AUDIMUS - SISTEMA DE RECONHECIMENTO DE FALA CONTÍNUA PARA O PORTUGUÊS EUROPEU

João P. Neto, Ciro Martins, Hugo Meinedo, and Luís B. Almeida

Instituto de Engenharia de Sistemas e Computadores (INESC)
Instituto Superior Técnico (IST)
R. Alves Redol, 9, 1000 Lisboa, Portugal
{jpn, cam, Hugo.Meinedo, lba}@inesc.pt
http://neural.inesc.pt/NN/RFC

Abstract. Neste artigo apresentamos o trabalho desenvolvido na realização do AUDIMUS, um sistema de reconhecimento de fala contínua para o Português europeu. Para a realização de um sistema deste tipo foi necessário recolher e desenvolver os diferentes componentes que são dependentes da língua, como sejam uma base de dados de fala e respectiva segmentação e etiquetagem, grandes quantidades de texto, dicionários de pronunciação e modelos de linguagem. De seguida treinaram-se os modelos acústicos baseados no perceptrão multicamada que interligados com um descodificador baseado nos modelos de Markov não-observáveis resultou num sistema híbrido MLP/HMM. O estágio actual de desenvolvimento deste sistema já nos permitiu obter um bom nível de desempenho, sendo, no entanto, ainda passível de melhorias. Por outro lado, o desenvolvimento de um sistema deste tipo permite-nos lançar novas linhas de investigação na procura de soluções tanto independentes da língua como outras onde se procura utilizar o conhecimento específico da língua.

1 Introdução

Nos últimos anos temos assistido a um desenvolvimento significativo de sistemas de reconhecimento de fala contínua independentes do orador para vocabulários largos. Este desenvolvimento tem-se verificado tanto ao nível dos principais institutos de investigação como nalgumas companhias que têm colocado no mercado um conjunto de produtos já extremamente interessantes. Este desenvolvimento, verificado inicialmente para o Inglês americano, tem sido estendido a outras línguas com enormes potencialidades de mercado o que, infelizmente, não é para já o caso do Português Europeu.

Neste artigo apresentaremos o trabalho por nós realizado no desenvolvimento de um sistema de reconhecimento de fala contínua para o Português Europeu. A nossa participação anterior em vários projectos, internacionais e nacionais, deu-nos uma experiência considerável no desenvolvimento deste tipo de sistemas, mas realizados para a língua inglesa, e nas bases de dados utilizadas pela comunidade científica internacional. A nossa tarefa consistiu, portanto, na portabilidade desses sistemas para o Português.

Tratava-se de uma tarefa à partida difícil dada a inexistência dos recursos necessários. Começámos pelo desenvolvimento e recolha de uma base de dados de fala e texto apropriada, designada como BD-PUBLICO [1]. Associados a esta base de dados foram desenvolvidos um dicionário de pronunciações e um modelo de linguagem, assim como, técnicas de segmentação e de etiquetagem automáticas do sinal de fala [2].

Após a existência destes recursos linguísticos desenvolvemos modelos acústicos baseados no perceptrão multicamada (*Multilayer Perceptron* - MLP) que interligados com um descodificador baseado nos modelos de Markov não-observáveis (*Hidden Markov models* - HMMs) permitiu-nos obter um sistema de reconhecimento híbrido MLP/HMM [3]. A avaliação inicial do sistema mostrou-nos uma taxa de erro ao nível da palavra de cerca de 15% para um vocabulário de 5.000 palavras e modelos de linguagem baseados em trigramas [4]. Trata-se de uma primeira iteração que necessita de novos realinhamentos e de uma actualização dos seus componentes linguísticos.

No entanto, este desenvolvimento permitiu-nos verificar que existem um conjunto de condições específicas do Português que necessitam de um tratamento diferente do apontado para outras línguas, nomeadamente para a língua inglesa. É nesse sentido que se tem vindo a estudar e desenvolver diferentes modelos de linguagem que resultam de uma análise morfológica das palavras [5]. Por outro lado, a introdução de informação silábica permite uma melhoria efectiva dos modelos acústicos [6]. A aplicação de técnicas de adaptação ao orador tanto supervisionadas como não supervisionadas já permitiu uma melhoria do desempenho do sistema à qual pretendemos juntar as introduzidas por um processo de aprendizagem automática de novas pronunciações.

