

## **Treinamento multi-estilo e adaptação de modelos via MAP para reconhecimento de fala em ambientes ruidosos**

**Thales Antonio Fernandes Valerio**

**Dezembro/2011**

# Treinamento multi-estilo e adaptação de modelos via MAP para reconhecimento de fala em ambientes ruidosos

THALES ANTONIO FERNANDES VALERIO

Dissertação apresentada ao Instituto Nacional de Telecomunicações, como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica.

Orientador: PROF. DR. CARLOS ALBERTO YNOGUTI

Santa Rita do Sapucaí  
2011

Valério, Thales Antonio Fernandes

V213t

Treinamento multi-estilo e adaptação de modelos via MAP para reconhecimentos de fala em ambientes ruidosos. / Thales Antonio Fernandes Valério. – Santa Rita do Sapucaí, 2011.

98 p.

Orientador: Prof. PhD. Carlos Alberto Ynoguti

Dissertação de Mestrado – Engenharia de Telecomunicações – Instituto Nacional de Telecomunicações – INATEL.

Inclui bibliografia e anexo.

1. Reconhecimento de fala 2. Modelos ocultos de Markov 3. Compensação de modelo baseada em ruído 4. Adaptação 5. Treinamento adaptativo. Engenharia de Telecomunicações. I. Ynoguti, Carlos Alberto . II. Instituto Nacional de Telecomunicações – INATEL. III. Título.

CDU 621.39

Dissertação defendida e aprovada em 16 de dezembro de 2011, pela comissão julgadora:

Prof. Dr. Aldebaro Barreto da Rocha Klautau Júnior - UFPA

Prof. Dr. José Marcos Câmara Brito - Inatel

Prof. Dr. Carlos Alberto Ynoguti - Inatel (Orientador)

---

Prof. Dr. Luciano Leonel Mendes - Inatel  
Coordenador do Curso de Mestrado

– “Mastery of language affords remarkable power”

– Frantz Fanon

# Agradecimentos

Ao Professor Carlos Alberto Ynoguti pela excelente orientação fornecida durante a elaboração deste trabalho.

A todos os colegas, professores e funcionários do Departamento de Telecomunicações do Instituto Nacional de Telecomunicações pela amizade.

# Índice

Lista de Figuras	vi
Lista de Abreviaturas e Siglas	xiii
<b>1 Introdução</b>	<b>1</b>
1.1 Objetivos e contribuições desta tese . . . . .	3
1.2 Organização do trabalho . . . . .	4
<b>2 Os efeitos do ruído no sinal de fala</b>	<b>5</b>
2.1 Modelo do ambiente . . . . .	5
2.2 Efeito do ruído na distribuição de probabilidade do sinal de voz . . . . .	8
<b>3 Métodos para reconhecimento robusto de fala</b>	<b>10</b>
3.1 Treinamento multi-estilo . . . . .	10
3.2 Adaptação e normalização . . . . .	11
3.3 Treinamento adaptativo . . . . .	12
3.4 Adaptação de modelo utilizando MAP . . . . .	12
<b>4 Resultados Experimentais</b>	<b>15</b>
4.1 Base de dados . . . . .	15
4.1.1 Corpus de fala limpa . . . . .	15
4.1.2 Corpus de fala ruidosa . . . . .	16
4.2 Sistema de reconhecimento de fala . . . . .	16
4.3 Métrica para avaliação de desempenho . . . . .	17
4.4 Resultados . . . . .	17
4.4.1 O sistema base . . . . .	18
4.4.2 Sistema treinado com locuções ruidosas . . . . .	19
4.4.3 Sistemas adaptados com locuções ruidosas . . . . .	22
<b>5 Conclusões</b>	<b>28</b>
5.1 Trabalhos Futuros . . . . .	30
5.2 Considerações Finais . . . . .	30

---

<b>A Resultados para sistema treinado com ruído</b>	<b>32</b>
<b>B Resultados para sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando multi-estilo</b>	<b>36</b>
<b>C Resultados para sistema treinado com locuções limpas adaptado utilizando multi-estilo</b>	<b>53</b>
<b>D Artigo resultante da tese</b>	<b>70</b>
<b>Bibliografia</b>	<b>75</b>



# Lista de Figuras

1.1	Modelagem em blocos dos ruídos e distorções que afetam a locução.	3
2.1	Modelagem em blocos dos ruídos e distorções que afetam a locução.	6
2.2	Efeito aditivo do ruído, linha sólida é a distribuição resultante e a linha pontilhada é a distribuição Gaussiana. . . . .	8
4.1	Modelo de Markov para um fone. . . . .	16
4.2	Gráfico comparativo entre o sistema treinado com locuções ruidosas (Tabela 4.6) e o sistema adaptado com locuções ruidosas (Tabela 4.14) para o caso do ruído aeroporto . . . . .	24
4.3	Gráfico comparativo entre o sistema treinado com locuções ruidosas (Tabela 4.6) e o sistema adaptado com locuções ruidosas (Tabela 4.14) para o caso do ruído balbulcio . . . . .	25
4.4	Gráfico comparativo entre o sistema treinado com locuções ruidosas (Tabela 4.6) e o sistema adaptado com locuções ruidosas (Tabela 4.14) para o caso do ruído restaurante . . . . .	26
4.5	Gráfico comparativo entre o sistema treinado com locuções ruidosas (Tabela 4.6) e o sistema adaptado com locuções ruidosas (Tabela 4.14) para o caso do ruído rua . . . . .	26
4.6	Gráfico comparativo entre o sistema treinado com locuções ruidosas (Tabela 4.6) e o sistema adaptado com locuções ruidosas (Tabela 4.14) para o caso do ruído metrô . . . . .	27
4.7	Gráfico comparativo entre o sistema treinado com locuções ruidosas (Tabela 4.6) e o sistema adaptado com locuções ruidosas (Tabela 4.14) para o caso do ruído trem . . . . .	27

# Lista de Tabelas

4.1	Taxa de acertos de um sistema treinado com locuções limpas e testado por um corpus contaminado por ruído. Este mesmo sistema obteve uma taxa de acertos de 75% para locuções limpas. . . . .	18
4.2	Taxa de acertos de um sistema treinado com locuções ruidosas (SNR 20) para os tipos de ruídos apresentados nas colunas da tabela testados com locuções limpas . . . . .	18
4.3	Taxa de acertos de um sistema treinado com locuções ruidosas com SNR 20 e 15 para os tipos de ruídos apresentados nas colunas da tabela testados com locuções limpas . . . . .	18
4.4	Taxa de acertos de um sistema treinado com todos os tipos de ruído com valores de SNR apresentados nas colunas da tabela, testados com locuções limpas. . . . .	19
4.5	Taxa de acertos de um sistema treinado com locuções contaminadas com todos os ruídos e todos os níveis de ruído (SNR) e testado por um corpus contaminado por ruído onde cada coluna representa a SNR do material de teste. . . . .	19
4.6	Taxa de acertos de um sistema treinado com todos os tipos de ruído mas com os níveis de ruído somente de 15 dB e 20 dB e testado com locuções ruidosas. . . . .	20
4.7	Taxa de acertos de um sistema treinado e testado com locuções do mesmo tipo que o ruído descrito. Como antes, apenas locuções com SNR = 15 dB e SNR = 20 dB foram utilizados para treinar o sistema. . . . .	20
4.8	Análise do erro introduzido ao não identificar corretamente o ruído presente no áudio ruidoso. . . . .	21
4.9	Análise global do erro introduzido ao não identificar corretamente o ruído presente no áudio ruidoso. . . . .	21
4.10	Taxa de acertos de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções limpas e com $\alpha$ variando de 0,1 a 0,9	22

4.11	Taxa de acertos de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções limpas e com $\alpha$ variando de 0,01 a 0,09 . . . . .	23
4.12	Taxa de acertos de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções ruidosas de aeroporto e com $\alpha$ variando de 0,01 a 0,09 . . . . .	23
4.13	Taxa de acertos de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções ruidosas de aeroporto e com $\alpha$ variando de 0,1 a 0,9 . . . . .	23
4.14	Precisão de palavras resultantes de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído indicado, testado com locuções ruidosas e com valores ótimos de $\alpha$ . . . . .	24
4.15	Comparação entre os sistemas resultantes das Tabelas 4.1, 4.6 e 4.14.	24
A.1	Taxa de acertos de palavras resultantes de um sistema treinado com ruído de aeroporto com SNR = 20 e 15 dB . . . . .	32
A.2	Taxa de acertos de palavras resultantes de um sistema treinado com ruído de balbulcio com SNR = 20 e 15 dB. . . . .	32
A.3	Taxa de acertos de palavras resultantes de um sistema treinado com ruído de carro com SNR = 20 e 15 dB. . . . .	33
A.4	Taxa de acertos de palavras resultantes de um sistema treinado com ruído de exposição com SNR = 20 e 15 dB. . . . .	33
A.5	Taxa de acertos de palavras resultantes de um sistema treinado com ruído de restaurante com SNR = 20 e 15 dB. . . . .	33
A.6	Taxa de acertos de palavras resultantes de um sistema treinado com ruído de rua com SNR = 20 e 15 dB. . . . .	34
A.7	Taxa de acertos de palavras resultantes de um sistema treinado com ruído de metrô com SNR = 20 e 15 dB. . . . .	34
A.8	Taxa de acertos de palavras resultantes de um sistema treinado com ruído de trem com SNR = 20 e 15 dB. . . . .	34
A.9	Taxa de acertos de palavras resultantes de um sistema treinado com todos os tipos de ruído mas com os níveis de ruído de 10 dB, 15 dB e 20 dB e testado por um corpus contaminado. . . . .	35
A.10	Taxa de acertos de palavras resultantes de um sistema treinado com todos os tipos de ruído mas com os níveis de ruído de 5 dB, 10 dB, 15 dB e 20 dB e testado por um corpus contaminado. . .	35

A.11	Taxa de acertos de palavras resultantes de um sistema treinado com todos os tipos de ruído mas com os níveis de ruído de 0 dB, 5 dB, 10 dB, 15 dB e 20 dB e testado por um corpus contaminado.	35
B.1	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,1 a 0,9 . . . . .	37
B.2	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,01 a 0,09 . . . . .	38
B.3	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de balbulcio, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,1 a 0,9 . . . . .	39
B.4	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de balbulcio, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,01 a 0,09 . . . . .	40
B.5	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de carro, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,1 a 0,9 . . . . .	41
B.6	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de carro, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,01 a 0,09 . . . . .	42
B.7	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de exposição, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,1 a 0,9 . . . . .	43
B.8	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de exposição, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,01 a 0,09 . . . . .	44
B.9	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de restaurante, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,1 a 0,9 . . . . .	45

B.10	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com $\text{SNR} = 20$ e $15$ dB adaptado utilizando modelos fonéticos corrompidos com ruído de restaurante, testado com locuções limpas e ruidosas e com $\alpha$ variando de $0,01$ a $0,09$ . . . . .	46
B.11	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com $\text{SNR} = 20$ e $15$ dB adaptado utilizando modelos fonéticos corrompidos com ruído de rua, testado com locuções limpas e ruidosas e com $\alpha$ variando de $0,1$ a $0,9$ . . . . .	47
B.12	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com $\text{SNR} = 20$ e $15$ dB adaptado utilizando modelos fonéticos corrompidos com ruído de rua, testado com locuções limpas e ruidosas e com $\alpha$ variando de $0,01$ a $0,09$ . . . . .	48
B.13	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com $\text{SNR} = 20$ e $15$ dB adaptado utilizando modelos fonéticos corrompidos com ruído de metrô, testado com locuções limpas e ruidosas e com $\alpha$ variando de $0,1$ a $0,9$ . . . . .	49
B.14	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com $\text{SNR} = 20$ e $15$ dB adaptado utilizando modelos fonéticos corrompidos com ruído de metrô, testado com locuções limpas e ruidosas e com $\alpha$ variando de $0,01$ a $0,09$ . . . . .	50
B.15	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com $\text{SNR} = 20$ e $15$ dB adaptado utilizando modelos fonéticos corrompidos com ruído de trem, testado com locuções limpas e ruidosas e com $\alpha$ variando de $0,1$ a $0,9$ . . . . .	51
B.16	Taxa de acertos de palavras de um sistema treinado com todos os ruídos com $\text{SNR} = 20$ e $15$ dB adaptado utilizando modelos fonéticos corrompidos com ruído de trem, testado com locuções limpas e ruidosas e com $\alpha$ variando de $0,01$ a $0,09$ . . . . .	52
C.1	Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções limpas e ruidosas e com $\alpha$ variando de $0,1$ a $0,9$ . . . . .	54
C.2	Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções limpas e ruidosas e com $\alpha$ variando de $0,01$ a $0,09$ . . . . .	55

C.3	Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de balbulcio, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,1 a 0,9 . . . . .	56
C.4	Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de balbulcio, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,01 a 0,09 . . . . .	57
C.5	Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de carro, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,1 a 0,9 . . . . .	58
C.6	Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de carro, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,01 a 0,09 . . . . .	59
C.7	Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de exposição, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,1 a 0,9 . . . . .	60
C.8	Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de exposição, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,01 a 0,09 . . . . .	61
C.9	Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de restaurante, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,1 a 0,9 . . . . .	62
C.10	Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de restaurante, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,01 a 0,09 . . . . .	63
C.11	Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de rua, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,1 a 0,9 . . . . .	64
C.12	Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de rua, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,01 a 0,09 . . . . .	65

---

C.13 Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de metrô, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,1 a 0,9 . . . . .	66
C.14 Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de metrô, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,01 a 0,09 . . . . .	67
C.15 Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de trem, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,1 a 0,9 . . . . .	68
C.16 Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de trem, testado com locuções limpas e ruidosas e com $\alpha$ variando de 0,01 a 0,09 . . . . .	69





## Lista de Abreviaturas e Siglas

<b>ASR</b>	<i>Automatic Speech Recognition</i> - reconhecimento automático de fala
<b>CDCN</b>	<i>Codeword-Dependent Cepstral Normalization</i> - normalização cesptral dependente da palavra código
<b>CMN</b>	<i>Cepstral Mean Normalization</i> - normalização por média cepstral
<b>DCT</b>	<i>Discrete Cosine Transform</i> - transformada discreta de cosseno
<b>DFT</b>	<i>Discrete Fourier Transform</i> - transformada discreta de Fourier
<b>EIH</b>	<i>Ensemble-Interval Histogram</i> - histograma de intervalo conjunto
<b>HMM</b>	<i>Hidden Markov Model</i> - modelo oculto de Markov
<b>LDA</b>	<i>Linear Discriminant Analysis</i> - análise linear discriminativa
<b>LPC</b>	<i>Linear Predictive Coding</i> - codificação linear preditiva
<b>MAP</b>	<i>Maximum a Posteriori</i>
<b>MFCC</b>	<i>Mel-Frequency Cepstral Coefficients</i> -coeficientes cepstrais baseados na escala de frequência Mel
<b>NIST</b>	<i>National Institute of Standards and Technology</i> - instituto nacional de padrões e tecnologia
<b>PLP</b>	<i>Perceptually-based Linear Prediction</i> - predição linear baseada em perceptualidade
<b>PMC</b>	<i>Parallel Model Combination</i> - combinação paralelo de modelos

<b>RASTA</b>	<i>RelAtive SpecTrAl</i> - técnica spectral relativa
<b>RATZ</b>	<i>MultivaRiate-GAussian-Based CepsTral NormaliZa-tion</i> - normalização cestral baseada em distribuição de Gaussiana multimodais
<b>SLP</b>	<i>Synchronous Linear Prediction</i> - predição linear síncrona
<b>SNR</b>	<i>Signal-to-Noise Ratio</i> - relação sinal ruído
<b>VTs</b>	<i>Vector Taylor Series</i> - vetores da série de Taylor
<b>WER</b>	<i>Word Error Rate</i> - taxa de erro de palavras

# Resumo

A precisão dos sistemas de reconhecimento de voz degrada severamente quando operam em ambientes ruidosos. Uma das causas apontadas na literatura para este fato é o descasamento acústico entre o ambiente em que as locuções de treinamento foram gravadas e aquele no qual o sistema de reconhecimento de fala opera. A partir do modelamento matemático da influência do ambiente sobre o sinal de fala, foram avaliadas duas formas de melhorar a taxa de acertos de um sistema de reconhecimento automático de fala nestes ambientes: o treinamento multi-estilo, onde usa-se material corrompido por ruídos variados, mas de baixa intensidade para treinar o sistema e adaptação dos modelos acústicos através do método de máximo a posteriori. Esta combinação elevou a taxa de acertos do sistema em média 23,67%.

Palavras-chave: reconhecimento de fala, modelos ocultos de Markov, compensação de modelo baseada em ruído, adaptação, treinamento adaptativo.

# Abstract

The accuracy of speech recognition systems degrades severely when operating in noisy environments. One of the causes mentioned in the literature for this is the mismatch between the acoustic environment in which the training utterances were recorded, and the one experienced by the automatic speech recognizer. From the mathematical modeling of the influence of environment on the speech signal, we analysed two ways to improve the accuracy rate of a system of automatic speech recognition in these environments: the multi-style training, which uses materials corrupted by varied noise , but low intensity to train the system and subsequent adaptation of acoustic models using the method of maximum posteriori. This combination increased the accuracy rate of the system on average 23,67%.

Keywords: speech recognition; noise robustness; hidden Markov models; model-based noise compensation; adaptation; adaptive training.

# Capítulo 1

## Introdução

A possibilidade de uma forma de interação com máquinas e sistemas através da fala está cada vez mais se tornando realidade. *Call centers* automáticos, celulares, carros e outros aparelhos operados por voz já começam a aparecer no mercado. A maioria dos sistemas operacionais também provêm sistemas de reconhecimento de fala para a realização de certas tarefas. Esta tecnologia deve permitir uma redução de custos em operadoras de call centers e uma interface mais natural dos usuários com vários sistemas computacionais.

Entretanto, há ainda vários problemas que devem ser resolvidos para que este cenário se torne realidade: a precisão de sistemas de reconhecimento automático de fala é ainda muito aquém dos níveis humanos: nós cometemos erros a uma taxa de menos de um centésimo de um por cento [1], ao reconhecer sequências de dígitos, enquanto que nos melhores sistemas automáticos a taxa de erros apenas avançou de 0,72% para 0,55% na última década [2]. Para tarefas mais complexas a diferença aumenta: por exemplo, nas transcrições de conversas telefônicas, a taxa de erro de palavras cometida por pessoas é de cerca de 4%, enquanto que nos melhores sistemas de transcrição automática as taxas são três vezes piores.

Por que as pessoas desempenham melhor esta tarefa? Por um lado, somos capazes de focar a atenção nos sinais de interesse, excluindo os demais. Isto melhora consideravelmente a relação sinal/ruído percebida (captação do sinal de fala, extração de parâmetros acústicos). Quando o sinal de voz é momentaneamente degradado por um ruído de alta intensidade por exemplo, temos a capacidade de recuperar a informação perdida pela análise do contexto do discurso antes e depois da interrupção (modelo de linguagem). Por fim, fazemos certo uso de leitura labial quando isso é possível. Isto é mais aparente em pessoas com deficiência auditiva, mas inconscientemente, todas as pessoas se aproveitam desta informação visual, principalmente em ambientes com alto nível de ruído (reconhecimento multi-modal).

Por outro lado, os sistemas automáticos têm seu desempenho afetado por vários fatores, incluindo: a grande variabilidade intra e inter locutor, variabilidades dialetais, modelagem pobre da fala espontânea, limitações fundamentais tanto na extração de parâmetros característicos da voz como também no seu enquadramento estatístico.

Um problema importante é a robustez ao ruído: após décadas de pesquisas sobre a robustez em face ao ruído, este ainda permanece um grande desafio [3]. Uma das razões para a pouca precisão em ambientes ruidosos é uma incompatibilidade entre as condições originais dos dados usados para treinar o reconhecedor e do ambiente ruidoso no qual este sistema é testado. Enquanto reconhecimento humano de fala degrada apenas ligeiramente com a presença do ruído, a taxa de erro dos sistemas ASRs aumentam dramaticamente.

Há muitos desafios na construção de um reconhecedor de fala que seja robusto ao ruído do ambiente: o ruído é imprevisível e tem uma variedade de propriedades: aditivo como por exemplo, ruídos de carros e ventiladores; convolucional, por exemplo, diferentes microfones; não-estacionário, por exemplo, outras pessoas falando, cliques do teclado ou o bater de portas.

O custo computacional também é um fator importante nesta equação [4], pois sistemas automáticos de reconhecimento de fala são bastante exigentes em poder de processamento e memória. O uso em sistemas embarcados limita bastante as possibilidades do que pode ser feito. Desta forma, muitos sistemas optam por fazer o reconhecimento em servidores remotos, ficando a cargo do terminal remoto apenas a coleta do sinal de fala, possivelmente com algum pré-processamento para reduzir a taxa de transmissão e mitigar o efeito do ruído.

Os métodos de reconhecimento robusto de fala podem ser divididos em 3 classes: representações robustas de locuções, compensação da representação da locução ruidosa e adaptação dos modelos de fala limpa para o ambiente ruidoso.

- **Representações robustas de locuções:** se a fala é representada com uma parametrização que é pouco afetada pelo ruído, pode-se assumir que o descompasso entre as condições de treino e as condições de reconhecimento pode ser ignorado. O objetivo é procurar por características de fala que sejam imunes ao ruído. Uma suposição comum nos métodos para parametrização das locuções é que o sinal de fala é considerado independente do ruído.

Entre os vários métodos podem ser citados: aplicação de filtragem cepstral (liftering) [5], métodos baseados em modelos auditivos [6][7][8][9], cepstrum na escala Mel [8][10], parametrizações discriminativas [11][12], remoção de variações lentas [10][13][14][15][16] e inclusão dos parâmetros derivados de tempo (delta-mel e delta-delta-mel) [17][18].

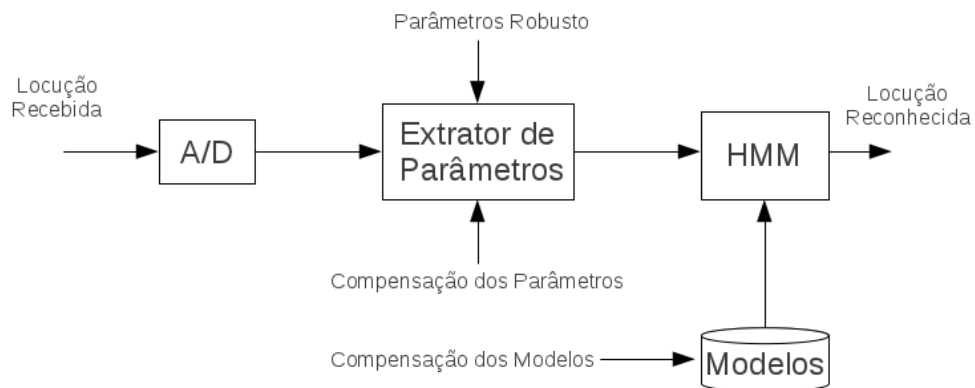
- **Compensação da representação da locução ruidosa:** O objetivo neste caso é reduzir o ruído captado pelo sistema de reconhecimento, e utilizar um sistema de reconhecimento treinado com locuções limpas (com baixo nível de ruído).

Nesta categoria encontram-se os seguintes métodos: mapeamento de parâmetros [19][20][21], subtração espectral [22], aprimoramento estatístico [23] e compensação com base em modelos de fala limpas [24][25][26][27].

- **Adaptação dos modelos de fala limpa para o ambiente ruidoso:** Neste caso, um sistema treinado para locuções limpas é adaptado para as condições de ruído da locução sendo processada.

Esta técnica permite portanto adaptar as matrizes de emissão dos HMM's (Hidden Markov Model) para dar conta do grau em que o ruído afetará as suas médias e variâncias. No entanto, melhoria dos resultados geralmente vem com um custo computacional significativo. Entre os métodos que utilizam esta abordagem estão: decomposição HMM [28][29], filtragem de Wiener dependentes do estado [30] e adaptação estatística de HMMs [10].

A Figura 1.1 mostra uma visão esquemática de onde esses métodos atuam.



**Figura 1.1:** Modelagem em blocos dos ruídos e distorções que afetam a locução.

## 1.1 Objetivos e contribuições desta tese

O objetivo principal deste trabalho é analisar dois métodos de adaptação dos modelos de reconhecimento de fala contínua, independente de locutor e com vo-

cabulário médio operando em condições ruidosas. Para isto, duas abordagens foram realizadas:

- Inicialmente foi utilizado o método de treinamento multi-estilo, que consiste em selecionar material ruidoso com vários tipos de ruído e com variadas relações sinal/ruído para treinar o sistema;
- Em seguida foi realizada uma adaptação utilizando o método de máximo a posteriori para melhorar o casamento acústico entre o sistema anterior (treinado para vários tipos de ruído) e o ambiente acústico atual.

Os resultados experimentais mostram que há ganho em fazer este procedimento em duas etapas.

## 1.2 Organização do trabalho

Esta dissertação está organizada da seguinte maneira: no Capítulo 2 é feita uma análise do efeito produzido pela contaminação do sistema de reconhecimento de fala por ruído; no Capítulo 3 são discutidos dois métodos para melhorar o desempenho dos sistemas de reconhecimento de fala utilizados nos experimentos executados nesta tese: o treinamento em multi-estilo e a adaptação usando o método MAP; no Capítulo 4 são apresentados os resultados e as análises dos experimentos e finalmente o capítulo seguinte apresenta as conclusões e sugestões para a continuação das pesquisas.



## Capítulo 2

# Os efeitos do ruído no sinal de fala

O mundo em que vivemos está cheio de sons de diferentes intensidades e de diferentes fontes de origem. Pode haver pessoas falando ao fundo, alguém pode bater a porta ou o condicionador de ar pode começar a fazer barulho sem aviso prévio. Além dos ruídos de fundo, temos de lidar com os ruídos feitos pelos próprios locutores, como os estalos labiais e sons sem significados. O ruído pode também ser oriundo do próprio dispositivo de entrada, como o microfone e ruídos de conversão Analógico/Digital. Se o reconhecimento de voz é incorporado em dispositivos móveis, como telefones celulares, o espectro de ruídos pode variar significativamente porque o usuário se movimenta.

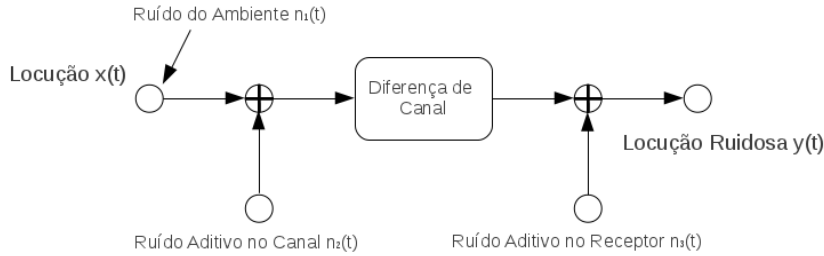
Outro fator importante a ser considerado é o canal no qual o voz se propaga: este adiciona uma filtragem passa-faixa não linear ao sinal, contribuindo para a sua distorção. Além disso, o tipo e posicionamento do microfone também interfere na qualidade do sinal recebido.

Desta forma, a implementação de um sistema de reconhecimento automático de fala que seja robusto ao ruído deve necessariamente passar pela modelagem deste sobre o sinal de fala. Este capítulo trata deste assunto.

### 2.1 Modelo do ambiente

Não é possível nomear e descrever todos as distorções que um reconhecedor de fala pode encontrar, pois o ruído é inerentemente imprevisível. O sinal de voz pode sofrer distorções aditivas, por exemplo ruído de fundo; distorções convolutivas, devido à resposta em frequência do microfone e do canal em que se propaga e mesmo no módulo de captação do sistema de reconhecimento de fala. Isto é

resumido no modelo mostrado na Figura 2.1 [31].



**Figura 2.1:** Modelagem em blocos dos ruídos e distorções que afetam a locução.

Este modelo [32] pode ser representado como:

$$y(t) = [\{(x(t) + n_1(t)) * h_{microfone}(t) + n_2(t)\} * h_{canal}(t)] + n_3(t) \quad (2.1)$$

onde  $y(t)$  é o sinal corrompido resultante,  $x(t)$  é o sinal de fala limpo,  $n_1(t)$  é o ruído do ambiente,  $h_{microfone}(t)$  é a resposta impulsiva do microfone,  $n_2(t)$  e  $h_{canal}(t)$  são respectivamente o ruído aditivo e a resposta impulsiva do canal de transmissão e  $n_3(t)$  é o ruído presente no entrada do receptor. Como é muito difícil gerar um modelo de estresse específico para o ruído presente (efeito Lombard [33]), este é ignorado.

Este modelo é bastante complexo. E, na prática, não é fácil isolar a contribuição de cada um destes fatores. Desta forma, torna-se necessário simplificá-lo. Uma aproximação frequentemente usada é combinar todas as fontes de ruído aditivo em uma única variável  $n(t)$ , e todas as fontes de ruídos convolucionais na variável  $h(t)$ . Ao fazer isto obtém-se um modelo padrão para ambientes ruidosos no domínio do tempo como mostrado abaixo:

$$y(t) = x(t) * h(t) + n(t) \quad (2.2)$$

Os sinais de voz são geralmente analisados no domínio da frequência para isso é utilizado o método de análise de tempo curto. Este método consiste em janelar o sinal e tirar a transformada discreta de Fourier (DFT) com  $2 \cdot K$  pontos e em seguida elevar ao quadrado o módulo da magnitude. Partindo da Equação (2.2) obtem-se:

$$\begin{aligned} |Y(f_k)|^2 &= |X(f_k)|^2 |H(f_k)|^2 + |N(f_k)|^2 + 2\text{Re}\{X(f_k)H(f_k)N^*(f_k)\} \\ |Y(f_k)|^2 &= |X(f_k)|^2 |H(f_k)|^2 + |N(f_k)|^2 + 2|X(f_k)||H(f_k)||N(f_k)| \cos(\theta) \end{aligned} \quad (2.3)$$

onde  $k = 0, 1, \dots, K$ , ou seja, do domínio espectral só é utilizado as frequências positivas e  $\theta$  é o ângulo entre o sinal de fala filtrado e o ruído para o feixe  $k$ . O valor esperado da terceira parcela da soma na Equação (2.3) é zero, pois  $x(t)$  e  $n(t)$  são estatisticamente independentes. Na prática este termo não é zero para um determinado quadro, esta operação pode ser considerada nula ao efetuar a média de uma faixa de frequências, como é o caso quando é computado o *mel-cepstrum*.

