

Estratégias para a criação de classes de regressão para a técnica de adaptação a locutor baseada em MLLR

Daniela Barude Fernandes e Carlos Alberto Ynoguti

Resumo—Neste trabalho exploram-se diversas alternativas para a formação das classes de regressão para o algoritmo MLLR, usado no processo de adaptação ao locutor. Foram testados métodos baseados em classificação fonética e em medidas de distância, variando-se também o número de classes de regressão. Os resultados mostram que, para o cenário estudado, com apenas 4 locuções de adaptação, a melhor abordagem é utilizar apenas 3 classes de regressão: vogais, consoantes e silêncio.

Palavras-Chave—Adaptação ao Locutor, MLLR, Classes de Regressão.

Abstract—In this work some alternatives for regression classes construction for the MLLR adaptation algorithm are tested. Methods based on phonetic classification and based on distance metrics were tested. Also, the influence of the number of regression classes on the error rate were checked. The experimental results showed that, for the scenario under analysis, with only 4 adaptation utterances, the better approach is to use only 3 regression classes: vowels, consonants and the silence.

Keywords—Speaker Adaptation, MLLR, Regression Classes.

I. INTRODUÇÃO

Sistemas de reconhecimento automático de fala podem ter um desempenho ruim para um dado locutor quando suas características acústicas não estão bem representadas na base de dados usadas para treinamento. Obviamente pode-se treinar um sistema para ser utilizado por um único locutor, entretanto esta não é uma abordagem prática dada a enorme quantidade de material de treinamento necessária.

Alternativamente, várias técnicas têm sido propostas para tentar minimizar este problema. Dentre estas podem-se destacar as de normalização de locutor [1] e as de adaptação ao locutor [2].

Na normalização de locutor, todos os locutores usados no re-treinamento do sistema são normalizados, iterativamente, em relação a um locutor médio. O objetivo principal é tentar normalizar as representações paramétricas do sinal de voz, de modo a reduzir os efeitos causados pela variabilidade da fala entre diferentes locutores.

A ideia de adaptação ao locutor, foco deste trabalho, consiste basicamente em transformar um sistema independente de locutor (SI) em um sistema dependente de locutor (SD), usando uma quantidade de dados menor do que a que seria usada para treinar um SD por completo [2]. A maioria das técnicas de adaptação usam como ponto de partida os HMMs

Daniela Barude Fernandes e Carlos Alberto Ynoguti, Instituto Nacional de Telecomunicações - Inatel, Santa Rita do Sapucaí, MG, Brasil, E-mails: daniela.fernandes@inatel.br, ynoguti@inatel.br.

de um SI e com algumas informações do locutor a ser adaptado, transformam um ou mais parâmetros dos HMMs originais, conseguindo melhorar o desempenho do SI para esse locutor. A Figura 1 abaixo ilustra este processo.

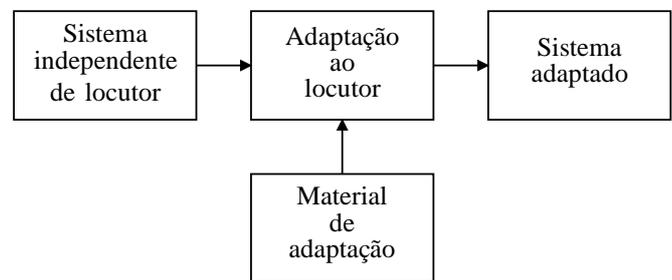


Fig. 1. Princípio básico do processo de adaptação ao locutor.

As técnicas de adaptação ao locutor aplicadas utilizando-se os HMMs podem ser divididas em três famílias [2]:

- MAP (do inglês “Maximum a Posteriori”)
- Agrupamento de Locutores
- Transformações Lineares

Pode-se citar como exemplo da família MAP, a técnica que lhe deu o nome: MAP; da família de agrupamento de locutores, a técnica Eigenvoices; e por fim, da família de transformações lineares, a MLLR, Regressão Linear de Máxima Verossimilhança (do inglês, Maximum Likelihood Linear Regression), que foi utilizada neste trabalho.