Encontramo-nos num estágio onde as diferentes peças se encontram interligadas e a funcionar correctamente sendo agora necessário o seu aperfeiçoamento. Por outro lado, começámos a lançar uma série de novas linhas de investigação que irão, certamente, contribuir para a melhoria do sistema final.

2 Desenvolvimento de um sistema de reconhecimento de fala contínua para uma nova língua

O transporte de um sistema de reconhecimento de fala contínua para uma nova língua é uma tarefa extremamente difícil. Este problema tem sido alvo de um grande interesse de investigação dado que existe uma enorme procura destes sistemas para outras línguas além do Inglês americano.

O primeiro passo nesta tarefa é identificar as diferentes etapas que são dependentes da língua. Os diferentes componentes, como sejam, a base de dados, o conjunto de fones/fonemas, o dicionário de pronunciação, o modelo da linguagem e a segmentação inicial e etiquetagem da base de dados de treino são significativamente dependentes da língua. E se quisermos ser um pouco mais precisos alguns deles são mesmo dependentes da tarefa.

Da base de dados de fala esperamos que seja representativa em termos dos oradores e da variabilidade das unidades fonéticas, dado tratar-se de uma com-

ponente crucial já que é a partir dela que se geram os modelos acústicos. No entanto, o nosso sistema irá depender das condições acústicas em que a nossa base de dados foi recolhida, ou seja, se se trata de fala telefónica, fala limpa, etc..

O conjunto de fones/fonemas é dependente da língua e encontramos mesmo diferentes conjuntos de fones/fonemas e diferentes classes de classificação para a mesma língua. O dicionário de pronunciação depende da língua e do conjunto de fones/fonemas, mas depende, essencialmente, do vocabulário. Para tornar o léxico independente da tarefa necessitamos de um número elevado de palavras no vocabulário o que, normalmente, introduz dificuldades adicionais. O modelo de linguagem é o componente mais pesado e é extremamente dependente da língua. Novamente para ser independente da tarefa necessitamos de uma grande quantidade de textos para uma estimação robusta do modelo de linguagem, o que torna a obtenção deste modelo difícil. Obviamente que o treino e desempenho do nosso sistema dependerá da qualidade e disponibilidade de cada um destes componentes.

3 Base de dados BD-PUBLICO

O desenvolvimento de um sistema de reconhecimento de fala contínua independente do orador e com um vocabulário largo necessita de uma base de dados de fala e de texto com conteúdo e tamanho adequados. No nosso caso esta base de dados não existia e foi necessário realizar todo um trabalho de definição, selecção e recolha que conduziu ao desenvolvimento de uma nova base de dados para o Português Europeu e denominada BD-PUBLICO [1].

Para esta base de dados escolhemos como fonte os textos do jornal PÚBLICO. Trata-se de um dos melhores jornais diários do Português Europeu com uma larga cobertura de assuntos. Numa primeira fase recolhemos 6 meses do jornal o que nos deu um total aproximado de 11 milhões de palavras com cerca de 158 mil palavras diferentes. Como descobrimos mais tarde, esta quantidade de textos era insuficiente para uma estimação robusta dos modelos de linguagem para a tarefa associada ao nosso sistema. Nesse sentido recolhemos mais textos do jornal PÚBLICO que nos levaram a um total de 46 milhões de palavras.

Os oradores foram seleccionados de entre os estudantes do Instituto Superior Técnico (IST). Trata-se de uma população jovem de falantes do Português mas com uma enorme variabilidade de sotaques dada a diversidade da sua origem. As gravações decorreram no INESC (Lisboa), numa "câmara insonorizada" e usando um microfone montado na secretária para a recolha do sinal. As gravações decorreram no período compreendido entre Abril e Novembro de 1997.

Na constituição desta base de dados foi pedido aos oradores que lessem um conjunto de frases extraídas em blocos de parágrafos dos textos do jornal. Desta forma estamos impondo uma tarefa de ditado, mas num modo independente do orador, dado o número de oradores e a quantidade de dados que estamos recolhendo para cada orador.