Utilizando um banco composto de  $M$  filtros, obtém-se a relação de energias para cada filtro através de:

$$|Y(f_i)|^2 \approx |X(f_i)|^2 |H(f_i)|^2 + |N(f_i)|^2 \quad (2.4)$$

A equação acima é o modelo do ambiente no domínio linear-espectral. Para obter um modelo log-espectral, aplicamos o logaritmo natural:

$$\begin{aligned} \ln |Y(f_i)|^2 &\approx \ln(|X(f_i)|^2 |H(f_i)|^2 + |N(f_i)|^2) \\ \ln |Y(f_i)|^2 &\approx \ln(|X(f_i)|^2 |H(f_i)|^2 (1 + \frac{|N(f_i)|^2}{|X(f_i)|^2 |H(f_i)|^2})) \\ \ln |Y(f_i)|^2 &\approx \ln(|X(f_i)|^2 |H(f_i)|^2) + \ln(1 + \exp(\ln \frac{|N(f_i)|^2}{|X(f_i)|^2 |H(f_i)|^2})) \\ \ln |Y(f_i)|^2 &\approx \ln |X(f_i)|^2 + \ln |H(f_i)|^2 + \ln(1 + \exp(\ln |N(f_i)|^2 - \ln |X(f_i)|^2 - \ln |H(f_i)|^2)) \end{aligned} \quad (2.5)$$

Como a maioria dos sistemas de reconhecimento de fala são representados no domínio cepstral, é útil ver os efeitos do ruído aditivo e da distorção do canal diretamente neste domínio.

$$\begin{aligned} x^c &= C(\ln |X(f_0)|^2 \ln |X(f_1)|^2 \cdots \ln |X(f_M)|^2) \\ h^c &= C(\ln |H(f_0)|^2 \ln |H(f_1)|^2 \cdots \ln |H(f_M)|^2) \\ n^c &= C(\ln |N(f_0)|^2 \ln |N(f_1)|^2 \cdots \ln |N(f_M)|^2) \\ y^c &= C(\ln |Y(f_0)|^2 \ln |Y(f_1)|^2 \cdots \ln |Y(f_M)|^2) \end{aligned} \quad (2.6)$$

onde  $C$  é a Transformada discreta de cosseno (DCT - *Discrete Cosine Transform*) e  $f_i$  com  $i = 0, 1, 2, \dots, M$  correspondem às frequências centrais do banco de filtros

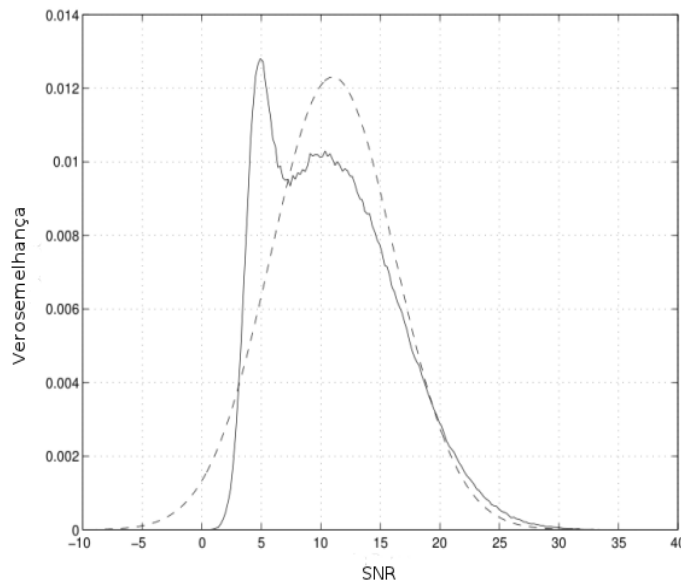
utilizados para o cálculo dos *mel-cepstrums*.

Finalmente combinando as Equações em (2.6) chegamos a

$$y^c = x^c + h^c + C \ln(1 + \exp(C^{-1}(n^c - x^c - h^c))) \quad (2.7)$$

## 2.2 Efeito do ruído na distribuição de probabilidade do sinal de voz

É difícil visualizar os efeitos do ruído na distribuição do sinal de fala [31], tendo em conta as não linearidades envolvidas. Entretanto, uma análise simplificada pode ser feita utilizando o seguinte cenário idealizado: uma distribuição de locução ruidosa com uma única dimensão no domínio log-espectral e ruído aditivo. Para este caso essa distribuição toma a forma mostrada na Figura 2.2 [31][32].



**Figura 2.2:** Efeito aditivo do ruído, linha sólida é a distribuição resultante e a linha pontilhada é a distribuição Gaussiana.

Para um certo conjunto de condições ruidosas, o vetor do sinal corrompido resultante ( $y^c$ ) é uma função altamente não linear do sinal de fala limpa ( $x^c$ ). Em termos gerais, uma pequena quantidade de ruído aditivo gaussiano acrescenta uma moda estatística a mais na matriz de emissão do modelo de locução. À medida que a relação sinal ruído cai, a moda do ruído começa a dominar e as médias da locução ruidosa deslocam-se para mais perto do ruído e suas variâncias encolhem. Em casos extremos, o ruído domina e há pouca ou nenhuma informação

restante do sinal de fala. O efeito final é que as funções de distribuição de probabilidades estimadas com dados limpos tornam-se muito pobres em relação às distribuições de dados observadas no ambiente ruidoso, e portanto as taxas de erros de reconhecimento aumentam rapidamente. O efeito do ruído convolucional sobre a distribuição utilizando o ambiente acústico simplificado consiste na mudança da posição das médias da voz limpa. Obviamente a distribuição real da fala corrompida é muito mais complexa pois os elementos dos vetores de características da locução são correlacionados uns com os outros e nem sempre têm distribuição Gaussiana, mas este modelo já fornece algumas pistas da influência do ruído sobre o sinal de fala.

# Capítulo 3

## Métodos para reconhecimento robusto de fala

Neste capítulo são mostrados os dois métodos utilizados neste trabalho para lidar com o reconhecimento robusto de fala: o treinamento multi-estilo e a adaptação de sistemas.

### 3.1 Treinamento multi-estilo

Quando um sistema é treinado somente com locuções limpas as informações contidas no sinal de voz são maximizadas. Entretanto, seu desempenho cai bastante quando este tem que lidar com locuções corrompidas por ruído.

Por outro lado, sistemas treinados com locuções ruidosas têm em geral desempenho melhor nestas condições. À primeira vista, esta parece ser a estratégia óbvia quando se quer um bom desempenho sob condições de ruído. Entretanto, as coisas não são bem assim: o sistema deve ser treinado com o mesmo tipo de ruído e com a mesma relação sinal/ruído do sinal a ser reconhecido, o que evidentemente não é realizável na prática, principalmente quando o sistema de reconhecimento experimenta diferentes tipos de ruído (em terminais móveis por exemplo).

Uma alternativa seria retreinar o sistema a cada novo cenário acústico, mas isto tem um custo computacional alto, o que o torna inviável para a maior parte dos casos.

Por outro lado, é interessante conhecer o desempenho de sistemas treinados com o mesmo tipo e intensidade do ruído presente na locução a ser conhecida para obter uma estimativa do limite superior da taxa de acertos que pode ser obtida.

Também pode ser verificado o que acontece quando testamos os sistemas com tipos diferentes de ruídos: espera-se uma degradação de desempenho neste caso. A questão agora é determinar o quanto este descasamento afeta a taxa de acertos do sistema.

Uma hipótese razoável a ser testada é a de que se treinarmos um sistema com vários tipos de ruídos, em várias relações sinal/ruído, este teria um desempenho razoavelmente bom em qualquer situação, pois a priori teria conhecimento de todas as condições que pudessem ocorrer. Esta é a filosofia do treinamento multi-estilo ou multi-condição.

Os resultados de nossos experimentos mostram que as coisas não são bem assim: é verdade que treinar o sistema com vários tipos de ruído ajuda o sistema a ter um melhor desempenho, pois ele passa a conhecê-los. Entretanto, locuções com baixa relação sinal/ruído fizeram diminuir a taxa de acertos do sistema. A hipótese levantada é a de que nestas, a informação de voz é deteriorada pelo ruído, e o sistema passa a modelar o ruído e não mais a fala.

## 3.2 Adaptação e normalização

Apesar da quantidade de dados usados para treinar os modelos acústicos e os esforços para produzir sistemas independentes de locutor, ainda há degradação quando fatores ou condições de teste não são representados nos dados de treinamento. Por exemplo, estes fatores podem ser novos locutores, diferentes sotaques, ou um ruído ambiente diferente daquele para o qual o sistema foi treinado.

Na literatura são citadas duas formas de mitigar este problema: a adaptação dos modelos e a normalização dos parâmetros acústicos.

- Adaptação: coletam-se amostras do ruído durante os intervalos em que o locutor não está falando e estes são utilizados para fazer uma adaptação dos parâmetros do sistema de reconhecimento ao ruído específico que está corrompendo a fala naquele momento. No caso de sistemas baseados em HMMs (como é o caso deste trabalho), em geral atualizam-se as médias das misturas de gaussianas, pois estas têm uma influência maior na taxa de acertos, como visto no Capítulo 2.
- Normalização: neste caso os parâmetros acústicos são adaptados para cada tipo de ruído, na tentativa de fazer com que a locução ruidosa seja “ouvida” pelo sistema de reconhecimento como uma locução sem ruído. Em outras palavras, tenta-se fazer com que todas as locuções (de treinamento e teste) estejam em conformidade com uma medida padrão, usando-se, por exemplo uma normalização por média cepstral (CMN).

Neste trabalho optou-se por explorar a primeira técnica: a de adaptação dos modelos, a qual será descrita com maiores detalhes a seguir.

### 3.3 Treinamento adaptativo

O treinamento adaptativo é uma técnica poderosa para a construção de sistemas de reconhecimento de voz em dados de treinamento não homogêneos.

Tem por base a utilização de uma ou mais transformações para cada locutor e ambiente acústico durante o treinamento do modelo acústico. Essas transformações têm por objetivo remover as variabilidades indesejadas (as mudanças de locutores e mudanças no ambiente), permitindo estimar um modelo canônico.

Assim, durante o treinamento dois conjuntos de modelos são gerados: um modelo canônico definido para a variabilidade dos dados de voz, e um conjunto de transformações para representar a variabilidade indesejada.

O modelo canônico representa a variabilidade da fala dos dados de treinamento, que é independente do locutor e das condições acústicas.

Um conjunto de transformações  $T$  representa as variabilidades indesejadas tais como locutores diferentes e ou ambiente acústico. Normalmente, uma transformação linear dos modelos acústicos é usada para representar a condição acústica de um determinado bloco homogêneo e adaptar o modelo canônico a essa condição acústica especial.

A técnica mais utilizada para promover as transformações  $T$  é a de máximo a posteriori (MAP), descrita a seguir.

### 3.4 Adaptação de modelo utilizando MAP

Ao invés de hipotetizar uma forma de transformação para representar as diferenças entre os locutores e os ambientes acústicos, é possível a utilização de abordagens estatísticas para obter estimativas dos parâmetros robustos.

Uma abordagem comum é o MAP, muitas vezes referida como adaptação Bayesiana. Esta envolve o uso de conhecimento a priori sobre a distribuição dos parâmetros do modelo. Portanto, se é possível saber como os parâmetros do modelo tendem a ser antes de observar todos os dados de adaptação, é razoável supor que seja possível fazer um bom uso dos dados limitados de adaptação para obter uma estimativa MAP robusta.

Esta abordagem foi utilizada com sucesso na área de reconhecimento de locutor [34], onde o modelo canônico é gerado a partir de vários locutores e depois é gerado um modelo específico para cada locutor a partir deste modelo canônico, com poucos dados de treinamento.



Neste trabalho, o modelo canônico é representado pelos HMMs treinados com locuções limpas (ou ruidosas) e as transformações  $T$  são utilizadas para adaptar este modelo canônico às condições de ruído específicas para a locução que está sendo reconhecida.

Primeiro é determinado o alinhamento probabilístico do ruído no modelo canônico. Ou seja, dado uma amostra do ruído qual o valor da probabilidade para a Gaussiana  $i$  que compõe a mistura da função densidade de probabilidade.

$$Pr(i|x_t) = \frac{w_i p_i(x_t)}{\sum_{j=1}^M w_j p_j(x_t)} \quad (3.1)$$

Depois, com este resultado ( $Pr(i|x_t)$ ) é calculado o peso (3.2), a média (3.3) e a variância (3.4) para essa amostra do ruído do ambiente.

$$n_i = \sum_{t=1}^T Pr(i|x_t) \quad (3.2)$$

$$E_i(x) = \frac{1}{n_i} \sum_{t=1}^T Pr(i|x_t) x_t \quad (3.3)$$

$$E_i(x^2) = \frac{1}{n_i} \sum_{t=1}^T Pr(i|x_t) x_t^2 \quad (3.4)$$

Finalmente, essas novas estatísticas a partir dos dados de treinamento para o ruído do ambiente são usados para atualizar o modelo canônico velho.

$$\hat{w}_i = \alpha_i^w \frac{n_i}{T} + (1 - \alpha_i^w) w_i \quad (3.5)$$

$$\hat{\mu}_i = \alpha_i^m E_i(x) + (1 - \alpha_i^m) \mu_i \quad (3.6)$$

$$\hat{\sigma}_i^2 = \alpha_i^v E_i(x^2) + (1 - \alpha_i^v)(\sigma_i^2 + \mu_i^2) - \hat{\mu}_i^2 \quad (3.7)$$

Os coeficientes de adaptação que controlam o equilíbrio entre o modelo canônico velho e novo são  $\hat{w}_i$  para o peso,  $\hat{\mu}_i$  para a média e  $\hat{\sigma}_i^2$  para a variância para componente de mistura da Gaussiana  $i$  que compõe o modelo. Portanto da modelagem acima, se a probabilidade de ocupação de um componente da gaussiana é pequeno, então a estimativa da média MAP irá permanecer perto da média do modelo canônico.

# Capítulo 4

## Resultados Experimentais

Neste capítulo são apresentados os resultados experimentais que formam a base empírica sobre a qual foram tiradas as conclusões deste trabalho.

Antes porém são descritos a base de dados, o aparato experimental, bem como a metodologia para avaliação do desempenho de reconhecimento.

### 4.1 Base de dados

Como este trabalho tem foco no problema de reconhecimento robusto de fala e na quantificação da diferença de desempenho devido ao descasamento acústico entre o material de treinamento e o de teste, optou-se por utilizar uma base de dados de fala com locuções obtidas em um ambiente silencioso, e a partir desta gerar versões ruidosas pela mistura de ruídos diversos e em diversas relações sinal/ruído.

#### 4.1.1 Corpus de fala limpa

O corpus de fala é composto de 40 locutores adultos sendo 20 homens e 20 mulheres [35]. Cada um desses locutores registrou 40 sentenças foneticamente balanceadas, retiradas de [36]. Desta forma, este corpus possui 1600 locuções.

Foram utilizados 16 locutores de cada sexo para treinar os sistemas (1200 locuções), e os restantes foram utilizados para os testes de desempenho (400 locuções).

Nestas sentenças foram identificadas 694 palavras diferentes, o que caracteriza o problema como reconhecimento de fala contínua, independente de locutor e com vocabulário de médio porte. Todos os arquivos de áudio foram gravados em um ambiente de baixo ruído, a uma taxa de amostragem de 11025 Hz e codificados em PCM linear com 16 bits por amostra. Para este trabalho, a taxa

de amostragem destes sinais foi reduzida para 8 kHz por causa da base de dados de ruídos utilizada.

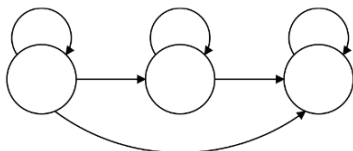
### 4.1.2 Corpus de fala ruidosa

Para gerar as versões corrompidas por ruído dessas gravações foram utilizados os ruídos do banco de dados Aurora [37]. Estes compreendem ruído de aeroporto, balbucio, carro, exposições, restaurante, rua, metrô e trem. Para cada locução limpa outras seis versões de sinais corrompidos foram criados combinando artificialmente os ruídos com os seguintes valores de relação sinal ruído (SNRs): 20, 15, 10, 5, 0 e -5 dB.

## 4.2 Sistema de reconhecimento de fala

Para os testes foi utilizado um software desenvolvido em [35][38], baseado em modelos ocultos de Markov contínuos.

Este sistema utiliza fonemas independentes de contexto como unidades fundamentais, cada um deles modelado como uma cadeia de Markov de 3 estados, como mostrado na Figura 4.1. Para esta tese os parâmetros acústicos escolhidos foram os MFCC com dimensão igual a 12, juntamente com seus derivados primeiro e segundo, resultando em vetores característicos de tamanho igual a 36 dimensões. Cada estado foi projetado com uma mistura de 10 distribuições Gaussianas multidimensionais com a matriz de covariância na forma diagonal. A transcrição fonética de cada palavra do vocabulário foi realizada por 36 modelos de fonema independente de contexto e um modelo de linguagem bigrama também foi utilizado para melhorar o desempenho do reconhecimento.



**Figura 4.1:** Modelo de Markov para um fonema.

Este sistema usa o One Pass [39] como algoritmo de busca.

## 4.3 Métrica para avaliação de desempenho

A precisão do reconhecimento de um sistema de reconhecimento de voz pode ser determinada por comparação entre a transcrição da hipótese (reconhecida pelo sistema) com a transcrição de referência (sentença correta). Existem várias métricas para medir esta precisão, sendo que as mais utilizadas são:

- **Taxa de erros de frases:** número de sentenças corretamente reconhecidas dividido pelo número total de sentenças.
- **Taxa de erros de palavras:** nesta, as sequências de palavras são comparadas usando um algoritmo de alinhamento dinâmico baseado em cadeias de palavras de modo a determinar o número de erros por deleção (D), erros por substituição (S) e os erros por inserção (I). Esta taxa é então calculada da seguinte forma:

$$WER = \frac{(D + S + I)}{N} * 100\% \quad (4.1)$$

onde  $N$  é o número de palavras na transcrição de referência.

Neste trabalho optou-se por utilizar a taxa de acertos de palavras (100% – WER, *Word Error Rate*) como métrica para o desempenho dos sistemas avaliados. A ferramenta *sc-lite* [40] foi utilizada para esta finalidade.

*Observação:* A presença de alguns valores negativos para a taxa de acertos pode causar estranheza. Isto é devido ao fato de que a sentença de referência tinha menos palavras do que a frase reconhecida, ou seja, houve um número muito grande de erros por inserção. Ao ser feito o cômputo dos erros de deleção, substituição e de inserção pelo *sc-lite* obtêm-se estes valores negativos.

## 4.4 Resultados

Os testes foram realizados em três etapas:

- Inicialmente foi treinado um sistema somente com locuções limpas (sem ruído), para estabelecer um desempenho padrão, o qual foi utilizado como base para comparar os resultados dos testes posteriores;
- a seguir foram realizadas diversas tentativas usando locuções corrompidas com ruído para treinar o sistema;
- por fim foram utilizadas apenas amostras do ruído para adaptar sistemas já treinados.

#### 4.4.1 O sistema base

O desempenho de referência foi obtido com um sistema treinado com locuções limpas. Este apresentou uma taxa de acertos de palavras de 75% quando testado com locuções limpas. Entretanto, seu desempenho caiu drasticamente quando testado com locuções ruidosas (o que era esperado devido ao descasamento acústico entre os materiais de treinamento e teste), como mostram os resultados da Tabela 4.1:

**Tabela 4.1:** Taxa de acertos de um sistema treinado com locuções limpas e testado por um corpus contaminado por ruído. Este mesmo sistema obteve uma taxa de acertos de 75% para locuções limpas.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
Aeroporto	41,4%	31,6%	18,2%	0,4%	-4,1%	-12,5%
Balbuicio	48,2%	31,5%	14,5%	-5,2%	-7,6%	-0,9%
Carro	55,7%	41,3%	22,3%	2,7%	2,5%	3,8%
Exposições	31,1%	10,2%	-0,7%	-7,5%	-6,3%	-2,8%
Restaurante	38,1%	19,5%	6,9%	-5,4%	-9,4%	-0,4%
Rua	41,8%	23,8%	19,2%	7,0%	-3,1%	1,7%
Metrô	27,7%	17,6%	-0,4%	-6,5%	-3,1%	0,8%
Trem	64,8%	55,7%	37,3%	18,1%	5,3%	3,9%

Esta hipótese de descasamento acústico é corroborada quando treinamos um sistema com locuções ruidosas e fazemos testes com locuções limpas, como mostram os resultados das Tabelas 4.2, 4.3 e 4.4

**Tabela 4.2:** Taxa de acertos de um sistema treinado com locuções ruidosas (SNR 20) para os tipos de ruídos apresentados nas colunas da tabela testados com locuções limpas

Aeroporto	Balbuicio	Carro	Exposição	Restaurante	Metrô	Rua	Trem
11,90%	11,90%	11,90%	12,90%	10,70%	13,50%	11,90%	13,20%

**Tabela 4.3:** Taxa de acertos de um sistema treinado com locuções ruidosas com SNR 20 e 15 para os tipos de ruídos apresentados nas colunas da tabela testados com locuções limpas

Aeroporto	Balbuicio	Carro	Exposição	Restaurante	Metrô	Rua	Trem
11,00%	11,60%	10,70%	9,40%	12,60%	12,30%	11,90%	13,80%

Na verdade, estes resultados são até surpreendentes, pois mostram que um sistema treinado para reconhecer locuções ruidosas tem seu desempenho drasticamente reduzido quando é testado com locuções limpas! Uma hipótese que pode ser feita neste sentido é a de que, ao misturar o ruído com a fala, parte da

**Tabela 4.4:** Taxa de acertos de um sistema treinado com todos os tipos de ruído com valores de SNR apresentados nas colunas da tabela, testados com locuções limpas.

SNR 20 e 15	SNR 20 15 10	SNR 20 15 10 5	SNR 20 15 10 5 0	SNR 20 15 10 5 0 -5
12,30%	11,6%	11,0%	8,5%	9,70%

informação desta é destruída pelo ruído. Os resultados a seguir corroboram com esta hipótese.

#### 4.4.2 Sistema treinado com locuções ruidosas

Dada a hipótese de que o descasamento acústico entre os dados de treinamento e teste é que causa a queda de desempenho do sistema, uma estratégia possível para lidar com esse problema seria treinar o sistema com locuções ruidosas.

A primeira tentativa foi construir um sistema com locuções contaminadas com todos os tipos de ruído e com várias relações sinal/ruído. Desta forma o sistema seria mais robusto, já que foi exposto a uma grande variedade de situações. A Tabela 4.5 mostra os resultados para este teste.

**Tabela 4.5:** Taxa de acertos de um sistema treinado com locuções contaminadas com todos os ruídos e todos os níveis de ruído (SNR) e testado por um corpus contaminado por ruído onde cada coluna representa a SNR do material de teste.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
Aeroporto	52,8%	49,7%	42,9%	22,6%	7,1%	-11,2%
Balucio	49,4%	44,0%	33,1%	16,5%	-1,1%	-0,9%
Carro	52,1%	61,5%	53,9%	21,1%	-1,0%	3,2%
Exposições	61,5%	56,2%	33,1%	14,7%	3,3%	2,2%
Restaurante	51,3%	52,0%	41,7%	17,7%	2,6%	-1,8%
Rua	43,5%	35,5%	31,9%	21,9%	-0,3%	3,1%
Metrô	54,8%	50,6%	35,8%	16,7%	4,7%	0,8%
Trem	58,7%	66,9%	63,5%	55,9%	25,6%	-1,3%

Pode ser visto que, em geral, os resultados são melhores do que no caso anterior (Tabela 4.1), mas não para todas as situações. Por exemplo, o teste com o ruído de carro com SNR = 20 dB apresentou um resultado pior quando comparado com o teste anterior.

Uma possível explicação para este resultado é que as locuções com uma SNR baixa perderam as informações da fala. Portanto, o sistema é treinado principalmente com o ruído e não mais com o sinal de fala. Desta forma, o sistema de reconhecimento de voz passaria a modelar o ruído ao invés da fala..

Para testar essa hipótese, um outro conjunto de teste foi realizado, agora usando apenas as locuções com altas SNRs (20 dB e 15 dB). Os resultados destes

testes são apresentados na Tabela 4.6.

**Tabela 4.6:** Taxa de acertos de um sistema treinado com todos os tipos de ruído mas com os níveis de ruído somente de 15 dB e 20 dB e testado com locuções ruidosas.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
Aeroporto	75,3%	69,4%	53,9%	24,0%	4,2%	-3,3%
Balbuicio	73,8%	62,0%	41,7%	14,5%	0,5%	4,6%
Carro	68,5%	68,2%	48,5%	15,2%	-0,8%	1,9%
Exposições	73,4%	59,4%	31,6%	5,1%	-0,4%	1,6%
Restaurante	68,5%	63,1%	38,7%	14,2%	-1,7%	0,5%
Rua	64,4%	51,2%	46,0%	31,6%	5,4%	1,7%
Metrô	69,2%	64,8%	40,0%	8,1%	0,6%	1,7%
Trem	75,6%	76,1%	64,4%	44,6%	13,5%	-0,4%

Os resultados destes testes mostram uma grande melhoria em comparação com os anteriores. Este resultado parece confirmar a hipótese de que o sistema começa a perder informações sobre o sinal da fala. Um ponto interessante é que o desempenho é melhor mesmo para sinais altamente degradados (com baixas SNRs), o que corrobora ainda mais a hipótese acima.

A próxima questão a ser respondida é: há ganho em treinar um sistema para um tipo específico de ruído? Intuitivamente, esta abordagem levará a um desempenho melhor devido a uma melhor adequação acústica entre as condições de treinamento e as de teste.

Para verificar isso, foram gerados 8 sistemas, cada qual especializado em um tipo de ruído específico, treinados com locuções de SNR 15 dB e 20 dB. Cada sistema foi testado com locuções corrompidas com o mesmo tipo de ruído. Os resultados são mostrados na Tabela 4.7.

**Tabela 4.7:** Taxa de acertos de um sistema treinado e testado com locuções do mesmo tipo que o ruído descrito. Como antes, apenas locuções com SNR = 15 dB e SNR = 20 dB foram utilizados para treinar o sistema.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
Aeroporto	77,5%	73,6%	56,3%	28,8%	4,4%	-1,1%
Balbuicio	70,1%	68,7%	56,1%	15,9%	2,5%	2,6%
Carro	71,8%	72,4%	56,3%	15,5%	0,6%	3,0%
Exposições	75,8%	70,3%	48,8%	14,5%	-4,1%	-0,3%
Restaurante	70,0%	63,5%	49,6%	21,0%	1,0%	-0,2%
Rua	63,0%	53,6%	39,1%	24,3%	3,1%	-0,2%
Metrô	75,3%	71,1%	48,1%	13,6%	1,4%	3,1%
Trem	70,2%	75,9%	68,6%	52,4%	14,4%	-4,0%

Os resultados mostram que, em geral, o desempenho desses sistemas é melhor quando comparado a um sistema treinado com todos os ruídos (Tabela 4.6).



Apesar de um melhor desempenho, há dois inconvenientes nesta abordagem: a necessidade de um sistema identificador de ruídos, e a possibilidade deste identificar de forma equivocada o ruído que se apresenta. Para verificar a influência deste erro no desempenho final do reconhecedor foram feitos diversos testes com os mesmos 8 sistemas treinados anteriormente, mas com locuções contaminadas por um ruído diferente daquele para o qual o sistema foi treinado.

Testes exaustivos foram realizados, cobrindo todas as combinações possíveis, e os resultados encontram-se no Anexo A. Para facilitar a análise, os dados foram compilados na Tabela 4.8.

Para estes testes foram usadas locuções com SNR igual a 20 dB. Na primeira coluna estão os valores para um sistema treinado e testado com locuções com o mesmo tipo de ruído, a coluna seguinte é a média dos resultados de uma simulação do que aconteceria se o sistema identificasse de forma errada o tipo de ruído que contamina a locução, e a terceira coluna é a diferença entre a primeira e a segunda colunas.

A Tabela 4.9 apresenta a média global da Tabela 4.8. Em cima destes resultados pode-se observar que o erro de não identificar corretamente acarreta um decréscimo médio de 41.38 % na taxa de acertos.

**Tabela 4.8:** *Análise do erro introduzido ao não identificar corretamente o ruído presente no áudio ruidoso.*

	mesmo ruído	ruído diferente	diferença
Aeroporto	77,5%	49,84%	27,66%
Balbuçio	70,1%	47,77%	22,33%
Carro	71,8%	41,84%	29,96%
Exposições	75,8%	25,97%	49,83%
Restaurante	70,0%	52,11%	17,89%
Rua	63,0%	54,72%	8,28%
Metrô	75,3%	21,06%	53,7%
Trem	70,2%	37,14%	33,06%

**Tabela 4.9:** *Análise global do erro introduzido ao não identificar corretamente o ruído presente no áudio ruidoso.*

	mesmo ruído	ruído diferente	diferença
Locuções ruidosas	71,71%	30,33%	41,38%

**Conclusão:** Estes últimos resultados mostram que é mais vantajoso usar um sistema treinado com todos os tipos de ruídos do que com ruídos específicos, pois seu desempenho é um pouco inferior (em média), e não cai drasticamente se porventura o sistema de identificação do ruído cometer um erro.

### 4.4.3 Sistemas adaptados com locuções ruidosas

A abordagem de usar locuções corrompidas com ruído para treinar os sistemas de reconhecimento de fala funciona bem para os tipos de ruídos considerados. Entretanto, este desempenho cai quando temos ruídos diferentes daqueles apresentados durante a fase de treinamento. Obviamente é impossível treinar um sistema com todos os tipos de ruído que possam ocorrer, então alternativas devem ser encontradas.

Neste trabalho propõe-se adaptar os parâmetros já treinados usando apenas amostras do ruído usando a técnica de adaptação MAP, mostrada no capítulo anterior. Esta abordagem permitiria ter um único sistema que poderia ser modificado de acordo com o tipo e nível de ruído presente junto com o sinal de voz, utilizando os trechos em que o locutor não está falando.

