

ANÁLISE COMPARATIVA DE PARÂMETROS EM SISTEMAS DE RECONHECIMENTO DE VOZ

AMARO A. DE LIMA, MARCOS S. FRANCISCO, SERGIO L. NETTO E FERNANDO G.V. RESENDE JR.

PEE/COPPE/DEL/EE, Universidade Federal do Rio de Janeiro

Cx. P. 68504, Rio de Janeiro, RJ

21945-970, Brasil

{ amaro, salazar, sergioln, gil } @lps.ufrj.br

RESUMO

Análises comparativas de diversos parâmetros em sistemas de reconhecimento de palavras isoladas são apresentadas. A influência da variação dos seguintes parâmetros em um sistema baseado em modelos escondidos de Markov discretos é avaliada: tamanho do *codebook*, número de estados, número de coeficientes cepstrais por segmento, tamanho da superposição das janelas, tamanho das janelas. Resultados obtidos através de diferentes algoritmos de extração de coeficientes mel-cepstrais também são mostrados a partir de um sistema com ajuste temporal dinâmico. Nas simulações com o sistema baseado em modelos de Markov os índices de acerto obtidos são 98,75% e 86,25% para sistemas dependentes e independentes do locutor, respectivamente. Nas simulações com ajuste temporal dinâmico o índice de acerto foi de 92,7% para um sistema dependente do locutor.

Palavras Chaves — Reconhecimento de voz, HMM, DTW, análise comparativa de parâmetros.

1. INTRODUÇÃO

O reconhecimento de voz é um campo que tem crescido muito recentemente. O interesse vem tanto de questões teóricas ainda sem resposta, como da infinidade de aplicações. Embora a literatura seja rica na descrição de algoritmos [1,2], a quantidade de trabalhos que mostram a influência dos diversos parâmetros envolvidos num sistema de reconhecimento de voz é relativamente pequena.

Neste artigo é feita uma análise comparativa dos parâmetros de sistemas típicos de reconhecimento de voz. A influência de cada parâmetro é obtida a partir de resultados com sistemas baseados em HMMs (*hidden Markov models*) e ajuste temporal dinâmico, ou DTW (*dynamic time warping*) [1,2].

Na implementação com HMMs, diversos resultados são apresentados variando-se tamanho do *codebook*, número de estados, número de coeficientes cepstrais por segmento, tamanho da superposição das janelas e tamanho das janelas. Os percentuais de acerto numa aplicação de reconhecimento de dez dígitos falados isoladamente para a melhor escolha dos

parâmetros mencionados acima são 98,75% e 86,25% para sistemas dependentes e independentes do locutor, respectivamente.

Os algoritmos de extração de coeficientes mel-cepstrais descritos em [1-3] são implementados e seus resultados são comparados usando DTW. Resultados com e sem a extração do valor médio de cada coeficiente são apresentados. A utilização do algoritmo LPC-MEL [3] (*linear prediction coefficients to mel-cepstral coefficients*) implica em uma melhoria significativa na performance do sistema, em comparação com os resultados obtidos com algoritmos MFCC (*mel frequency cepstral coefficients*) descritos em [1,2].

Este artigo é organizado da seguinte forma. Na Seção 2 os resultados baseados na implementação usando HMMs são analisados. Na Seção 3 algoritmos de extração de coeficientes mel-cepstrais são comparados a partir de uma implementação de DTW. A Seção 4 apresenta conclusões e direções para futuros trabalhos.

2. ANÁLISE COMPARATIVA BASEADA EM IMPLEMENTAÇÃO COM HMM DISCRETO

Existem diversos fatores envolvidos num processo de reconhecimento de fala. Nesta seção, são mostrados resultados onde é feita uma análise da influência de alguns destes fatores no reconhecimento de dígitos isolados (0 a 9).

2.1 Descrição do Sistema

A implementação desse sistema está descrita com detalhes em [4]. O sistema é constituído por quatro etapas: detecção de extremos; extração de parâmetros; quantização vetorial; classificação do sinal.

A detecção de extremos é realizada pelo algoritmo apresentado em [5,6]. O vetor de parâmetros extraído em cada segmento é composto por coeficientes cepstrais, delta cepstrais, energia e delta energia. A quantização vetorial é feita pelo algoritmo LBG com *centroid splitting* [1,2]. Os modelos HMM têm estrutura *left-right*.

