

Análise Exploratória de Dados Lingüísticos para uma Modelagem Linear Robusta da Duração Segmental da Fala

Edmilson S. Morais e Fábio Violaro

Faculdade de Engenharia Elétrica e de Computação - FEEC - UNICAMP
Campinas, SP, Brasil

{emorais,fabio}@decom.fee.unicamp.br

***Abstract.** In this work a new method based on exploratory analysis of linguistic data is proposed in order to improve the performance of regression methods for predicting segmental duration of speech. The improved prediction method must be able to perform robust mappings from a linguistic feature space (syntactical, prosodical and phonological features) into a continuous interval of phoneme durations. The proposed method can be usefull for both purely linguistic analysis and Text-To-Speech synthesis systems. The proposed method was evaluated using a database from the Toshiba Speech Technology Laboratory in Cambridge, UK.*

***Resumo.** O objetivo deste trabalho é apresentar um novo método para análise exploratória de dados lingüísticos visando à estimação de modelos para a predição da duração segmental da fala. Este novo método de predição deverá realizar mapeamentos robustos entre vetores de um espaço simbólico de características lingüísticas (sintáticas, prosódicas e fonológicas) em um intervalo numérico, limitado e contínuo de durações segmentais de fones. O método proposto pode ser útil tanto para análises puramente lingüísticas quanto para sistemas de síntese de fala do tipo TTS, Text-To-Speech. Na avaliação do método proposto foi utilizada uma base de dados de propriedade do Laboratório de Tecnologias da Fala da Toshiba em Cambridge, UK.*

1. Introdução

Um bom sistema de síntese de fala a partir de texto (TTS - *Text to Speech Synthesis*) deve ser capaz de sintetizar fala com um alto grau de naturalidade. Para que isto seja possível torna-se imperativo uma boa modelagem tanto das características acústico-segmentais quanto dos aspectos que governam os contornos prosódicos. Em sistemas TTS, os contornos prosódicos da fala são caracterizados principalmente pela duração segmental e pelo contorno entonacional.

Em um sistema TTS, a modelagem e estimação da estrutura duracional da fala se dá a partir de um conjunto de atributos lingüísticos extraídos da sentença que se deseja sintetizar. Alguns dos fatores que geralmente dificultam a estimação da estrutura duracional da fala são:

- Elevado número de fatores lingüísticos
- Complexa interação entre os fatores lingüísticos
- Elevado número de possíveis combinações entre os fatores lingüísticos
- Ausência de teorias lingüísticas mais elaboradas que expliquem a estrutura rítmica da fala

Apesar de existirem várias abordagens para a modelagem da estrutura duracional da fala [3, 4, 5, 7, 8, 9], este trabalho ficará restrito a uma modelagem bastante simplificada e baseada apenas na predição da duração segmental de fones.

Várias técnicas têm sido apresentadas para a predição da duração segmental da fala. Algumas das mais usuais são: Árvores de regressão (*CART – Classification and Regression Trees*) [6], Soma de Produtos (*SoP – Sum of Products*) [7] e Regressores lineares (*QMTI – Quantification Method Type I*) [5, 1].

Este trabalho introduz uma nova técnica, baseada em algoritmos genéticos – GA (*Genetic Algorithm*) [2, 11], para auxiliar na otimização dos modelos de regressão lineares do tipo QMTI. Esta nova técnica permite a identificação de quais fatores lingüísticos são realmente relevantes para os modelos de predição e também permite a identificação de quais fones podem ser modelados conjuntamente (fones que apresentam características duracionais semelhantes) para a maximização do desempenho e robustez dos modelos.

A seção 2 deste artigo apresenta a formulação do problema de predição da duração segmental da fala. Uma descrição da base de dados utilizada neste artigo é apresentada na seção 3. O método proposto para análise exploratória dos dados lingüísticos e para estimação dos modelos de regressão encontra-se na seção 4. A seção 5 apresenta os resultados de alguns experimentos realizados e, finalmente, a seção 6 conclui este trabalho com algumas considerações e propostas para trabalhos futuros.

2. Formulação do Problema

O problema de predição da duração segmental de fones pode ser compreendido como um mapeamento entre o espaço simbólico de *fatores* lingüísticos e o intervalo de durações segmentais dos fones.