Segundo esta técnica, o modelo adaptado é formado pela combinação linear entre o modelo original (HMM já treinado) e uma estimativa de um novo modelo, gerado a partir de amostras de treinamento (amostras do ruído no nosso caso), como mostrado na Equação (3.5) e reproduzida abaixo:

$$\hat{w}_i = \alpha_i^w \frac{n_i}{T} + (1 - \alpha_i^w)w_i$$

O parâmetro  $\alpha$  controla o quanto o modelo original incorpora as novas informações (ruído): quanto maior, mais o modelo final é modificado em direção à estimativa do ruído. Como não há na literatura um procedimento para escolha do valor ótimo para  $\alpha$ , este foi escolhido empiricamente por varredura.

Partindo de um modelo treinado com locuções limpas, fez-se uma varredura inicial de  $\alpha = 0,1$  a  $\alpha = 0,9$  para ruído de aeroporto, e reconhecimento de locuções limpas. Os resultados são mostrados na Tabela 4.10

**Tabela 4.10:** Taxa de acertos de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções limpas e com  $\alpha$  variando de 0,1 a 0,9

$\alpha$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Limpas	75,90%	73,40%	74,70%	75,30%	68,40%	61,70%	55,40%	36,70%	3,20%

Como o melhor valor obtido foi com  $\alpha = 0,1$ , foram verificados valores ainda menores, entre  $\alpha = 0,01$  e  $\alpha = 0,09$ . De fato, os melhores resultados foram conseguidos com  $\alpha$  entre 0,05 e 0,07. Os resultados destes testes aparecem na Tabela 4.11

Entretanto, quando testado com sinais ruidosos, este procedimento não funcionou bem. Nas Tabelas 4.12 e 4.13 são mostrados os resultados realizados com locuções corrompidas pelo ruído aeroporto para ilustrar esta questão:

**Tabela 4.11:** Taxa de acertos de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções limpas e com  $\alpha$  variando de 0,01 a 0,09

$\alpha$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
Limpo	77,90%	77,90%	77,90%	77,90%	78,90%	78,90%	78,90%	78,20%	76,30%

**Tabela 4.12:** Taxa de acertos de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções ruidosas de aeroporto e com  $\alpha$  variando de 0,01 a 0,09

$\alpha$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
snr20	43,8%	44,8%	46,4%	47,0%	47,90%	48,6%	49,8%	50,2%	51,4%
snr15	29,3%	32,2%	31,5%	35,0%	34,70%	38,2%	38,80%	41,0%	42,3%
snr10	16,1%	17,0%	17,0%	18,3%	21,50%	20,2%	22,40%	22,4%	24,0%
snr5	-3,5%	-1,9%	0,00%	1,30%	5,4%	6,90%	6,3%	5,40%	7,3%
snr0	-5,4%	-3,2%	-2,2%	-0,60%	-0,3%	0,30%	0,6%	0,9%	2,5%
snr-5	-12,3%	-14,2%	-13,2%	-14,50%	-14,8%	-14,8%	-15,5%	-15,8%	-14,5%

**Tabela 4.13:** Taxa de acertos de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções ruidosas de aeroporto e com  $\alpha$  variando de 0,1 a 0,9

$\alpha$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
snr20	51,3%	50,9%	52,2%	50,9%	51,9%	43,7%	36,7%	32,9%	12,3%
snr15	44,0%	45,6%	43,0%	41,1%	40,5%	41,1%	31,3%	17,1%	13,3%
snr10	24,1%	28,5%	26,9%	28,2%	29,1%	31,3%	22,5%	17,1%	3,8%
snr5	5,7%	12,3%	12,7%	15,2%	11,1%	8,5%	8,20%	1,3%	-7,0%
snr0	1,3%	4,4%	4,1%	5,1%	7,3%	3,5%	0,6%	-1,3%	-5,7%
snr-5	-15,5%	-17,7%	-15,8%	-16,8%	-13,9%	-16,1%	-13,9%	-14,2%	-12%

Com o fracasso desta abordagem, tentou-se então adaptar um sistema já treinado com locuções ruidosas. Para isso escolheu-se o sistema obtido com o treinamento com locuções ruidosas corrompidas com todos os tipos de ruídos e SNRs 20 dB e 15 dB, pois foi aquele de melhor desempenho até então. Os resultados destes testes estão resumidos na Tabela 4.14. Desta vez conseguiu-se taxas de acertos consistentemente superiores ao melhor sistema até então (compare os resultados acima com aqueles obtidos na Tabela 4.6).

Após essa série de experimentos uma última análise a ser feita é a comparação entre os sistemas resultantes das Tabelas 4.1, 4.6 e 4.14. Ao fazer a diferença média entre as Tabelas 4.1 e 4.6 tem-se um ganho médio de 20,72% na taxa de acertos para o sistema treinado com SNR 20 dB e 15 dB, quando comparado com o sistema treinado apenas com locuções limpas. Adicionando a etapa de adaptação, o ganho salta para 23,67%. Com isso, pode-se concluir que a adaptação conseguiu

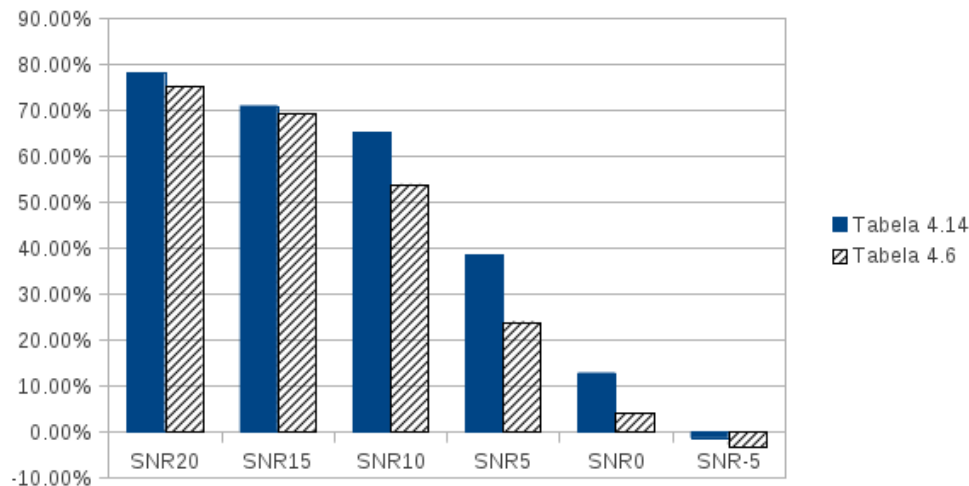
**Tabela 4.14:** Precisão de palavras resultantes de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído indicado, testado com locuções ruidosas e com valores ótimos de  $\alpha$

	SNR20	$\alpha$	SNR15	$\alpha$	SNR10	$\alpha$	SNR5	$\alpha$	SNR0	$\alpha$	SNR-5	$\alpha$
Aeroporto	78,2%	0,04	71,0%	0,02	65,3%	0,2	38,5%	0,2	12,9%	0,4	-1,3%	0,04
Balbuçio	78,5%	0,01	58,7%	0,01	44,5%	0,05	15,1%	0,03	1,9%	0,01	3,2%	0,07
Carro	37,9%	0,01	45,4%	0,04	41,6%	0,07	19,2%	0,06	-0,3%	0,07	0,3%	0,06
Exposições	34,1%	0,04	31,5%	0,05	8,2%	0,05	-3,5%	0,1	-9,1%	0,01	-0,6%	0,01
Restaurante	69,7%	0,01	65,6%	0,01	41,6%	0,2	17,4%	0,09	1,9%	0,2	0,6%	0,02
Rua	65,9%	0,01	54,3%	0,05	42,9%	0,01	35%	0,05	10,1%	0,07	4,1%	0,2
Metrô	71,6%	0,04	60,9%	0,01	43,2%	0,2	17,7%	0,1	-1,3%	0,02	1,6%	0,03
Trem	75,1%	0,01	78,5%	0,02	69,1%	0,2	55,2%	0,5	21,1%	0,6	0,3%	0,01

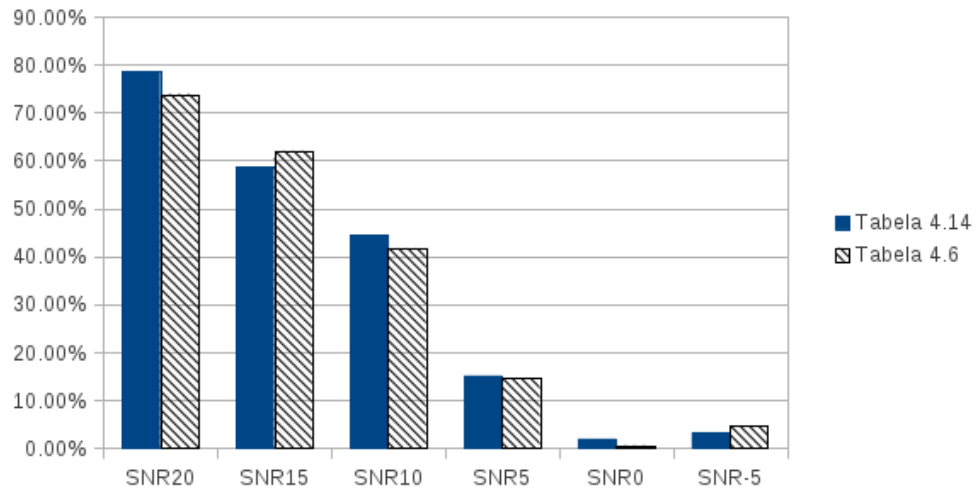
**Tabela 4.15:** Comparação entre os sistemas resultantes das Tabelas 4.1, 4.6 e 4.14.

Tabela	1 e 6	1 e 16	6 e 16
Ganho	20,72%	23,67%	2,95%

uma melhora média de 2,95% na taxa de acertos em relação ao sistema treinado com a técnica de multi-estilo.



**Figura 4.2:** Gráfico comparativo entre o sistema treinado com locuções ruidosas (Tabela 4.6) e o sistema adaptado com locuções ruidosas (Tabela 4.14) para o caso do ruído aeroporto

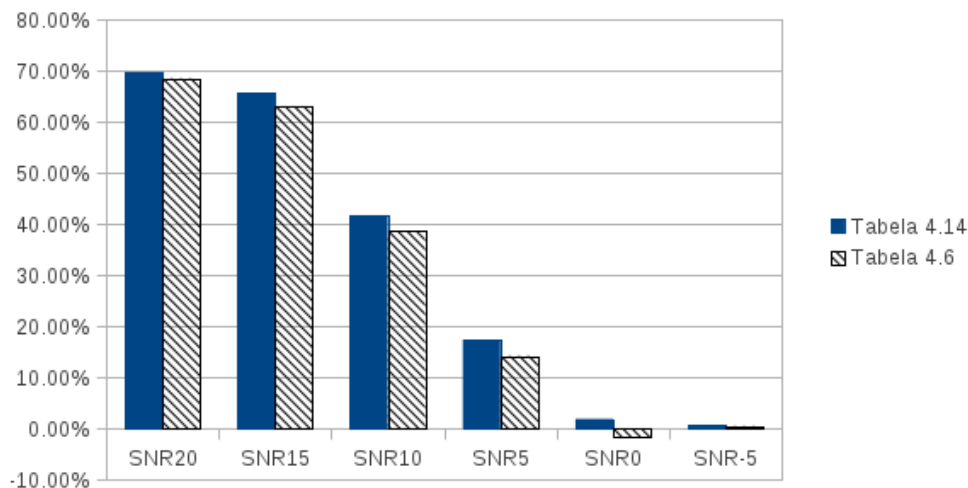


**Figura 4.3:** Gráfico comparativo entre o sistema treinado com locuções ruidosas (Tabela 4.6) e o sistema adaptado com locuções ruidosas (Tabela 4.14) para o caso do ruído balbulcio

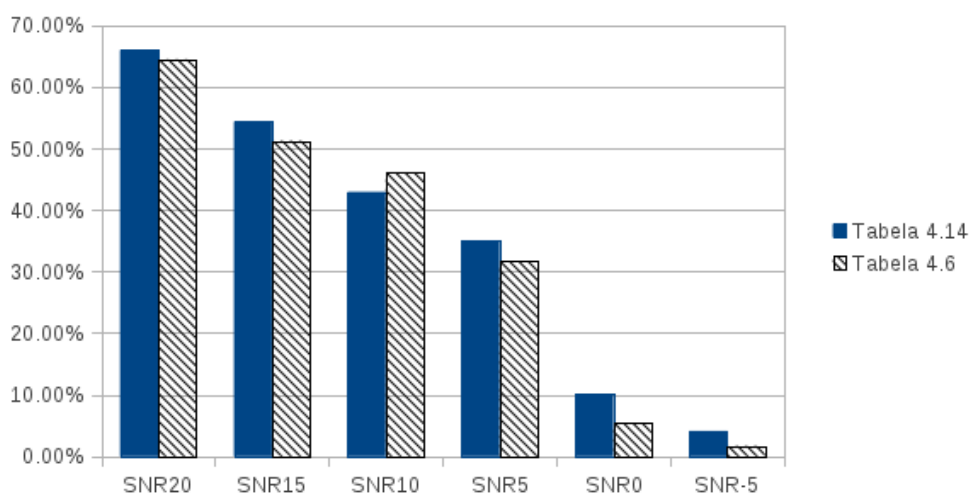
Uma ressalva a ser feita é que para calcular estes valores foram descartados os resultados dos sistemas treinados e adaptados com ruídos de carro e exposição, pois estes apresentaram um comportamento bastante ruim com estas técnicas. Estudos mais aprofundados devem ser feitos para investigar as causas deste comportamento.

A questão ainda em aberto é porque este procedimento funciona bem a partir de um sistema treinado já com locuções corrompidas com ruído mas não para locuções sem ruído. Uma hipótese levantada é que esta relação não é linear como proposto pelo modelo MAP (3.5). Se os parâmetros do sistema ruidoso estão mais próximos do modelo ideal para cada tipo de ruído, o erro causado pela aproximação linear do MAP seriam menores, explicando o melhor desempenho neste caso. Entretanto, esta ainda é uma questão que precisa ser investigada.

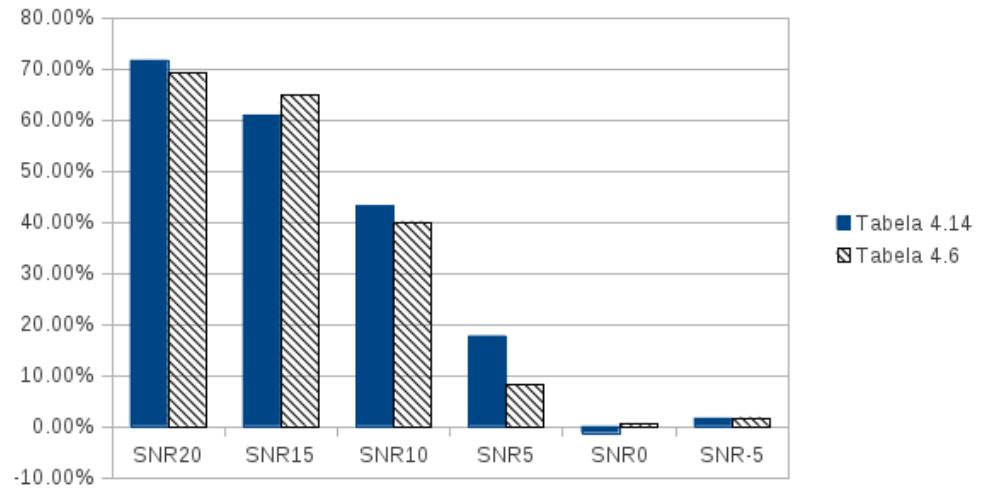
Outra questão que precisa uma maior investigação são os valores de  $\alpha$  ótimos encontrados através do processo de varredura do próprio  $\alpha$  no último teste. É preciso verificar se existe alguma correlação entre o formato do espectro do ruído e a sua intensidade com o valor de  $\alpha$ .



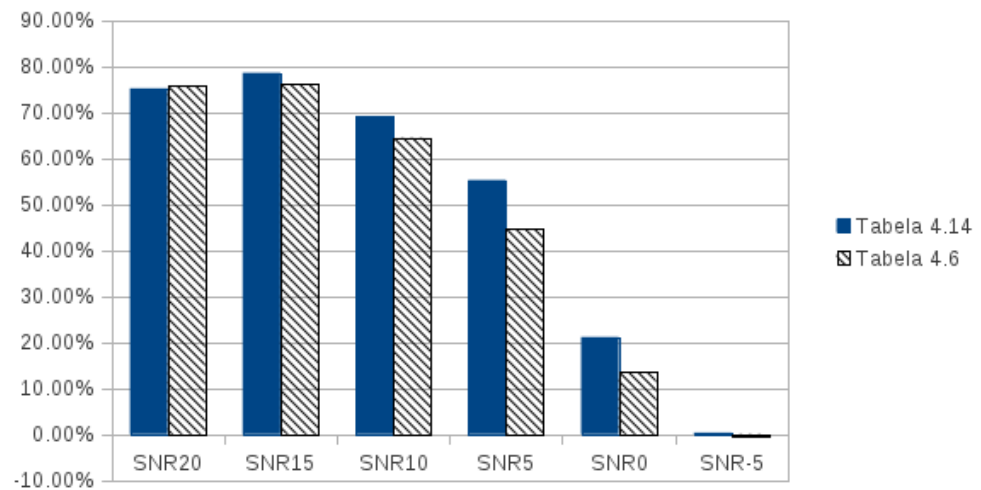
**Figura 4.4:** Gráfico comparativo entre o sistema treinado com locuções ruidosas (Tabela 4.6) e o sistema adaptado com locuções ruidosas (Tabela 4.14) para o caso do ruído restaurante



**Figura 4.5:** Gráfico comparativo entre o sistema treinado com locuções ruidosas (Tabela 4.6) e o sistema adaptado com locuções ruidosas (Tabela 4.14) para o caso do ruído rua



**Figura 4.6:** Gráfico comparativo entre o sistema treinado com locuções ruidosas (Tabela 4.6) e o sistema adaptado com locuções ruidosas (Tabela 4.14) para o caso do ruído metrô



**Figura 4.7:** Gráfico comparativo entre o sistema treinado com locuções ruidosas (Tabela 4.6) e o sistema adaptado com locuções ruidosas (Tabela 4.14) para o caso do ruído trem

# Capítulo 5

## Conclusões

Neste trabalho, um estudo sobre a adequação do material de treinamento para construir um robusto ASR foi realizado a partir de uma referência de base proposta de um sistema treinado com locuções limpas. O desempenho deste sistema proposto caiu drasticamente quando testado com locuções ruidosas, o que era esperado devido ao descasamento acústico entre os materiais de treinamento e teste.

Esta hipótese de descasamento acústico é corroborada quando foi treinado um sistema com locuções ruidosas e feitos testes com locuções limpas. Os resultados mostraram que um sistema treinado para reconhecer locuções ruidosas tem seu desempenho drasticamente reduzido quando é testado com locuções limpas. Por consequência é possível hipotetizar que ao misturar o ruído com a fala, parte da informação desta é destruída pelo ruído.

Portanto dada a hipótese de que o descasamento acústico entre os dados de treinamento e teste é que causa a queda de desempenho do sistema, uma estratégia possível para lidar com esse problema seria treinar o sistema com locuções ruidosas.

A primeira tentativa foi construir um sistema com locuções contaminadas com todos os tipos de ruído e com várias relações sinal/ruído. Desta forma o sistema seria mais robusto, já que foi exposto a uma grande variedade de situações. Os resultados dos teste mostraram que apenas uma parte desta hipótese é verdade. A medida que a relação sinal ruído cai, o modo estatístico do ruído começa a dominar. O efeito global é que as distribuições de probabilidade treinadas com dados limpos tornam-se estimativas muito pobres para as distribuições dos dados observados no ambiente ruidoso, resultando em um aumento rápido nas taxas de erro de reconhecimento.

Este resultado levou a um novo conjunto de testes, onde o sistema foi treinado com todos os tipos de ruído, mas apenas com locuções de SNR alta, ou seja, SNR



igual a 20 dB e 15 dB. Esta estratégia levou a uma grande melhora no desempenho do sistema em comparação com os anteriores. Este resultado parece confirmar a hipótese de que o sistema começa a perder informações sobre o sinal da fala. Um ponto interessante é que o desempenho é melhor mesmo para sinais altamente degradados (com baixas SNRs), o que corrobora ainda mais a hipótese acima.

Uma outra questão a ser respondida é se existe algum ganho pela formação de um sistema a ser utilizado para um tipo específico de ruído. Intuitivamente, esta abordagem leva a um melhor desempenho devido a uma melhor adequação acústica entre a formação e a condições de teste. No entanto, os resultados do teste mostraram que há um ganho pequeno na maioria dos casos, com queda de desempenho em alguns casos. Além disto há dois inconvenientes nesta abordagem: a necessidade de um sistema identificador de ruídos, e a possibilidade deste identificar de forma equivocada o ruído que se apresenta. Para verificar a influência deste erro no desempenho final do reconhecedor foram feitos diversos testes com os mesmos 8 sistemas treinados anteriormente, mas com locuções contaminadas por um ruído diferente daquele para o qual o sistema foi treinado. O desempenho do sistema caiu drasticamente nestas condições, mostrando que é mais vantajoso usar um sistema treinado com todos os tipos de ruídos do que com ruídos específicos, pois seu desempenho é um pouco inferior (em média), e não cai drasticamente se porventura o sistema de identificação do ruído cometer um erro.

A abordagem de usar locuções corrompidas com ruído para treinar os sistemas de reconhecimento de fala funciona bem para os tipos de ruídos considerados. Porém não é possível treinar um sistema com todos os tipos de ruído que possam ocorrer. Então propõe-se adaptar os parâmetros já treinados usando apenas amostras do ruído usando a técnica de adaptação MAP. Esta abordagem permitiria ter um único sistema que poderia ser modificado de acordo com o tipo de nível de ruído presente junto com o sinal de voz, utilizando trechos em que o locutor não está falando.

Partindo de um modelo treinado com locuções limpas, fez-se uma varredura inicial de  $\alpha$ . Entretanto, quando testado com sinais ruidosos, este procedimento não funcionou bem. Tentou-se então adaptar um sistema já treinado com locuções ruidosas. Para isso escolheu-se o sistema obtido com o treinamento com locuções ruidosas corrompidas com todos os tipos de ruídos e SNRs 20 dB e 15 dB, pois foi aquele de melhor desempenho até então. Desta vez conseguiu-se taxas de acertos superiores ao melhor sistema encontrado através do treinamento com locuções ruidosas.

## 5.1 Trabalhos Futuros

A fim de reduzir os efeitos da variabilidade ambiental, a eficiência do modo de treino multi-estilo é completada através da aplicação de algoritmos simples de adaptação utilizando pequenas quantidade de dados de adaptação. Este trabalho demonstrou que uma taxa de reconhecimento boa pode ser alcançada, mesmo em condições de degradação severa de ruído. Os resultados obtidos nesta tese podem ser ainda melhorados ao utilizarmos uma taxa de amostragem superior a 8 kHz.

A questão ainda em aberto é porque este procedimento funciona bem a partir de um sistema treinado já com locuções corrompidas com ruído mas não para locuções sem ruído. Uma hipótese levantada é que esta relação não é linear como proposto pelo modelo MAP (3.5). Se os parâmetros do sistema ruidoso estão mais próximos do modelo ideal para cada tipo de ruído, o erro causado pela aproximação linear do MAP seriam menores, explicando o melhor desempenho neste caso. Entretanto, esta ainda é uma questão que precisa ser investigada.

Outra questão que precisa uma maior investigação são os valores de  $\alpha$  ótimos encontrados através do processo de varredura. É preciso verificar se existe alguma correlação entre o formato do espectro do ruído e a sua intensidade com o valor deste parâmetro.

Estudos mais aprofundados devem ser feitos levando em consideração ambientes mais complexos e em situações reais de uso.

Outras técnicas de reconhecimento robusto também devem ser desenvolvidas, a fim de aumentar a robustez e reduzir a carga de complexidade global do sistema.

E por último é de suma importância tornar este sistema capaz de trabalho em tempo real.

## 5.2 Considerações Finais

O reconhecimento automático de fala (ASR) melhorou significativamente na última década, de tal forma que ele pode ser usado para transcrever fala em uma variedade de domínios, tais como aplicativos para celular, brinquedos, call center e softwares para computadores pessoais.

O objetivo de um sistema de reconhecimento é converter eficientemente e precisamente um sinal de fala em uma transcrição textual das palavras pronunciadas independente do dispositivo utilizado para gravar a locução, o sotaque do locutor, ou o ambiente acústico no qual o locutor está inserido, ou seja, o último objetivo, que ainda não foi alcançado, é ter um desempenho tão bom quanto o de um ouvinte humano.

Apesar de anos de pesquisa, os sistemas de reconhecimento de fala atuais ainda estão aquém do objetivo final que é de atuar como ou melhor que um ouvinte

humano. Porém isto não significa que é uma tarefa impossível. Veja o exemplo do Watson da IBM [41][42]: até o início de 2006 o reconhecimento natural da fala era uma tarefa com resultados bem limitados. Em apenas num curto espaço de tempo de três anos os pesquisadores da IBM foram capazes de desenvolver Watson que no início do ano 2011 venceu os dois melhores competidores que haviam participado do programa Jeopardy!. Watson é capaz de arrancar a resposta correta por si próprio, em outras palavras, ele executa mais do que os sites de busca como o Google e Bing que simplesmente apontam para um documento onde você pode encontrar uma possível resposta. Tecnólogos consideram este tipo de inteligência artificial como um santo graal, porque permitirá que as máquinas conversem mais naturalmente com as pessoas, permitindo portanto o usuário perguntar ao invés de digitar palavras-chave.

# Anexo A

## Resultados para sistema treinado com ruído

**Tabela A.1:** Taxa de acertos de palavras resultantes de um sistema treinado com ruído de aeroporto com SNR = 20 e 15 dB .