II. MLLR

A técnica de Regressão Linear de Máxima Verossimilhança (MLLR - Maximum Likelihood Linear Regression), quando utilizada para adaptar sistemas que usam o HMM, usa uma matriz de transformação para alterar os parâmetros [3]:

$$\hat{\lambda} = W\lambda \quad (1)$$

onde:

- W é a matriz de transformação
- $\hat{\lambda}$ são os parâmetros adaptados.
- λ são os parâmetros originais.

Idealmente todos os parâmetros deveriam ser transformados. Entretanto, em muitas aplicações, a quantidade de material disponível para a adaptação é limitada, tornando inviável esta abordagem [4][5][6]. Desta forma é preciso escolher qual ou quais sofrerão a adaptação, e quais serão deixados como estão.

Segundo [5], os parâmetros que causam o maior impacto na taxa de acertos são as médias das gaussianas: uma adaptação adequada das mesmas pode levar a uma grande melhoria no reconhecimento. As matrizes de covariância também influem nesta taxa, mas sua contribuição já é bem mais modesta. A alteração nas probabilidades de transição de estados e no peso das misturas quase não apresenta melhora no desempenho do sistema.

Embora a adaptação das matrizes de covariância leve a uma melhoria na taxa de acertos, é melhor deixá-las inalteradas do que modificá-las com as mesmas transformações utilizadas para as médias. Caso seja possível gerar uma matriz de transformação separada para as matrizes de covariância, isso pode trazer uma melhora no desempenho do sistema. Diante disso, optou-se neste trabalho por adaptar apenas as médias das gaussianas.

O vetor média para adaptação das componentes gaussianas é calculado multiplicando-se o vetor média original (proveniente de um SI, por exemplo) por uma matriz de transformação, da seguinte forma:

$$\hat{\mu}_s = W_s \varepsilon_s \quad (2)$$

onde:

- W_s é a matriz de transformação da componente gaussiana s .
- $\hat{\mu}_s$ é o vetor média adaptado.
- ε_s é o vetor média estendido para a componente gaussiana s . Esse vetor média estendido é na verdade o vetor média original mais um parâmetro de offset que é usado quando as condições de gravação do material de adaptação são diferentes.

Uma classe de regressão é um conjunto de médias das componentes gaussianas a serem adaptadas usando-se a mesma matriz de transformação. É importante ressaltar que uma componente gaussiana de um mesmo estado de um modelo HMM pode ter suas médias divididas em diferentes classes de regressão. Esta abordagem, embora interessante, não foi adotada neste trabalho.

A questão principal da técnica MLLR é determinar quantas classes de regressão deverão ser utilizadas e quais médias das componentes gaussianas irão pertencer a cada classe.

Teoricamente, o número de classes de regressão pode variar de uma única classe para todas as médias das componentes gaussianas ou uma classe para cada gaussiana. O que vai limitar essa quantidade de classes de regressão é o material disponível para adaptação.

Com esse objetivo em mente, faz-se nesse trabalho um estudo considerando duas formas de divisão das médias das componentes gaussianas em classes de regressão [5]:

- 1) Divisão por características fonéticas.
- 2) Divisão usando medidas de distância.

É importante observar que as classes fonéticas, bem como suas características acústicas, podem variar de um idioma para outro, e que este estudo foi feito especificamente para o Português Brasileiro. Desta forma, principalmente na divisão por características fonéticas, é possível que as conclusões aqui

apresentadas possam ser diferentes daquelas encontradas em outros trabalhos na literatura, para outras línguas.

A seguir será apresentado o aparato experimental usado para os testes.

III. APARATO EXPERIMENTAL

Nesta seção são apresentados o aparato experimental utilizado bem como os testes realizados e os resultados obtidos.