Cada *fator* lingüístico F_i será representado por meio de um vetor cujas respectivas posições serão denominadas *níveis* e cada um destes *níveis* corresponderá a uma determinada categoria lingüística. Por exemplo, se F_i representa o fator *CategoriaGramatical*, então:

$$\text{CategoriaGramatical} \Leftrightarrow F_i = \langle \text{Verb, Noun, Adjective, ...} \rangle \quad (2.1)$$

O *espaço fatorial* de todos os *fatores lingüísticos* disponíveis $F_1, F_2, F_3, \dots, F_N$ será denominado \mathbf{F} , e dado pelo seguinte produto cartesiano:

$$\mathbf{F} = F_1 \times F_2 \times F_3 \times \dots \times F_N \quad (2.2)$$

O *intervalo duracional* que compreende todas as possíveis durações assumidas pelos segmentos fonéticos, será representado por \mathbf{D} . O intervalo \mathbf{D} estará contido no eixo dos números reais \mathcal{R} e limitado por um valor mínimo e máximo de duração. Em outras palavras:

$$\mathbf{D} \subset \mathcal{R} \text{ e } \forall \text{ dur} \in \mathbf{D}, \text{dur}_{\min} \leq \text{dur} \leq \text{dur}_{\max} \quad (2.3)$$

Utilizando-se os termos \mathbf{F} e \mathbf{D} definidos acima, o problema de predição da duração segmental pode ser escrito como:

$$\text{DUR} : \mathbf{F} \rightarrow \mathbf{D} \quad (2.4)$$

No caso de um *modelo aditivo simples*, este mapeamento pode ser decomposto em N termos A_i , um para cada um dos N fatores F_i .

$$A_i : F_i \rightarrow \mathbf{D} \quad (2.5)$$

Combinando aditivamente todos os termos A_i , têm-se:

$$DUR : \mathbf{F} \rightarrow \mathbf{D} \Leftrightarrow \sum_{i=1}^N A_i(F_i) \quad (2.6)$$

Em um *modelo aditivo simples* (ou *modelo de regressão linear simples*), cada termo A_i é um *vetor de parâmetros*, cujo número de componentes é igual ao número de *níveis* do fator F_i . A contribuição de cada *vetor de parâmetros* A_i para o modelo $DUR(\mathbf{F})$ é denominada *efeito* do fator F_i .

3. Descrição dos Dados de Treinamento

A base de dados utilizada neste trabalho é de um locutor do sexo masculino e falante do inglês americano padrão. Esta base é de propriedade do laboratório de Tecnologias da Fala da Toshiba em Cambridge, UK. Cada exemplo de treinamento possui o seguinte formato:

$$O(i) = \langle phID, F_1, F_2, F_3, \dots, F_N, DurPhone \rangle \quad (3.1)$$

A Tabela 1 apresenta uma descrição dos 14 *fatores* lingüísticos utilizados

Tabela 1: Fatores utilizados para predição dos valores de duração segmental

Fatores	Descrição
F0 : phID	Identidade do fone a ser modelado.
F1 : PosInSyll	Posição do fone a ser modelado em relação à sílaba acentuada da palavra corrente.
F2 : PrevPh	Classe à qual o fone anterior pertence.
F3 : NextPh	Classe à qual o fone seguinte pertence.
F4 : NNNextPh	Classe à qual o fone seguinte ao seguinte pertence.
F5 : PoS	Rótulo sintático (<i>Part-of-Speech tagging</i>).
F6 : ACC	Nível acentual da palavra que contém o fone a ser modelado.
F7 : NSyll	Número de sílabas na palavra corrente.
F8 : DistEnd	Distância, em sílabas, até o final da palavra corrente.
F9 : DistStress	Distância até a próxima sílaba acentuada (da palavra seguinte).
F10 : NextPause	Distância, em sílabas, até a próxima pausa.
F11 : Prev pause	Distância, em sílabas, a partir da pausa anterior.
F12 : Chunk	Distância até o final do grupo acentual ao qual o fone a ser modelado pertence.
F13 : PosInWord	Posição, na palavra corrente, da sílaba que contém o fone a ser modelado.

Os possíveis *níveis* assumidos por cada um dos 14 *Fatores* da Tabela 1 estão descritos na Tabela 2.