Com esta nova base de dados de fala resulta um conjunto de treino com 100 oradores (50 masculinos e 50 femininos) num total de 8089 locuções (aproximadamente 22 horas de fala). Foram também recolhidos 2 conjuntos de teste, um de desenvolvimento e outro de avaliação, com 10 oradores cada (5 masculinos e 5 femininos) com 40 frases por orador. Estes conjuntos de teste confinam-se a um vocabulário de 5000 palavras (5K).

Deste trabalho resultou, portanto, uma nova base de dados (BD-PUBLICO) contendo quantidades apreciáveis de texto e de fala.

4 Sistema de reconhecimento híbrido HMM/MLP

No nosso trabalho utilizamos um sistema híbrido que combina as capacidades de modelamento temporal dos modelos de Markov não-observáveis com as capacidades de classificação dos perceptrões multicamada. Neste sistema híbrido MLP/HMM é usado um processo de Markov para modelar a natureza temporal do sinal de fala. O MLP é usado como modelo acústico no âmbito dos HMMs. O MLP estima a probabilidade *a posteriori* dos fones independentes do contexto que será utilizada no processo de Markov.

Ao sinal de fala aplica-se uma técnica de extracção de características baseada na análise PLP (*Perceptual Linear Prediction*). Desta fase de pré-processamento resulta uma trama com o logaritmo da energia e os coeficientes cepstrais PLP e as respectivas derivadas temporais de primeira e, eventualmente, de segunda ordem.

4.1 Modelos acústico-fonéticos

Ao longo do desenvolvimento do nosso sistema utilizámos duas estruturas diferentes para o MLP. Na primeira estrutura utilizámos como características de entrada 12 coeficientes PLP com o logaritmo da energia e as suas primeiras derivadas temporais. Assim tínhamos um vector de características com um total de 26 coeficientes. O MLP apresenta contexto acústico local na entrada através de uma janela de várias tramas. Usámos uma janela de contexto de 7 tramas (3 tramas de contexto à esquerda e à direita em torno da trama central). Devido a estas tramas que são adicionadas na entrada do MLP passamos a um total de 182 entradas. A rede resultante tem uma única camada escondida com 1.000 unidades e 39 saídas correspondentes a classes dos fones independentes do contexto (a rede totaliza cerca de 222.000 pesos). As classes dos fones são as mesmas do SAM_PA. Na segunda estrutura usámos 12 coeficientes PLP e as suas derivadas temporais de primeira e de segunda ordem. Usámos também a primeira e a segunda derivada temporal do logaritmo da energia mas não o próprio logaritmo da energia. Neste caso o vector de características tem um total de 38 coeficientes. Usámos uma janela de contexto de 9 tramas que resulta numa camada de entrada do MLP com 342 unidades. A camada escondida passou agora a ter 2.000 unidades e a camada de saída manteve as mesmas unidades (resultando num total de cerca de 764.000 pesos).

4.2 Dicionário de pronunciações

Da selecção dos diferentes conjuntos (treino e teste) a partir da base de dados BD-PUBLICO resultou uma lista de 27.833 palavras diferentes. Esta lista de palavras foi transcrita foneticamente por um sistema de regras gerando um conjunto inicial de pronunciações [2]. Neste trabalho utilizámos as 39 classes de fones do conjunto de fones do SAM_PA. O sistema de regras apresenta um conjunto de problemas conhecido. Devido a esses problemas o dicionário de pronunciações foi revisto manualmente por uma pessoa especializada em linguística resultando um dicionário com multipronunciações.

Estas pronunciações foram definidas com base em conhecimentos linguísticos da forma correcta de pronunciação de uma dada palavra na nossa língua. No entanto nos ficheiros de fala pertencentes à base de dados encontramos uma grande variedade de sotaques que tornam a correspondência entre o dicionário de pronunciações e os sinais de fala algo imprecisa.