A. Base de dados

Como material a ser gravado, foram escolhidas frases do trabalho realizado por Alcaim et. al. [7]. Neste, foram criadas 20 listas de 10 frases foneticamente balanceadas, segundo o português falado no Rio de Janeiro. Nestas listas, contou-se 694 palavras distintas.

Para as gravações foram selecionados 40 locutores adultos, sendo 20 homens e 20 mulheres. Cada locutor gravou 40 frases, o que faz com que esta base de dados tenha no total, 1600 locuções. As gravações foram realizadas em ambiente relativamente silencioso, com um microfone direcional de boa qualidade. A taxa de amostragem utilizada foi de 11,025 kHz e resolução de 16 bits [8]. Esta base será chamada neste trabalho de *Base de Treinamento*.

Adicionalmente, para verificar o ganho obtido com o sistema adaptado, foram selecionados mais 4 locutores, 1 do sexo feminino e 3 do sexo masculino, que gravaram 4 vezes todas as 200 frases [9]. Desta forma, esta segunda base de dados tem 800 locuções/locutor (3200 locuções ao todo). Esta segunda base será referida como *Base de Teste*.

A transcrição fonética de ambas as bases de dados foi feita manualmente para cada locução, utilizando programa de visualização gráfica do espectrograma e forma de onda do sinal, e fones de ouvido para audição da mesma. As sub-unidades utilizadas nesta tarefa (36 ao todo) são mostradas na Tabela I.

TABELA I

FONES UTILIZADOS NA TRANSCRIÇÃO FONÉTICA DAS LOCUÇÕES.

Fone	Exemplo	Fone	Exemplo
#	silêncio	g	g orila
a	a çafraão	j	j iló
an	maç ã	k	c achoeira
e	e levador	l	l eão
E	p e le	L	lh ama
en	s en ta	m	m ontanha
i	s i no	n	n évoa
y	fu i	N	i nh ame
in	p in to	p	p oente
o	b o lo	r	ce r a
O	b o la	rr	ce rr ado
on	s om bra	R	ca r ta
u	l u a	s	s apo
un	um	t	t empestade
b	b e la	T	t igela
d	d ádiva	v	v erão
D	d iferente	x	ch ave
f	f eira	z	z abumba

B. Sistema de reconhecimento

Para os testes foi utilizado um software desenvolvido em [8], baseado em modelos ocultos de Markov contínuos. Cada subunidade fonética neste sistema é modelado por um HMM de 3 estados, como mostrado na Figura 2.

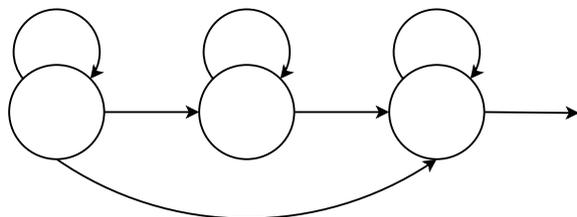


Fig. 2. Estrutura do HMM para uma subunidade fonética.

Este sistema utiliza fones independentes de contexto como unidades fundamentais, o One Pass [10] como algoritmo de busca, e parâmetros mel-cepstrais com suas respectivas derivadas primeira e segunda (parâmetros delta e delta-delta). Para este trabalho foram usadas misturas com 5 gaussianas para cada estado.

Finalmente, como modelo de linguagem, foi adotada uma gramática do tipo bigrama.

IV. EXPERIMENTOS E RESULTADOS

Conforme mencionado anteriormente, a adaptação ao locutor tem como objetivo melhorar o desempenho de um SI para um locutor específico usando algumas (poucas) amostras de voz desse locutor. Após a adaptação espera-se que o Sistema Adaptado (SA) fique com um desempenho melhor do que o sistema SI original.

Obviamente, o limite teórico para o desempenho do sistema adaptado é aquele conseguido pelo SD. Portanto, para efeito de comprovação da eficácia da adaptação MLLR foram testados quatro locutores adultos, sendo três homens (Locutores 1, 2 e 3) e uma mulher (Locutor 4). Esses locutores serão chamados daqui pra frente de locutores de referência.