Tabela 2: Níveis assumidos pelos 14 fatores da Tabela 1

Fatores	Níveis
F0	@,AR,ER,H,OR,Q,aa,ae,ai,au,b,ccc,ch,d,dh,dx,e,ei,f,g,i,ii,jh,k,l,m,nng,oi,oo,ou,p,r,s,sh,t,th,u,uh,uu,v,w,y,z,zh
F1	pre,mid,aft,non
F2	ShortVowel,LongVowel,Diphthong,VC1,VC2,VPlosive,UPlosive,Closure,UC,Sil,none
F3	ShortVowel,LongVowel,Diphthong,VC1,VC2,VPlosive,UPlosive,Closure,UC,Sil,none
F4	ShortVowel,LongVowel,Diphthong,VC1,VC2,VPlosive,UPlosive,Closure,UC,Sil,none
F5	n,nam,adj,adv,itf,deny,dig,pron2,vi,vs,vt,vb,NULL,w,pnc,nud,int,prep,freq
F6	deacc,acc,high
F7	0,1,2,3,4,6,7,8,9
F8	0,1,2,3,4,6,7,8,9

F9	0,1,2,3,4,6,7,8,9,none
F10	0,1,2,3,4,6,7,8,9
F11	0,1,2,3,4,6,7,8,9
F12	0,1,2,3,4,6,7,8,9
F13	start,middle,end

A Tabela 3 apresenta uma lista com todos os fonemas utilizados. Os fonemas se encontram em duas notações, a notação da Toshiba, que será a notação adotada neste artigo, e a notação IPA (*International Phonetic Alphabet*).

Tabela 3: Relação dos fones utilizados e suas respectivas representações em IPA

Toshiba	IPA	Exemplo	Toshiba	IPA	Exemplo	Toshiba	IPA	Exemplo	Toshiba	IPA	Exemplo
ii	i:	ease	ai	aɪ	rise	d	d	din	jh	dʒ	gin
i	ɪ	pit	au	aʊ	house	g	g	give	H	h	hit
e	ɛ	pet	oi	ɔɪ	noise	f	f	fin	m	m	mock
ae	æ	pat	ei	eɪ	raise	v	v	van	n	n	not
aa	ɑ:	calm	ou	ou	nose	s	s	sir	ng	N	doing
uh	ʌ	cut	AR	a(r)	far	z	z	zoo	l	L	left
oo	ɔ:	cause	OR	ɔ(r)	port	sh	ʃ	shin	r	R	right
uu	u:	lose	p	p	pin	zh	ʒ	measure	dx	r (flap)	writer
u	ʊ	put	t	t	tin	th	θ	thin	w	W	wasp
ER	ɚ	bird, mother	k	k	kin	dh	ð	this	y	J	yes
@	ə	allow	b	b	bin	ch	tʃ	chin	Q	ʔ (glottal stop)	
									ccc		oclusão nas plosivas

3.1 Dimensionalidade do espaço dos fatores lingüísticos

A partir da Tabela 2 pode ser verificado que se a identidade do *fone*, $phID$, também for considerada como um *fator*, o *espaço fatorial* $\mathbf{F} = F_0 \times F_1 \times F_2 \times F_3 \times \dots \times F_{14}$ terá uma dimensão da ordem de 10^{11}

$$Dim(\mathbf{F}) = 45 \cdot 4 \cdot 11 \cdot 11 \cdot 11 \cdot 19 \cdot 3 \cdot 10 \cdot 10 \cdot 11 \cdot 10 \cdot 10 \cdot 10 \cdot 3 = 4.5 \cdot 10^{13} \quad (3.2)$$

Como já foi dito na introdução deste artigo, este elevado número de possíveis combinações de *fatores* é um dos grandes desafios na modelagem robusta da estrutura duracional da fala.

4. Regressão Linear a partir de Variáveis Categóricas: Método QMTI

Para a construção de um modelo de regressão linear a partir de *fatores lingüísticos*, foi utilizado neste trabalho o método QMTI [5, 1]. Este método consiste nos seguintes passos:

- Mapear o *espaço* dos *fatores* lingüísticos em um *espaço fatorial* binário $\mathbf{F} \rightarrow \mathbf{B}$
- Utilizar o *modelo aditivo* para estimar $DUR(\mathbf{F}) \Leftrightarrow DUR(\mathbf{B}) = \sum_{i=1}^N A_i(B_i)$. Sendo B_i a representação binária equivalente de F_i .
- Determinar os mapeamentos A_i (coeficientes de regressão) empregando o método da Pseudo Inversa [10].