A base de dados utilizada para treino dos modelos e testes está descrita na Tabela 1. A base "Teste 1" é formada pelos mesmos locutores que fizeram o treinamento e a base "Teste 2" é formada por locutores diferentes daqueles que fizeram o treinamento.

Todos os sons foram gravados em mono, com 16 bits e frequência de amostragem de 11025 Hz com microfone Lucky DM-202.

Tabela 1 : Descrição das bases de dados utilizadas no treino e no teste.

		Locutores (H/M)	Palavras / dígito / pessoa	Total
Base 1	Treino	1	20	200
	Teste		40	400
Base 2	Treino	7/2	3	510
	Teste 1		2	340
	Teste 2	6/2	5	400

As simulações foram realizadas em um computador Pentium 100MHz com 40Mb de RAM em ambiente Windows 98 / Delphi 3 / Borland C 5.

2.2 Resultados para o sistema dependente do locutor

Os resultados são apresentados a partir da Tabela 2 até a Tabela 7. Em cada teste é alterada uma das variáveis e as demais são mantidas constantes. A Tabela 2 mostra a porcentagem de acerto de cada dígito (de 0 a 9), o tempo total de processamento (40 amostras por dígito), e o percentual médio de acerto para todos os dígitos em função do tamanho do codebook, mantendo as demais variáveis envolvidas constantes. Na Tabela 3, observa-se outro teste, no qual se varia o número de estados. Nas Tabelas 4, 5 e 6, são mostrados resultados em função do número de coeficientes cepstrais, tamanho da superposição e da janela, respectivamente. Todas as tabelas são relativas às 400 amostras de teste.

Em todos os testes, aparece o tempo gasto no reconhecimento de todas as 40 amostras de cada dígito, num total de 400 amostras. Percebeu-se que o tamanho do *codebook* e o tamanho da superposição são os fatores que mais afetam o tempo de processamento. E também, para 512 e 1024 centróides, o tempo médio gasto para reconhecer um dígito é de 180 ms e 330 ms, respectivamente. Esse tempo é da mesma ordem que o tamanho das palavras, pois a palavra de menor tamanho possui 220 ms.

Os tamanhos 128 e 256 para o *codebook* apresentaram resultados equivalentes do ponto de vista estatístico, mas o tempo gasto no processamento para 128 é bem inferior.

Para 7 ou 8 estados nos modelos HMM, verificou-se que dentre todas as seqüências de estados possíveis, alguns estados não possuíam ocorrências. Isso mostra que o número de estados deve ser menor que 7 para esta aplicação.

Para esta base de dados, a melhor taxa de acerto no reconhecimento foi de 98,75% para os seguintes valores: tamanho da janela: de 18 ms a 22 ms; superposição: 66% e 50%; número de coeficientes cepstrais: 12 a 14; tamanho do *codebook*: 128; número de estados: 3 e 4.

A Tabela 7 mostra dígitos reconhecidos em função dos dígitos apresentados. Percebe-se que o maior problema no reconhecimento é para os dígitos "três", "seis" e "sete".

Tabela 2 : Porcentagens de acerto no reconhecimento em função do tamanho do *codebook*. Para janela de 20 ms, superposição de 66%, 4 estados nos modelos HMM e 18 parâmetros.

Tam. Cbk	32	64	128	256	512	1024
0	97,5	97,5	100	97,5	92,5	95
1	85	87,5	85	87,5	92,5	95
2	97,5	100	100	100	95	95
3	97,5	95	100	95	95	90
4	90	72,5	97,5	95	95	97,5
5	100	100	100	100	100	100
6	92,5	92,5	95	92,5	95	90
7	87,5	90	95	100	92,5	100
8	100	97,5	100	100	97,5	100
9	77,5	100	95	92,5	85	92,5
Tempo(Seg)	17,3	20,9	28,1	42,2	72,2	132
Média	92,5	93,25	96,75	96	94	95,5

Tabela 3 : Porcentagens de acerto no reconhecimento em função do número de estados. Para janela de 20 ms, superposição de 66%, 128 centróides e 18 parâmetros.