De modo a verificar o ganho obtido pelas diversas alternativas de adaptação proposta, é necessário inicialmente estabelecer o desempenho do SI e do SD, o que é feito a seguir.

A. Desempenho dos sistemas independente e dependente de locutor

O sistema independente de locutor foi treinado com as 1600 locuções dos 40 locutores da Base de Treinamento, e cada sistema dependente de locutor foi treinado com 600 das 800 locuções da Base de Teste.

Os testes foram resultados com as 200 gravações restantes de cada locutor de referência (Base de Teste), e os resultados, tanto para o SI como para os SDs são mostrados na Tabela II.

Estabelecido o desempenho do SI e dos SDs, serão mostradas a seguir as várias alternativas testadas na criação das classes de regressão para o MLLR.

TABELA II

COMPARAÇÃO ENTRE OS SIS E OS SDs PARA OS QUATRO LOCUTORES DE REFERÊNCIA (TAXA DE ACERTOS EM %).

Locutor 1		Locutor 2		Locutor 3		Locutor 4	
SI	SD	SI	SD	SI	SD	SI	SD
63,30	87,50	90,10	95,10	73,30	94,80	65,20	92,90

B. Divisão apenas uma classe de regressão

Como primeira abordagem, todas as médias foram adaptadas usando a mesma matriz de transformação, ou seja, foi criada apenas uma classe de regressão. Para este teste, bem como para todos os demais a partir de agora, cada locutor adaptou o SI para si usando 4 locuções. Os resultados deste primeiro teste podem ser vistos na Tabela III.

TABELA III

DESEMPENHO DOS SISTEMAS ADAPTADOS USANDO APENAS UMA CLASSE DE REGRESSÃO. OS RESULTADOS DO SI E DOS SDs TAMBÉM SÃO MOSTRADOS PARA FINS DE COMPARAÇÃO.

Locutor 1	Locutor 2	Locutor 3	Locutor 4
SI = 63,30%	SI = 90,10%	SI = 73,30%	SI = 65,20%
SD = 87,50%	SD = 95,10%	SD = 94,80%	SD = 92,90%
67,80%	91,10%	78,80%	68,00%

Verifica-se que mesmo com apenas uma classe de regressão há melhoras no desempenho dos sistemas adaptados em relação àquele observado para o SI, mas este fica ainda abaixo daquele observado nos SDs, como esperado.

C. Divisão com 36 classes de regressão

Neste teste partiu-se para o outro extremo, com cada classe de regressão sendo formada por um único fone. Como foram utilizados 36 fones neste trabalho (ver Tabela I), foram geradas também 36 classes de regressão.

Os resultados obtidos nestes testes foram muito ruins, como mostrado na Tabela IV. Isto se deve provavelmente ao pouco material disponível para a adaptação.

TABELA IV

DESEMPENHO DOS SISTEMAS ADAPTADOS USANDO 36 CLASSES DE REGRESSÃO.

Locutor 1	Locutor 2	Locutor 3	Locutor 4
SI = 63,30%	SI = 90,10%	SI = 73,30%	SI = 65,20%
SD = 87,50%	SD = 95,10%	SD = 94,80%	SD = 92,90%
53,00	76,10	66,60	45,70

D. Divisão entre vogais, consoantes e silêncio

Como segundo teste, pensou-se em dividir os fones em vogais, consoantes e silêncio, formando assim, 3 classes de regressão. É uma divisão um tanto quanto simplista, mas forneceu bons resultados, como pode ser visto na Tabela V.

Nota-se uma melhora no desempenho para alguns locutores e uma piora para outros, quando comparam-se estes resultados com os da Tabela III.

TABELA V

DESEMPENHO DOS SISTEMAS ADAPTADOS USANDO COMO CLASSES DE REGRESSÃO, SILÊNCIO, VOGAIS E CONSOANTES.