O mapeamento do *espaço dos fatores* lingüísticos para o espaço *fatorial* binário, $F \rightarrow B$, utilizado pelo método QMTI, será explicado a seguir através do exemplo da Tabela 4: Se o *nível do fator F5*, que descreve a classe gramatical à qual pertence o fone cuja duração se deseja modelar for *adj*, então o vetor binário correspondente a este *fator* será:

Tabela 4: Representação binária dos fatores lingüísticos

Simbólico	F2	n	nam	adj	Adv	itf	deny	dig	pron2	vi	vs	...	pnc	nud	int	prep	freq
Númerico	B2	0	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0

Dois questões são fundamentais na construção dos modelos QMTI robustos. A primeira diz respeito ao número de modelos QMTI que devem ser utilizados, isto é, cada fone deve ser modelado separadamente ou alguns fones devem ser agrupados em um único modelo de regressão? A segunda diz respeito à topologia ótima do modelo, isto é, quais *fatores* devem ser utilizados em cada modelo de regressão? A seguir são apresentadas soluções baseadas em GA [2, 11], para estas duas questões.

4.1 Determinação da Topologia Ótima dos Modelos QMTI

Uma estimação robusta dos modelos QMTI exige, inicialmente, uma análise exploratória dos dados lingüísticos para a escolha da topologia ótima do modelo. Esta escolha da topologia ótima consiste na identificação de quais *fatores* são realmente relevantes na estimação das durações segmentais dos fones. Neste trabalho foi proposto um procedimento baseado em GA para a estimação destas topologias ótimas.

O algoritmo proposto inicia com uma *partição dos dados* em dados de treinamento (80% do total) e dados de validação (20% do total). Em seguida passa-se a um processo iterativo que se segue ao longo de várias *épocas*. Ao final deste processo iterativo obtém-se uma topologia para o modelo QMTI que será denominada *topologia intermediária*.

O algoritmo é operado várias vezes, e cada uma delas com diferentes partições dos dados entre dados de treinamento e validação. A cada operação é obtida uma nova *topologia intermediária* (que pode ou não ser diferente das obtidas anteriormente). Finalmente uma regra denominada *regra majoritária* é aplicada a todas as *topologias intermediárias* para a obtenção de uma única *topologia média*. Esta *topologia média* será considerada a *topologia ótima* para o modelo QMTI.

A seguir são apresentadas algumas considerações sobre a representação cromossômica dos indivíduos (topologias candidatas à topologia ótima), as operações realizadas a cada geração e a regra majoritária.

4.1.1 Representação Cromossômica dos Indivíduos (Topologias)

Utilizou-se uma representação binária bastante simples, em que cada cromossomo possui dimensão 14 (14 *fatores*). Alelos iguais a 1 indicam a presença do respectivo *fator* e alelos iguais a 0 indicam a ausência do respectivo *fator*. Por exemplo:

Tabela 5: Representação cromossômica das topologias

Cromossomo	0	1	1	1	1	1	0	0	0	1	1	1	0	0
Fatores selecionados	-	F2	F3	F4	F5	F6	-	-	-	F10	F11	F12	-	-

4.1.2 Épocas e Gerações

Cada *época* é constituída de várias gerações. As *épocas* são sempre iniciadas com uma nova população de indivíduos e finalizadas com a seleção de uma *topologia intermediária*. Em cada geração os indivíduos são modificados por operadores de *crossover* e mutação e em seguida são estimados modelos QMTI para cada um destes indivíduos, utilizando-se apenas os dados de treinamento. Estes modelos QMTI são avaliados através do cálculo do coeficiente de correlação de *Pearson* [10] entre as durações originais e as durações previstas, utilizando-se apenas os dados de validação. Os valores de *fitness* de cada indivíduo são derivados a partir dos coeficientes de correlação de *Pearson*. De acordo com os valores de *fitness* são realizadas seleções e reproduções de indivíduos para a geração seguinte.