4.3 Modelo de linguagem

Os textos iniciais existentes na base de dados BD-PUBLICO correspondem a aproximadamente 6 meses de edições *on-line* na WEB do jornal PÚBLICO (de 25 de Setembro de 1995 ao fim de Março de 1996 com 188 edições) num total de aproximadamente 11 milhões de palavras. Desses textos seleccionámos 80% para treino, 10% para desenvolvimento e 10% para avaliação. Dos textos de treino (aproximadamente 9 milhões de palavras o que é um número bastante limitado quando comparado com os utilizados noutras línguas [7]) estimámos um modelo de linguagem fechado de bigramas com *backoffs*. Com esta quantidade de textos só nos era possível estimar modelos de linguagem de bigramas mas, mesmo assim, com uma percentagem de cobertura muito reduzida dos pares de palavras das frases do conjunto de teste. O modelo de linguagem apresentava uma perplexidade de 229 para o conjunto de teste de desenvolvimento de 5K. Trata-se de um valor bastante elevado de perplexidade.

Os resultados preliminares da avaliação mostraram-nos que para o Português, ou pelo menos para o caso da base de dados BD-PUBLICO, necessitamos de grandes quantidades de texto para podermos estimar de uma forma robusta modelos de linguagem estocásticos. Assim, foi necessário recolher mais textos do jornal PÚBLICO.

Para esta tarefa contámos novamente com a colaboração do jornal PÚBLICO que disponibilizou as suas edições a partir de Abril de 96 até Março de 98 num total de 634 edições. No final passámos a contar com 822 edições diárias do jornal PÚBLICO num total de cerca de 46 milhões de palavras e 321 mil palavras diferentes.

Com estes textos gerámos novos modelos de linguagem utilizando o *CMU-Cambridge toolkit* [8]. Na Tabela 1 apresentamos um conjunto de valores extraídos desses modelos de linguagem.

Na primeira linha temos o modelo estimado inicialmente. Quando juntamos mais textos, conservando um modelo de bigramas (2a. linha), obtemos uma

Tipo	Textos	Perplexidade	% de Cobertura				Número de		
			1g	2g	3g	4g	2g	3g	4g
2-gramas	95-96	229	17	83			222K		
2-gramas	95-98	193	10	90			549K		
3-gramas	95-98	107	7	25	68		927K	5,3M	
4-gramas	95-98	103	8	25	27	40	818K	4,5M	9,4M

Table 1. Diferentes valores associados aos vários modelos de linguagem.

redução na perplexidade, uma percentagem de cobertura maior e um número maior de pares no modelo de linguagem (ambos os modelos foram estimados com um valor de *cutoff* de 1). O modelo de linguagem de trigramas (com *cutoff* de 1) trouxe uma diminuição acentuada na perplexidade. Quando evoluímos para um modelo de linguagem de quadrigramas houve apenas uma pequena diminuição na perplexidade e a partir da percentagem de cobertura podemos concluir que são ainda necessários mais textos para gerar de uma forma robusta um modelo desta ordem.

Estes novos textos dão-nos um bom ponto de partida em termos de estimação de modelos de linguagem estocásticos. No entanto, estes textos foram extraídos de uma única fonte, apesar de o jornal PÚBLICO apresentar uma grande variabilidade de assuntos e de colaboradores, o que de alguma forma pode condicionar os nossos modelos de linguagem.

4.4 Descodificação

Quando se trabalha com fala contínua e grandes vocabulários o processo de descodificação da sequência de palavras associada ao sinal acústico assume um papel extremamente importante. O descodificador por nós utilizado realiza uma estratégia de busca eficiente baseada na descodificação de *stacks*, utilizando estimativas das probabilidades fonéticas *à posteriori*, produzidas pelo modelo acústico (MLP/HMM) como base para um método de *poda* denominado *desactivação de fones* - que é um processo altamente eficiente de reduzir o esforço computacional, diminuindo o número total de hipóteses a considerar [9].

O algoritmo, com uma única passagem, está naturalmente influenciado pelo processamento assíncrono no tempo da sequência de palavras e pelo processamento síncrono no tempo da sequência de estados dos HMMs. Assim, é possível que a busca seja desacopulada do modelo de linguagem mantendo-se na mesma os benefícios computacionais do processamento síncrono no tempo. Mais ainda este descodificador implementa uma série de aproximações eficientes como alternativa ao uso das probabilidades exactas do modelo de linguagem.