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
aeroporto	77.5%	73.6%	56.3%	28.8%	4.4%	-1.1%
balbulcio	58.1%	46.1%	33.7%	6.9%	-6.6%	-2.4%
carro	55.7%	68.4%	49.2%	5.8%	-1.5%	1.3%
exposição	37.9%	18.4%	-3.2%	-7.4%	-5.1%	-1.9%
restaurante	52.9%	47.5%	35.8%	9.8%	-4.1%	-7.5%
rua	44.6%	34.1%	35.1%	27.4%	8.1%	0.3%
metrô	30.8%	25%	-0.5%	-6.8%	-1.1%	0.7%
trem	68.9%	75.8%	69.8%	53.7%	19.1%	-1%

**Tabela A.2:** Taxa de acertos de palavras resultantes de um sistema treinado com ruído de balbulcio com SNR = 20 e 15 dB.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
aeroporto	61.7%	65.9%	50.2%	25.3%	6.5%	-9.4%
balbulcio	70.1%	68.7%	56.1%	15.9%	2.5%	2.6%
carro	55.2%	70.3%	55.6%	8.6%	-2.4%	1.2%
exposição	38.4%	20.2%	0.3%	-9.7%	-7.3%	-5.2%
restaurante	50.3%	52.2%	42.6%	11.5%	-3.3%	-5.9%
rua	43.4%	35.9%	33.4%	26.5%	-0.6%	0%
metrô	29.5%	24.8%	-0.3%	-9.8%	-2.7%	0.5%
trem	55.9%	64.1%	68.9%	58.5%	24%	-4.4%

**Tabela A.3:** Taxa de acertos de palavras resultantes de um sistema treinado com ruído de carro com SNR = 20 e 15 dB.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
aeroporto	41.5%	41.8%	31.9%	9.9%	-9%	-23.5%
balbulcio	38%	30.8%	18.5%	-5.3%	-19.4%	-23.1%
carro	71.8%	72.4%	56.3%	15.5%	0.6%	3%
exposição	41.7%	28.7%	3.9%	-14.5%	-16.3%	-20.1%
restaurante	33.1%	23.6%	12.1%	-9.9%	-26.7%	-31.2%
rua	47.3%	34%	22.1%	15.1%	-16.9%	-4.7%
metrô	38.4%	36.1%	8.7%	-10.5%	-8.5%	-1.6%
trem	52.9%	62.9%	63.3%	50.8%	22.1%	-7.7%

**Tabela A.4:** Taxa de acertos de palavras resultantes de um sistema treinado com ruído de exposição com SNR = 20 e 15 dB.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
aeroporto	6.9%	2.3%	-7.3%	-33.1%	-54.7%	-53.5%
balbulcio	12.7%	8.5%	-9.1%	-36%	-47.6%	-46%
carro	35.3%	31.7%	21.5%	5%	-11.4%	-2.7%
exposição	75.8%	70.3%	48.8%	14.5%	-4.1%	-0.3%
restaurante	13.1%	10.3%	-10.1%	-31.1%	-51.8%	-55%
rua	51.9%	36.5%	-1.8%	-24.9%	-60.5%	-2.7%
metrô	44.5%	43.1%	33.1%	9.6%	-7.7%	-3.5%
trem	17.4%	18.2%	15.7%	1.2%	-21.4%	-22.2%

**Tabela A.5:** Taxa de acertos de palavras resultantes de um sistema treinado com ruído de restaurante com SNR = 20 e 15 dB.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
aeroporto	68.8%	72.4%	60.4%	36%	7.2%	-10.4%
balbulcio	65.5%	51.7%	42%	17.3%	-0.8%	0.4%
carro	53.6%	62%	53.1%	6.8%	-1.2%	1.3%
exposição	40.5%	25.7%	3.1%	-5.9%	-4.7%	-1.9%
restaurante	70%	63.5%	49.6%	21%	1%	-0.2%
rua	48.7%	39.7%	34.5%	22.6%	-2.1%	0.6%
metrô	29.5%	28.7%	0.3%	-6%	-2.7%	-0.1%
trem	58.2%	66.7%	61.1%	47.5%	17.3%	-2.1%

**Tabela A.6:** Taxa de acertos de palavras resultantes de um sistema treinado com ruído de rua com  $SNR = 20$  e  $15$  dB.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
aeroporto	64.2%	59.3%	45.8%	13.4%	-2.3%	-15.2%
balbulcio	45.4%	40.5%	29.5%	2.5%	-10.9%	-18.2%
carro	56%	67.2%	46.1%	12.3%	-4.1%	0.6%
exposição	67.4%	50%	26.6%	4.2%	-7.1%	-4.1%
restaurante	37.7%	26.7%	20.3%	-3.4%	-15.6%	-20.7%
rua	63%	53.6%	39.1%	24.3%	3.1%	-0.2%
metrô	45.3%	44.9%	22.7%	-0.6%	-4.3%	-1.2%
trem	67.1%	72.3%	62.7%	45.6%	15.6%	-2.8%

**Tabela A.7:** Taxa de acertos de palavras resultantes de um sistema treinado com ruído de metrô com  $SNR = 20$  e  $15$  dB.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
aeroporto	4.2%	-3%	-14.5%	-46.1%	-51.2%	-48.6%
balbulcio	10.2%	4.2%	-10.6%	-38.5%	-40.5%	-38.6%
carro	28.4%	17%	-0.6%	-15.3%	-19.6%	-16.3%
exposição	53.2%	37.4%	11.1%	-7.3%	-16.3%	-12.2%
restaurante	3%	-5.1%	-20.2%	-37%	-52.3%	-57.5%
rua	25.2%	15.2%	-9.4%	-23.3%	-63.1%	-15%
metrô	75.3%	71.1%	48.1%	13.6%	1.4%	3.1%
trem	23.2%	23.5%	15%	3%	-25.2%	-22.1%

**Tabela A.8:** Taxa de acertos de palavras resultantes de um sistema treinado com ruído de trem com  $SNR = 20$  e  $15$  dB.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
aeroporto	43.8%	41.6%	27.5%	0.5%	-14%	-26.9%
balbulcio	40%	30.8%	20%	-9.6%	-22.7%	-26.7%
carro	49.2%	48.9%	35.6%	1%	-7.2%	-0.5%
exposição	31.1%	9.5%	-13.7%	-22.3%	-17.7%	-20.3%
restaurante	28.6%	18.3%	6.7%	-15.3%	-28.8%	-32.7%
rua	39.8%	27.3%	23.5%	16.6%	-10.5%	-5.4%
metrô	27.5%	19.8%	-7.7%	-18.8%	-10%	-2.7%
trem	70.2%	75.9%	68.6%	52.4%	14.4%	-4%

**Tabela A.9:** Taxa de acertos de palavras resultantes de um sistema treinado com todos os tipos de ruído mas com os níveis de ruído de 10 dB, 15 dB e 20 dB e testado por um corpus contaminado.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
aeroporto	72.9%	68.6%	57.7%	30.5%	7%	-1.8%
balbulcio	67.4%	62.8%	46.4%	17.6%	4.2%	4.5%
carro	68.5%	69.9%	50.5%	18.1%	0.5%	2.8%
exposição	73.1%	65.9%	34%	9.6%	-0.6%	1%
restaurante	67.1%	64.5%	44.9%	15.3%	0.9%	1.1%
rua	59.7%	50.5%	38.5%	34.4%	5.4%	0.8%
metrô	67.7%	64%	46%	9.2%	1.3%	2.1%
trem	72.1%	75.6%	69.1%	54.1%	17.2%	0%

**Tabela A.10:** Taxa de acertos de palavras resultantes de um sistema treinado com todos os tipos de ruído mas com os níveis de ruído de 5 dB, 10 dB, 15 dB e 20 dB e testado por um corpus contaminado.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
aeroporto	67.5%	69.4%	57%	35.2%	11.6%	-2.4%
balbulcio	64%	60.1%	42.6%	19.8%	3%	2.6%
carro	55.9%	71.1%	62.2%	22.1%	-3.4%	1.3%
exposição	67.4%	63.2%	39.9%	13.2%	-0.3%	0.1%
restaurante	61.1%	62.2%	48.1%	21.3%	-0.4%	-1.8%
rua	54%	48.3%	40.7%	35.4%	6.9%	0%
metrô	64%	62.7%	46.5%	16.3%	0.9%	2%
trem	67.3%	70.1%	66.2%	53.1%	19.8%	-0.4%

**Tabela A.11:** Taxa de acertos de palavras resultantes de um sistema treinado com todos os tipos de ruído mas com os níveis de ruído de 0 dB, 5 dB, 10 dB, 15 dB e 20 dB e testado por um corpus contaminado.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
aeroporto	60.5%	61.7%	51.7%	32.2%	7.9%	-7.2%
balbulcio	53.4%	49.7%	35.7%	16.7%	0.8%	-1.3%
carro	48.9%	58.7%	53.5%	20.9%	0.9%	0.9%
exposição	59.5%	52.9%	33.8%	13.8%	-0.6%	-0.4%
restaurante	46.1%	50.2%	39.8%	16.2%	-1.1%	-6.2%
rua	42.8%	34.3%	36.5%	28.4%	5.8%	-1.1%
metrô	51.7%	51.1%	34.5%	11.4%	2.2%	1.1%
trem	55%	65%	60.5%	53.4%	27.6%	-1.4%

## Anexo B

Resultados para sistema treinado  
com todos os ruídos com SNR =  
20 e 15 dB adaptado utilizando  
multi-estilo



**Tabela B.1:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	64.7%	62.1%	60.6%	55.8%	51.1%	47.6%	40.7%	26.5%	-1.6%
aeroporto snr 20	75.4%	67.5%	62.5%	59%	52.4%	46.1%	40.4%	28.4%	6.3%
aeroporto snr 15	69.7%	64.4%	60.9%	64%	61.5%	52.7%	42%	34.4%	15.5%
aeroporto snr 10	64.4%	65.3%	57.7%	57.1%	52.7%	46.4%	40.4%	29%	10.7%
aeroporto snr 05	37.2%	38.5%	38.5%	37.5%	33.4%	32.2%	28.1%	19.6%	6.9%
aeroporto snr 00	8.2%	11%	10.4%	12.9%	10.7%	6.6%	4.1%	-0.3%	-2.8%
aeroporto snr -5	-6%	-8.8%	-10.7%	-10.7%	-14.5%	-13.9%	-13.6%	-13.6%	-15.1%
balbulcio snr 20	71%	67.2%	61.8%	59%	53%	44.5%	38.2%	32.5%	5%
balbulcio snr 15	55.2%	49.2%	48.3%	42.9%	37.9%	33.4%	21.5%	16.1%	4.1%
balbulcio snr 10	43.2%	37.2%	36.3%	30.3%	29.3%	24.3%	15.8%	7.9%	-7.9%
balbulcio snr 05	12.3%	10.4%	7.3%	6.6%	8.2%	9.8%	6.3%	-1.3%	-12.6%
balbulcio snr 00	1.3%	-0.6%	-0.9%	-1.6%	-5.7%	-6.9%	-4.1%	-6.9%	-9.5%
balbulcio snr -5	3.5%	2.8%	2.5%	0%	-2.8%	-5%	-6.6%	-8.8%	-12%
carro snr 20	62.1%	63.7%	59.6%	56.5%	51.4%	47.9%	40.4%	29%	10.1%
carro snr 15	70.3%	68.8%	67.8%	65.9%	66.2%	59.3%	55.8%	44.8%	28.4%
carro snr 10	51.1%	51.7%	54.6%	55.5%	57.4%	50.5%	47.9%	36.9%	21.8%
carro snr 05	19.2%	24%	24.3%	25.9%	23%	24%	18.3%	12.6%	6.3%
carro snr 00	0.6%	-0.3%	1.3%	-0.9%	-2.2%	-2.8%	-4.4%	-6.6%	-9.8%
carro snr -5	4.7%	2.2%	2.5%	2.2%	0.9%	0.6%	0.6%	-0.9%	-0.9%
exposição snr 20	74.8%	74.4%	73.5%	71.9%	71.3%	67.8%	53%	41.6%	24.9%
exposição snr 15	60.6%	64.4%	62.8%	64.4%	58%	55.2%	48.9%	40.1%	16.4%
exposição snr 10	34.4%	34.7%	36.3%	40.7%	32.5%	29.7%	28.7%	23.7%	8.8%
exposição snr 05	5.7%	8.5%	4.4%	5.7%	4.7%	1.9%	-0.6%	-1.6%	-6%
exposição snr 00	-0.6%	0.3%	0.9%	0.6%	2.2%	-0.9%	-1.6%	1.3%	-1.9%
exposição snr -5	1.9%	2.5%	2.2%	0%	0.3%	1.3%	2.2%	2.5%	1.9%
restaurante snr 20	71.3%	61.8%	57.7%	54.3%	49.8%	45.7%	37.2%	30%	-1.3%
restaurante snr 15	64.4%	62.5%	53%	49.2%	43.8%	43.2%	34.4%	25.9%	8.8%
restaurante snr 10	37.2%	41%	37.9%	34.1%	27.4%	23.7%	18.6%	6.3%	0.6%
restaurante snr 05	13.9%	12%	12.9%	13.6%	15.8%	13.2%	6.9%	1.6%	-6.3%
restaurante snr 00	-1.6%	1.9%	0%	-0.6%	0.9%	-0.6%	-6.3%	-10.7%	-15.1%
restaurante snr -5	-2.8%	-5.7%	-7.3%	-8.5%	-11.4%	-11.7%	-12.3%	-15.8%	-21.1%
rua snr 20	67.8%	63.7%	61.2%	61.2%	61.8%	58%	50.8%	37.5%	20.8%
rua snr 15	53.6%	51.7%	53.6%	51.1%	46.1%	44.5%	37.9%	30.3%	12.6%
rua snr 10	39.7%	35.6%	36.9%	34.7%	27.4%	24.9%	20.8%	23.3%	2.8%
rua snr 05	35.3%	29.3%	26.8%	25.2%	24.3%	20.5%	12.6%	8.8%	1.6%
rua snr 00	8.2%	5.4%	4.1%	4.4%	2.8%	4.1%	-1.9%	-6.3%	-9.8%
rua snr -5	0.6%	0.9%	0.6%	-0.6%	-1.6%	-1.3%	-2.2%	-1.6%	-1.6%
metrô snr 20	69.4%	69.7%	68.5%	65.3%	63.4%	53.6%	46.4%	30.3%	4.7%
metrô snr 15	61.2%	59.3%	60.6%	60.6%	57.7%	52.7%	41.3%	32.8%	15.1%
metrô snr 10	38.2%	39.7%	38.5%	38.5%	39.7%	39.4%	44.5%	16.4%	2.5%
metrô snr 05	9.8%	9.1%	11.7%	12.6%	12.9%	7.6%	3.5%	3.2%	-6%
metrô snr 00	0%	0.9%	0.3%	1.9%	-0.9%	0%	0.6%	-3.5%	-1.9%
metrô snr -5	2.8%	1.9%	1.6%	2.2%	0.3%	0.9%	0.9%	0.9%	0.3%
trem snr 20	70%	70%	67.2%	65.3%	59.3%	56.5%	47.3%	36%	7.9%
trem snr 15	78.2%	72.9%	70.3%	69.7%	69.1%	64%	57.7%	42.9%	18.3%
trem snr 10	65.9%	64.7%	64.4%	61.2%	63.1%	59.6%	55.5%	42.6%	21.8%
trem snr 05	47%	50.2%	53.9%	51.7%	49.2%	53.6%	51.1%	45.1%	22.7%
trem snr 00	16.7%	16.7%	17%	18%	20.5%	18%	22.1%	13.9%	5.4%
trem snr -5	-0.9%	-2.2%	-3.2%	-7.9%	-8.2%	-6.6%	-9.1%	-10.4%	-12%

**Tabela B.2:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	64.7%	63.4%	63.4%	62.1%	62.5%	62.8%	63.7%	64.4%	64.7%
aeroporto snr 20	77.6%	77%	77.6%	78.2%	77.9%	77%	77.6%	77%	76%
aeroporto snr 15	70.3%	71%	70.7%	70%	70%	69.4%	70%	70.3%	70.3%
aeroporto snr 10	56.5%	57.4%	58.7%	59%	60.9%	61.2%	61.5%	64.4%	64%
aeroporto snr 05	28.1%	28.4%	29%	30%	32.5%	32.2%	31.9%	35.3%	35%
aeroporto snr 00	4.1%	3.2%	5%	6%	6.6%	6.6%	7.3%	7.9%	7.3%
aeroporto snr -5	-3.5%	-2.8%	-1.6%	-1.3%	-2.2%	-3.5%	-4.1%	-4.1%	-4.7%
balbulcio snr 20	78.5%	77.3%	77%	77.3%	75.7%	75.7%	75.1%	72.9%	71.9%
balbulcio snr 15	59%	59.9%	59.3%	55.2%	55.2%	55.5%	57.1%	56.8%	56.2%
balbulcio snr 10	40.7%	43.5%	44.2%	43.8%	43.2%	41.6%	39.7%	39.7%	41%
balbulcio snr 05	13.9%	13.9%	13.6%	13.2%	14.2%	13.6%	13.9%	13.2%	12.6%
balbulcio snr 00	0.9%	-0.9%	-1.6%	-0.6%	0.3%	0%	0%	0.3%	-0.3%
balbulcio snr -5	2.5%	2.2%	3.5%	3.2%	3.5%	3.8%	3.8%	3.8%	2.8%
carro snr 20	66.9%	66.2%	65.3%	64.7%	63.4%	63.4%	62.5%	61.5%	61.8%
carro snr 15	69.7%	70.7%	71.6%	71%	71.6%	69.7%	69.7%	69.1%	69.1%
carro snr 10	49.2%	48.3%	48.9%	49.5%	49.8%	49.5%	50.5%	50.8%	52.1%
carro snr 05	19.2%	19.2%	19.9%	20.2%	20.8%	20.8%	19.6%	19.9%	19.9%
carro snr 00	3.2%	1.9%	3.2%	2.8%	1.9%	2.2%	1.6%	1.3%	0.9%
carro snr -5	2.8%	3.5%	4.1%	4.1%	3.8%	3.8%	4.1%	4.1%	4.7%
exposição snr 20	71.3%	72.2%	73.5%	73.2%	72.9%	73.8%	73.5%	74.8%	74.8%
exposição snr 15	62.1%	62.5%	62.5%	62.8%	62.5%	61.8%	61.8%	60.6%	61.5%
exposição snr 10	31.5%	32.8%	34.1%	34.1%	35.6%	35.6%	35%	35%	35%
exposição snr 05	6.3%	5.7%	4.7%	5.7%	5.7%	4.1%	4.7%	4.7%	5.4%
exposição snr 00	-1.9%	-1.6%	-1.9%	-1.9%	-0.9%	-0.9%	-1.6%	-1.6%	0%
exposição snr -5	2.2%	2.2%	2.2%	2.2%	2.2%	2.5%	2.2%	1.6%	1.6%
restaurante snr 20	70.7%	69.4%	69.1%	69.4%	69.4%	71%	70.3%	71%	70.3%
restaurante snr 15	65.6%	62.8%	63.4%	62.5%	61.8%	64%	65%	64.4%	63.7%
restaurante snr 10	33.4%	34.7%	34.7%	35.6%	37.5%	37.9%	39.1%	37.9%	36.3%
restaurante snr 05	13.9%	12.6%	13.2%	13.9%	13.2%	12.3%	13.2%	13.2%	13.9%
restaurante snr 00	-2.8%	-3.2%	-2.2%	-1.3%	-0.6%	-2.2%	-1.9%	-1.6%	-2.2%
restaurante snr -5	-0.3%	-0.3%	0.3%	-0.6%	-1.9%	-2.8%	-2.5%	-2.2%	-2.8%
rua snr 20	65.9%	65%	65.6%	66.2%	66.2%	64.4%	65.6%	65.3%	67.5%
rua snr 15	49.8%	52.4%	51.1%	50.2%	52.1%	53.3%	53.3%	52.7%	53.6%
rua snr 10	43.5%	41.6%	37.5%	36.9%	37.9%	39.1%	38.2%	37.2%	39.4%
rua snr 05	32.8%	32.2%	33.1%	34.1%	34.7%	35.3%	36.3%	36.3%	35.6%
rua snr 00	8.2%	7.6%	7.3%	9.1%	8.2%	8.2%	7.3%	7.6%	7.9%
rua snr -5	2.5%	1.9%	1.9%	1.3%	1.3%	0.9%	1.3%	1.9%	1.3%
metrô snr 20	71.9%	69.1%	69.7%	70.3%	72.6%	72.9%	69.7%	70%	69.7%
metrô snr 15	61.2%	61.8%	62.5%	60.3%	59.6%	60.9%	58.7%	59%	60.9%
metrô snr 10	36%	36%	33.4%	36.6%	39.1%	39.4%	39.4%	38.8%	37.9%
metrô snr 05	6%	6%	6.9%	8.2%	8.5%	10.4%	10.4%	10.7%	9.1%
metrô snr 00	-0.6%	-1.9%	-1.9%	-1.9%	-1.6%	-0.9%	-0.9%	-0.6%	0.6%
metrô snr -5	1.3%	1.3%	1.9%	1.6%	1.9%	2.5%	2.8%	2.5%	2.8%
trem snr 20	75.1%	75.1%	74.8%	73.2%	71%	70.3%	69.7%	69.7%	69.4%
trem snr 15	77.9%	77.9%	77.9%	79.2%	78.5%	78.2%	79.2%	79.2%	79.2%
trem snr 10	65.3%	65.3%	65.9%	65.6%	65.6%	65.6%	66.2%	66.2%	65.9%
trem snr 05	45.1%	47.6%	47.6%	47.9%	49.2%	47%	47%	46.4%	47.6%
trem snr 00	12.9%	13.6%	14.5%	14.8%	15.8%	16.1%	16.7%	14.5%	14.8%
trem snr -5	0%	-0.9%	-1.3%	-0.3%	0%	-0.9%	-0.6%	-0.9%	-0.9%

**Tabela B.3:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de balbulcio, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	64.7%	64%	59.3%	53%	45.4%	39.1%	35.3%	18.6%	-0.6%
aeroporto snr 20	71.9%	64%	55.8%	48.6%	48.3%	42.9%	33.8%	19.2%	-1.9%
aeroporto snr 15	69.4%	65%	60.6%	51.4%	46.4%	42.6%	38.5%	15.5%	3.8%
aeroporto snr 10	60.6%	58.7%	58%	50.5%	44.8%	31.5%	24%	15.5%	-5%
aeroporto snr 05	31.9%	32.5%	32.8%	29.3%	25.6%	29%	20.8%	6%	-0.6%
aeroporto snr 00	8.2%	7.9%	6.9%	8.2%	8.2%	3.8%	1.6%	-2.2%	-10.4%
aeroporto snr -5	-6.3%	-6.9%	-12%	-15.1%	-21.8%	-20.8%	-22.7%	-25.2%	-22.7%
balbulcio snr 20	74.4%	70%	65.6%	58.4%	54.3%	44.2%	34.7%	24.9%	-0.6%
balbulcio snr 15	55.8%	53.3%	44.8%	37.9%	32.8%	27.4%	20.2%	18.3%	-0.6%
balbulcio snr 10	41.3%	36.6%	29%	28.7%	26.8%	23.7%	13.9%	8.2%	-6%
balbulcio snr 05	10.1%	8.5%	12%	11.7%	9.1%	8.8%	4.7%	-4.1%	-11.4%
balbulcio snr 00	0%	0%	0%	-1.9%	-2.5%	-5%	-7.3%	-9.5%	-11.7%
balbulcio snr -5	2.5%	3.2%	2.2%	1.3%	-0.9%	-2.2%	-3.5%	-6%	-9.8%
carro snr 20	62.8%	62.1%	60.3%	57.1%	54.3%	51.1%	34.4%	24.9%	9.8%
carro snr 15	68.5%	70%	65.6%	62.5%	56.5%	52.4%	48.3%	36%	22.7%
carro snr 10	52.1%	53.6%	52.7%	52.4%	52.1%	48.9%	41%	33.4%	14.2%
carro snr 05	21.1%	23.7%	24.9%	24%	23.7%	19.6%	17.7%	12.9%	8.5%
carro snr 00	1.6%	-1.6%	-2.2%	0%	-1.6%	-0.6%	-5%	-8.2%	-10.1%
carro snr -5	4.4%	2.8%	2.8%	2.8%	4.1%	2.5%	1.6%	0%	-1.3%
exposição snr 20	71.6%	71.3%	72.9%	68.1%	61.5%	54.6%	48.9%	35.6%	13.6%
exposição snr 15	61.5%	59.3%	59%	56.2%	51.7%	43.8%	41.6%	27.1%	5.7%
exposição snr 10	32.8%	29.7%	29%	31.2%	28.7%	27.1%	23.7%	14.2%	-2.2%
exposição snr 05	6.9%	9.8%	10.1%	2.8%	3.5%	0%	-3.8%	-8.2%	-8.5%
exposição snr 00	-0.9%	0.3%	0%	-0.6%	-2.2%	-2.5%	-3.2%	-6%	-9.5%
exposição snr -5	1.6%	1.6%	0.3%	-2.2%	0.3%	0.3%	0%	-1.6%	0.9%
restaurante snr 20	67.8%	61.8%	51.4%	48.9%	39.4%	32.2%	28.1%	16.1%	-4.4%
restaurante snr 15	60.9%	59.3%	52.4%	43.2%	34.1%	26.2%	24.3%	18.6%	2.5%
restaurante snr 10	36%	37.9%	38.5%	29.7%	30%	24.3%	17.7%	9.5%	-1.3%
restaurante snr 05	12.9%	11%	10.1%	11.4%	10.4%	8.5%	4.4%	-7.3%	-14.2%
restaurante snr 00	-2.8%	0.9%	-0.6%	-1.6%	-2.2%	-5%	-5%	-12.6%	-20.5%
restaurante snr -5	-0.9%	-6.9%	-6.6%	-8.8%	-10.1%	-10.7%	-15.5%	-16.7%	-20.5%
rua snr 20	64.7%	58.7%	59.6%	57.4%	53.3%	48.3%	41.6%	31.2%	5.4%
rua snr 15	53%	49.8%	50.5%	47%	41%	31.9%	31.5%	23.3%	2.8%
rua snr 10	42.9%	36.6%	31.2%	27.1%	20.8%	18.6%	10.1%	9.5%	0%
rua snr 05	33.1%	30.3%	25.9%	24.9%	18.6%	12.9%	12.6%	5.7%	-1.3%
rua snr 00	7.3%	5%	1.9%	-0.3%	-1.9%	-5.4%	-7.3%	-13.2%	-14.5%
rua snr -5	1.6%	0.9%	0%	-1.3%	-2.5%	-3.5%	-4.1%	-4.7%	-4.1%
metrô snr 20	69.1%	68.5%	67.5%	60.9%	59.6%	46.7%	41.3%	21.8%	4.1%
metrô snr 15	61.5%	61.8%	58.7%	54.9%	52.1%	44.8%	42.9%	24%	12.3%
metrô snr 10	39.7%	42%	39.4%	39.1%	31.9%	30%	24.9%	17%	3.2%
metrô snr 05	10.4%	9.5%	13.9%	11.7%	10.4%	8.5%	3.2%	0.3%	-2.5%
metrô snr 00	0%	-2.8%	-1.9%	-2.8%	-4.1%	-2.8%	-4.4%	-4.7%	-6.6%
metrô snr -5	1.6%	0%	-0.3%	-0.3%	-0.6%	-0.3%	-0.9%	-0.6%	0.3%
trem snr 20	69.4%	67.5%	62.8%	63.4%	60.3%	48.9%	43.8%	23.3%	-0.6%
trem snr 15	77.3%	76%	71%	67.2%	67.8%	61.2%	53.3%	35.6%	8.8%
trem snr 10	67.2%	64%	60.9%	62.5%	57.4%	55.8%	52.1%	42.6%	20.8%
trem snr 05	45.1%	45.1%	47%	49.8%	44.8%	48.6%	45.1%	31.9%	18%
trem snr 00	16.1%	14.2%	15.8%	16.7%	18%	18%	18%	12%	0%
trem snr -5	0%	-1.3%	-4.4%	-4.4%	-5%	-8.2%	-11.4%	-11.4%	-13.2%

**Tabela B.4:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de balbulcio, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	64.4%	63.7%	63.4%	63.4%	62.8%	62.8%	63.7%	64%	64%
aeroporto snr 20	77%	77%	76.3%	77.6%	75.7%	75.1%	75.1%	74.4%	73.8%
aeroporto snr 15	70%	69.7%	69.4%	69.4%	69.4%	68.5%	68.8%	69.1%	69.1%
aeroporto snr 10	54.9%	55.5%	58.7%	59.3%	58.4%	59%	60.9%	61.8%	61.5%
aeroporto snr 05	26.8%	28.4%	28.1%	27.1%	28.4%	31.2%	31.5%	32.5%	32.2%
aeroporto snr 00	3.8%	4.1%	5.7%	5.7%	6.3%	7.6%	7.6%	7.9%	8.8%
aeroporto snr -5	-4.1%	-4.1%	-4.7%	-4.7%	-4.7%	-5%	-6.3%	-6.6%	-6.3%
balbulcio snr 20	78.5%	77%	77%	77.6%	78.2%	77.9%	77%	75.4%	74.4%
balbulcio snr 15	58.7%	58.4%	57.7%	56.5%	55.8%	57.4%	56.5%	57.1%	56.2%
balbulcio snr 10	40.7%	42.6%	42.6%	42.9%	44.5%	43.8%	42.9%	42%	39.1%
balbulcio snr 05	13.2%	13.9%	15.1%	13.9%	13.9%	12.9%	12.6%	12.3%	10.7%
balbulcio snr 00	1.9%	0.6%	0.6%	0.6%	0.3%	0.3%	0.6%	0.3%	0.3%
balbulcio snr -5	2.2%	2.2%	2.5%	2.5%	2.2%	2.8%	3.2%	2.8%	2.2%
carro snr 20	66.9%	65.3%	65%	63.4%	64.4%	63.7%	62.8%	61.8%	62.1%
carro snr 15	69.4%	70%	71.6%	71.6%	71.6%	69.4%	69.4%	69.1%	68.5%
carro snr 10	49.8%	47.9%	49.5%	49.2%	49.2%	49.8%	50.2%	50.8%	49.8%
carro snr 05	19.2%	19.6%	20.2%	18%	18.6%	18.9%	19.9%	21.5%	21.8%
carro snr 00	3.5%	3.2%	2.5%	1.9%	1.3%	2.2%	1.6%	0.9%	1.3%
carro snr -5	2.8%	3.5%	4.1%	4.4%	3.8%	3.8%	3.5%	3.5%	4.4%
exposição snr 20	71.3%	71.9%	71.9%	73.2%	73.5%	73.2%	73.8%	73.5%	73.2%
exposição snr 15	61.8%	62.1%	61.8%	61.5%	61.8%	62.1%	62.5%	62.1%	61.2%
exposição snr 10	31.9%	31.9%	32.8%	33.4%	32.8%	32.8%	32.2%	31.5%	32.2%
exposição snr 05	5.7%	5.7%	6%	6.6%	5.4%	6%	6%	6.3%	6.6%
exposição snr 00	-1.9%	-1.9%	-2.2%	-2.8%	-2.2%	-1.6%	-1.6%	-0.9%	-0.9%
exposição snr -5	1.9%	1.9%	2.2%	1.6%	2.8%	1.9%	1.9%	1.6%	1.9%
restaurante snr 20	69.7%	70%	69.1%	69.1%	69.1%	69.1%	68.5%	69.7%	68.5%
restaurante snr 15	65.6%	62.8%	61.2%	62.5%	61.8%	63.7%	60.9%	60.9%	60.6%
restaurante snr 10	35%	34.4%	34.1%	33.8%	32.8%	34.1%	35.6%	36%	37.2%
restaurante snr 05	13.9%	11.7%	10.1%	11%	12.6%	12.6%	12.9%	12.6%	11.7%
restaurante snr 00	-2.2%	-2.2%	-0.9%	-0.9%	-0.6%	-1.6%	-1.9%	-2.2%	-2.2%
restaurante snr -5	-0.6%	0%	0.3%	-0.9%	-1.6%	-1.9%	-2.2%	-1.3%	-0.9%
rua snr 20	65%	65.6%	65.3%	64.7%	65.9%	65.9%	66.6%	65.3%	64.7%
rua snr 15	49.8%	52.1%	50.8%	49.8%	50.8%	51.7%	52.7%	52.1%	52.7%
rua snr 10	43.2%	39.7%	41.6%	37.9%	37.5%	36.9%	38.8%	39.1%	38.5%
rua snr 05	33.1%	32.5%	32.8%	33.8%	33.1%	32.2%	31.9%	32.2%	33.4%
rua snr 00	8.2%	8.2%	7.6%	8.8%	7.9%	7.6%	7.9%	7.6%	7.6%
rua snr -5	2.8%	2.2%	2.2%	1.9%	0.9%	1.3%	1.3%	0.9%	0.6%
metrô snr 20	71.6%	69.1%	68.8%	68.8%	69.1%	69.7%	68.8%	68.5%	68.1%
metrô snr 15	61.2%	62.5%	61.2%	60.9%	60.3%	59.9%	59.3%	59.3%	59.6%
metrô snr 10	36%	34.1%	37.5%	37.2%	39.4%	40.1%	39.1%	39.1%	39.4%
metrô snr 05	8.5%	8.2%	8.2%	10.1%	10.1%	9.8%	9.5%	10.4%	9.5%
metrô snr 00	-0.6%	-1.3%	-1.3%	-0.9%	-1.6%	-0.6%	-1.6%	-0.3%	-1.3%
metrô snr -5	1.3%	0.9%	0.9%	1.3%	0.9%	0.9%	1.3%	1.6%	1.9%
trem snr 20	75.4%	74.4%	73.8%	72.9%	72.2%	70.7%	70%	69.4%	69.4%
trem snr 15	77.9%	77.9%	78.2%	78.5%	79.2%	78.9%	78.5%	76.7%	77.3%
trem snr 10	65.3%	65.6%	65.9%	65.9%	66.2%	66.2%	66.9%	66.9%	67.2%
trem snr 05	45.4%	46.1%	47.9%	47%	46.4%	43.8%	43.8%	45.7%	46.1%
trem snr 00	13.2%	13.6%	15.1%	15.1%	15.1%	16.4%	16.4%	16.1%	16.1%
trem snr -5	0.3%	-0.6%	-0.3%	-0.6%	-0.9%	-0.3%	-0.3%	0%	-0.3%