4.1.3 Regra majoritária

Esta regra constrói a topologia considerada ótima para o modelo QMTI por meio de uma média majoritária das topologias intermediárias. A regra majoritária adotada seleciona para a topologia ótima somente aqueles *fatores* lingüísticos que ocorrerem em mais de 50% das topologias intermediárias.

4.2 Clusterização de Segmentos Fonéticos

Na tentativa de lidar com o problema de falta de dados de treinamento, foi proposto e implementado um procedimento baseado em algoritmo genético (GA) para determinar automaticamente classes de fones que possam ser modelados conjuntamente. Este agrupamento foi implementado através de um procedimento de clusterização hierárquica do tipo *Top-Down*. Além de ser do tipo *Top-Down*, a clusterização adotada apenas permite que cada cluster seja dividido (quando dividido) em somente dois clusters filhos. Este tipo de clusterização inicia com todos os fones em um único cluster e em seguida divide este cluster gerando dois clusters filhos. A seguir cada um destes dois clusters filhos são divididos gerando um total de quatro clusters. Este processo de divisão/duplicação é repetido até que se tenha um único fone por cluster nos nós terminais da árvore.

A seguir são apresentadas algumas considerações sobre a representação cromossômica dos indivíduos (representação binária para as possíveis divisões dos clusters) e sobre a *função-objetivo* utilizada.

4.2.1 Representação

A Tabela 6 apresenta a representação binária adotada para as possíveis divisões de cada clusters em dois clusters filhos:

Tabela 6: Representação cromossômica das possíveis divisões dos clusters

Fone	1	2	3	4	...	40	41	42	43	45
Cromossomo	0	1	1	1	...	0	0	1	1	1

Se o alelo da *i-ésima* posição for igual a 0, então o *i-ésimo* fone deve ser classificado na classe *L*; caso a *i-ésima* posição seja igual a 1, então o *i-ésimo* fone deve ser classificado na classe *R* (considerando-se que cada cluster será dividido em apenas 2 clusters filhos, *L-Left* e *R-Right*).

4.2.2 Função-Objetivo

Para cada divisão do $Cluster_i$ nos clusters filhos, $Cluster_i^L$ e $Cluster_i^R$, foi utilizada a seguinte função-objetivo:

$$\begin{aligned} FunçãoObjetivo(Cluster_i^L, Cluster_i^R) &= \frac{(1 - Corr(Cluster_i^L)) + (1 - Corr(Cluster_i^R))}{2} \quad (4.1) \\ &= 1 - \frac{Corr(Cluster_i^L) + Corr(Cluster_i^R)}{2} \end{aligned}$$

sendo $Corr(Cluster_i^L)$ e $Corr(Cluster_i^R)$ os índices de correlação de *Pearson* para os $Cluster_i^L$ e $Cluster_i^R$, respectivamente.

Segundo esta definição, a função-objetivo $FunçãoObjetivo(Cluster_i^L, Cluster_i^R)$ assumirá valores menores quando a soma dos índices de correlação para os clusters *L* e *R* forem altos. A função de *fitness* será definida a partir da função-objetivo e favorecerá a divisão dos clusters em clusters filhos que apresentem altos valores de correlação de *Pearson*.

5. Resultados

5.1 Topologias ótimas: considerando-se um modelo QMTI por fone

A Figura 1 mostra as topologias ótimas para os modelos QMTI, obtidas com o uso do algoritmo proposto e considerando-se um modelo por fone. Na Figura 1 os fones (modelos QMTI) estão representados ao longo da abscissa (eixo horizontal) e os *fatores* lingüísticos estão representados ao longo da ordenada (eixo vertical). A cor preta indica que o *fator* correspondente deve fazer parte do modelo e a cor branca indica que tal *fator* não é relevante para o modelo de predição. Por exemplo, para o fone /zh/, tem-se:

$$DUR(zh) = a_1 \cdot posSyll + a_3 \cdot next\ ph$$



Figura 1: Topologia dos 45 modelos lineares gerados para cada um dos fones.

Pode ser observado da Figura 1 que o *fator* responsável pela identidade fonética *phID* não aparece em nenhuma das topologias. Isto acontece porque cada fone é modelado separadamente e por isso a identidade fonética torna-se redundante.