5 Avaliação do sistema

O sistema que estamos a desenvolver deve ser independente do orador. No entanto, dada a grande quantidade de dados de fala presente na base de dados

BD-PUBLICO e dadas as limitações dos nossos recursos computacionais tivemos que procurar sistemas mais pequenos que nos permitissem realizar rapidamente o desenvolvimento pretendido.

A primeira limitação imposta foi a de dividir os oradores de acordo com o seu género criando, assim, dois modelos acústicos diferentes, para os oradores masculinos e para os oradores femininos. Passámos, assim, a ter para treino de cada um dos modelos cerca de 4.000 locuções de 50 oradores. O alinhamento inicial foi baseado num sistema básico desenvolvido anteriormente a partir de um sistema para o Inglês [2]. O treino dos modelos acústicos foi realizado com o MLP de dimensão maior referido anteriormente o que nos conduziu a um resultado da taxa de erro ao nível da palavra de 15,3%. Realizando o treino com o modelo acústico baseado no MLP de dimensões menores obtivemos uma taxa de 17,0% o que se traduz numa degradação do nível de desempenho, mas obtido com um modelo acústico com um número reduzido de parâmetros.

Torna-se evidente que estes resultados podem ainda ser melhorados, nomeadamente com a utilização de técnicas de adaptação ao orador supervisionada ou não-supervisionada. Por outro lado temos trabalhado noutras vias de investigação cujos resultados pensamos vir a incorporar neste nosso sistema permitindo-nos uma melhoria do nível de desempenho do sistema.

6 Adaptação ao orador incremental e não-supervisionada

Em trabalhos anteriores estudámos o uso da adaptação ao orador incremental e não-supervisionada para melhorar o desempenho de um sistema híbrido de reconhecimento da fala contínua independente do orador desenvolvido para o Inglês americano [10]. O modo incremental e não-supervisionado possibilita a incorporação da adaptação ao orador num sistema independente do orador em tempo real sem alterar, do ponto de vista do utilizador, a forma como o sistema funciona.

Esta técnica de adaptação ao orador é baseada numa arquitectura que utiliza uma Rede Linear de Entrada (RLE) para transformar os vectores de características de entrada específicos do orador (tipicamente coeficientes cepstrais PLP) para o sistema independente do orador (SI). Foi inicialmente apresentada em [11] onde a avaliação e comparação de diferentes arquitecturas para adaptação ao orador no contexto de sistemas híbridos MLP/HMM e RNN/HMM foi realizada. De entre as técnicas apresentadas a Rede Linear de Entrada mostrou ter o melhor desempenho quando comparada com várias outras alternativas.

No âmbito do nosso sistema *AUDIMUS* avaliámos esta técnica usando o conjunto de teste de desenvolvimento. Dado que estamos a realizar adaptação ao orador incremental não-supervisionada o sistema só pode usar informação extraída dos dados de teste que já tenha reconhecido, não existindo, portanto, conjuntos de treino/adaptação e de teste separados. O processo de adaptação e de avaliação é efectuado incrementalmente sobre o mesmo conjunto de dados. Nesta experiência de adaptação ao orador partimos de um sistema independente do orador com um resultado de 17,6%, resultado este inferior ao melhor obtido e

reportado anteriormente. Após o processo de adaptação obtivemos uma taxa de erro ao nível da palavra de 14,5% usando o modelo de linguagem de trigramas, o que representa uma melhoria quando comparado com os 17,6% do sistema antes da adaptação. Trata-se de um bom resultado dado que esta melhoria foi conseguida sem exigências adicionais sobre os oradores.