**Tabela B.5:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de carro, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	32.2%	31.9%	28.4%	24%	17.7%	9.1%	6.9%	-1.9%	-4.1%
aeroporto snr 20	36.3%	37.9%	32.5%	18.3%	14.5%	7.3%	-7.3%	-16.1%	-16.7%
aeroporto snr 15	47.3%	39.4%	36.3%	24.6%	19.9%	8.5%	-13.6%	-21.5%	-19.6%
aeroporto snr 10	32.8%	23.7%	17%	12.6%	1.9%	-8.8%	-11.7%	-23.3%	-27.1%
aeroporto snr 05	12.6%	10.1%	3.2%	0%	-6.6%	-16.1%	-20.2%	-22.1%	-18.9%
aeroporto snr 00	-2.8%	-1.9%	-2.2%	-3.8%	-14.5%	-13.6%	-21.1%	-25.2%	-21.5%
aeroporto snr -5	-6.3%	-14.2%	-12%	-18.9%	-18.6%	-22.1%	-18.3%	-20.8%	-10.4%
balbulcio snr 20	37.9%	37.2%	28.1%	23%	14.5%	6%	-1.6%	-13.9%	-20.5%
balbulcio snr 15	30.6%	26.2%	18.9%	13.2%	8.5%	1.6%	-7.6%	-15.5%	-19.9%
balbulcio snr 10	20.2%	15.5%	9.8%	10.1%	4.1%	-0.3%	-14.2%	-24.6%	-18.6%
balbulcio snr 05	-2.8%	-12%	-6%	-12.6%	-16.4%	-23.7%	-23.7%	-33.4%	-29%
balbulcio snr 00	-8.5%	-11%	-16.7%	-23%	-24.3%	-29%	-33.1%	-32.5%	-27.4%
balbulcio snr -5	-2.5%	-6%	-8.8%	-15.5%	-18%	-24%	-26.5%	-27.8%	-20.2%
carro snr 20	33.8%	33.8%	29.7%	26.5%	23.7%	13.2%	-3.2%	-8.2%	-13.6%
carro snr 15	43.5%	40.1%	33.4%	27.8%	20.5%	14.5%	5.4%	-3.8%	-13.9%
carro snr 10	41%	37.2%	35.6%	29%	25.6%	16.7%	0.3%	-6%	-12%
carro snr 05	17.7%	18.6%	13.6%	10.1%	12%	4.7%	-1.3%	-10.1%	-16.1%
carro snr 00	-1.3%	-2.8%	-2.8%	-3.8%	-8.2%	-9.8%	-13.2%	-14.2%	-13.6%
carro snr -5	0%	-0.6%	-3.8%	-5.4%	-6.6%	-9.1%	-13.9%	-13.2%	-17.7%
exposição snr 20	41%	36.3%	33.4%	34.7%	32.8%	32.5%	20.5%	0.6%	-2.5%
exposição snr 15	32.5%	33.4%	30.6%	29.3%	25.6%	19.6%	10.1%	0.9%	-7.6%
exposição snr 10	8.2%	9.8%	8.2%	0.9%	0.3%	-0.6%	-9.5%	-5.7%	-8.5%
exposição snr 05	-4.7%	-2.8%	-3.8%	1.3%	-4.1%	-6.6%	-16.1%	-11.7%	-9.8%
exposição snr 00	-6%	-8.2%	-6.9%	-7.3%	-8.8%	-12%	-10.1%	-5.4%	-6.3%
exposição snr -5	-0.9%	-0.3%	-1.6%	-2.8%	-7.3%	-11.4%	-13.2%	-8.2%	-4.7%
restaurante snr 20	23.7%	19.2%	17.4%	14.8%	8.2%	-3.8%	-4.7%	-16.7%	-14.5%
restaurante snr 15	24.3%	19.2%	13.6%	4.1%	3.8%	-8.2%	-14.2%	-19.2%	-15.1%
restaurante snr 10	21.1%	18%	2.8%	-1.9%	-6%	-13.2%	-15.5%	-14.8%	-18.3%
restaurante snr 05	-1.3%	-3.5%	-9.8%	-19.9%	-21.5%	-24.3%	-24%	-26.8%	-19.9%
restaurante snr 00	-6.3%	-10.7%	-14.2%	-17%	-19.9%	-22.4%	-28.7%	-21.8%	-17.4%
restaurante snr -5	-9.5%	-14.5%	-16.1%	-24.6%	-29.3%	-33.8%	-35.3%	-33.8%	-20.8%
rua snr 20	36.6%	36.3%	24.9%	21.5%	13.2%	7.9%	1.9%	-6.6%	-12.6%
rua snr 15	32.5%	30.6%	25.6%	18.6%	13.6%	3.2%	-7.3%	-7.9%	-9.5%
rua snr 10	19.9%	13.6%	9.8%	3.5%	-1.6%	-6.3%	-13.6%	-16.4%	-16.7%
rua snr 05	16.1%	12.3%	7.6%	1.6%	-3.8%	-10.1%	-19.6%	-19.6%	-19.2%
rua snr 00	0.6%	-1.3%	-8.8%	-14.2%	-19.9%	-25.2%	-31.5%	-24%	-16.1%
rua snr -5	-2.8%	-3.2%	-7.9%	-10.1%	-15.1%	-15.5%	-20.8%	-23.7%	-13.6%
metrô snr 20	36.9%	38.5%	38.2%	34.1%	31.5%	18%	11.4%	2.2%	-8.5%
metrô snr 15	34.7%	33.8%	34.1%	31.2%	23%	23%	22.4%	6.6%	-2.8%
metrô snr 10	23.7%	24.9%	21.5%	23%	17.7%	8.5%	5%	-2.2%	-6%
metrô snr 05	-0.9%	-2.8%	-4.4%	-4.7%	-8.8%	-7.3%	-5.7%	-4.7%	-3.8%
metrô snr 00	-6.3%	-6.6%	-9.1%	-9.1%	-9.1%	-10.4%	-11%	-7.3%	-5.4%
metrô snr -5	1.9%	2.5%	0.6%	-0.6%	-4.4%	-5.4%	-1.6%	-4.1%	-0.3%
trem snr 20	44.2%	40.4%	36%	28.1%	19.2%	15.1%	2.8%	-7.9%	-13.2%
trem snr 15	45.7%	43.8%	45.7%	43.2%	38.8%	33.8%	4.7%	-0.6%	-12.3%
trem snr 10	44.8%	44.2%	45.4%	40.7%	35.6%	28.4%	13.6%	-5%	-10.7%
trem snr 05	40.7%	40.1%	38.2%	35%	33.8%	22.4%	11.4%	-2.5%	-14.5%
trem snr 00	13.9%	12.3%	15.1%	14.8%	13.9%	10.1%	9.8%	-0.6%	-14.5%
trem snr -5	-5%	-8.2%	-12%	-10.1%	-13.6%	-16.4%	-17%	-20.2%	-18%

**Tabela B.6:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de carro, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	35.6%	35%	38.8%	38.5%	37.9%	36.3%	34.7%	35%	35%
aeroporto snr 20	45.7%	46.1%	44.8%	45.7%	43.8%	43.2%	41.6%	38.5%	37.5%
aeroporto snr 15	45.1%	47.3%	49.2%	50.8%	49.8%	49.2%	48.3%	47.3%	46.4%
aeroporto snr 10	34.1%	35.3%	34.7%	35%	36.6%	36.9%	35.3%	33.4%	33.8%
aeroporto snr 05	11.7%	11.7%	11%	11%	12.3%	12.3%	12.9%	13.9%	13.6%
aeroporto snr 00	-0.6%	-0.6%	-0.6%	-0.9%	-1.3%	-1.6%	-2.5%	-2.5%	-2.8%
aeroporto snr -5	-4.1%	-5%	-4.4%	-3.8%	-6.6%	-5.7%	-5%	-5%	-6.3%
balbulcio snr 20	42.9%	42.3%	41.6%	41%	41.6%	41.3%	40.4%	40.1%	38.8%
balbulcio snr 15	30%	29.3%	29%	28.4%	29%	31.9%	30.6%	29%	30%
balbulcio snr 10	20.5%	20.2%	19.6%	21.1%	23%	22.4%	23%	21.5%	19.9%
balbulcio snr 05	-0.9%	0.6%	2.2%	1.9%	0.9%	0.3%	-0.3%	-2.2%	-1.3%
balbulcio snr 00	-6%	-7.3%	-7.6%	-7.6%	-6.6%	-6%	-6.9%	-7.9%	-9.1%
balbulcio snr -5	-0.6%	-0.3%	-1.9%	-2.2%	-1.9%	-1.3%	-1.3%	-0.6%	-1.6%
carro snr 20	37.9%	37.5%	36%	36.3%	33.1%	34.4%	35.3%	35%	35%
carro snr 15	43.8%	42.6%	43.8%	45.4%	44.8%	44.2%	44.5%	44.8%	44.2%
carro snr 10	37.5%	37.2%	38.8%	39.1%	40.4%	41.6%	41.6%	41.3%	41.3%
carro snr 05	15.8%	13.2%	13.2%	17%	18%	19.2%	17.4%	17.4%	17.4%
carro snr 00	-3.8%	-2.2%	-1.3%	-0.9%	-1.3%	-0.6%	-0.3%	-0.3%	-1.9%
carro snr -5	0%	0%	0%	0%	-0.3%	0.3%	0%	0%	0%
exposição snr 20	34.4%	37.2%	37.9%	38.5%	37.5%	40.4%	41%	42.6%	42.3%
exposição snr 15	22.4%	22.7%	26.5%	29.3%	30%	29.3%	32.2%	31.9%	32.8%
exposição snr 10	4.1%	5%	6%	5.7%	7.9%	10.1%	11.7%	13.2%	10.4%
exposição snr 05	-5%	-3.5%	-4.7%	-6.9%	-6.9%	-6.6%	-4.1%	-4.1%	-4.7%
exposição snr 00	-7.3%	-6.6%	-6%	-6.3%	-6%	-6.3%	-7.6%	-6.9%	-6.3%
exposição snr -5	0%	-0.3%	-0.3%	0%	-0.6%	-0.3%	-0.6%	-0.6%	-1.3%
restaurante snr 20	30.3%	27.1%	26.8%	27.4%	26.8%	26.2%	26.8%	25.2%	24%
restaurante snr 15	30%	27.8%	28.4%	26.8%	27.1%	24.9%	25.9%	26.5%	25.6%
restaurante snr 10	24%	23.3%	23%	21.8%	21.1%	21.8%	22.4%	19.2%	21.1%
restaurante snr 05	0.3%	-0.9%	0%	2.2%	0.3%	0.6%	-0.3%	0%	0%
restaurante snr 00	-5%	-5.4%	-4.4%	-4.7%	-5.4%	-5.4%	-7.3%	-7.3%	-6.3%
restaurante snr -5	-7.9%	-7.9%	-8.2%	-8.5%	-9.1%	-9.1%	-8.8%	-9.8%	-9.5%
rua snr 20	37.9%	37.5%	36%	35.3%	34.1%	37.2%	36.9%	36.3%	36.3%
rua snr 15	32.8%	31.9%	34.1%	32.5%	34.1%	32.2%	33.8%	31.9%	35%
rua snr 10	18%	24%	23.3%	21.8%	19.6%	18.9%	21.8%	21.1%	20.8%
rua snr 05	14.8%	15.8%	15.5%	14.8%	14.2%	15.5%	15.8%	17%	16.1%
rua snr 00	4.4%	4.7%	4.7%	4.4%	3.8%	2.2%	1.6%	1.3%	1.6%
rua snr -5	0.3%	0.6%	0.6%	-0.6%	-1.3%	-2.8%	-2.2%	-1.6%	-2.8%
metrô snr 20	37.9%	37.2%	36.9%	37.9%	36.9%	37.2%	35.6%	35.6%	36%
metrô snr 15	32.8%	33.1%	32.5%	36.3%	35.3%	35%	35%	35.6%	35.6%
metrô snr 10	18.6%	22.1%	23.3%	22.1%	22.4%	22.4%	20.5%	23%	24.6%
metrô snr 05	-5.4%	-5.7%	-5%	-3.5%	-1.9%	-2.5%	-1.3%	-0.9%	-1.3%
metrô snr 00	-5.7%	-6%	-6.6%	-6.3%	-6%	-5.4%	-4.4%	-5.7%	-6.3%
metrô snr -5	0.9%	1.6%	1.6%	1.6%	1.6%	1.3%	1.6%	1.9%	1.9%
trem snr 20	45.7%	44.8%	45.7%	44.8%	44.5%	43.8%	43.5%	44.5%	43.5%
trem snr 15	46.7%	46.4%	46.7%	46.7%	44.8%	45.4%	45.4%	46.4%	47.3%
trem snr 10	43.5%	43.2%	44.5%	45.4%	45.1%	43.5%	43.8%	44.8%	44.5%
trem snr 05	36.9%	37.5%	38.5%	39.4%	38.2%	37.9%	39.4%	39.1%	40.7%
trem snr 00	8.8%	9.5%	9.8%	10.1%	9.5%	9.1%	10.4%	10.7%	12.6%
trem snr -5	-6%	-5.7%	-5.7%	-6.3%	-6.6%	-6%	-6.3%	-6.6%	-6%

**Tabela B.7:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de exposição, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	31.5%	29.3%	28.1%	23.3%	9.8%	0.6%	-8.2%	-13.6%	-5%
aeroporto snr 20	40.1%	36.9%	37.9%	28.4%	16.1%	4.4%	-1.6%	-7.6%	-6.3%
aeroporto snr 15	44.2%	39.7%	34.1%	32.5%	22.4%	7.6%	-1.6%	-14.2%	-13.6%
aeroporto snr 10	28.4%	22.7%	25.2%	21.5%	9.5%	1.9%	-6.3%	-19.6%	-14.5%
aeroporto snr 05	16.4%	12.6%	12%	2.5%	-3.8%	-8.8%	-17.4%	-16.4%	-8.8%
aeroporto snr 00	0%	2.5%	0.9%	-2.5%	-5%	-7.6%	-17%	-18%	-9.5%
aeroporto snr -5	-8.2%	-9.5%	-8.2%	-5.7%	-11.4%	-11.4%	-16.1%	-13.9%	-6.9%
balbulcio snr 20	37.9%	32.5%	33.1%	30%	20.2%	4.7%	-6%	-11%	-19.2%
balbulcio snr 15	30.6%	27.4%	26.5%	18.3%	17%	5%	-2.5%	-14.8%	-16.1%
balbulcio snr 10	23.3%	21.8%	24%	19.6%	5.7%	-8.5%	-7.6%	-18.9%	-22.7%
balbulcio snr 05	4.1%	2.5%	-6.3%	-6.6%	-5.7%	-9.5%	-11%	-26.5%	-20.2%
balbulcio snr 00	-5.4%	-10.1%	-11%	-13.2%	-15.5%	-15.1%	-18.6%	-26.5%	-23.7%
balbulcio snr -5	-4.7%	-6.9%	-8.8%	-7.6%	-9.8%	-12.3%	-18.3%	-22.1%	-21.8%
carro snr 20	39.7%	39.7%	36.3%	30.9%	22.4%	15.5%	7.9%	-9.5%	-7.6%
carro snr 15	49.8%	45.1%	41.3%	38.2%	28.1%	15.8%	1.9%	-4.4%	-6%
carro snr 10	42.9%	39.7%	45.1%	35.3%	24.6%	19.9%	2.8%	-5.4%	-6.3%
carro snr 05	18%	24.3%	17.4%	16.4%	4.1%	-4.1%	-7.9%	-12.3%	-4.7%
carro snr 00	-1.6%	-3.2%	-3.5%	-1.9%	-6%	-9.5%	-14.8%	-13.2%	-10.4%
carro snr -5	-0.3%	-0.6%	-1.3%	-2.2%	-2.8%	-3.8%	-6.6%	-3.5%	-0.6%
exposição snr 20	29.7%	22.4%	22.4%	16.1%	11.7%	6.6%	-3.5%	-9.5%	-12%
exposição snr 15	28.7%	20.2%	16.7%	11%	11.4%	5.4%	-3.2%	-4.4%	-12%
exposição snr 10	2.2%	0%	-4.1%	-6%	-10.1%	-16.7%	-18%	-18%	-18%
exposição snr 05	-3.5%	-3.8%	-3.8%	-4.4%	-4.7%	-8.8%	-13.2%	-22.1%	-14.8%
exposição snr 00	-12.6%	-22.4%	-17.4%	-18%	-18.3%	-18.3%	-21.8%	-27.8%	-19.6%
exposição snr -5	-7.6%	-14.8%	-14.8%	-13.9%	-14.2%	-19.2%	-18.3%	-19.2%	-21.8%
restaurante snr 20	28.4%	29.7%	22.4%	19.2%	16.7%	6%	-1.6%	-13.9%	-16.1%
restaurante snr 15	29.3%	28.1%	26.8%	19.2%	16.4%	4.7%	-5%	-15.8%	-21.5%
restaurante snr 10	22.4%	16.4%	12.9%	3.8%	-4.1%	-6.9%	-11.4%	-19.6%	-17.7%
restaurante snr 05	-0.9%	-0.9%	-1.6%	-7.9%	-11%	-11.7%	-15.8%	-23.7%	-27.4%
restaurante snr 00	-8.5%	-7.3%	-9.8%	-9.1%	-11%	-13.2%	-19.9%	-21.1%	-24.9%
restaurante snr -5	-12.9%	-13.2%	-15.5%	-16.1%	-17.7%	-17.7%	-18.9%	-22.7%	-24.3%
rua snr 20	27.8%	19.9%	18%	15.8%	12.9%	0.6%	-1.9%	-8.5%	-8.2%
rua snr 15	25.2%	22.7%	12.6%	9.1%	10.4%	-1.6%	-1.6%	-11%	-9.1%
rua snr 10	19.6%	12.6%	6.6%	1.9%	-6.9%	-8.2%	-18.3%	-18.9%	-18.6%
rua snr 05	14.8%	11.7%	6.3%	3.8%	2.2%	-8.2%	-12.9%	-23%	-15.8%
rua snr 00	-1.6%	-2.5%	-5.4%	-8.8%	-17%	-22.7%	-26.5%	-22.4%	-6.6%
rua snr -5	-5%	-4.1%	-6.3%	-7.3%	-9.8%	-11.7%	-12.6%	-12%	-6%
metrô snr 20	37.5%	31.9%	25.9%	26.5%	17.4%	8.5%	9.8%	-0.6%	-12.3%
metrô snr 15	39.1%	34.7%	30%	23.3%	16.4%	13.2%	5.4%	0%	-11.4%
metrô snr 10	25.6%	24.9%	23.3%	18.9%	13.2%	15.5%	-1.3%	-7.9%	-9.5%
metrô snr 05	5%	9.5%	6%	-1.6%	-0.6%	-8.5%	-8.8%	-13.6%	-15.1%
metrô snr 00	-4.1%	-12.3%	-10.4%	-12.3%	-17.7%	-17.7%	-16.7%	-19.6%	-7.3%
metrô snr -5	2.5%	-0.6%	-3.8%	-6.3%	-10.4%	-12%	-15.1%	-12.3%	-5.4%
trem snr 20	44.5%	41.3%	38.2%	30.3%	19.2%	4.7%	-2.2%	-13.9%	-9.5%
trem snr 15	48.6%	50.2%	49.5%	42.3%	35.6%	12.9%	0.6%	-5.7%	-10.1%
trem snr 10	44.8%	44.5%	40.7%	39.4%	31.9%	15.1%	-0.6%	-12%	-13.9%
trem snr 05	40.7%	38.2%	39.7%	39.1%	32.8%	17.4%	-2.5%	-9.8%	-8.5%
trem snr 00	13.6%	17%	12%	12.9%	7.6%	3.5%	-7.3%	-14.5%	-6.9%
trem snr -5	-3.8%	-8.2%	-8.8%	-9.5%	-10.1%	-12.6%	-12.3%	-5.7%	-1.6%

**Tabela B.8:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de exposição, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	33.8%	35%	35%	34.7%	35%	33.1%	33.1%	33.4%	33.1%
aeroporto snr 20	46.7%	44.5%	44.2%	42.6%	43.2%	42%	41.3%	41.6%	40.1%
aeroporto snr 15	49.2%	48.6%	50.2%	49.5%	48.9%	47.9%	47%	45.4%	45.7%
aeroporto snr 10	34.1%	33.8%	33.1%	32.8%	31.2%	29.3%	30.6%	31.5%	28.7%
aeroporto snr 05	13.6%	11%	12.9%	12.3%	11.7%	13.6%	13.2%	14.2%	13.9%
aeroporto snr 00	-0.3%	-0.3%	-0.6%	-1.9%	-0.9%	0.3%	-0.3%	-0.6%	-1.3%
aeroporto snr -5	-4.1%	-4.7%	-4.7%	-3.8%	-4.4%	-5.4%	-6.3%	-7.9%	-7.6%
balbulcio snr 20	43.2%	40.4%	40.7%	40.4%	39.7%	40.1%	36.6%	36.9%	36.9%
balbulcio snr 15	32.5%	31.2%	30.6%	32.8%	30.9%	34.1%	31.5%	32.2%	31.2%
balbulcio snr 10	24.6%	24.3%	25.2%	24.6%	24%	24.6%	22.4%	21.8%	21.8%
balbulcio snr 05	1.3%	4.4%	2.2%	3.2%	1.6%	5%	5.4%	5%	3.8%
balbulcio snr 00	-6.6%	-6%	-5.4%	-5.4%	-5%	-5%	-5.7%	-4.7%	-5%
balbulcio snr -5	-0.9%	-1.6%	-2.2%	-2.8%	-3.2%	-3.8%	-3.5%	-4.4%	-5%
carro snr 20	38.8%	39.4%	37.9%	38.8%	40.4%	40.1%	40.4%	39.1%	37.9%
carro snr 15	44.2%	46.1%	47.6%	48.3%	48.6%	48.9%	50.2%	49.5%	50.5%
carro snr 10	39.1%	40.7%	40.7%	41.3%	40.7%	41%	40.4%	41.6%	41.6%
carro snr 05	14.2%	18.9%	19.6%	18%	17.7%	17%	15.8%	17.4%	19.2%
carro snr 00	-1.3%	-0.3%	0.3%	1.6%	0.3%	0%	-0.6%	0.6%	-0.3%
carro snr -5	0.6%	0.6%	0.9%	0.6%	0%	0%	-0.6%	-0.3%	-0.6%
exposição snr 20	31.2%	31.2%	33.1%	34.1%	33.8%	31.5%	32.2%	32.2%	31.9%
exposição snr 15	24.3%	27.4%	29%	28.4%	31.5%	30.6%	30.6%	30.3%	28.7%
exposição snr 10	3.5%	3.5%	3.8%	5.4%	8.2%	8.2%	6.6%	4.4%	3.2%
exposição snr 05	-5%	-6.9%	-5.4%	-5.7%	-7.3%	-6%	-6.6%	-6.3%	-8.5%
exposição snr 00	-9.1%	-9.1%	-10.4%	-10.7%	-10.7%	-12%	-11.7%	-10.7%	-13.6%
exposição snr -5	-0.6%	-0.6%	-2.2%	-2.8%	-4.4%	-5.7%	-6.6%	-6.6%	-7.6%
restaurante snr 20	29.7%	31.2%	28.1%	27.1%	27.4%	27.8%	27.4%	29%	28.1%
restaurante snr 15	29%	28.7%	28.4%	26.5%	28.1%	28.4%	26.8%	26.5%	27.1%
restaurante snr 10	23%	22.1%	24.9%	25.2%	25.9%	24.3%	23.7%	23%	22.1%
restaurante snr 05	0.9%	2.8%	2.5%	0.3%	-0.6%	-1.3%	-2.8%	-1.3%	-0.9%
restaurante snr 00	-4.4%	-4.7%	-4.7%	-5%	-7.6%	-7.6%	-8.2%	-8.5%	-9.8%
restaurante snr -5	-6.9%	-6.3%	-7.3%	-9.1%	-8.8%	-9.5%	-9.8%	-11%	-12%
rua snr 20	35.3%	33.4%	32.2%	32.2%	31.2%	29%	28.4%	27.1%	26.5%
rua snr 15	31.2%	32.8%	33.1%	30.6%	28.7%	31.2%	31.2%	28.1%	26.5%
rua snr 10	23.7%	23.7%	21.8%	20.8%	19.2%	20.2%	18.3%	17.4%	18.9%
rua snr 05	15.5%	13.2%	11%	13.9%	12.9%	14.5%	13.6%	15.1%	14.5%
rua snr 00	4.7%	4.1%	4.1%	3.8%	3.8%	2.2%	1.9%	-0.9%	-0.9%
rua snr -5	0.9%	-0.9%	-2.5%	-2.8%	-3.5%	-4.1%	-2.5%	-3.8%	-4.1%
metrô snr 20	37.9%	37.5%	40.4%	39.1%	39.4%	39.4%	39.1%	38.8%	37.9%
metrô snr 15	30.9%	34.1%	35.3%	33.1%	33.4%	37.2%	36%	36%	37.5%
metrô snr 10	20.8%	19.6%	17.4%	19.9%	25.9%	24.6%	25.2%	25.6%	24.9%
metrô snr 05	-5%	-5.4%	-0.9%	-2.2%	0.9%	4.4%	3.2%	4.7%	4.4%
metrô snr 00	-5.4%	-5.4%	-6%	-7.3%	-5.7%	-7.9%	-8.2%	-6%	-6.3%
metrô snr -5	0.6%	0.6%	0.6%	1.3%	1.3%	0.9%	1.3%	1.9%	1.9%
trem snr 20	45.4%	44.8%	45.4%	45.1%	44.2%	43.8%	44.5%	43.8%	44.2%
trem snr 15	47.6%	47.6%	48.3%	47%	47.3%	45.7%	46.4%	46.7%	47.9%
trem snr 10	45.7%	46.4%	44.2%	45.7%	46.4%	46.1%	46.7%	47.3%	46.4%
trem snr 05	38.5%	37.5%	36.6%	37.5%	39.1%	37.9%	38.8%	39.1%	38.5%
trem snr 00	6%	6.9%	7.6%	8.8%	9.1%	10.1%	12%	12.3%	13.2%
trem snr -5	-5.7%	-6.3%	-3.8%	-4.1%	-4.7%	-4.4%	-4.1%	-3.8%	-4.1%



**Tabela B.9:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de restaurante, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	65.3%	65%	61.2%	58.7%	52.1%	44.2%	36.9%	28.4%	9.8%
aeroporto snr 20	76.3%	63.4%	58%	55.5%	50.8%	42.9%	36.6%	25.2%	6.6%
aeroporto snr 15	68.8%	65.9%	63.1%	59%	53%	45.1%	40.1%	22.7%	10.7%
aeroporto snr 10	62.1%	62.5%	61.2%	55.5%	51.7%	39.7%	29%	14.2%	4.4%
aeroporto snr 05	36.9%	36.3%	32.5%	31.2%	28.7%	24.6%	16.1%	7.6%	-3.8%
aeroporto snr 00	9.1%	13.2%	12.3%	12.6%	8.2%	5%	0.9%	-4.4%	-9.1%
aeroporto snr -5	-3.5%	-4.4%	-7.9%	-11.7%	-16.4%	-14.5%	-15.8%	-20.8%	-22.4%
balbulcio snr 20	71.9%	65.9%	60.9%	54.6%	52.7%	45.4%	39.7%	26.5%	12.9%
balbulcio snr 15	56.2%	47.9%	45.7%	42%	30.6%	28.1%	20.2%	18.9%	6%
balbulcio snr 10	42.6%	37.9%	35.6%	30.3%	26.5%	28.4%	13.6%	11.7%	-0.3%
balbulcio snr 05	12.3%	11.4%	9.5%	8.2%	11%	5.4%	1.3%	-7.3%	-15.1%
balbulcio snr 00	0.6%	-1.3%	-0.6%	-2.5%	-2.5%	-5.7%	-8.8%	-12%	-12.9%
balbulcio snr -5	2.8%	2.8%	2.5%	-1.3%	-3.8%	-5.4%	-9.5%	-8.8%	-15.5%
carro snr 20	60.9%	61.2%	61.8%	59.3%	61.5%	55.8%	47.3%	35%	18.9%
carro snr 15	70.7%	68.8%	69.4%	69.1%	67.2%	57.4%	49.2%	44.2%	33.4%
carro snr 10	50.2%	51.4%	52.4%	51.7%	55.2%	54.3%	43.8%	34.1%	18.9%
carro snr 05	21.1%	23%	29%	27.1%	30%	24.3%	22.7%	18.9%	8.5%
carro snr 00	0.9%	-0.9%	-0.3%	-1.9%	-1.6%	-1.9%	-3.5%	-4.4%	-6.9%
carro snr -5	4.4%	3.2%	3.8%	3.5%	3.8%	3.2%	1.6%	-0.3%	-2.8%
exposição snr 20	73.2%	73.5%	71.9%	72.6%	66.9%	60.3%	57.4%	48.6%	25.6%
exposição snr 15	61.8%	60.9%	61.8%	57.1%	57.7%	53.6%	43.8%	28.7%	18%
exposição snr 10	34.7%	34.4%	30.9%	29.3%	31.2%	34.4%	33.1%	19.2%	13.6%
exposição snr 05	6%	6.9%	5.4%	11.4%	8.8%	6.9%	5.4%	0%	-3.8%
exposição snr 00	0%	-0.3%	-0.6%	1.6%	-1.3%	-1.3%	-0.3%	-1.9%	-5.4%
exposição snr -5	1.6%	2.2%	0.3%	-0.6%	0.3%	-0.6%	-1.3%	-0.9%	-2.5%
restaurante snr 20	68.1%	58.7%	54.6%	52.7%	48.3%	40.1%	36%	19.6%	3.2%
restaurante snr 15	65.3%	57.4%	48.9%	46.1%	42.9%	37.5%	33.4%	25.2%	6.9%
restaurante snr 10	34.1%	41.6%	36.9%	34.4%	31.9%	26.2%	20.2%	14.2%	6.3%
restaurante snr 05	17.4%	13.9%	11%	11%	10.7%	5.7%	6.9%	5.4%	-6%
restaurante snr 00	-1.6%	1.9%	-0.6%	-1.9%	-4.1%	-2.8%	-2.2%	-9.5%	-17.7%
restaurante snr -5	-3.5%	-8.8%	-8.5%	-12%	-11.4%	-14.2%	-14.2%	-13.9%	-18.9%
rua snr 20	64.7%	59.9%	58.7%	58.4%	52.4%	49.2%	37.5%	28.4%	15.1%
rua snr 15	52.4%	53.3%	50.8%	50.8%	44.5%	39.1%	36.9%	25.9%	8.2%
rua snr 10	38.2%	40.7%	31.2%	24.6%	25.2%	18.6%	16.4%	5.7%	-2.2%
rua snr 05	31.9%	26.8%	26.2%	23.3%	20.8%	17.4%	6%	-0.3%	-5%
rua snr 00	7.9%	6.6%	2.5%	2.5%	-1.6%	-6.6%	-11%	-18.3%	-20.8%
rua snr -5	0.6%	-1.3%	-2.2%	-1.9%	-2.2%	-1.9%	-3.8%	-3.8%	-8.8%
metrô snr 20	72.6%	69.4%	66.9%	65.3%	58.7%	51.1%	45.1%	33.1%	8.2%
metrô snr 15	62.1%	59.3%	57.4%	53.6%	50.8%	48.9%	40.4%	29.7%	11.4%
metrô snr 10	38.2%	42.3%	41.6%	40.4%	39.4%	36.6%	37.9%	22.1%	1.6%
metrô snr 05	10.1%	12.6%	16.1%	15.1%	11.4%	10.7%	5.7%	4.4%	-8.2%
metrô snr 00	0.3%	-1.3%	-2.8%	-2.8%	-3.2%	-1.9%	-2.8%	-6.6%	-8.5%
metrô snr -5	2.2%	1.6%	0.9%	1.6%	1.9%	1.3%	0.3%	1.3%	-2.8%
trem snr 20	70.7%	70.3%	69.4%	66.6%	61.5%	51.7%	41.3%	30.3%	15.5%
trem snr 15	76%	75.4%	73.2%	71.6%	69.4%	63.7%	52.7%	36.3%	7.6%
trem snr 10	66.9%	65%	63.1%	61.2%	62.5%	57.1%	46.1%	37.5%	12%
trem snr 05	43.8%	43.8%	42%	44.8%	47%	51.1%	45.1%	37.9%	12.9%
trem snr 00	14.8%	15.8%	15.1%	17.7%	19.2%	20.5%	17%	11.7%	3.5%
trem snr -5	-0.6%	-2.5%	-3.5%	-4.7%	-4.1%	-7.6%	-10.1%	-5.7%	-6.6%