5.2 Um modelo de predição por classe de fones

Para determinar quais fones devem ser agrupados (e modelados conjuntamente), foram gerados modelos QMTI, com topologia ótima segundo o método descrito na seção 4.1, para cada um dos clusters da árvore clusterização descrita na seção 4.2. Os clusters selecionados foram os que apresentaram um coeficiente de correlação de *Pearson* maior que o somatório dos coeficientes de correlação de *Pearson* de seus dois clusters filhos.

Os clusters selecionados, ou seja, aqueles que apresentaram um melhor resultado segundo a análise descrita acima, encontram-se na Tabela 7:

Tabela 7: Classes de fones modelados conjuntamente

C01	C02	C03	C04	C05	C06	C07	C08	C09
<i>aa,sh</i>	<i>Q,b,ng</i>	<i>H,z</i>	<i>AR,OR</i>	<i>dh,dx</i>	<i>ccc,u,v,w,y</i>	<i>g,i,z</i>	<i>ae,ii</i>	<i>ai,oi</i>

A Figura 2 mostra as topologias obtidas para os modelos QMTI, considerando-se os clusters selecionados.



Figura 2: Topologia para os modelos QMTI após a clusterização de alguns fones. Como resultado final obteve-se 24 classes com apenas um único fone, seis classes com dois fones, duas classes com três fones e uma classe com cinco fones.

Diferentemente do que acontece para os fones modelados separadamente, pode-se verificar pela Figura 2 que classes de fones com mais de um fone requerem a identidade fonética *phID* como um dos *fatores* lingüísticos para seus modelos QMTI.

5.3 Avaliação do Método Proposto

A Figura 3 mostra o desempenho do método proposto, para o caso de um modelo QMTI para cada classe de fones (conforme Figura 2), quando comparado com árvores de regressão (RT) e com modelos QMTI com topologias escolhidas de acordo com o método estatístico de ANOVA (Análise de Variância para Múltiplas Variáveis) [10].

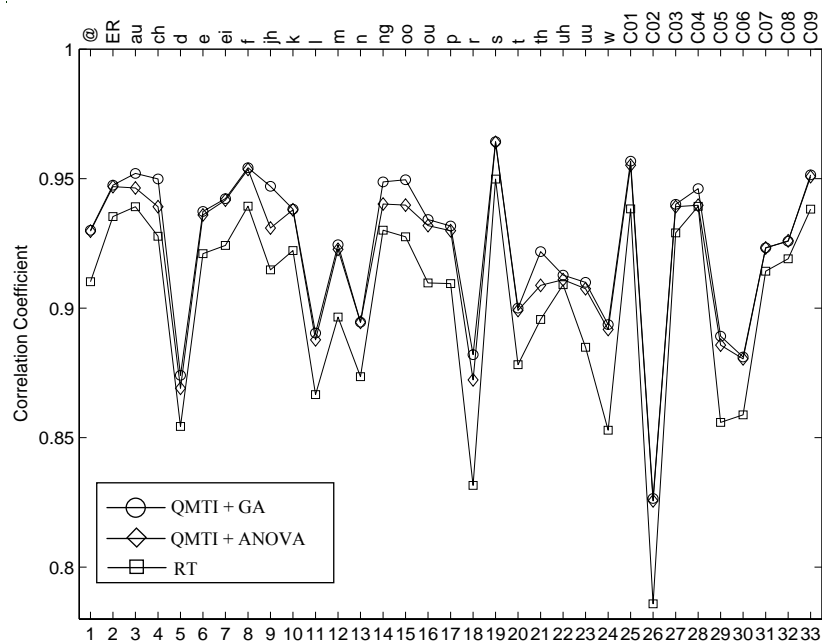


Figura 3: Desempenho dos modelos QMTI + GA, RT e QMTI + ANOVA.

6. Considerações Finais

Este trabalho apresentou um novo método para a análise e predição da duração segmental da fala. Este novo método de predição é capaz de realizar mapeamentos robustos entre vetores de um espaço simbólico de características lingüísticas em um intervalo numérico, limitado e contínuo de durações segmentais (durações dos fones).

O método proposto utiliza Métodos de clusterização hierárquica + GA para identificar quais fones devem ser modelados conjuntamente, e em seguida utiliza QMTI + GA para a obtenção da topologia ótima dos modelos de regressão (modelos de predição da duração segmental). Na verdade, o método proposto consiste em uma nova técnica para realizar uma análise exploratória dos dados lingüísticos disponíveis, e determinar quantos e quais são os estimadores ótimos para a duração segmental.