Núm. Estados	3	4	5	6	7	8
0	97,5	100	97,5	100	97,5	100
1	90	85	82,5	77,5	92,5	85
2	100	100	97,5	100	100	100
3	95	100	97,5	92,5	95	95
4	95	97,5	97,5	95	95	85
5	100	100	100	100	100	100
6	95	95	92,5	97,5	95	92,5
7	92,5	95	95	92,5	92,5	97,5
8	100	100	100	100	100	100
9	97,5	95	90	95	92,5	100
Tempo(Seg)	27,6	28,1	28,7	29,5	30,3	31,2
Média	96,25	96,75	95	95	96	95,5

Tabela 4 : Porcentagens de acerto no reconhecimento em função do número de coeficientes. Para janela de 20 ms, superposição de 66%, 128 centróides e 4 estados nos modelos HMM.

N.	8	9	10	11	12	13	14	15	16
0	100	95	97,5	97,5	100	100	100	100	100
1	85	87,5	95	90	100	100	100	100	100
2	100	97,5	97,5	97,5	100	97,5	100	100	100
3	100	100	95	100	95	95	100	97,5	97,5
4	97,5	90	100	100	100	97,5	100	97,5	100
5	100	100	100	100	100	100	100	100	100
6	95	97,5	90	87,5	95	95	92,5	90	90
7	95	90	92,5	90	100	100	97,5	92,5	90
8	100	100	100	100	97,5	100	100	100	100
9	95	95	97,5	95	95	97,5	97,5	100	100
T(s)	28,1	30,7	33,2	35,6	37,7	40,4	43,1	45,6	47,9
Med.	96,75	95,25	96,5	95,75	98,25	98,25	98,75	97,75	97,75

Tabela 5 : Porcentagens de acerto no reconhecimento em função do tamanho da superposição. Para janela de 20 ms, 128 centróides, 4 estados nos modelos HMM e 18 parâmetros.

Sup.	0%	33%	50%	66%
0	100	97,5	100	100
1	97,5	97,5	100	100
2	100	97,5	97,5	100
3	92,5	97,5	100	100
4	97,5	97,5	100	100
5	100	100	100	100
6	82,5	92,5	97,5	92,5
7	90	92,5	97,5	97,5
8	100	100	95	100
9	100	100	97,5	97,5
Tempo(Seg)	16,2	23,9	30,5	43,1
Média	96	97,25	98,5	98,75

Tabela 6 : Porcentagens de acerto no reconhecimento em função do tamanho da janela. Para janela de 20 ms, superposição de 66%, 4 estados nos modelos HMM e 18 parâmetros.

Tam. Janela	15 ms	18 ms	20 ms	22 ms	25 ms
0	100	100	100	100	100
1	95	100	100	100	100
2	97,5	100	100	97,5	97,5
3	100	97,5	100	92,5	97,5
4	97,5	100	100	100	100
5	100	100	100	100	100
6	92,5	92,5	92,5	97,5	87,5
7	87,5	95	97,5	100	95
8	97,5	97,5	100	97,5	100
9	100	100	97,5	100	100
Tempo(Seg)	55,2	48,3	43,1	40,9	37,1
Média	96,75	98,25	98,75	98,5	97,75

Tabela 7 : Tabela de confusão para a melhor taxa de acerto encontrada no teste. Janela de 20 ms, superposição de 66%, 128 centroides, 4 estados nos modelos HMM e 18 parâmetros.

		Dígito reconhecido (Base 1)										Média	
		0	1	2	3	4	5	6	7	8	9		
Dígito apresentado	0	40										100	
	1		40									100	
	2			40								100	
	3				40							100	
	4					40						100	
	5						40					100	
	6							2			37	1	92,5
	7								1			39	97,5
	8										40		100
	9											39	97,5
											98,75		

Tabela 8 : Porcentagens de acerto no reconhecimento em função do tamanho do *codebook*. Para janela de 20 ms, superposição de 66%, 4 estados nos modelos HMM e 30 parâmetros.

Tam.	128		256		512	
	Tes1	Tes2	Tes1	Tes2	Tes1	Tes2
0	100	77,5	100	92,5	94,44	95
1	66,66	27,5	88,89	40	100	47,5
2	100	67,5	100	77,5	100	60
3	100	87,5	83,33	65	83,33	62,5
4	100	90	100	100	100	95
5	100	100	100	97,5	100	90
6	88,89	82,5	77,78	52,5	100	70
7	100	77,5	100	90	100	85
8	94,44	90	100	90	100	95
9	83,33	72,5	94,44	77,5	100	70
T(Seg)	28,7	70,4	43,5	104,4	74,1	180
Média	93,33	77,25	94,44	78,25	97,78	77

Tabela 9 : Porcentagens de acerto no reconhecimento em função do número de estados. Para janela de 20 ms, superposição de 66%, 256 centróides e 30 parâmetros.