Locutor 1	Locutor 2	Locutor 3	Locutor 4
SI = 63,30%	SI = 90,10%	SI = 73,30%	SI = 65,20%
SD = 87,50%	SD = 95,10%	SD = 94,80%	SD = 92,90%
67,20%	89,90%	80,10%	71,60%

E. Divisão por classes fonéticas

Nesta abordagem foram criadas classes de regressão agrupando os fones de acordo com a sua classe fonética. As classes fonéticas utilizadas foram: vogais orais, vogais nasais, plosivas, fricativas, laterais, vibrantes e nasais. O silêncio foi considerado como uma classe separada. Os fones pertencentes a cada classe são mostrados na Tabela VI.

TABELA VI

DIVISÃO FONÉTICA LEVANDO-SE EM CONTA O FONE SILÊNCIO.

Divisão 1		
Classe	Classe Fonética	Fones em cada Classe
1	Silêncio	01 fone - #
2	Vogais Orais	08 fones - a, e, E, i, y, o, O, u.
3	Vogais Nasais	05 fones - an, en, in, on, un
4	Plosivas	08 fones - p, t, T, k, b, d, D, g
5	Fricativas	06 fones - f, s, x, v, z, j
6	Laterais	02 fones - l, L
7	Nasais	03 fones - n, m, N
8	Vibrantes	03 fones - r, rr, R

Os resultados destes testes são mostrados na Tabela VII.

TABELA VII

DESEMPENHO DOS SISTEMAS ADAPTADOS USANDO COMO CLASSES DE REGRESSÃO AS CLASSES FONÉTICAS, SILÊNCIO, VOGAIS ORAIS, VOGAIS NASAIS, PLOSIVAS, FRICATIVAS, LATERAIS, VIBRANTES E NASAIS.

Locutor 1	Locutor 2	Locutor 3	Locutor 4
SI = 63,30%	SI = 90,10%	SI = 73,30%	SI = 65,20%
SD = 87,50%	SD = 95,10%	SD = 94,80%	SD = 92,90%
64,20%	88,50%	77,00%	69,10%

F. Divisão baseada na fonética acústica

As classes são definidas segundo uma classificação baseada na fonética acústica [11], onde é levada em conta a configuração do trato vocal. Os fones colocados em cada classe segundo este critério são mostrados na Tabela VIII. Observe que neste caso o fone silêncio não foi adaptado.

E os resultados dos testes usando-se esta abordagem são mostrados na Tabela IX

Verifica-se que, em geral, os resultados obtidos com esta abordagem foram um pouco inferiores àqueles obtidos com a abordagem anterior.

G. Divisão baseada em medidas de distância

A criação de classes usando a classificação fonética levam em conta similaridades em relação à forma de produção dos sons da fala. Entretanto isto não se reflete diretamente na

TABELA VIII

DIVISÃO DAS CLASSES BASEADA NA FONÉTICA ACÚSTICA.

Divisão 2		
Classe	Classe Fonética	Fones em cada Classe
1	Vogais Anteriores	06 fones - i, y, e, E, en, in
2	Vogais Médias	02 fones - a, an
3	Vogais Posteriores	05 fones - o, O, u, on, un
4	Consoantes Labiais	05 fones - p, b, m, f, v
5	Consoantes Médias	07 fones - t, T, d, D, n, s, z
6	Consoantes Posteriores	05 fones - k, g, N, x, j
7	Laterais	02 fones - l, L
8	Vibrantes	03 fones - r, rr, R

TABELA IX

DESEMPENHO DOS SISTEMAS ADAPTADOS USANDO CLASSES DE REGRESSÃO BASEADAS NA FONÉTICA ACÚSTICA.