**Tabela B.10:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de restaurante, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	64.4%	64%	63.4%	63.4%	63.4%	63.4%	63.4%	63.1%	65%
aeroporto snr 20	77%	76%	77.3%	77%	77%	78.2%	77.9%	77%	76%
aeroporto snr 15	71.6%	70.3%	69.7%	70.3%	68.8%	70.3%	70.3%	69.1%	69.7%
aeroporto snr 10	55.2%	57.1%	58%	58.4%	60.9%	61.8%	60.9%	61.5%	63.7%
aeroporto snr 05	28.4%	29.7%	30.9%	31.2%	31.2%	32.8%	31.5%	33.4%	34.4%
aeroporto snr 00	3.5%	5.4%	5%	6.9%	7.3%	7.9%	8.5%	8.2%	8.8%
aeroporto snr -5	-4.1%	-2.8%	-3.5%	-3.2%	-3.8%	-2.8%	-2.2%	-4.1%	-4.4%
balbulcio snr 20	78.5%	77%	77.3%	77.6%	76.3%	75.1%	74.4%	73.8%	72.2%
balbulcio snr 15	58.7%	58%	56.8%	56.5%	57.4%	57.4%	57.1%	58%	57.7%
balbulcio snr 10	40.1%	43.5%	46.1%	45.7%	45.1%	44.2%	43.2%	42.9%	42.6%
balbulcio snr 05	14.2%	13.6%	12.3%	11%	13.9%	12.6%	12.3%	12.3%	11%
balbulcio snr 00	1.6%	0%	0.3%	0.9%	0.9%	0.3%	0.9%	1.3%	1.3%
balbulcio snr -5	2.2%	2.2%	2.2%	1.9%	1.6%	1.6%	2.2%	2.8%	3.2%
carro snr 20	67.2%	66.2%	65.6%	64.7%	64%	64.7%	62.5%	61.5%	61.5%
carro snr 15	69.7%	69.1%	69.7%	69.7%	68.8%	69.4%	69.7%	69.7%	71.3%
carro snr 10	49.5%	48.6%	48.9%	49.5%	49.5%	49.8%	49.2%	49.5%	50.5%
carro snr 05	19.2%	20.2%	20.8%	19.2%	19.9%	21.1%	22.1%	20.8%	21.1%
carro snr 00	3.2%	3.5%	2.8%	1.9%	2.2%	1.3%	1.6%	0.9%	0.6%
carro snr -5	3.5%	3.5%	4.7%	4.1%	3.8%	3.5%	3.8%	3.8%	4.1%
exposição snr 20	71.6%	71.6%	73.5%	73.2%	73.5%	73.8%	74.1%	72.6%	72.9%
exposição snr 15	60.9%	59.6%	60.9%	61.5%	61.8%	61.2%	61.5%	61.5%	61.8%
exposição snr 10	32.2%	32.2%	33.1%	32.8%	32.8%	33.4%	32.5%	32.2%	32.8%
exposição snr 05	6%	5.7%	6.3%	6.9%	6.6%	5.7%	5%	5%	5%
exposição snr 00	-2.2%	-1.9%	-1.9%	-2.5%	-0.9%	-0.3%	-0.6%	-0.6%	-0.6%
exposição snr -5	1.6%	1.9%	1.9%	2.2%	1.3%	1.3%	1.3%	1.3%	1.3%
restaurante snr 20	69.7%	69.4%	68.8%	68.8%	67.8%	68.8%	68.1%	67.8%	68.1%
restaurante snr 15	65.6%	62.8%	63.4%	63.4%	63.4%	63.1%	63.1%	63.1%	64%
restaurante snr 10	34.4%	35.3%	35%	36%	35.6%	37.2%	36.9%	36%	35%
restaurante snr 05	13.9%	11.4%	11.4%	10.4%	14.2%	15.5%	15.5%	17%	17.4%
restaurante snr 00	-2.5%	-2.2%	-0.9%	-0.3%	-1.6%	-1.9%	-1.9%	-1.9%	-2.2%
restaurante snr -5	-0.3%	0.6%	-0.9%	-1.9%	-2.2%	-2.2%	-1.9%	-2.5%	-2.5%
rua snr 20	65.6%	65%	65.6%	65.3%	66.2%	65.6%	64.7%	65.3%	65.3%
rua snr 15	50.5%	50.2%	51.7%	51.7%	54.3%	54.6%	52.4%	53.3%	52.7%
rua snr 10	43.8%	41%	40.1%	38.2%	39.1%	41.6%	39.7%	39.1%	38.8%
rua snr 05	32.5%	33.8%	33.8%	34.7%	32.8%	30.6%	30.9%	31.5%	31.9%
rua snr 00	8.2%	7.6%	8.5%	9.5%	8.5%	7.6%	7.3%	7.3%	8.2%
rua snr -5	2.5%	2.2%	0.9%	0.9%	1.3%	0.9%	1.6%	1.3%	0.6%
metrô snr 20	71.6%	69.1%	69.7%	70.3%	70%	69.7%	68.8%	71.6%	72.6%
metrô snr 15	59.3%	61.5%	59.9%	59.6%	59.3%	59.3%	61.5%	61.5%	61.5%
metrô snr 10	34.7%	36.6%	36%	35.6%	36.9%	37.9%	38.2%	41%	41%
metrô snr 05	8.5%	8.8%	10.1%	10.7%	10.4%	10.4%	9.5%	10.4%	10.1%
metrô snr 00	-1.3%	-1.6%	-0.9%	-1.6%	-1.9%	-1.6%	-1.3%	-0.9%	0.3%
metrô snr -5	1.3%	1.3%	0.9%	1.3%	1.3%	0.9%	0.9%	1.3%	1.3%
trem snr 20	75.4%	76%	74.8%	73.5%	71.9%	71.6%	71.6%	71.9%	71%
trem snr 15	77.9%	78.5%	79.2%	79.5%	78.9%	78.9%	78.5%	78.5%	77.9%
trem snr 10	65%	65.6%	65.6%	65.6%	65.9%	66.6%	66.6%	66.6%	66.2%
trem snr 05	45.1%	47.3%	47.9%	45.1%	44.8%	45.4%	43.2%	43.2%	43.5%
trem snr 00	13.2%	14.2%	14.8%	15.1%	15.5%	17%	16.4%	15.5%	15.5%
trem snr -5	0%	-0.6%	0%	0%	0%	-0.6%	-0.3%	-0.9%	-0.6%

**Tabela B.11:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de rua, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	65.3%	63.1%	61.5%	57.7%	50.5%	44.5%	32.5%	19.6%	-4.4%
aeroporto snr 20	70.7%	62.8%	58%	51.1%	50.8%	42.6%	33.4%	19.6%	-5.4%
aeroporto snr 15	66.9%	63.1%	61.8%	56.8%	49.5%	42.3%	32.8%	24.6%	-1.9%
aeroporto snr 10	59.6%	58.7%	56.2%	53%	47.3%	35.3%	22.1%	7.3%	-2.5%
aeroporto snr 05	30.3%	31.9%	33.8%	32.2%	26.8%	19.9%	12.3%	5.4%	-11.7%
aeroporto snr 00	6.9%	6.6%	6.6%	4.7%	2.8%	3.2%	-3.5%	-7.9%	-16.1%
aeroporto snr -5	-7.9%	-8.5%	-10.1%	-13.6%	-16.7%	-19.6%	-19.6%	-18.3%	-21.1%
balbulcio snr 20	70%	68.1%	64.4%	54.6%	46.7%	36.6%	30.3%	24.3%	0.9%
balbulcio snr 15	53.6%	48.3%	43.8%	40.4%	37.2%	33.4%	25.6%	16.1%	3.8%
balbulcio snr 10	37.5%	33.4%	29.7%	28.7%	22.4%	20.2%	11.4%	-0.3%	-15.1%
balbulcio snr 05	11%	8.8%	7.3%	2.2%	0%	-2.2%	-9.1%	-11.4%	-19.9%
balbulcio snr 00	0%	-1.6%	-3.8%	-4.7%	-6.9%	-11%	-17%	-21.8%	-24.6%
balbulcio snr -5	2.2%	1.6%	-1.3%	-2.5%	-4.1%	-5%	-8.8%	-13.6%	-16.4%
carro snr 20	63.4%	64.7%	61.8%	57.1%	53.9%	51.7%	44.5%	30%	13.6%
carro snr 15	66.9%	66.2%	62.8%	61.8%	62.5%	59.9%	51.7%	44.8%	22.1%
carro snr 10	52.7%	50.2%	50.5%	46.1%	44.2%	44.8%	43.2%	35%	18%
carro snr 05	20.8%	22.4%	22.7%	20.5%	20.2%	19.6%	16.4%	13.9%	7.9%
carro snr 00	3.5%	3.5%	5.4%	3.8%	2.5%	1.6%	-3.8%	-4.7%	-9.5%
carro snr -5	4.4%	3.5%	3.2%	3.8%	4.4%	2.2%	1.9%	1.6%	-1.3%
exposição snr 20	71.9%	67.5%	67.2%	68.8%	68.1%	62.8%	54.9%	40.7%	14.5%
exposição snr 15	59.3%	59.3%	57.1%	50.2%	47.9%	43.2%	40.1%	29%	9.8%
exposição snr 10	32.5%	31.2%	28.1%	31.9%	30%	26.2%	19.6%	9.8%	2.8%
exposição snr 05	5.7%	6.9%	5%	4.1%	3.2%	1.6%	0.9%	-0.6%	-4.7%
exposição snr 00	0.9%	-0.3%	0.9%	0.9%	0%	-0.9%	-3.5%	-2.2%	-4.4%
exposição snr -5	0.9%	1.3%	0.3%	0.9%	0.3%	-0.9%	-1.9%	-0.9%	-0.6%
restaurante snr 20	66.6%	60.3%	52.7%	48.6%	44.8%	38.8%	28.4%	13.6%	-4.4%
restaurante snr 15	62.8%	61.2%	48.9%	40.7%	36.3%	32.5%	24.6%	10.7%	-4.7%
restaurante snr 10	30.3%	32.5%	27.8%	24.9%	21.1%	15.8%	4.7%	-3.2%	-11.4%
restaurante snr 05	12.6%	10.4%	5.4%	5%	1.6%	0.9%	-5%	-11.4%	-21.8%
restaurante snr 00	-1.3%	-1.9%	-2.8%	-4.7%	-5.4%	-11%	-12%	-19.2%	-31.9%
restaurante snr -5	-1.9%	-6.9%	-8.8%	-11.4%	-12.3%	-17%	-20.2%	-23%	-26.8%
rua snr 20	63.4%	57.7%	55.5%	52.4%	48.6%	42.3%	38.2%	24%	3.5%
rua snr 15	54.3%	52.4%	47%	48.9%	43.8%	36.3%	29.3%	19.9%	7.9%
rua snr 10	38.5%	35.3%	29%	29.3%	24.6%	22.1%	17%	13.2%	0%
rua snr 05	30.9%	27.4%	20.8%	18.3%	18.6%	12%	4.4%	-1.9%	-8.2%
rua snr 00	10.1%	6%	5.4%	2.2%	1.6%	-1.6%	-4.1%	-6%	-15.1%
rua snr -5	2.8%	4.1%	1.9%	0.3%	-2.2%	-2.8%	-5%	-5.4%	-7.6%
metrô snr 20	68.1%	66.6%	63.1%	57.7%	57.7%	46.7%	39.4%	26.5%	0.9%
metrô snr 15	61.8%	58.4%	48.6%	47%	42%	37.5%	38.5%	27.4%	1.3%
metrô snr 10	33.8%	38.2%	37.5%	35%	39.4%	37.5%	26.8%	12%	3.5%
metrô snr 05	12.3%	15.8%	15.5%	15.1%	9.1%	3.8%	2.8%	-2.8%	-5.4%
metrô snr 00	-0.9%	-1.9%	-2.5%	-2.2%	-1.9%	-4.4%	-4.7%	-5.4%	-9.8%
metrô snr -5	1.3%	1.9%	1.3%	0.6%	0.6%	-1.3%	-2.2%	-1.6%	-5%
trem snr 20	72.9%	71.3%	66.6%	65.3%	61.5%	60.3%	48.9%	33.8%	0.3%
trem snr 15	76.3%	72.2%	69.1%	66.6%	67.2%	64%	50.5%	39.4%	16.4%
trem snr 10	68.5%	66.6%	65%	61.5%	59%	56.8%	48.3%	38.8%	11.4%
trem snr 05	46.4%	49.8%	46.7%	46.4%	47.9%	47.6%	45.4%	36.3%	5.7%
trem snr 00	18.9%	18.9%	17%	18.6%	20.5%	22.4%	16.7%	12.3%	1.3%
trem snr -5	-0.9%	-0.9%	-2.2%	-3.2%	-4.1%	-2.8%	-6%	-11%	-14.5%

**Tabela B.12:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de rua, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	63.7%	63.7%	63.1%	63.4%	63.7%	64%	64%	64%	64.4%
aeroporto snr 20	77.6%	75.7%	74.1%	72.9%	73.5%	73.5%	72.2%	72.6%	72.6%
aeroporto snr 15	69.7%	70.3%	69.7%	69.4%	70.3%	68.1%	68.5%	69.1%	66.6%
aeroporto snr 10	55.2%	55.5%	55.8%	57.7%	57.4%	58%	59.3%	59.6%	59.3%
aeroporto snr 05	25.6%	27.8%	26.8%	27.1%	29.3%	29.3%	29.3%	29%	28.4%
aeroporto snr 00	4.4%	4.7%	4.1%	4.4%	5.7%	6%	6.9%	6.9%	7.6%
aeroporto snr -5	-2.2%	-4.4%	-4.4%	-4.7%	-5%	-5.4%	-5.4%	-6.9%	-7.3%
balbulcio snr 20	78.2%	77%	75.1%	75.4%	74.4%	73.8%	73.5%	72.6%	71%
balbulcio snr 15	59.3%	57.7%	57.7%	57.4%	56.5%	56.8%	55.8%	55.5%	55.2%
balbulcio snr 10	40.1%	40.1%	41%	40.7%	40.4%	40.7%	41%	38.2%	36.9%
balbulcio snr 05	13.9%	13.2%	13.9%	12.3%	11.4%	14.2%	14.2%	12.6%	12%
balbulcio snr 00	1.9%	1.6%	1.6%	1.9%	1.6%	0.9%	1.3%	-0.3%	-0.9%
balbulcio snr -5	2.5%	2.5%	2.5%	2.5%	1.9%	2.2%	2.5%	2.5%	1.9%
carro snr 20	65.9%	66.2%	65.9%	64%	64%	62.8%	63.1%	63.4%	63.4%
carro snr 15	68.5%	69.4%	69.7%	69.1%	68.1%	67.2%	68.8%	68.5%	67.8%
carro snr 10	47.6%	49.8%	50.2%	48.9%	50.5%	50.5%	50.8%	51.4%	51.7%
carro snr 05	18.9%	18.3%	18.3%	19.9%	20.2%	19.6%	21.5%	20.8%	21.1%
carro snr 00	2.8%	2.2%	1.9%	0.9%	1.9%	2.2%	4.1%	4.1%	3.5%
carro snr -5	3.5%	4.1%	4.4%	3.8%	4.4%	4.1%	3.8%	4.4%	4.4%
exposição snr 20	71%	73.2%	74.8%	74.8%	72.9%	72.6%	71.6%	71.6%	72.2%
exposição snr 15	62.8%	63.1%	61.8%	62.1%	60.9%	62.1%	62.1%	62.1%	60.3%
exposição snr 10	31.2%	32.5%	31.5%	31.9%	32.5%	31.5%	31.2%	32.5%	32.2%
exposição snr 05	4.1%	5%	4.1%	4.4%	5%	4.7%	3.5%	4.4%	5%
exposição snr 00	-1.9%	-2.2%	-1.3%	-0.6%	-1.9%	-1.6%	-0.3%	0.3%	0.3%
exposição snr -5	0.6%	0.3%	1.3%	0.3%	0.6%	0.6%	0%	0.3%	0.3%
restaurante snr 20	68.1%	69.1%	68.1%	67.8%	67.8%	68.5%	68.1%	67.5%	66.9%
restaurante snr 15	66.2%	66.2%	63.7%	63.7%	63.7%	63.4%	62.8%	62.8%	63.4%
restaurante snr 10	35.3%	33.4%	33.4%	33.1%	33.4%	34.1%	33.8%	34.1%	31.2%
restaurante snr 05	13.2%	11.7%	12%	12.3%	13.9%	12.9%	12%	13.6%	12.6%
restaurante snr 00	-2.8%	-2.8%	-2.5%	-2.8%	-3.2%	-2.2%	-2.5%	-2.5%	-2.2%
restaurante snr -5	-0.3%	0.9%	0.3%	-0.6%	-1.9%	-1.9%	-1.6%	-1.3%	-1.3%
rua snr 20	65.9%	65.9%	65%	65.9%	64.7%	63.1%	63.4%	62.5%	63.1%
rua snr 15	49.8%	50.2%	52.4%	53%	54.3%	54.3%	53%	52.1%	51.7%
rua snr 10	42.9%	40.7%	39.7%	40.1%	40.4%	41%	41.3%	39.7%	41%
rua snr 05	32.2%	31.9%	33.8%	34.7%	35%	32.8%	31.9%	32.5%	32.2%
rua snr 00	7.9%	8.8%	8.8%	8.5%	8.2%	7.9%	10.1%	9.8%	10.1%
rua snr -5	2.5%	2.5%	2.2%	2.5%	1.9%	2.5%	1.9%	2.5%	2.2%
metrô snr 20	70.3%	68.8%	68.5%	69.1%	69.1%	69.7%	68.5%	68.8%	68.5%
metrô snr 15	59.3%	59.9%	60.6%	59.9%	62.1%	61.2%	59.6%	59.9%	61.8%
metrô snr 10	33.4%	33.4%	35.3%	34.4%	33.8%	33.4%	33.1%	32.8%	32.5%
metrô snr 05	8.8%	10.7%	9.8%	10.1%	10.7%	11%	12.3%	13.2%	11.7%
metrô snr 00	-1.3%	-0.6%	-0.9%	-1.3%	-1.3%	-0.6%	-1.6%	-2.2%	-0.9%
metrô snr -5	0.9%	0.6%	0.3%	1.6%	1.3%	1.3%	1.3%	1.3%	1.6%
trem snr 20	74.8%	76%	73.8%	73.2%	73.2%	71.9%	72.6%	72.6%	72.9%
trem snr 15	78.2%	78.5%	77.9%	78.2%	78.2%	78.2%	77.9%	77.9%	77%
trem snr 10	65%	65.6%	65.6%	65.9%	66.6%	66.9%	66.6%	67.2%	67.5%
trem snr 05	45.4%	46.1%	45.7%	47.9%	47.3%	45.4%	46.1%	44.8%	46.4%
trem snr 00	12.9%	13.6%	14.8%	17%	18.6%	19.6%	19.9%	19.6%	18.9%
trem snr -5	0%	0%	-0.3%	-0.3%	0%	0.3%	0%	-0.6%	-0.9%

**Tabela B.13:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de metrô, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	65.9%	65.6%	63.7%	57.4%	47.9%	42.9%	34.7%	2.5%	-10.7%
aeroporto snr 20	67.8%	68.1%	64%	54.3%	45.4%	30.9%	17.7%	-5%	-27.1%
aeroporto snr 15	68.8%	65.9%	59.6%	54.3%	42.6%	27.4%	11.7%	-13.2%	-23.3%
aeroporto snr 10	57.4%	61.8%	53.3%	49.5%	40.4%	22.1%	1.9%	-20.2%	-28.4%
aeroporto snr 05	25.6%	24.9%	24.3%	21.5%	10.1%	-1.3%	-16.4%	-24.6%	-25.6%
aeroporto snr 00	4.4%	6.6%	6.6%	6%	1.3%	-8.2%	-17.7%	-26.8%	-25.9%
aeroporto snr -5	-6%	-8.2%	-12%	-12.9%	-17%	-23.3%	-23.7%	-21.8%	-16.4%
balbulcio snr 20	70.3%	62.5%	57.4%	49.8%	38.2%	31.5%	14.2%	-5.4%	-26.2%
balbulcio snr 15	55.2%	50.8%	42.6%	33.8%	30.9%	27.8%	11%	-6.6%	-22.4%
balbulcio snr 10	36.6%	33.4%	32.8%	27.4%	20.8%	8.5%	-2.8%	-20.8%	-36%
balbulcio snr 05	12.6%	7.9%	4.1%	5.4%	-2.8%	-2.5%	-16.1%	-23.7%	-27.4%
balbulcio snr 00	2.2%	0%	-0.9%	-2.2%	-5.4%	-13.6%	-18.6%	-28.7%	-27.8%
balbulcio snr -5	2.2%	2.2%	0.3%	0%	-4.1%	-7.3%	-14.8%	-23.3%	-29.3%
carro snr 20	69.7%	66.6%	71%	64%	57.1%	48.6%	31.9%	10.1%	-13.6%
carro snr 15	64.7%	63.4%	66.6%	63.4%	58.4%	53.9%	36.3%	2.5%	-12.3%
carro snr 10	46.4%	49.5%	48.9%	45.1%	44.2%	36%	25.9%	-1.9%	-15.5%
carro snr 05	19.9%	19.2%	22.4%	20.8%	25.9%	17.4%	12.9%	-6.6%	-14.5%
carro snr 00	5.7%	5.4%	5.7%	4.7%	2.2%	2.2%	-1.3%	-7.6%	-16.7%
carro snr -5	4.1%	5.7%	5%	6.3%	5.7%	3.8%	0.6%	-5.4%	-7.9%
exposição snr 20	75.1%	75.7%	71.9%	63.4%	61.5%	52.1%	41%	23.7%	0%
exposição snr 15	57.1%	53.9%	49.2%	41.3%	36%	30.6%	25.6%	12%	-6.9%
exposição snr 10	34.4%	27.1%	25.6%	21.1%	23%	17%	10.7%	-7.6%	-18.3%
exposição snr 05	7.6%	8.2%	7.9%	8.5%	6.6%	5.4%	0%	-10.4%	-20.2%
exposição snr 00	0.9%	1.3%	0%	-4.1%	-2.5%	-3.5%	-6.9%	-10.4%	-14.8%
exposição snr -5	1.6%	0.6%	0.6%	1.6%	-1.3%	-3.5%	-5.7%	-11.7%	-17.7%
restaurante snr 20	64.4%	59.3%	53%	48.3%	42.3%	27.8%	14.8%	-9.1%	-30%
restaurante snr 15	64.4%	55.5%	47.6%	40.1%	32.5%	19.9%	9.8%	-5%	-24.3%
restaurante snr 10	33.1%	31.2%	27.1%	22.1%	18%	5.4%	-3.2%	-16.7%	-31.2%
restaurante snr 05	8.2%	5%	1.6%	-2.8%	-5.7%	-6.9%	-16.7%	-21.1%	-28.1%
restaurante snr 00	-2.8%	-2.5%	-5%	-6.9%	-10.4%	-13.6%	-20.2%	-28.4%	-30.3%
restaurante snr -5	-6.6%	-6.9%	-10.1%	-11%	-12.6%	-19.9%	-25.2%	-33.8%	-30.6%
rua snr 20	60.3%	56.2%	54.6%	46.7%	40.4%	26.8%	19.6%	2.2%	-15.8%
rua snr 15	51.7%	51.1%	47.3%	37.5%	32.2%	24.6%	15.5%	-6%	-15.8%
rua snr 10	43.8%	39.4%	36.3%	24.6%	15.1%	10.1%	-10.7%	-18.6%	-20.2%
rua snr 05	26.8%	26.5%	22.7%	17%	13.2%	0.3%	-11.4%	-20.5%	-24.9%
rua snr 00	6.6%	8.5%	6.6%	-2.5%	-8.5%	-10.7%	-16.1%	-11.7%	-7.6%
rua snr -5	5%	5%	4.4%	1.9%	-0.9%	0%	-6%	-15.8%	-14.8%
metrô snr 20	68.1%	67.5%	67.2%	60.9%	53.9%	42.3%	27.4%	11%	-8.2%
metrô snr 15	58.4%	57.4%	56.8%	47.6%	36.6%	31.5%	19.2%	8.5%	-12%
metrô snr 10	40.7%	43.2%	42.3%	36.6%	30%	31.5%	17.7%	-2.5%	-10.4%
metrô snr 05	17.7%	13.9%	14.2%	12.3%	11%	10.7%	3.5%	-3.5%	-12.6%
metrô snr 00	-5.4%	-2.5%	-2.8%	-1.3%	-6.9%	-5.4%	-6.6%	-10.1%	-16.1%
metrô snr -5	0.9%	-0.6%	-1.3%	-0.9%	-2.8%	-1.9%	-4.7%	-10.7%	-17%
trem snr 20	75.1%	72.2%	69.7%	66.2%	60.6%	42%	27.4%	9.1%	-12.3%
trem snr 15	77.3%	74.8%	75.4%	68.1%	63.1%	54.6%	32.2%	13.2%	-10.1%
trem snr 10	66.2%	69.7%	69.7%	68.8%	59.3%	40.7%	19.6%	4.4%	-16.4%
trem snr 05	43.8%	44.2%	39.7%	38.8%	38.2%	32.5%	8.5%	-10.4%	-14.5%
trem snr 00	13.6%	12.3%	14.8%	12.3%	15.1%	10.7%	-2.5%	-13.2%	-11%
trem snr -5	-0.6%	0%	0.3%	0.3%	-1.3%	-11.4%	-16.4%	-7.3%	-2.2%