	3		4		5		6	
	Tes1	Tes2	Tes1	Tes2	Tes1	Tes2	Tes1	Tes2
0	94,44	90	100	92,5	100	80	100	87,5
1	88,89	32,5	88,89	40	100	37,5	88,89	35
2	100	70	100	77,5	100	72,5	100	60
3	83,33	62,5	83,33	65	77,78	62,5	88,89	70
4	100	100	100	100	100	100	100	100
5	100	100	100	97,5	100	100	100	100
6	94,44	77,5	77,78	52,5	88,89	72,5	83,33	80
7	100	85	100	90	100	90	100	90
8	100	90	100	90	100	90	100	90
9	100	77,5	94,44	77,5	100	82,5	100	75
T(Seg)	42,9	103	43,5	104,4	44,1	105,9	44,6	106
M.	96,11	78,5	94,44	78,25	96,67	78,75	96,11	78,75

Tabela 10 : Porcentagens de acerto no reconhecimento em função do número de coeficientes. Para janela de 20 ms, superposição de 66%, 256 centróides e 5 estados nos modelos HMM.

	14		16		18		20	
	Tes1	Tes2	Tes1	Tes2	Tes1	Tes2	Tes1	Tes2
0	100	80	100	92,5	94,44	90	100	95
1	100	37,5	94,44	37,5	88,89	45	100	50
2	100	72,5	100	80	100	72,5	100	65
3	77,78	62,5	94,44	77,5	88,89	70	72,22	72,5
4	100	100	100	100	100	100	100	92,5
5	100	100	100	97,5	100	97,5	100	92,5
6	88,89	72,5	88,89	75	100	85	100	65
7	100	90	100	77,5	100	87,5	100	80
8	100	90	100	90	100	100	100	100
9	100	82,5	100	87,5	100	87,5	100	85
T(Seg)	44,1	105,9	48,7	116,8	54,5	130,8	59,1	141,7
M.	96,67	78,75	97,78	81,5	97,22	83,5	97,22	79,75

2.3 Resultados para o sistema independente do locutor

Em cada teste é alterada uma das variáveis e as demais são mantidas constantes da mesma forma que na Seção 2.2. A Tabela 8 mostra a taxa de acerto em função do tamanho do *codebook*, mantendo as demais variáveis envolvidas constantes. Na Tabela 9, observa-se outro teste, no qual se varia o número de estados. Nas Tabelas 10 e 11, são mostrados resultados em função do número de coeficientes cepstrais e do tamanho da superposição entre as janelas, respectivamente.

Os três tamanhos para o *codebook* apresentaram resultados equivalentes do ponto de vista estatístico, mas o tempo gasto no reconhecimento e no treinamento cresce significativamente com o número de centróides. Então, optou-se por utilizar nos próximos testes o valor de 256.

Na Tabela 9, observa-se que a taxa de acerto e o tempo de processamento são praticamente constantes. Porém, para 6 estados nos modelos HMM, verificou-se que dentre todas as seqüências de estados possíveis, alguns estados possuíam poucas ou nenhuma ocorrência. Então, nos próximos testes, optou-se por 5 estados.

Em relação a Tabela 10, um aumento de 2 coeficientes proporciona um aumento de cerca de 10% no tempo de processamento. Tanto para a base de treino quanto para a base de teste "Teste 1", a taxa de acerto foi a mesma. Porém, o melhor resultado obtido para a base "Teste 2" foi com 18 coeficientes, sendo o número total de parâmetros por segmento igual a 38 (18 coeficientes cepstrais, 18 delta cepstrais, 1 coeficiente de energia e 1 de delta energia), mas o resultado foi bem próximo para 16 coeficientes, sendo o número total de parâmetros por segmento igual a 34.

Analisando a superposição entre as janelas, a taxa de acerto média com a base "Teste 2" é melhor quando diminui-se a superposição. Entretanto, para a base "Teste 1", ocorre o contrário, o melhor resultado é para 66% de superposição.

Tabela 11 : Porcentagens de acerto no reconhecimento em função do tamanho da superposição. Para janela de 20 ms, 256 centróides, 5 estados nos modelos HMM e 38 parâmetros.