7 Utilização de análise morfológica na estimação de modelos de linguagem

Para garantir uma representatividade e generalização aceitáveis os actuais sistemas de reconhecimento de fala contínua têm de operar sobre vocabulários de grande dimensão, o que necessariamente aumenta o espaço de procura e diminui a eficácia dos mesmos. Isto é particularmente verdade para línguas altamente flexivas, como o Português, o Alemão, etc. Neste caso, mesmo utilizando vocabulários de grande dimensão torna-se impossível representar todas as palavras de uma determinada língua, dando-se assim origem à ocorrência das palavras designadas por OOVs (*Out-Of-Vocabulary*) durante a fase de reconhecimento. Por outro lado, para obter estimativas fiáveis dos parâmetros do modelo de linguagem torna-se necessário utilizar corpos de texto de grande dimensão, corpos esses que nem sempre se encontram disponíveis.

Comparando a língua portuguesa com a língua inglesa podemos facilmente constatar que existem grandes diferenças a nível lexical, sobretudo originadas pelo elevado número de flexões existentes na língua portuguesa. Efectuar uma decomposição morfológica completa seria o ideal em termos de redução do vocabulário. Contudo, seria necessário possuir os conhecimentos linguísticos suficientes para efectuar uma decomposição automática e exaustiva de todo o vocabulário. Por outro lado, tornar-se-ia demasiado complicado o processo de composição, isto é, encontrar um mecanismo para efectuar a composição dos morfemas no final da tarefa de reconhecimento. Dado que na língua portuguesa os verbos constituem a classe de palavras mais variável, o método aqui em estudo é baseado apenas na decomposição morfológica dos verbos regulares. A unidade base inerente ao processo de reconhecimento passa a ser o morfema e não a palavra por si só.

Para as várias experiências realizadas utilizou-se a base de dados BD-PUBLICO, a qual, na sua versão inicial, era constituída por cerca de 11 milhões de palavras num total de cerca de 158 mil palavras diferentes. Numa primeira fase procedeu-se a uma decomposição das conjugações verbais na forma de radical/sufixo. Da lista de 158 mil palavras classificadas morfologicamente constatámos que 35% das mesmas resultavam de conjugações verbais. Após o processo de decomposição morfológica das conjugações verbais, passámos dessa lista para uma lista com cerca 112 mil morfemas (entre palavras, radicais e sufixos), o que corresponde a uma redução na dimensão do vocabulário de cerca de 29%. Nos diversos testes que se efectuaram, analisámos o referido método de decomposição morfológica segundo diferentes critérios. Numa primeira fase geraram-se dois modelos de bigramas (utilizando o método recursivo de *backoff*) com o objec-

tivo de avaliar os requisitos de memória, taxa de palavras OOV e perplexidade do novo modelo de linguagem resultante do processo de decomposição morfológica. Numa segunda fase pretendeu-se avaliar o novo modelo em termos do tempo de processamento e taxa de erro inerentes à tarefa do reconhecimento.

Nos testes efectuados obtiveram-se reduções de 28% nos requisitos de memória, uma pequena redução no número de OOVs e uma redução significativa no valor da perplexidade do modelo de linguagem (uma redução de 229 para 85, isto é, aproximadamente 63%). Obtiveram-se ainda reduções no tempo de processamento, embora à custa de uma ligeira degradação no que se refere à taxa de erro [5].

8 Introdução de informação silábica para melhoria dos modelos acústicos

Embora ainda não sejam conhecidas as unidades básicas empregues pelo sistema humano de percepção da fala, existem pesquisas nos campos da psico-acústica e da psico-linguística onde se sugere que a informação associada a intervalos de tempo da ordem da duração média da sílaba deverá ser muito importante para a compreensão humana da fala, especialmente em condições adversas [12]. Pensa-se que a sílaba pode constituir uma unidade natural, sendo também normalmente aceite que as fronteiras silábicas são mais precisas que as dos fonemas. Por outro lado, sendo a sílaba uma forma mais natural de modelar as palavras, a incorporação da informação sobre as fronteiras silábicas no processo de descodificação pode eliminar possíveis redundâncias, quer na computação quer no armazenamento, baixando efectivamente a taxa de erro. Além disto a sílaba pode providenciar um meio para expressar algumas características de longa duração da fala como efeitos de co-articulação, e informação prosódia.

Neste trabalho desenvolvemos diferentes métodos para detectar as fronteiras silábicas no âmbito da fala contínua [6]. Estes métodos foram inicialmente realizados com um subconjunto da base de dados EUROM.1 SAM Portuguesa [13] decorrendo actualmente o trabalho de desenvolvimento no âmbito da base de dados BD-PUBLICO.