**Tabela B.14:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de metrô, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	63.4%	64.4%	64%	65.6%	65.6%	65%	65.6%	66.6%	66.9%
aeroporto snr 20	75.7%	75.7%	75.4%	75.7%	74.8%	72.6%	72.6%	70.7%	68.1%
aeroporto snr 15	70.3%	70.3%	69.7%	69.4%	70.7%	71%	69.1%	68.5%	68.8%
aeroporto snr 10	54.6%	53.6%	55.8%	55.8%	57.1%	56.8%	58%	58.7%	58.7%
aeroporto snr 05	24.6%	24.3%	23%	24.3%	25.2%	24.6%	24.6%	24%	25.9%
aeroporto snr 00	5%	3.8%	5%	3.8%	4.4%	5%	4.7%	4.7%	4.4%
aeroporto snr -5	-3.5%	-3.2%	-1.9%	-1.9%	-1.6%	-2.5%	-2.8%	-3.2%	-4.7%
balbulcio snr 20	77.6%	76%	77.3%	77.9%	77%	76%	72.9%	72.2%	71.6%
balbulcio snr 15	59%	57.1%	57.1%	55.8%	55.5%	55.2%	56.5%	56.5%	56.5%
balbulcio snr 10	41.3%	41.6%	41%	40.7%	42%	42.6%	42%	38.8%	36.9%
balbulcio snr 05	12.3%	11.7%	11%	10.4%	9.5%	10.4%	11.4%	11.7%	11%
balbulcio snr 00	2.8%	3.5%	2.2%	0.9%	0.6%	1.6%	1.3%	1.3%	2.8%
balbulcio snr -5	2.2%	1.9%	2.5%	2.5%	2.2%	2.2%	2.8%	1.9%	1.9%
carro snr 20	68.5%	66.9%	66.9%	66.2%	65.9%	67.2%	66.9%	67.5%	68.1%
carro snr 15	67.8%	68.8%	68.8%	67.8%	67.8%	68.5%	65.9%	64.7%	65.3%
carro snr 10	46.7%	45.7%	47.3%	47.9%	47.6%	48.3%	48.3%	47.3%	46.1%
carro snr 05	18.3%	19.9%	20.5%	19.9%	20.5%	21.5%	20.2%	20.5%	19.9%
carro snr 00	2.5%	3.2%	2.5%	3.2%	2.8%	3.8%	4.7%	4.7%	4.4%
carro snr -5	3.5%	4.4%	4.4%	4.7%	4.4%	5%	4.4%	4.7%	4.7%
exposição snr 20	70.3%	73.2%	72.6%	72.2%	72.9%	74.1%	74.8%	73.8%	75.4%
exposição snr 15	63.1%	60.9%	61.2%	62.1%	60.6%	58.7%	58.4%	58.7%	56.5%
exposição snr 10	31.9%	31.9%	30.9%	31.2%	31.5%	30.9%	31.5%	32.2%	33.8%
exposição snr 05	4.1%	1.9%	3.8%	3.2%	5.4%	6.6%	6%	6.6%	6.9%
exposição snr 00	-3.8%	-1.3%	-0.3%	-0.6%	0.3%	0.6%	0.9%	0.9%	0.6%
exposição snr -5	1.3%	0.9%	0.9%	1.6%	1.6%	0.9%	0.6%	1.9%	1.3%
restaurante snr 20	68.8%	66.6%	68.1%	66.6%	66.6%	65.6%	65%	65.6%	65%
restaurante snr 15	66.6%	65.6%	61.2%	62.1%	62.5%	62.1%	65%	65.6%	64.7%
restaurante snr 10	34.7%	33.8%	32.2%	29%	30%	31.5%	30.6%	32.2%	33.4%
restaurante snr 05	13.9%	10.7%	12%	12.9%	14.2%	12.6%	10.7%	9.1%	8.5%
restaurante snr 00	-2.5%	-2.5%	-3.2%	-2.5%	-2.2%	-2.8%	-2.8%	-1.9%	-3.5%
restaurante snr -5	0%	-0.6%	-0.6%	-1.3%	-2.2%	-2.8%	-4.4%	-4.1%	-6%
rua snr 20	66.6%	64%	63.7%	64%	62.5%	62.5%	62.1%	60.6%	61.5%
rua snr 15	50.5%	51.1%	49.2%	50.2%	49.2%	49.5%	51.4%	51.4%	51.7%
rua snr 10	40.1%	40.7%	41.6%	42.6%	42.9%	43.2%	44.5%	44.5%	44.8%
rua snr 05	31.5%	31.9%	30.9%	31.9%	31.2%	30.6%	30.9%	30.6%	28.1%
rua snr 00	8.2%	7.3%	7.3%	7.3%	7.3%	7.3%	6.9%	6.6%	6.6%
rua snr -5	2.8%	3.2%	4.4%	5.7%	5.4%	5.4%	5.4%	5%	4.7%
metrô snr 20	71%	70.3%	70.3%	71.6%	70%	68.5%	69.1%	69.4%	69.4%
metrô snr 15	60.9%	60.9%	58.4%	59.6%	59.3%	58.7%	58.7%	59%	59.3%
metrô snr 10	34.7%	37.9%	38.2%	35.3%	37.5%	36%	37.5%	40.1%	39.4%
metrô snr 05	8.5%	11.4%	10.1%	12.6%	12.6%	13.9%	15.1%	14.5%	16.4%
metrô snr 00	-2.5%	-1.3%	-1.9%	-2.5%	-3.5%	-5%	-5.7%	-5.4%	-4.4%
metrô snr -5	0.9%	1.3%	1.6%	0.6%	0.6%	0.6%	0.9%	1.6%	1.3%
trem snr 20	76.3%	76.7%	76.7%	75.7%	76%	76.3%	75.1%	74.1%	75.1%
trem snr 15	78.2%	78.9%	78.9%	78.5%	78.5%	78.5%	75.7%	76.3%	78.2%
trem snr 10	64.7%	65.6%	66.2%	66.2%	67.2%	66.9%	66.9%	65.6%	65.3%
trem snr 05	46.1%	44.8%	44.8%	44.2%	44.8%	42.9%	43.2%	42.6%	43.2%
trem snr 00	12.3%	13.2%	12.6%	13.2%	13.9%	14.2%	13.9%	14.2%	14.2%
trem snr -5	0.6%	-0.3%	-0.6%	-1.3%	-1.6%	-0.3%	0%	-0.9%	-0.9%

**Tabela B.15:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de trem, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	65.3%	58.7%	57.1%	52.4%	49.2%	40.4%	30.9%	7.3%	-9.8%
aeroporto snr 20	68.8%	63.4%	46.1%	43.5%	37.5%	37.2%	29.3%	12.9%	-4.7%
aeroporto snr 15	69.7%	62.8%	54.3%	46.7%	42.9%	37.2%	25.9%	13.6%	-6.3%
aeroporto snr 10	60.3%	59%	54.6%	46.1%	41.6%	37.2%	22.1%	5.4%	-6.3%
aeroporto snr 05	33.1%	31.5%	32.5%	32.5%	27.4%	18.6%	15.8%	6.9%	-12.6%
aeroporto snr 00	7.6%	4.7%	7.9%	6.6%	4.1%	-1.3%	-4.1%	-9.1%	-16.7%
aeroporto snr -5	-7.9%	-9.1%	-13.2%	-14.5%	-18.9%	-20.5%	-20.2%	-18%	-18.3%
balbulcio snr 20	70.3%	65%	60.6%	53.6%	50.5%	38.8%	24%	10.4%	-3.8%
balbulcio snr 15	55.8%	46.7%	43.8%	32.2%	29.3%	23%	16.7%	3.8%	-8.2%
balbulcio snr 10	36.9%	33.1%	27.1%	26.2%	18.3%	14.2%	5%	0.3%	-14.8%
balbulcio snr 05	9.5%	7.3%	2.5%	1.6%	0.9%	-1.3%	-1.3%	-8.5%	-17.4%
balbulcio snr 00	0.3%	-1.6%	-4.1%	-6.9%	-10.4%	-12%	-17.7%	-18.3%	-18%
balbulcio snr -5	1.9%	0.3%	-2.8%	-2.5%	-5.4%	-10.7%	-11.7%	-13.6%	-15.1%
carro snr 20	60.9%	60.9%	60.9%	58.7%	55.8%	46.4%	40.4%	27.8%	5%
carro snr 15	66.6%	65.9%	66.6%	64%	57.4%	52.7%	46.7%	37.5%	14.8%
carro snr 10	50.8%	51.1%	46.4%	50.2%	49.5%	42.6%	39.7%	33.1%	11.7%
carro snr 05	23%	23%	21.5%	22.4%	27.1%	23.7%	17.4%	12%	4.4%
carro snr 00	1.9%	0.6%	1.6%	0.3%	0.6%	0%	-3.8%	-6.3%	-10.7%
carro snr -5	4.4%	2.2%	3.5%	3.8%	2.5%	2.2%	0.6%	0.6%	-6.9%
exposição snr 20	72.9%	72.6%	72.2%	72.2%	65.3%	60.6%	51.4%	37.5%	17.4%
exposição snr 15	61.5%	60.3%	53.3%	51.4%	56.5%	47%	41.3%	31.5%	15.8%
exposição snr 10	31.2%	30.9%	28.7%	28.1%	29.7%	26.2%	23.7%	17.4%	-1.6%
exposição snr 05	3.5%	7.9%	2.2%	5.4%	2.5%	-1.3%	-1.6%	-4.1%	-1.6%
exposição snr 00	-1.6%	-1.6%	-3.5%	-1.6%	-1.6%	-4.1%	-4.4%	-4.1%	-5%
exposição snr -5	1.9%	0.3%	-0.3%	-0.9%	-1.9%	-0.9%	-0.9%	-1.9%	-2.2%
restaurante snr 20	64%	52.1%	50.2%	42.9%	38.2%	31.9%	20.2%	6.6%	-6%
restaurante snr 15	64.7%	60.9%	53.9%	34.7%	32.5%	21.8%	14.8%	9.5%	-11.7%
restaurante snr 10	36.9%	37.2%	25.9%	21.5%	20.8%	12.3%	8.8%	2.5%	-14.5%
restaurante snr 05	14.2%	9.1%	8.2%	7.3%	5.4%	0%	-3.5%	-7.3%	-16.1%
restaurante snr 00	-1.6%	-2.5%	-3.8%	-5%	-7.6%	-9.8%	-11.4%	-18.6%	-22.4%
restaurante snr -5	-1.9%	-6.3%	-8.5%	-9.5%	-13.2%	-18%	-21.8%	-24.6%	-25.6%
rua snr 20	63.7%	63.1%	63.1%	58%	55.5%	47.9%	38.2%	34.7%	4.7%
rua snr 15	50.5%	51.7%	40.7%	35%	37.5%	35%	27.1%	21.1%	4.1%
rua snr 10	42%	38.5%	29.7%	28.1%	22.7%	16.7%	7.9%	2.5%	-6.3%
rua snr 05	35.3%	33.1%	27.4%	18.3%	17.4%	11%	6.6%	-2.2%	-11.4%
rua snr 00	8.8%	2.5%	0.6%	-0.9%	0%	-7.3%	-8.2%	-9.5%	-15.1%
rua snr -5	1.3%	0.3%	0%	-1.3%	-1.6%	-2.2%	-7.3%	-7.6%	-9.1%
metrô snr 20	67.8%	64.7%	69.4%	63.1%	53.3%	51.1%	38.8%	23.3%	7.3%
metrô snr 15	60.3%	57.7%	57.7%	57.7%	53.9%	43.8%	36%	23.7%	4.4%
metrô snr 10	36.6%	34.7%	36.9%	33.4%	27.1%	21.8%	23%	10.1%	-0.6%
metrô snr 05	9.5%	11%	10.4%	7.9%	-1.6%	-2.8%	-6%	-3.2%	-10.1%
metrô snr 00	-1.9%	-1.3%	-2.8%	-2.2%	-2.2%	-5%	-7.9%	-7.9%	-4.1%
metrô snr -5	2.2%	1.6%	1.6%	0.9%	0%	0.6%	1.6%	0%	-0.6%
trem snr 20	70.3%	67.5%	64.4%	57.1%	50.5%	44.8%	37.5%	18.6%	1.9%
trem snr 15	78.5%	69.7%	67.2%	67.5%	64%	57.1%	47.6%	32.2%	10.4%
trem snr 10	67.5%	69.1%	63.1%	62.1%	61.8%	53%	46.1%	38.8%	13.6%
trem snr 05	45.7%	47.6%	51.4%	51.4%	55.2%	54.6%	42.3%	24%	10.1%
trem snr 00	16.4%	16.1%	18%	19.2%	19.6%	21.1%	18.3%	10.4%	4.7%
trem snr -5	-2.5%	-2.2%	-5%	-3.5%	-6.3%	-6.9%	-9.8%	-13.2%	-13.6%

**Tabela B.16:** Taxa de acertos de palavras de um sistema treinado com todos os ruídos com SNR = 20 e 15 dB adaptado utilizando modelos fonéticos corrompidos com ruído de trem, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	64.4%	63.7%	63.1%	62.8%	62.1%	62.8%	64.7%	64.7%	65.9%
aeroporto snr 20	77.6%	76.7%	77.9%	77.6%	76%	72.2%	72.9%	72.2%	69.7%
aeroporto snr 15	69.7%	69.7%	70.3%	70%	70%	70.3%	70%	70.3%	69.1%
aeroporto snr 10	55.5%	55.8%	59%	57.7%	59.3%	59.6%	59.9%	61.2%	61.2%
aeroporto snr 05	26.2%	27.4%	28.7%	28.7%	28.4%	29.3%	30.6%	32.2%	33.1%
aeroporto snr 00	4.4%	3.8%	5%	5.7%	6.3%	6.3%	6%	6%	7.3%
aeroporto snr -5	-4.7%	-4.1%	-2.8%	-2.8%	-4.7%	-3.5%	-4.7%	-5.4%	-7.3%
balbulcio snr 20	78.5%	76.7%	76.3%	74.8%	74.8%	73.5%	72.2%	69.4%	69.4%
balbulcio snr 15	59.3%	59.3%	56.5%	56.2%	57.7%	58.4%	59%	56.8%	56.8%
balbulcio snr 10	40.4%	41%	40.7%	39.4%	39.4%	37.9%	39.7%	39.1%	38.2%
balbulcio snr 05	14.5%	13.6%	11.7%	13.2%	11.4%	11.4%	12.3%	10.7%	9.8%
balbulcio snr 00	1.6%	1.3%	0.3%	-0.3%	0%	0.6%	0.3%	-0.6%	-1.3%
balbulcio snr -5	2.2%	1.9%	1.9%	2.2%	2.5%	2.5%	2.2%	1.9%	1.6%
carro snr 20	65.6%	65.6%	65.6%	64.7%	64.4%	65.6%	65.3%	63.1%	62.1%
carro snr 15	69.4%	68.8%	70.3%	70.3%	69.4%	70%	67.8%	67.5%	67.8%
carro snr 10	48.3%	47.6%	47.9%	50.5%	49.8%	50.8%	50.8%	51.1%	51.1%
carro snr 05	18.9%	19.2%	20.2%	19.9%	20.5%	20.8%	21.5%	21.8%	21.5%
carro snr 00	2.8%	2.5%	1.6%	0.3%	2.8%	2.2%	1.6%	0.6%	0.6%
carro snr -5	2.8%	4.1%	4.4%	3.8%	3.8%	3.8%	4.1%	4.1%	4.1%
exposição snr 20	71.3%	73.5%	74.1%	73.8%	73.8%	75.4%	74.4%	73.5%	72.2%
exposição snr 15	62.1%	62.1%	62.5%	62.5%	63.1%	63.4%	63.1%	61.2%	62.1%
exposição snr 10	32.5%	30.9%	33.1%	33.8%	34.4%	32.2%	32.5%	32.8%	32.5%
exposição snr 05	6.3%	5.7%	5%	5.4%	5%	5%	4.7%	5.4%	4.7%
exposição snr 00	-1.9%	-1.9%	-2.5%	-1.9%	-1.3%	-0.6%	-0.6%	0%	-0.9%
exposição snr -5	1.3%	1.6%	1.6%	1.6%	1.6%	1.3%	1.3%	0.9%	0.9%
restaurante snr 20	69.7%	69.1%	69.4%	68.1%	68.1%	67.2%	65%	65.6%	63.1%
restaurante snr 15	65.9%	63.1%	62.1%	62.5%	62.5%	66.2%	65.3%	65.3%	64.7%
restaurante snr 10	35.3%	34.4%	35%	33.4%	33.8%	33.1%	33.8%	36.6%	36.9%
restaurante snr 05	13.9%	11%	11%	12%	12.9%	13.6%	13.2%	12.6%	13.6%
restaurante snr 00	-2.8%	-3.2%	-1.3%	-1.9%	-2.5%	-3.2%	-2.5%	-2.5%	-1.3%
restaurante snr -5	0%	-0.3%	0%	-0.3%	-0.9%	-1.3%	-1.6%	-1.3%	-1.6%
rua snr 20	66.2%	64.4%	64.4%	64.4%	65%	65.9%	64%	65.3%	65%
rua snr 15	49.8%	49.5%	52.1%	52.1%	52.4%	52.7%	51.1%	50.5%	50.8%
rua snr 10	42.9%	40.1%	40.4%	40.4%	39.1%	40.1%	40.4%	40.4%	39.4%
rua snr 05	34.1%	34.4%	35%	34.4%	34.4%	35%	36%	35.6%	34.7%
rua snr 00	9.1%	8.8%	8.5%	9.5%	8.5%	7.6%	8.2%	7.9%	8.2%
rua snr -5	2.5%	2.2%	1.9%	1.9%	2.2%	2.2%	2.8%	1.9%	1.3%
metrô snr 20	72.2%	70.7%	71.3%	67.8%	67.2%	68.1%	68.1%	71.3%	71.3%
metrô snr 15	59.3%	61.2%	61.5%	62.8%	62.1%	60.6%	57.1%	57.7%	60.9%
metrô snr 10	35.6%	33.4%	35.6%	36.3%	33.4%	33.8%	35%	35.6%	36%
metrô snr 05	8.2%	5.7%	8.2%	7.6%	7.9%	9.1%	9.1%	9.1%	10.1%
metrô snr 00	-0.9%	-1.6%	-0.9%	-0.6%	-1.6%	-0.9%	-0.6%	-0.9%	-0.9%
metrô snr -5	0.9%	0.9%	1.6%	0.9%	1.3%	1.3%	1.6%	1.9%	2.2%
trem snr 20	75.1%	74.8%	72.6%	71.9%	71.6%	71.3%	71.3%	71%	70.7%
trem snr 15	78.2%	78.5%	78.2%	77.3%	77.6%	77.9%	78.5%	78.5%	77.9%
trem snr 10	65.3%	65.6%	65.6%	66.2%	66.9%	66.9%	66.6%	67.5%	67.8%
trem snr 05	45.1%	47.6%	46.7%	47.9%	47.9%	48.3%	45.1%	46.1%	46.4%
trem snr 00	13.2%	14.2%	16.1%	18%	19.2%	18.6%	18.6%	18.9%	18.3%
trem snr -5	0.3%	-0.3%	-0.3%	-0.9%	-1.6%	-1.3%	-1.6%	-1.6%	-2.2%



## Anexo C

Resultados para sistema treinado  
com locuções limpas adaptado  
utilizando multi-estilo

**Tabela C.1:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	75.9%	73.4%	74.7%	75.3%	68.4%	61.7%	55.4%	36.7%	3.2%
aeroporto snr 20	51.3%	50.9%	52.2%	50.9%	51.9%	43.7%	36.7%	32.9%	12.3%
aeroporto snr 15	44%	45.6%	43%	41.1%	40.5%	41.1%	31.3%	17.1%	13.3%
aeroporto snr 10	24.1%	28.5%	26.9%	28.2%	29.1%	31.3%	22.5%	17.1%	3.8%
aeroporto snr 05	5.7%	12.3%	12.7%	15.2%	11.1%	8.5%	8.2%	1.3%	-7%
aeroporto snr 00	1.3%	4.4%	4.1%	5.1%	7.3%	3.5%	0.6%	-1.3%	-5.7%
aeroporto snr -5	-15.5%	-17.7%	-15.8%	-16.8%	-13.9%	-16.1%	-13.9%	-14.2%	-12%
balbulcio snr 20	57.6%	57.6%	58.9%	54.4%	49.4%	45.3%	40.8%	28.8%	7.3%
balbulcio snr 15	38.3%	37%	39.2%	36.4%	30.1%	31%	29.4%	19.9%	4.4%
balbulcio snr 10	22.5%	19.6%	22.8%	19%	17.4%	16.8%	14.9%	5.7%	-2.2%
balbulcio snr 05	-3.5%	-0.3%	0.6%	2.8%	5.7%	4.4%	-1.9%	-4.4%	-15.8%
balbulcio snr 00	-7.9%	-7.6%	-7.9%	-10.4%	-12%	-11.4%	-14.6%	-15.2%	-19%
balbulcio snr -5	-4.1%	-3.5%	-4.1%	-3.2%	-5.7%	-9.2%	-11.4%	-15.2%	-12.3%
carro snr 20	57.6%	57.9%	60.8%	59.2%	55.7%	53.8%	48.7%	37.3%	13.6%
carro snr 15	45.6%	48.4%	50.9%	49.1%	50%	46.2%	42.4%	33.2%	14.9%
carro snr 10	28.2%	32%	34.5%	33.2%	32.3%	31.3%	33.5%	25.3%	10.4%
carro snr 05	6.3%	7%	8.2%	9.5%	9.5%	9.5%	10.1%	3.8%	-0.9%
carro snr 00	1.3%	2.8%	4.4%	5.7%	3.8%	3.5%	2.2%	-2.8%	-6%
carro snr -5	6%	6.3%	6%	5.4%	4.1%	5.4%	4.7%	4.7%	3.2%
exposição snr 20	26.9%	30.4%	29.4%	32.3%	33.5%	26.6%	22.2%	18.4%	2.2%
exposição snr 15	12%	9.5%	10.8%	9.2%	9.2%	14.2%	12.7%	9.2%	-4.4%
exposição snr 10	0.6%	-4.1%	2.5%	-2.8%	1.9%	-0.6%	0.9%	-3.8%	-7%
exposição snr 05	-8.5%	-4.1%	-6.3%	-3.8%	-4.4%	-8.2%	-8.5%	-13.3%	-14.6%
exposição snr 00	-10.4%	-7.9%	-7.6%	-5.4%	-2.8%	-7.6%	-7.3%	-12.3%	-12%
exposição snr -5	-7.9%	-5.7%	-6%	-7.9%	-6%	-4.7%	-3.5%	-3.8%	-4.7%
restaurante snr 20	39.2%	38.3%	38.6%	38.9%	37%	37%	30.4%	19.6%	2.2%
restaurante snr 15	19.9%	23.7%	27.2%	27.8%	23.4%	19%	17.1%	10.4%	-1.9%
restaurante snr 10	13%	13.3%	13.6%	14.6%	9.8%	8.2%	7.9%	2.2%	-9.2%
restaurante snr 05	-4.7%	-1.6%	-2.2%	-3.2%	-1.3%	-5.1%	-10.4%	-11.7%	-22.8%
restaurante snr 00	-8.2%	-12%	-14.6%	-15.5%	-14.9%	-17.4%	-13%	-11.4%	-18.7%
restaurante snr -5	-6.6%	-7.3%	-10.4%	-16.8%	-17.1%	-20.6%	-20.9%	-26.6%	-26.6%
rua snr 20	41.5%	38.6%	37%	38.3%	37.7%	37.7%	32.3%	25%	2.8%
rua snr 15	30.7%	27.8%	26.6%	25.3%	29.1%	24.1%	19.9%	19%	0%
rua snr 10	21.5%	23.4%	25%	25%	23.7%	19.6%	13.3%	8.9%	-3.8%
rua snr 05	8.9%	13.6%	13%	13%	11.4%	12%	10.4%	9.8%	1.6%
rua snr 00	1.3%	0%	-1.3%	-0.9%	-3.2%	-2.2%	-3.8%	-7.6%	-6.6%
rua snr -5	3.2%	1.6%	0.9%	1.3%	1.3%	0.3%	1.6%	-0.9%	-1.6%
metrô snr 20	33.5%	33.5%	38.3%	36.1%	34.8%	29.7%	22.8%	21.2%	-6.3%
metrô snr 15	24.4%	23.7%	25%	25.9%	22.8%	22.5%	19%	10.8%	-4.7%
metrô snr 10	1.3%	1.9%	4.1%	5.1%	6.3%	7.9%	1.9%	0.9%	-11.4%
metrô snr 05	-5.1%	-3.5%	-5.1%	-2.8%	-3.5%	-2.5%	-6.3%	-7.9%	-14.2%
metrô snr 00	-0.3%	-2.5%	-4.1%	-4.1%	-5.7%	-6%	-5.7%	-7.9%	-3.8%
metrô snr -5	0.9%	0.9%	0.3%	1.3%	0.6%	0.9%	0.3%	-0.9%	1.6%
trem snr 20	65.8%	69.3%	66.8%	67.7%	65.2%	59.8%	50.6%	49.1%	28.5%
trem snr 15	57.9%	64.2%	63%	62.7%	58.9%	57.9%	57%	47.8%	24.4%
trem snr 10	42.4%	41.8%	39.6%	37.7%	37.7%	35.1%	32.6%	28.2%	18.4%
trem snr 05	22.5%	26.3%	27.8%	25%	26.6%	23.4%	22.5%	20.9%	8.5%
trem snr 00	6%	5.4%	5.7%	7.3%	5.4%	7%	7.3%	3.5%	0.6%
trem snr -5	6%	4.7%	3.8%	4.1%	2.5%	1.3%	0%	1.3%	-4.4%

**Tabela C.2:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de aeroporto, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	77.9%	77.9%	77.9%	77.9%	78.9%	78.9%	78.9%	78.2%	76.3%
aeroporto snr 20	43.8%	44.8%	46.4%	47%	47.9%	48.6%	49.8%	50.2%	51.4%
aeroporto snr 15	29.3%	32.2%	31.5%	35%	34.7%	38.2%	38.8%	41%	42.3%
aeroporto snr 10	16.1%	17%	17%	18.3%	21.5%	20.2%	22.4%	22.4%	24%
aeroporto snr 05	-3.5%	-1.9%	0%	1.3%	5.4%	6.9%	6.3%	5.4%	7.3%
aeroporto snr 00	-5.4%	-3.2%	-2.2%	-0.6%	-0.3%	0.3%	0.6%	0.9%	2.5%
aeroporto snr -5	-12.3%	-14.2%	-13.2%	-14.5%	-14.8%	-14.8%	-15.5%	-15.8%	-14.5%
balbulcio snr 20	52.4%	51.1%	51.7%	52.4%	53.3%	53.3%	54.6%	55.2%	57.4%
balbulcio snr 15	34.1%	34.4%	34.4%	39.4%	37.9%	36%	37.5%	38.5%	38.5%
balbulcio snr 10	12.6%	15.1%	17%	18.3%	18%	21.1%	21.8%	22.1%	23.7%
balbulcio snr 05	-4.7%	-5%	-4.4%	-4.4%	-4.4%	-2.8%	-3.8%	-3.2%	-3.2%
balbulcio snr 00	-8.8%	-7.6%	-8.2%	-8.2%	-9.8%	-10.1%	-10.4%	-9.8%	-8.5%
balbulcio snr -5	-1.3%	-0.9%	-1.3%	-2.8%	-2.5%	-1.9%	-1.9%	-1.9%	-1.9%
carro snr 20	56.2%	56.8%	56.8%	55.5%	54.6%	56.2%	55.8%	56.8%	56.5%
carro snr 15	41.6%	42.3%	41.6%	41.6%	44.8%	45.4%	46.7%	47%	46.1%
carro snr 10	17.7%	20.5%	23%	24.3%	24%	24.6%	23%	24%	27.1%
carro snr 05	2.5%	1.9%	5.4%	6%	5.7%	5.7%	5.4%	6%	5.7%
carro snr 00	2.5%	1.9%	1.3%	1.9%	2.5%	2.8%	2.5%	1.9%	1.6%
carro snr -5	4.4%	4.4%	4.7%	4.4%	5.4%	5.7%	5.7%	5.4%	6%
exposição snr 20	27.8%	31.9%	32.2%	32.2%	30.6%	28.7%	28.4%	27.8%	27.1%
exposição snr 15	7.6%	7.9%	6%	6.6%	8.2%	9.5%	9.1%	10.1%	12%
exposição snr 10	1.6%	4.1%	2.5%	3.8%	0.6%	-0.9%	1.3%	0.3%	1.9%
exposição snr 05	-7.3%	-9.1%	-9.8%	-9.8%	-11.7%	-11.4%	-10.7%	-9.8%	-7.6%
exposição snr 00	-6.9%	-9.1%	-9.8%	-10.7%	-10.1%	-11%	-10.7%	-10.4%	-10.4%
exposição snr -5	-5.4%	-6%	-7.3%	-6.9%	-5.4%	-5%	-5.7%	-7.9%	-6.3%
restaurante snr 20	40.1%	40.1%	40.7%	39.1%	38.2%	38.2%	38.5%	38.2%	39.1%
restaurante snr 15	17.4%	16.7%	14.8%	16.1%	16.1%	16.1%	19.9%	19.6%	21.5%
restaurante snr 10	3.8%	6.6%	8.5%	6.6%	8.8%	10.4%	10.4%	10.4%	10.7%
restaurante snr 05	-5%	-5%	-4.1%	-5.7%	-6.3%	-6.3%	-6.6%	-6.3%	-5.4%
restaurante snr 00	-7.9%	-8.2%	-8.5%	-9.8%	-10.7%	-9.8%	-7.9%	-8.5%	-8.2%
restaurante snr -5	-1.9%	-2.5%	-3.5%	-2.8%	-3.5%	-3.8%	-4.4%	-5.7%	-7.3%
rua snr 20	40.7%	42%	45.7%	46.4%	47.6%	49.8%	49.2%	45.7%	44.2%
rua snr 15	28.4%	27.4%	27.1%	29%	29.7%	30.9%	30%	30.9%	32.2%
rua snr 10	20.5%	19.6%	20.8%	21.1%	21.8%	24.6%	24.3%	23.7%	23.3%
rua snr 05	7.9%	8.2%	7.3%	8.8%	9.5%	8.5%	8.5%	9.5%	8.8%
rua snr 00	-1.6%	-1.3%	-0.3%	0.9%	2.5%	1.6%	0.3%	1.9%	1.9%
rua snr -5	1.3%	1.3%	2.8%	2.5%	2.5%	2.2%	2.5%	2.8%	2.8%
metrô snr 20	29%	30.6%	30.3%	31.5%	30.9%	34.7%	36%	34.4%	34.4%
metrô snr 15	22.4%	20.8%	20.8%	21.1%	20.2%	20.5%	21.1%	21.1%	24.3%
metrô snr 10	0%	0%	-0.6%	-0.3%	-0.6%	0%	-0.6%	0.9%	0.9%
metrô snr 05	-5%	-5.4%	-5%	-5.4%	-5%	-5%	-6.3%	-6.3%	-5.7%
metrô snr 00	-4.1%	-4.1%	-3.8%	-3.2%	-2.5%	-2.2%	-1.6%	-1.6%	-0.6%
metrô snr -5	0.3%	0.6%	0%	0.9%	1.6%	1.6%	1.3%	1.3%	1.6%
trem snr 20	63.7%	65%	65%	65.9%	65.9%	66.2%	65.3%	64.7%	66.6%
trem snr 15	53%	53.3%	53.3%	54.6%	56.5%	56.8%	56.8%	57.1%	58%
trem snr 10	39.7%	41.6%	43.2%	44.8%	44.8%	44.5%	44.8%	42.3%	42.6%
trem snr 05	17%	16.1%	16.4%	16.1%	17.7%	18.9%	19.2%	20.2%	21.5%
trem snr 00	6.9%	7.3%	7.6%	6.9%	6.6%	6.6%	6.3%	5.7%	6%
trem snr -5	5.7%	5.7%	6%	6.3%	6%	5.7%	5.7%	6%	6.3%