	0%		33%		50%		66%	
	Tes1	Tes2	Tes1	Tes2	Tes1	Tes2	Tes1	Tes2
0	100	90	94,44	90	100	92,5	94,44	90
1	83,33	60	88,89	60	100	57,5	88,89	45
2	100	85	100	65	100	55	100	72,5
3	88,89	80	94,44	62,5	77,78	57,5	88,89	70
4	100	100	100	100	100	92,5	100	100
5	100	100	100	100	100	97,5	100	97,5
6	83,33	65	88,89	85	83,33	77,5	100	85
7	100	92,5	100	87,5	100	87,5	100	87,5
8	100	95	100	97,5	100	100	100	100
9	94,44	95	100	97,5	94,44	85	100	87,5
T(Seg)	19,4	46,3	29,3	70,8	38,1	91,3	54,5	130,8
M.	95	86,25	96,67	84,5	95,56	80,25	97,22	83,5

As Tabelas 12 e 13 mostram dígitos reconhecidos em função dos dígitos apresentados. Assim como no caso dependente do locutor, percebe-se que os maiores problemas no reconhecimento são observados para os dígitos "dois", "três", "seis" e "sete". Outro problema é em relação ao dígito "um" ser confundido com o dígito "cinco".

Tabela 12 : Tabela de confusão para a melhor taxa de acerto encontrada no teste para a Base 2/Teste 1. Janela de 20 ms, superposição de 66%, 256 centróides, 5 estados no modelo HMM e 38 parâmetros.

	Dígito reconhecido (Base 2 - Teste 1)										Média	
	0	1	2	3	4	5	6	7	8	9		
Dígito apresentado	0	18										100
	1	1	15		1					1		83,33
	2			18								100
	3			1	16			1				88,89
	4					18						100
	5						18					100
	6			1	1		1	15				83,33
	7								18			100
	8									18		100
	9			1							17	94,44
												95

Tabela 13 : Tabela de confusão para a melhor taxa de acerto encontrada no teste para Base2/Teste2. Janela de 20 ms, superposição de 0%, 256 centróides, 5 estados no modelo HMM e 38 parâmetros.

	Dígito reconhecido (Base 2 - Teste 2)										Média	
	0	1	2	3	4	5	6	7	8	9		
Dígito apresentado	0	36							4			90
	1	1	24				12		1	2		60
	2	1		34	1			1	1	1	1	85
	3	1			32			2	3	2		80
	4					40						100
	5						40					100
	6				3		3	26	7	1		65
	7	2						1	37			92,5
	8					2				38		95
	9					2					38	95
												86,25

2.4 Resumo dos sistemas correspondentes aos melhores resultados

Todos estes resultados foram obtidos com a utilização de um algoritmo detecção de extremos semi-automático, e os melhores resultados para as duas bases de dados utilizadas nesse trabalho podem ser vistos na tabela abaixo.

Tabela 14 : Melhores resultados para a análise com HMM.

Base	1	2
Locutor	Dependente	Independente
Tamanho da janela	18 ms a 22 ms	20 ms
Superposição	66% e 50%	0%
Número de coeficientes	12 a 14	18
Tamanho do <i>codebook</i>	128 e 256	256
Número de estados:	3 e 4	3,4 e 5
Taxa de acerto (treino)	100%	99,62%
Taxa de acerto (teste 1)	98,75%	95%
Taxa de acerto (teste 2)	-	86,25%
Tempo médio por dígito	108 ms	115 ms

3. ANÁLISE COMPARATIVA BASEADA EM IMPLEMENTAÇÃO COM DTW

Nesta seção, a performance de algoritmos de extração de coeficientes mel-cepstrais é avaliada a partir de resultados obtidos no reconhecimento de dígitos isolados (0 a 9) com DTW para um locutor.

3.1 Descrição do Sistema

O sistema implementado é apresentado em [7]. O algoritmo utilizado para detectar o início e fim das elocuições é descrito em [5]. No treinamento são gravados 3 ou 5 sinais para cada dígito, enquanto o teste é realizado com 100 sinais por dígito. Todos os sons foram gravados em mono, 16 bits e com taxa de amostragem de 8KHz, através de um microfone comum (modelo Leardship com pedestal). Janelas de Hamming de 20ms com superposição de 50% são aplicadas em todos os sinais.