Locutor 1	Locutor 2	Locutor 3	Locutor 4
SI = 63,30%	SI = 90,10%	SI = 73,30%	SI = 65,20%
SD = 87,50%	SD = 95,10%	SD = 94,80%	SD = 92,90%
62,60%	87,10%	77,80%	65,30%

proximidade das características acústicas dos mesmos. Desta forma pensou-se em criar as classes de regressão baseadas nas distâncias entre as médias das gaussianas de cada fone, independentemente de sua classe fonética.

O agrupamento dos fones em classes foi feito usando o algoritmo LBG para quantização vetorial [12]. Foram testadas duas métricas para a determinação das classes: a distância euclidiana e a distância Bhattacharya [13]. Para efeito de comparação com os testes anteriores foram geradas 8 classes para cada métrica. Também neste caso, o fone silêncio não foi adaptado.

Usando a distância euclidiana, foram geradas as classes mostradas na Tabela X.

TABELA X

DIVISÃO USANDO A DISTÂNCIA EUCLIDIANA.

Classe de Regressão	Fones em cada Classe
1	03 fones - x, T, s
2	05 fones - D, f, j, y, t
3	04 fones - i, N, in, en
4	04 fones - L, v, e, d
5	05 fones - E, a, r, n, R
6	04 fones - k, rr, p, un
7	05 fones - O, o, on, an,u
8	05 fones - m, b, g, l, z

Já com a distância Bhattacharya foram geradas as classes de regressão mostradas na Tabela XI.

Os resultados destes dois testes são mostrados na Tabela XII.

Observa-se que, à exceção do Locutor 1, a divisão das classes de regressão usando a distância Bhattacharya apresenta um resultado melhor do que aquela que utiliza a distância euclidiana.

H. Resumo dos resultados

De forma a visualizar melhor os resultados obtidos e poder fazer uma análise comparativa, todos os resultados acima são

TABELA XI
DIVISÃO USANDO A DISTÂNCIA BHATTACHARYA.

Classe de Regressão	Fones em cada Classe
1	04 fones - in, en, i, n
2	05 fones - o, Ô, on, u, un
3	04 fones - L, e, R, m
4	05 fones - x, s, T, D, j
5	03 fones - a, an, rr
6	03 fones - N, E, l
7	06 fones - f, p, v, z, b, d
8	05 fones - k, t, g, y, r

TABELA XII
DESEMPENHO DOS SISTEMAS ADAPTADOS USANDO CLASSES DE REGRESSÃO BASEADAS NA DISTÂNCIA EUCLIDIANA (DE) E NA DISTÂNCIA BHATTACHARYA (DB).

	Locutor 1	Locutor 2	Locutor 3	Locutor 4
	SI = 63,30% SD = 87,50%	SI = 90,10% SD = 95,10%	SI = 73,30% SD = 94,80%	SI = 65,20% SD = 92,90%
DE	62,90%	85,20%	78,20%	62,30%
DB	57,00%	87,10%	78,50%	65,60%

agrupados na Tabela XIII.

TABELA XIII
QUADRO COMPARATIVO DO DESEMPENHO DE TODOS OS SISTEMAS TESTADOS (TAXA DE ACERTOS EM %). NESTA, DIV1 É A DIVISÃO DE CLASSES USANDO A CLASSIFICAÇÃO FONÉTICA, E DIV2 É A DIVISÃO QUE USOU COMO CRITÉRIOS OS CONCEITOS DE FONÉTICA ACÚSTICA.

Classes	Locutor 1	Locutor 2	Locutor 3	Locutor 4
	SI = 63,30 SD = 87,50	SI = 90,10 SD = 95,10	SI = 73,30 SD = 94,80	SI = 65,20 SD = 92,90
1	67,80	91,10	78,80	68,00
36	53,00	76,10	66,60	45,70
3	67,20	89,90	80,10	71,60
8 (Div1)	64,20	88,50	77,00	69,10
8 (Div2)	62,60	87,10	77,80	65,30
DE	62,90	85,20	78,20	62,30
DB	57,00	87,10	78,50	65,60

Como conclusão geral deste estudo, pode-se dizer que, para a quantidade de material de adaptação utilizado (4 frases), a melhor escolha é criar apenas 3 classes de regressão: vogais, consoantes e silêncio. Obviamente este resultado pode mudar, dependendo da quantidade de material disponível para a adaptação.

