usando modelos HMM treinados para reconhecimento.
- **Estratégia 3**: Descodificador usando modelos HMM treinados para reconhecimento e um Alinhador usando modelos HMM treinados para alinhamento.

A descodificação e posterior alinhamento para cada uma destas estratégias realizou-se a partir do reticulado de transcrições múltiplas obtidas pelo módulo GTA utilizando as regras descritas em 2.2. Foram utilizados os mesmos indicadores para

avaliar o alinhamento mas, desta feita, utilizando apenas as "zonas" onde a descodificação realizada coincide com a transcrição manual do *corpus* de treino.

	Precisão	**< 10ms**	**Percentil 90%**
Estratégia 1	85,32 %	77,39 %	20 ms
Estratégia 2	85,78 %	44,02 %	29 ms
Estratégia 3	85,78 %	78,05 %	19 ms

Tabela 4. Resultados em função das estratégias de anotação

Estes valores confirmam que a estratégia 3, que combina o melhor descodificador com o melhor alinhador, produz os melhores resultados globais.

4.2 Experiência de Anotação Fonética

Definida a melhor estratégia de transcrição e segmentação fez-se uma experiência de anotação agora sobre os *corpora* de teste 1 e 2 por forma a aferir o desempenho global do sistema desenvolvido na presença de material desconhecido.

Regras	**Precisão**	
	Teste 1	**Teste 2**
Canónica	73,99 %	76,95 %
Sandhi	77,11 %	79,37 %
Redução vocálica e pronúncia alternativa	85,08 %	84,52 %

Tabela 5. Resultados de transcrição dos *corpora* de teste para diferentes conjuntos de regras

Os resultados obtidos mostram que após efectuada a transcrição canónica dos *corpora* de teste a precisão na descodificação é melhor no *corpus* de teste 2. Supõe-se que tal se deva ao facto destas frases terem sido "construídas" por forma a compensar desequilíbrios fonéticos existentes noutras fracções do *corpus*. Este procedimento deu origem a frases pouco usuais o que levou os locutores a pronunciarem-nas de forma mais cuidada. Este comportamento explica a maior aproximação entre a transcrição fonética de referência e a canónica obtida no módulo G2P.

Outra consideração a fazer acerca dos resultados obtidos relaciona-se com o facto do incremento de precisão do descodificador no *corpus* de teste 1 ser maior do que no *corpus* de teste 2. Tal comportamento tem a seguinte explicação: À medida que se vão aplicando os diferentes conjuntos de regras, um crescente número de transcrições alternativas implicará uma maior dependência da precisão do descodificador. Uma experiência de descodificação realizada sobre os *corpora* de teste com gramática ergódica, conduziu a precisões de 48,15% e 42,21% respectivamente no *corpus* de teste 1 e 2. A conjugação destes dois factores explica o maior incremento verificado nos resultados referentes ao *corpus* de teste 1.

Os resultados obtidos mostram também que a aplicação sucessiva de conjunto de regras com os fundamentos descritos em 2.2 conduzem a uma melhoria significativa na precisão de transcrição. O mesmo se pode verificar com os indicadores de alinhamento:

Regras	Alinhamento			
	Teste 1		Teste 2	
	< 10 ms	**90 %**	**< 10 ms**	**90 %**
Canónica	74,68 %	24 ms	75,18 %	25 ms
Sandhi	75,04 %	23 ms	75,41 %	24 ms
Redução vocálica e pronúncia alternativa	78,76 %	19 ms	77,27 %	22 ms

Tabela 6. Resultados de alinhamento nos *corpora* de teste

Verifica-se que as sucessivas transcrições fonéticas produzidas na descodificação influenciam os resultados de alinhamento. Isto dever-se-á à melhor semelhança acústica reflectida também numa melhor aproximação à transcrição de referência.

A comparação directa destes resultados de alinhamento com os obtidos em outros sistemas não é possível, na medida em que não se partilha o mesmo *corpus* de teste. Todavia, os resultados obtidos para 90% dos casos são da mesma ordem de grandeza que os conseguidos em [4] e mesmo em [2] onde era utilizada uma modelação mais complexa.

As transições entre classes de fones que possuem maiores erros de alinhamento (40-90 ms) são: vogais → vogais, oclusivas surdas → oclusivas sonoras, oclusivas sonoras → vogais nasais, consoantes nasais → fricativas não vozeadas, liquidas → fricativas não vozeadas e vogais nasais → vogais.