Os resultados obtidos mostram que usando um MLP com parâmetros de entrada PLP se consegue obter taxas de 92,7% de acertos com 14,8% de inserções, demonstrando, assim, que é possível segmentar silabicamente um sinal de fala com precisão relativamente elevada.

9 Conclusões

Neste artigo apresentámos o trabalho por nós realizado no desenvolvimento de um sistema de reconhecimento de fala contínua para o Português Europeu. Tratava-se de uma tarefa à partida difícil dada a inexistência dos recursos necessários. Os resultados já obtidos mostram um nível de desempenho aceitável havendo, no entanto, ainda possibilidade de melhorias apesar da dificuldade da tarefa associada à base de dados BD-PUBLICO.

O desenvolvimento de um sistema deste tipo deverá ser um processo iterativo onde em cada passo teremos de melhorar alguns dos componentes resultando numa melhoria do sistema global. Daí resulta uma nova possibilidade para mais desenvolvimentos nos componentes individuais repetindo-se novamente o processo.

10 Agradecimentos

Este trabalho foi parcialmente financiado através do projecto SPRACH (LTR 20077) e do projecto PRAXIS 1654. Um agradecimento ao jornal PÚBLICO por disponibilizar os textos das suas edições.

References

1. J. Neto, C. Martins, H. Meinedo and L. Almeida, *The Design of a Large Vocabulary Speech Corpus for Portuguese*, in Proceedings of EUROSPEECH 97, Rhodes, Greece, 1997.
2. J. Neto, C. Martins and L. Almeida, *The Development of a Speaker Independent Continuous Speech recognizer for Portuguese*, in Proceedings EUROSPEECH 97, Rhodes, Greece, 1997.
3. H. Bourlard and N. Morgan, *Connectionist Speech Recognition - A Hybrid Approach*, Kluwer Academic Press, 1994.
4. J. Neto, C. Martins and L. Almeida, *A large vocabulary continuous speech recognition hybrid system for the Portuguese language*, in Proceedings ICSLP 98, Sydney, Australia, 1998.
5. C. Martins, J. Neto and L. Almeida, *Using partial morphological analysis in language modeling estimation for large vocabulary Portuguese speech recognition*, in Proceedings EUROSPEECH 99, Budapest, Hungary, 1999.
6. H. Meinedo, J. Neto and L. Almeida, *Syllable onset detection applied to the Portuguese language*, in Proceedings EUROSPEECH 99, Budapest, Hungary, 1999.
7. L. Lamel, M. Decker and J. L. Gauvain, *Issues in large Vocabulary, Multilingual Speech Recognition*, in Proceedings EUROSPEECH 95, Madrid, Spain, 1995.
8. P. Clarkson and R. Rosenfeld, *Statistical Language Modeling using the CMU-Cambridge Toolkit*, in Proceedings EUROSPEECH 97, Rhodes, Greece, 1997.
9. S. Renals and M. Hochberg, *Efficient search using posterior phone probability estimates*, in Proceedings ICASSP 95, Detroit, USA, 1995.
10. J. Neto, C. Martins and L. Almeida, *An Incremental Speaker-Adaptation Technique for Hybrid HMM-MLP Recognizer*, Proceedings ICSLP 96, Philadelphia, USA, 1996.
11. J. Neto, L. Almeida, M. Hochberg, C. Martins, L. Nunes, S. Renals and T. Robinson, *Speaker-Adaptation For Hybrid HMM-ANN Continuous Speech Recognition System*, in Proceedings EUROSPEECH 95, Madrid, Spain, 1995.
12. S. Greenberg, *On the origins of speech intelligibility*, in Proceedings of the ESCA Workshop on Robust Speech Recognition for Unknown Communication Channels, pp. 23-32, Pont-a-Mousson, France, 1997.
13. C. Ribeiro, I. Trancoso and M. Viana, *EUROM.1 Portuguese Database*, Technical Report, ESPRIT Project 6819 SAM-A, 1993.