V. CONCLUSÕES

Este trabalho apresentou um estudo sobre diversas alternativas para se criar as classes de regressão para o algoritmo MLLR de adaptação ao locutor.

Para a adaptação foram usadas apenas 4 locuções, o que é bastante razoável supor para aplicações em que não haja material abundante à disposição. Neste cenário, os melhores resultados foram obtidos usando-se apenas 3 classes de regressão: vogais, consoantes e silêncio.

Obviamente a quantidade de material de adaptação e o número de classes de regressão estão diretamente relacionados. Para se conseguir um SA com desempenho superior ao SI deve existir um compromisso entre esses dois parâmetros.

Outra conclusão importante é que determinados fones influenciam mais no processo de adaptação que outros. A adaptação de um SI utilizando-se a mesma quantidade de material de adaptação, o mesmo número de frases, e ainda o mesmo número de classes de regressão, apresentou resultados diferentes, conforme mostrado na Tabela XIII.

Algumas sugestões para trabalhos futuros são:

- Comparação das medidas de distância utilizadas considerando mais classes de regressão;
- Um estudo do material de adaptação na escolha dos fones que têm maior influência no processo de adaptação.

REFERÊNCIAS

- [1] LEE, L. and ROSE, R., A Frequency Warping Approach to Speaker Normalization. IEEE Transactions on Communications, COM-28(1), pages 49-60, January 1980.
- [2] WOODLAND, P., Speaker Adaptation: Techniques and Challenges. Proceedings IEEE Automatic Speech Recognition and Understanding Workshop, pages 85-90, Colorado 2000.
- [3] LEGGETTER, C.J. and WOODLAND, P.C., Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Markov Models. Computer Speech and Language, volume 09 (no 02): pages 171-185, April 1995.
- [4] LEGGETTER, C.J. and WOODLAND, P.C., Speaker Adaptation of HMM's Using Linear Regression. Cambridge University, Technical Report, TR.181, June, 1994.
- [5] LEGGETTER, C.J., Improved Acoustic Modelling for HMM's using Linear Transformations. Ph.D. Thesis, Cambridge University, 1995.
- [6] LEGGETTER, C.J. and WOODLAND, P.C., Flexible Speaker Adaptation for Large Vocabulary Speech Recognition. Proceedings EUROSPEECH95, pages 1155-1158, 1995.
- [7] ALCAIM, A., SOLEWICZ, J. A., MORAES, J. A., Frequência de ocorrência dos fones e lista de frases foneticamente balanceadas no português falado no Rio de Janeiro. Revista da Sociedade Brasileira de Telecomunicações, 7(1): p 23-41. Dezembro, 1992.
- [8] YNOGUTI, C. A., Reconhecimento de Fala Contínua Usando Modelos Ocultos de Markov. Tese de Doutorado, Universidade Estadual de Campinas, Maio 1999.
- [9] FERNANDES, D. B. Adaptação ao locutor usando a técnica MLLR. Dissertação de Mestrado, Inatel, 2010.
- [10] NEY, H. The use of a one-stage dynamic programming algorithm for connected word recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-32(2), april, 1984.
- [11] CALLOU, D. e LEITE, Y. Iniciação à fonética e à fonologia. Rio de Janeiro : Jorge Zahar, 1995.
- [12] LINDE, Y., BUZO, A., GRAY, R. M. An algorithm for vector quantizer design. IEEE Transactions on Communications. COM-28(1). January, 1980.
- [13] KAILATH, T., The Divergence and Bhattacharya Distance Measures in Signal Selection, IEEE Transactions on Communication Technology, 15-1, February 1967, pp. 52-60.