As dificuldades com vogais, vogais nasais e liquidas, são conhecidas de outros trabalhos [1] [2] [10] e situam-se justamente no tipo de transições onde surgem maiores problemas de coerência entre anotadores humanos. Existem outras transições com erros superiores a 40 ms que não foram consideradas importantes devido à sua reduzida frequência de ocorrência e à pequena dimensão do *corpus* utilizado.

5 Conclusões e Trabalhos Futuros

A difícil tarefa de substituir a anotação manual por métodos automáticos foi abordada na perspectiva de aplicação deste anotador automático à anotação de *corpora* transcritos ortograficamente. Tentou-se colmatar a discrepância entre a transcrição grafema-fone rígida (transcrição fonética larga) e o que na realidade foi dito pelo falante (transcrição fonética estreita), através da geração (por regra) de transcrições fonéticas alternativas, deixando que um descodificador fonético escolha a mais adequada.

Os resultados obtidos nas experiências de anotação ilustram de imediato a possibilidade de obter melhorias relevantes utilizando esta ferramenta, comparativamente a um processo mais tradicional de geração de uma única transcrição para alinhamento.

O trabalho futuro incidirá sobre a investigação da escolha das regras a utilizar na produção de transcrições fonéticas alternativas. Uma possibilidade será a inferência automática de regras com base na comparação dos resultados de transcrição do módulo de conversão grafema-fone e os dados anotados manualmente, não

descurando, claro está, uma validação ou interpretação fonético-linguística das mesmas.

A ferramenta desenvolvida será utilizada para segmentar outros *corpora* de fala permitindo assim a construção de melhores modelos para a sua própria optimização.

Um outro campo de trabalho futuro será o da atribuição de informação probabilística, ou pesos associados às regras de geração de transcrições fonéticas alternativas, sendo previsível um melhor desempenho do sistema. Os algoritmos utilizados [8] [9] poderão tratar correctamente esta informação do ponto de vista do reconhecimento de fala, pelo que a sua utilização seria simples. Apesar da obtenção da informação probabilística ser problemática, grandes quantidades de material transcrito poderão ser utilizadas, ou, na impossibilidade de seguir por esta via, poder-se-ão utilizar métodos iterativos baseados no algoritmo EM (*expectation maximization*).

Referências

1 Andrej Ljolje, Julia Hirschberg, Jan P.H. van Santen, “Automatic Speech Segmentation for Concatenative Inventory Selection”, Progress In Speech Synthesis, pag. 304-311

2 Andrej Ljolje, Michael D.Riley, "Automatic Segmentation of Speech for TTS", Proceedings EUROSPEECH 93, Volume 2, pag. 1445-1448, Setembro 1993, Berlim

3 C. M. Ribeiro, I. M. Trancoso, M. C. Viana, “EUROM1 Portuguese Database”, Relatório D6, SAM-A/INESC, Novembro 1993.

4 Colin W. Wightman, David T. Talkin, “The Aligner: Text-to-Speech Alignment Using Markov Models”, Progress In Speech Synthesis, pag. 313-323

5 Emmanuel Roche e Yves Schabes (Eds), "Finite State Language Processing", MIT Press, 1997.

6 Luís C. Oliveira, Maria C. Viana, Isabel M. Trancoso, “A rule-based text-to-speech system for Portuguese”, actas Int.Conf. on Acoustic Speech and Signal Processing, Vol 2, Pag 73-76, São Francisco, Março 1992

7 Luís C. Oliveira, “Síntese de Fala a Partir de Texto”, Tese de Doutoramento, IST - UTL, Outubro de 1996

8 Mehryar Mohri, Fernando Pereira e Michael Riley, "Weighted Automata in Text and Language Processing", actas ECAI 96 Workshop, 1996.

9 Mehryar Mohri, "Finite-State Transducers in Language and Speech Processing", Computational Linguistics, 23:2, 1997.

10 Pedro Carvalho, Isabel Trancoso, Luis Oliveira, “Automatic Segment Alignment for Concatenative Speech Synthesis in Portuguese”, actas 10th Portuguese Conference on Pattern Recognition, RECPAD98, Fevereiro 1998, IST/UTL Lisboa

Os autores gostariam de agradecer a colaboração da Drª. Isabel Mascarenhas e Drª. Céu Viana pela sua contribuição indispensável para a realização deste trabalho.

Este trabalho foi desenvolvimento no âmbito das Bolsas de Doutoramento de Pedro Carvalho, Rui Amaral. e Diamantino Caseiro. Bolsas PRAXIS XXI/BD/4526/94 - 15844/98 - 15836/98, respectivamente.