Os vários experimentos realizados para validar a técnica proposta comprovam a superioridade do método proposto quando comparados a RT (*Regression Trees*) e a modelos QMTI com topologias escolhidas de acordo com a técnica estatística ANOVA. Uma versão estendida deste artigo pode ser encontrada em [12].

Como proposta para trabalhos futuros, sugere-se:

- Utilizar modelos de regressão que incluam interações entre *fatores*, por exemplo:

$$DUR(\mathbf{F}) = \sum_{i=1}^N A_i(F_i) + \sum_{i=1}^J \sum_j^K A_{ij}(F_i \times F_j) \quad (6.1)$$

sendo A_{ij} uma *matriz de parâmetros*, cujo número de componentes é igual ao número de *níveis* dado pelo *fatorial* $F_i \times F_j$. Os índices J e K pertencem aos intervalos $1 \leq J \leq N$ e $1 \leq K \leq N$.

- Ao invés de modelar segmentos fonéticos, tentar modelar os segmentos silábicos como proposto por Campbell [8].
- Tentar construir um modelo superposicional do ritmo duracional da fala, combinando a técnica aqui proposta com os conceitos de “*Grupo Acentual*” e unidade “*VV*” introduzidos por Barbosa [9].

Agradecimentos

Os autores gostariam de agradecer ao Laboratório de Tecnologias da Fala da Toshiba em Cambridge, UK, pela base de dados utilizada neste trabalho, e em particular as valiosas sugestões da Dra. Kate Knill.

Referências Bibliográficas

- [1] HAYASHI, C., “On the Quantification of Qualitative Data from the Mathematico-Statistical Point of View”, *Annals of the Institute of Statistical Mathematics*, Vol. 2, 1950.
- [2] Thomaz, A., Ferreira, B., et al, “AGBIN - Toolbox para Algoritmos Genéticos”, <http://www.dca.fee.unicamp.br/~vonzuben/courses/ia707.html>, Unicamp, Brasil, 2004.
- [3] SPROAT, Richard. “Multilingual Text-To-Speech Synthesis”. *The Bell Labs Approach*, Kluwer Academic Publishers, Boston, USA, 1998.
- [4] SANTEN, J. V.; SPROAT, R. et al. “Progress in Speech Synthesis”, Springer-Verlag, New York, USA, 1997.
- [5] HANS., C., SAGISAKA, Y., “Analysis of Segmental Duration for Thai Speech Synthesis”, *Speech Prosody 2004*, Nara, Japan, March 23-26, 2004.
- [6] BREIMAN, L., FRIEDMAN, J., and OLSHEN, R., “Classification and Regression Trees”, Wadsworth and Brooks, Pacific Grove, CA, USA, 1984.
- [7] VAN SANTEN, J., “Assignment of Segmental Duration in Text-to-Speech Synthesis”, *Computer, Speech and Language*, 8, 1994.
- [8] CAMPBELL, W. N., ISARD, S., D., “Segment Durations in a Syllable Frame”, *Journal of Phonetics*, Special Issue on Speech Synthesis, Vol. 19(1), 37-48.
- [9] BARBOSA, PLÍNIO, A., “Explaining the Brazilian Portuguese Resistance to Stress Shift With a Coupled-Oscillator Model of Speech Rhythm Production”. *Cadernos de Estudos Linguísticos*, Instituto de Estudos da Linguagem, UNICAMP, Num. 43, p. 71 – 92. , Jul./Dez. 2002.
- [10] JOBSON, J., D., “Applied Multivariate Data Analysis Volume I: Regression and Experimental Design”, Springer Verlag, New York, USA, 1991.
- [11] Mitchell, M., “An Introduction to Genetic Algorithms”, The MIT Press, Cambridge, Massachusetts, USA, 1998.
- [12] Morais, E. S., “Technical Report: Exploratory Analysis of Linguistic Data to Optimal Modelling of the Segmental Duration of American English”, Laboratório de Pesquisa em Processamento Digital de Fala – LPDF - <http://www.fee.unicamp.br/FEEC-nova/index2.htm>, Março de 2005.