Duas implementações de algoritmos de extração de MFCC baseados em [1] e [2], assim como o algoritmo de extração LPC-MEL [3] são avaliadas. Na extração de coeficientes MFCC são utilizados 20 filtros triangulares e FFT com 1024 pontos [8]. Nas implementações apresentadas, para cada segmento são extraídos 8 (ou 15) coeficientes mel-cepstrais, 8 (ou 15) coeficientes delta-mel-cepstrais, energia e delta energia, totalizando 18 (ou 32) parâmetros por segmento.

3.2 Resultados

A Tabela 15 apresenta os resultados obtidos. Cada coluna corresponde a um algoritmo de extração de coeficientes mel-cepstrais. Cada linha está relacionada a quantidade de coeficientes extraídos assim como ao número de sinais de treino, que são utilizados como centróides, por dígito. Por exemplo, na terceira linha são mostrados os percentuais de acerto para os três algoritmos de extração usando 32 parâmetros por segmento e 3 sinais de treinamento por dígito. A retirada do nível médio é feita subtraindo-se de cada coeficiente o valor médio deste, obtido a partir dos vetores extraídos em todos os segmentos. Somente na primeira linha da Tabela 15 os resultados são apresentados sem a retirada do nível médio. O algoritmo LPC-MEL obteve melhor performance do que ambos algoritmos de extração de MFCC.

Tabela 15 : Percentuais de acerto no reconhecimento de dígitos para três implementações de algoritmos de extração de coeficientes mel-cepstrais.

	MFCC[1]	MFCC[2]	LPC-MEL
8 coef. + 3 sinais/dígito (sem retirar o nível médio)	53,6%	80,8%	86,1%
8 coef. + 3 sinais/dígito	63,5%	80,5%	88,8%
15 coef. + 3 sinais/dígito	76,6%	83,5%	90,5%
15 coef. + 5 sinais/dígito	73,3%	84,0%	92,7%

4. CONCLUSÕES

Análises comparativas de parâmetros em sistemas de reconhecimento de voz são apresentadas. Resultados obtidos com um sistema baseado em HMMs demonstram a influência dos valores do tamanho do *codebook* e do segmento, do percentual de superposição e do número de estados e de coeficientes, no desempenho do sistema. Um resumo dos parâmetros que geraram os melhores resultados para esta aplicação é apresentado na Tabela 14. Três implementações de algoritmos de extração de coeficientes mel-cepstrais são analisadas usando DTW. Nos resultados obtidos para esta aplicação, o algoritmo LPC-MEL apresenta melhor performance que os algoritmos MFCC. A implementação de outros algoritmos de extração de coeficientes mel-cepstrais, assim como de coeficientes perceptuais de predição linear são temas de pesquisa em andamento.

5. REFERÊNCIAS

- [1] J. R. Deller, J. G. Proakis, J. H. Hansen, *Discrete Time Processing of Speech Signals*, New York: Macmillan, 1993.
- [2] L. R. Rabiner, B. Juang, *Fundamentals on Speech Recognition*, New Jersey: Prentice Hall, 1996.
- [3] T. Fukada, K. Tokuda, S. Imai, "An Adaptive Algorithm for Mel-Cepstral Analysis of Speech", in *Proc. ICASSP-92*, pp. I-137-I-140, 1992.
- [4] M. Salazar, "Sistema de Reconhecimento de Palavras Isoladas em Tempo Real", *Projeto Final de Curso – Dep. de Eng. Eletrônica*, Dezembro 1999. Universidade Federal do Rio de Janeiro.
- [5] L. R. Rabiner, M. R. Sambur, "An Algorithm for determining the endpoints of isolated utterances", *Bell System Technical Journal*, vol. 54, pp. 297-315, Fevereiro 1975.
- [6] E. J. Caprini, L. G. R. Carneiro, "Sistema de Reconhecimento Automático de Voz", *Projeto Final de Curso – Dep. de Eng. Eletrônica*, Outubro 1998. Universidade Federal do Rio de Janeiro.
- [7] A. A. Lima, "Reconhecimento de Dígitos Isolados usando DTW", *Projeto Final de Curso – Dep. de Eng. Eletrônica*, Setembro 1999. Universidade Federal do Rio de Janeiro.
- [8] C. Espain, R. Fernandes, "Isolated Digit Recognition in Konkani", *Relatório técnico. Fac. de eng. da Universidade do Porto e Goa University – Dep. Of English*.