**Tabela C.3:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de balbulcio, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	39.2%	39.2%	38%	35.4%	31.3%	30.1%	26.3%	18.7%	-2.8%
aeroporto snr 20	29.1%	33.2%	30.1%	27.5%	18.7%	17.1%	7.9%	0.3%	-11.7%
aeroporto snr 15	25.3%	25.6%	25.3%	24.7%	20.9%	14.2%	9.8%	-1.3%	-4.1%
aeroporto snr 10	11.4%	9.2%	8.5%	13.6%	14.2%	6%	-4.4%	-9.5%	-14.2%
aeroporto snr 05	2.2%	-0.6%	2.2%	2.8%	-1.3%	-3.8%	-7.6%	-14.2%	-15.2%
aeroporto snr 00	-0.6%	-4.7%	-8.2%	-7%	-11.7%	-12.3%	-14.9%	-14.9%	-16.8%
aeroporto snr -5	0.6%	-4.1%	-3.8%	-6.6%	-10.1%	-13.9%	-12.7%	-18%	-14.6%
balbulcio snr 20	26.6%	25.3%	25.3%	25.3%	24.1%	16.5%	14.9%	6.6%	-3.2%
balbulcio snr 15	20.6%	21.8%	20.9%	21.8%	18.4%	13.3%	8.5%	2.8%	-8.2%
balbulcio snr 10	7%	7.9%	7.9%	4.7%	1.6%	-3.5%	-7.6%	-10.8%	-13%
balbulcio snr 05	-10.4%	-7.9%	-12.3%	-14.9%	-19.9%	-18.4%	-17.4%	-20.9%	-20.6%
balbulcio snr 00	-4.4%	-8.9%	-14.9%	-16.8%	-18%	-20.3%	-22.8%	-23.4%	-23.7%
balbulcio snr -5	0%	-0.3%	-6%	-10.8%	-16.8%	-16.5%	-19.3%	-20.6%	-18.7%
carro snr 20	38.6%	35.4%	32.6%	28.8%	28.2%	23.1%	19.3%	9.8%	-7.3%
carro snr 15	33.2%	36.1%	38.6%	38.3%	37%	32.9%	19.3%	6.6%	-2.5%
carro snr 10	23.7%	23.1%	24.4%	20.6%	22.5%	15.8%	13.3%	7.3%	-7.6%
carro snr 05	0.6%	3.8%	4.1%	2.2%	-0.9%	2.2%	0.6%	-4.1%	-14.6%
carro snr 00	-1.3%	-0.9%	-0.9%	-0.6%	-2.8%	-2.2%	-3.5%	-7.6%	-9.2%
carro snr -5	0.3%	0.3%	0.9%	0%	0.9%	1.3%	-1.3%	-1.6%	-3.2%
exposição snr 20	13.6%	22.8%	21.2%	19%	18%	15.8%	8.5%	3.5%	-2.5%
exposição snr 15	13.6%	14.9%	13.6%	13.6%	14.2%	10.8%	7.3%	-4.4%	-2.8%
exposição snr 10	-2.2%	-8.5%	-5.1%	-4.4%	-4.1%	-7%	-7.3%	-8.2%	-11.1%
exposição snr 05	-10.4%	-7%	-11.1%	-12%	-11.4%	-13.6%	-11.1%	-14.2%	-8.9%
exposição snr 00	-9.2%	-7.6%	-8.5%	-3.2%	-5.4%	-4.7%	-4.4%	-4.4%	-4.4%
exposição snr -5	-5.4%	-9.2%	-5.7%	-3.8%	-4.7%	-3.5%	-1.9%	-1.6%	-2.2%
restaurante snr 20	19.9%	22.8%	24.1%	21.8%	13.6%	11.4%	4.7%	-4.7%	-8.9%
restaurante snr 15	11.7%	10.1%	10.1%	9.2%	5.1%	1.6%	-1.6%	-10.8%	-13%
restaurante snr 10	-0.3%	4.4%	3.8%	-0.9%	-3.2%	-2.8%	-10.4%	-9.8%	-11.7%
restaurante snr 05	-11.1%	-11.4%	-10.4%	-10.4%	-12.3%	-19.3%	-19.6%	-20.3%	-19%
restaurante snr 00	-5.4%	-7.9%	-10.4%	-14.2%	-18.4%	-22.5%	-20.6%	-18%	-16.8%
restaurante snr -5	-2.5%	-6.6%	-9.8%	-15.8%	-21.8%	-23.4%	-25%	-24.1%	-22.8%
rua snr 20	29.7%	21.8%	24.4%	20.6%	17.7%	19.3%	14.9%	4.4%	-2.5%
rua snr 15	12%	12%	14.2%	14.9%	8.5%	6%	1.3%	-7%	-9.8%
rua snr 10	7.3%	7.6%	2.5%	-0.6%	4.7%	-2.8%	-5.1%	-10.1%	-16.5%
rua snr 05	-0.6%	2.2%	0.6%	0.9%	2.8%	4.7%	1.3%	-3.5%	-7.3%
rua snr 00	-3.5%	-3.2%	-8.9%	-17.4%	-19.3%	-21.2%	-20.9%	-20.6%	-20.6%
rua snr -5	0.6%	0.6%	1.3%	0%	-1.9%	-1.6%	-0.9%	-3.2%	-3.5%
metrô snr 20	21.2%	20.9%	15.5%	19.9%	16.8%	16.8%	14.9%	2.5%	-7.9%
metrô snr 15	17.1%	15.8%	17.1%	17.1%	10.8%	7%	7%	2.2%	-6.3%
metrô snr 10	-6.3%	-3.2%	-0.3%	1.3%	-0.6%	-3.8%	-4.4%	-7%	-7.9%
metrô snr 05	-9.5%	-13%	-13.3%	-15.2%	-16.5%	-16.1%	-9.2%	-12%	-11.1%
metrô snr 00	-2.8%	-7%	-8.5%	-4.7%	-6%	-3.5%	-4.1%	-6%	-3.8%
metrô snr -5	0.3%	0.3%	0.3%	0.3%	0.6%	0.3%	-0.3%	-1.6%	0.3%
trem snr 20	45.9%	44.3%	44.9%	43%	38.3%	35.8%	31.3%	18%	7.6%
trem snr 15	39.2%	37.3%	39.2%	42.1%	37.3%	33.5%	29.4%	19.3%	5.7%
trem snr 10	25.9%	27.2%	31%	29.1%	27.8%	25.3%	19%	12%	0.9%
trem snr 05	14.6%	15.8%	16.1%	20.9%	19%	19.9%	18.4%	16.5%	6.6%
trem snr 00	1.3%	2.2%	3.5%	6%	4.7%	5.1%	4.4%	2.5%	0%
trem snr -5	1.6%	1.3%	0%	0.6%	0%	-2.8%	-1.9%	-7.6%	-8.5%

**Tabela C.4:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de balbulcio, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	38.5%	39.4%	40.7%	39.4%	39.7%	39.4%	39.7%	39.4%	38.8%
aeroporto snr 20	20.8%	21.8%	22.4%	24%	23.7%	24%	25.6%	29%	28.7%
aeroporto snr 15	16.1%	18%	19.2%	20.2%	21.5%	21.5%	23.7%	24%	23.7%
aeroporto snr 10	5%	6.6%	8.2%	7.9%	8.2%	7.3%	7.3%	9.1%	11%
aeroporto snr 05	-0.3%	0%	0.6%	0%	-0.6%	-0.3%	0.3%	0.6%	2.8%
aeroporto snr 00	-0.3%	-0.9%	-0.3%	0.9%	0%	0.3%	0%	-0.3%	-0.3%
aeroporto snr -5	2.5%	1.9%	1.3%	-0.3%	-0.3%	0.6%	0.9%	0%	0.3%
balbulcio snr 20	25.2%	24.6%	26.8%	29.3%	28.7%	29%	30%	29.3%	26.5%
balbulcio snr 15	18%	17.4%	18.9%	20.8%	21.1%	22.1%	21.1%	22.1%	21.8%
balbulcio snr 10	2.5%	3.5%	5%	6%	6.6%	6.3%	5.7%	6.3%	6.3%
balbulcio snr 05	-8.5%	-8.8%	-8.2%	-8.5%	-10.1%	-11.4%	-10.4%	-9.5%	-8.8%
balbulcio snr 00	-1.6%	-2.2%	-2.5%	-3.5%	-3.2%	-3.8%	-3.5%	-4.4%	-3.5%
balbulcio snr -5	0%	0%	0.3%	0.3%	0.9%	0.3%	0.6%	0.6%	0.3%
carro snr 20	37.2%	34.4%	38.2%	39.1%	39.1%	39.7%	38.2%	39.1%	38.8%
carro snr 15	32.2%	32.8%	34.1%	35.6%	35.6%	33.8%	33.1%	32.2%	33.4%
carro snr 10	17.4%	18.9%	19.6%	19.9%	21.8%	22.7%	23.3%	23.3%	23.7%
carro snr 05	1.6%	1.9%	1.6%	2.5%	1.9%	2.8%	2.8%	2.5%	0.9%
carro snr 00	1.9%	1.6%	1.3%	0.3%	0%	-0.3%	-0.3%	0%	-1.9%
carro snr -5	1.3%	1.3%	1.3%	1.6%	0.9%	0.6%	0.3%	0.6%	0.6%
exposição snr 20	20.5%	19.6%	20.8%	21.5%	19.9%	17.7%	18.3%	17.4%	15.8%
exposição snr 15	7.9%	10.4%	10.4%	10.4%	10.4%	12.9%	12.3%	12.9%	10.4%
exposição snr 10	-3.5%	-2.8%	-3.5%	-2.8%	-3.2%	-2.8%	-0.9%	0.3%	-3.5%
exposição snr 05	-7.3%	-6.9%	-8.8%	-9.1%	-8.8%	-8.2%	-9.8%	-9.8%	-9.5%
exposição snr 00	-5%	-4.4%	-5.4%	-6.6%	-7.6%	-9.1%	-8.2%	-8.5%	-10.1%
exposição snr -5	-3.8%	-3.5%	-3.2%	-3.5%	-4.1%	-4.1%	-5.4%	-5.4%	-6%
restaurante snr 20	16.7%	16.7%	18%	17.7%	18.3%	18.6%	18.9%	19.6%	19.6%
restaurante snr 15	5.7%	5%	6.6%	6.9%	7.6%	9.5%	11.4%	12%	12%
restaurante snr 10	-4.1%	-3.2%	-2.5%	-3.2%	-0.6%	-1.3%	-0.6%	0.9%	2.2%
restaurante snr 05	-6.9%	-6.9%	-8.2%	-9.5%	-7.6%	-10.1%	-10.4%	-8.2%	-7.6%
restaurante snr 00	-1.3%	-3.2%	-1.9%	-1.6%	-2.8%	-3.8%	-5%	-5.7%	-5%
restaurante snr -5	-0.6%	-0.9%	-0.9%	-0.9%	-0.6%	-0.9%	-1.3%	-1.6%	-2.2%
rua snr 20	25.9%	28.1%	28.7%	27.1%	27.4%	29.3%	30.6%	30%	29.3%
rua snr 15	7.9%	8.8%	8.2%	8.2%	9.5%	8.2%	7.6%	10.7%	11.4%
rua snr 10	4.7%	5.4%	5.4%	7.6%	9.8%	7.6%	7.9%	6.9%	6.9%
rua snr 05	0.3%	-0.6%	0%	0.9%	1.6%	1.6%	0.6%	0.6%	0.9%
rua snr 00	-2.2%	-3.2%	-3.2%	-2.8%	-2.2%	-3.2%	-2.5%	-3.8%	-4.7%
rua snr -5	0.6%	0.6%	0.6%	0.6%	0.6%	0.6%	0.9%	0.9%	1.3%
metrô snr 20	19.9%	19.6%	20.5%	21.8%	21.8%	24.6%	24.6%	22.4%	23.7%
metrô snr 15	13.6%	16.1%	15.8%	16.1%	15.8%	18.3%	18.3%	18.3%	18.3%
metrô snr 10	0.3%	-3.8%	-3.8%	-2.8%	-2.5%	-2.2%	-4.7%	-2.8%	-4.4%
metrô snr 05	-7.6%	-6.3%	-7.6%	-9.8%	-8.5%	-9.1%	-8.2%	-7.3%	-6.6%
metrô snr 00	-8.8%	-7.6%	-5.7%	-5%	-6.6%	-6.9%	-4.4%	-4.7%	-4.7%
metrô snr -5	0.9%	-0.3%	0.6%	0.3%	0.9%	-0.3%	-2.2%	-1.6%	-0.3%
trem snr 20	42.3%	42.3%	42.6%	43.8%	44.8%	43.8%	44.2%	45.1%	45.7%
trem snr 15	35.6%	34.4%	35%	34.7%	34.4%	34.7%	36%	35.3%	37.9%
trem snr 10	24.3%	24%	24.6%	25.9%	25.2%	25.6%	27.1%	25.9%	25.9%
trem snr 05	11%	11%	12.3%	12.3%	12.3%	11.7%	11.7%	12%	13.9%
trem snr 00	-0.3%	0.6%	0.9%	1.6%	1.6%	1.6%	0.6%	0.6%	0.6%
trem snr -5	1.9%	2.5%	2.2%	1.9%	1.3%	1.6%	1.6%	1.3%	1.6%

**Tabela C.5:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de carro, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	39.6%	37.3%	38.9%	32.9%	25.6%	18.4%	6.6%	0.3%	-2.2%
aeroporto snr 20	28.5%	29.1%	23.7%	22.2%	11.7%	8.2%	4.7%	-13.6%	-16.5%
aeroporto snr 15	23.4%	23.7%	19.3%	18.7%	11.1%	2.5%	-0.6%	-10.8%	-12.3%
aeroporto snr 10	6.3%	4.7%	8.5%	6.6%	0.6%	-6.3%	-10.1%	-19.9%	-15.8%
aeroporto snr 05	-1.6%	1.6%	0.9%	-3.5%	-5.7%	-10.8%	-17.7%	-23.7%	-18.4%
aeroporto snr 00	-2.5%	-7.6%	-9.2%	-13.6%	-11.1%	-14.6%	-16.1%	-16.8%	-20.9%
aeroporto snr -5	0.6%	-2.2%	-7.9%	-12.3%	-10.1%	-12.7%	-13%	-10.4%	-12%
balbulcio snr 20	24.7%	22.2%	21.2%	16.8%	13.6%	7%	-7%	-10.4%	-15.8%
balbulcio snr 15	16.1%	19.3%	16.8%	10.1%	3.2%	-1.9%	-6%	-13.9%	-16.1%
balbulcio snr 10	5.1%	10.4%	5.4%	4.1%	-5.4%	-9.5%	-12.7%	-13.3%	-16.5%
balbulcio snr 05	-11.7%	-12.7%	-17.7%	-22.8%	-25.9%	-27.2%	-30.1%	-18%	-13.3%
balbulcio snr 00	-5.7%	-13%	-20.6%	-24.4%	-20.6%	-25.3%	-28.2%	-24.1%	-10.4%
balbulcio snr -5	0.9%	-0.9%	-8.9%	-13.9%	-21.8%	-22.2%	-27.5%	-25.9%	-12%
carro snr 20	37.7%	36.4%	35.1%	33.2%	31.3%	24.4%	17.4%	4.1%	-7%
carro snr 15	34.8%	39.9%	40.2%	35.4%	32.6%	22.8%	5.1%	-1.9%	-6%
carro snr 10	24.7%	25%	25%	22.8%	20.3%	17.1%	7.9%	2.2%	-4.4%
carro snr 05	1.6%	1.3%	3.5%	4.1%	0.6%	0.3%	-4.7%	-6%	-6.3%
carro snr 00	-0.6%	-2.8%	-4.1%	-3.5%	-5.1%	-6%	-8.5%	-10.4%	-9.8%
carro snr -5	0%	1.3%	0.6%	-0.3%	-1.9%	-4.1%	-4.1%	-7%	-6.6%
exposição snr 20	20.9%	20.6%	20.9%	20.3%	16.5%	9.8%	2.8%	-7%	-12.3%
exposição snr 15	12%	13.3%	16.1%	11.1%	9.8%	2.5%	-10.4%	-13%	-9.2%
exposição snr 10	-1.6%	-5.7%	-5.7%	-7.6%	-6.3%	-12%	-18.7%	-14.2%	-15.5%
exposição snr 05	-10.8%	-11.7%	-12%	-10.4%	-17.4%	-20.9%	-25.9%	-20.9%	-20.6%
exposição snr 00	-7.9%	-9.8%	-12%	-8.9%	-7.9%	-14.6%	-17.4%	-14.6%	-11.7%
exposição snr -5	-4.7%	-10.4%	-12.3%	-12.7%	-12%	-20.6%	-12.7%	-9.5%	-7%
restaurante snr 20	19%	21.5%	23.4%	15.2%	11.4%	-2.5%	-10.4%	-17.4%	-15.8%
restaurante snr 15	8.9%	9.2%	12.3%	8.2%	5.1%	-3.2%	-6.6%	-16.5%	-13.3%
restaurante snr 10	3.5%	1.9%	3.2%	-0.9%	-7%	-14.9%	-14.6%	-13.6%	-12%
restaurante snr 05	-9.8%	-13.3%	-16.8%	-12.7%	-16.5%	-13%	-22.8%	-18%	-16.5%
restaurante snr 00	-7.6%	-12%	-15.2%	-13.3%	-17.1%	-19.6%	-20.6%	-20.9%	-12%
restaurante snr -5	-3.2%	-4.4%	-13.3%	-17.1%	-29.7%	-26.3%	-29.4%	-25.9%	-13.3%
rua snr 20	28.5%	23.7%	23.1%	17.1%	13.9%	3.8%	-2.5%	-13.9%	-16.8%
rua snr 15	11.4%	10.8%	6.3%	0.9%	1.9%	-11.1%	-6.6%	-7%	-9.2%
rua snr 10	6.3%	7.6%	-1.3%	-5.1%	-9.2%	-12.3%	-15.2%	-16.8%	-12.3%
rua snr 05	2.5%	1.3%	0.9%	-4.4%	-6.3%	-7.6%	-9.8%	-13%	-7.9%
rua snr 00	-4.1%	-12.3%	-16.5%	-19.6%	-24.7%	-23.7%	-17.4%	-8.9%	-6%
rua snr -5	0.9%	-2.2%	-4.4%	-6%	-10.8%	-14.6%	-9.8%	-12.7%	-9.2%
metrô snr 20	19%	19.6%	22.5%	20.3%	15.2%	11.4%	-1.3%	-11.1%	-13%
metrô snr 15	16.5%	15.5%	12.3%	12.3%	-7%	-3.5%	-4.1%	-11.7%	-7.9%
metrô snr 10	-1.9%	-1.9%	-0.9%	-4.4%	-8.9%	-10.1%	-10.4%	-13%	-9.5%
metrô snr 05	-9.2%	-11.7%	-18%	-13.3%	-11.7%	-15.8%	-16.8%	-13.3%	-11.1%
metrô snr 00	-10.1%	-11.7%	-12%	-13%	-17.1%	-16.8%	-16.5%	-12%	-5.4%
metrô snr -5	-3.2%	-6.3%	-4.7%	-9.5%	-15.5%	-15.5%	-19.3%	-9.5%	-1.6%
trem snr 20	44.3%	44.9%	40.8%	39.6%	35.8%	30.4%	16.5%	3.2%	-6%
trem snr 15	36.7%	38.9%	38.9%	34.5%	33.9%	24.4%	10.8%	4.4%	-5.1%
trem snr 10	26.6%	31%	28.2%	28.2%	28.2%	22.8%	14.9%	5.4%	-6.6%
trem snr 05	15.8%	17.1%	18%	20.3%	17.1%	16.5%	11.7%	6.3%	0%
trem snr 00	0.9%	1.3%	2.5%	5.4%	7.6%	7.6%	4.4%	0.9%	-3.2%
trem snr -5	1.9%	-0.9%	-3.5%	-5.4%	-7.9%	-5.4%	-8.5%	-11.7%	-18.4%

**Tabela C.6:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de carro, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	37.5%	39.1%	37.5%	39.1%	38.5%	37.2%	37.2%	37.9%	39.7%
aeroporto snr 20	21.5%	21.8%	23.3%	23.7%	24.9%	25.2%	25.6%	26.5%	27.4%
aeroporto snr 15	16.4%	17.4%	16.7%	17.4%	18.9%	20.5%	20.2%	21.8%	21.8%
aeroporto snr 10	6.3%	4.1%	3.2%	5.7%	5.4%	5.4%	7.9%	8.5%	6.9%
aeroporto snr 05	-0.3%	0.3%	-0.3%	-0.9%	-0.9%	-2.5%	-1.3%	-0.3%	-0.9%
aeroporto snr 00	-0.6%	-0.6%	-1.3%	0.3%	0.3%	0%	-0.3%	-0.9%	-0.9%
aeroporto snr -5	2.5%	0.9%	0.9%	0.6%	0%	-0.9%	0%	-0.3%	0%
balbulcio snr 20	25.2%	24.9%	26.5%	27.4%	27.1%	26.8%	25.9%	25.2%	26.5%
balbulcio snr 15	18.3%	18.3%	19.2%	21.1%	22.1%	20.8%	20.8%	18.3%	18.3%
balbulcio snr 10	4.4%	4.1%	5.7%	7.3%	6.3%	6.6%	5%	4.7%	5.4%
balbulcio snr 05	-8.5%	-7.9%	-8.5%	-11.7%	-11%	-11.4%	-11%	-10.4%	-9.8%
balbulcio snr 00	-0.9%	-1.9%	-2.5%	-1.9%	-2.2%	-2.8%	-3.2%	-4.4%	-4.4%
balbulcio snr -5	0%	0%	-0.3%	0.3%	0.9%	0.9%	0.6%	0.6%	0.3%
carro snr 20	36.9%	38.5%	38.8%	41.3%	41.3%	40.4%	40.1%	38.8%	37.9%
carro snr 15	32.2%	33.8%	32.8%	33.4%	32.8%	34.1%	33.1%	34.7%	35%
carro snr 10	18%	18.6%	19.9%	20.5%	21.8%	23.7%	24.3%	23.7%	24.3%
carro snr 05	0%	2.5%	2.8%	1.9%	2.2%	1.3%	1.9%	0.9%	1.6%
carro snr 00	1.9%	0.9%	0%	-0.6%	-0.6%	0.9%	-0.9%	-0.6%	-0.3%
carro snr -5	1.3%	0.9%	1.6%	1.3%	1.3%	1.3%	0.6%	0.3%	0.3%
exposição snr 20	20.2%	21.1%	19.9%	18.9%	18%	20.5%	19.6%	21.5%	21.5%
exposição snr 15	9.1%	8.8%	12.9%	12.6%	12.9%	12.3%	11.7%	12.6%	12.3%
exposição snr 10	-2.5%	-2.5%	-4.1%	-2.8%	-1.3%	-1.3%	-4.1%	-1.9%	-3.2%
exposição snr 05	-7.6%	-8.5%	-6.9%	-9.1%	-8.8%	-11%	-8.5%	-11%	-10.7%
exposição snr 00	-4.4%	-4.7%	-5.4%	-7.3%	-5.4%	-6.3%	-6.9%	-7.6%	-10.7%
exposição snr -5	-2.2%	-3.8%	-3.8%	-2.8%	-2.2%	-4.1%	-5.4%	-4.1%	-4.1%
restaurante snr 20	16.4%	16.7%	17.4%	17.4%	16.7%	16.1%	19.6%	18.9%	19.2%
restaurante snr 15	6%	6.3%	6%	7.3%	6.6%	7.3%	6.6%	8.5%	9.5%
restaurante snr 10	-3.8%	-5%	-3.5%	-2.8%	-1.9%	-2.2%	0%	0.9%	1.3%
restaurante snr 05	-7.3%	-8.2%	-7.6%	-9.1%	-10.7%	-10.1%	-10.4%	-9.1%	-9.8%
restaurante snr 00	-1.3%	-1.9%	-0.6%	-2.2%	-3.5%	-3.8%	-5%	-5%	-5.4%
restaurante snr -5	-0.6%	-0.9%	-0.9%	-0.6%	-2.2%	-2.5%	-1.9%	-1.9%	-2.5%
rua snr 20	25.2%	28.7%	27.1%	29%	27.1%	29.7%	30.9%	28.4%	28.4%
rua snr 15	9.1%	10.1%	9.8%	12%	13.6%	13.2%	12.9%	12.9%	12.3%
rua snr 10	3.5%	4.7%	6.3%	6.9%	8.2%	6.9%	5.7%	4.7%	5.4%
rua snr 05	1.3%	1.9%	2.5%	3.5%	2.5%	3.8%	3.8%	3.5%	3.2%
rua snr 00	-2.5%	-3.2%	-3.2%	-2.2%	-2.2%	-3.2%	-4.4%	-1.9%	-2.8%
rua snr -5	0.9%	0.6%	0.9%	0.6%	0.6%	0.3%	0.9%	0.9%	1.3%
metrô snr 20	19.6%	20.5%	21.5%	21.5%	21.8%	21.5%	21.1%	20.8%	20.2%
metrô snr 15	12.6%	15.1%	16.7%	16.1%	15.8%	15.1%	15.5%	14.5%	14.5%
metrô snr 10	-0.3%	-2.5%	-2.5%	-3.5%	-2.2%	-5.7%	-4.1%	-2.8%	-2.8%
metrô snr 05	-8.5%	-6.6%	-7.9%	-6.6%	-6.6%	-8.5%	-10.1%	-6%	-7.9%
metrô snr 00	-8.8%	-8.5%	-7.3%	-7.9%	-7.6%	-8.2%	-7.9%	-9.5%	-11%
metrô snr -5	0%	0%	-0.6%	0.3%	-0.6%	-0.6%	-1.3%	-2.2%	-1.9%
trem snr 20	42%	42%	41.3%	41.6%	42.6%	43.2%	44.2%	44.5%	44.8%
trem snr 15	35.3%	35.3%	34.7%	35.6%	35%	35.6%	36.9%	37.5%	37.9%
trem snr 10	24.6%	23.7%	24.3%	24.9%	26.8%	28.1%	27.4%	26.8%	26.2%
trem snr 05	12%	11.7%	11.4%	11.4%	12.3%	12.9%	13.2%	15.1%	15.5%
trem snr 00	0%	1.6%	0.9%	1.6%	1.6%	1.3%	0.6%	0%	0%
trem snr -5	1.9%	1.9%	1.6%	0.6%	0.9%	1.3%	1.3%	1.6%	1.9%

**Tabela C.7:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de exposição, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	78.8%	78.8%	71.5%	66.5%	60.1%	44.9%	34.2%	13.9%	-9.5%
aeroporto snr 20	50%	44.6%	48.1%	45.9%	35.4%	32.6%	21.2%	-1.9%	-14.2%
aeroporto snr 15	38%	38.9%	38.3%	32.6%	27.2%	19.6%	11.4%	-11.7%	-15.2%
aeroporto snr 10	24.4%	22.5%	26.9%	21.5%	13.6%	7.3%	-2.8%	-7.6%	-24.1%
aeroporto snr 05	6%	11.7%	7.9%	5.7%	0%	-2.2%	-8.5%	-17.7%	-14.2%
aeroporto snr 00	-1.3%	0.6%	2.2%	-0.6%	-0.6%	-7.9%	-13%	-23.7%	-14.9%
aeroporto snr -5	-13.9%	-9.5%	-13.3%	-13%	-16.5%	-19.3%	-21.5%	-22.5%	-4.4%
balbulcio snr 20	54.4%	52.8%	50.3%	42.7%	38.6%	31%	19.3%	-1.3%	-18.4%
balbulcio snr 15	38%	34.2%	37.7%	28.2%	28.8%	13.9%	0.6%	-10.4%	-21.8%
balbulcio snr 10	23.7%	15.8%	15.2%	15.8%	7.3%	4.1%	-8.2%	-17.7%	-23.4%
balbulcio snr 05	-2.2%	1.3%	5.1%	0.6%	-2.2%	-8.2%	-11.4%	-18.4%	-19.9%
balbulcio snr 00	-7.3%	-9.8%	-7.9%	-7.3%	-10.4%	-14.6%	-19.9%	-21.2%	-19.9%
balbulcio snr -5	-3.2%	-2.8%	-2.8%	-2.8%	-4.7%	-7%	-14.6%	-25.3%	-20.6%
carro snr 20	55.7%	59.2%	52.8%	49.1%	46.8%	42.1%	33.9%	9.8%	-13%
carro snr 15	44.9%	42.7%	40.2%	41.5%	41.8%	28.8%	19.9%	6%	-15.5%
carro snr 10	24.4%	25.3%	25.6%	26.6%	26.6%	17.1%	7.3%	-0.6%	-13.6%
carro snr 05	6%	5.4%	6.3%	6%	7.3%	6.3%	2.5%	-3.2%	-7%
carro snr 00	0.3%	2.2%	3.8%	3.8%	4.4%	3.8%	-1.3%	-3.2%	-11.7%
carro snr -5	4.7%	5.1%	6%	7.3%	4.7%	3.2%	2.2%	-0.9%	-3.5%
exposição snr 20	35.4%	34.5%	35.1%	34.2%	33.2%	26.9%	15.5%	5.1%	-10.8%
exposição snr 15	13.9%	16.1%	16.1%	14.9%	18.4%	15.8%	6.6%	-7.3%	-14.2%
exposição snr 10	3.2%	0.3%	1.9%	2.2%	-1.3%	-2.2%	-4.7%	-9.8%	-14.9%
exposição snr 05	-6.3%	-1.9%	-2.2%	-3.8%	-7.6%	-8.2%	-10.1%	-9.8%	-15.8%
exposição snr 00	-8.9%	-7.6%	-6.3%	-9.8%	-9.2%	-9.2%	-12.3%	-10.8%	-16.5%
exposição snr -5	-5.7%	-6.3%	-7.6%	-8.5%	-9.8%	-11.4%	-13.6%	-13.9%	-20.6%
restaurante snr 20	36.7%	38.6%	34.2%	33.2%	29.7%	22.2%	10.1%	-10.8%	-25.9%
restaurante snr 15	22.2%	24.7%	21.8%	18.4%	17.1%	11.1%	-0.3%	-14.2%	-26.9%
restaurante snr 10	12.3%	7%	7%	6.3%	1.9%	0.6%	-7.6%	-22.2%	-24.4%
restaurante snr 05	-3.5%	-3.8%	-3.5%	-5.7%	-9.5%	-9.5%	-19%	-29.1%	-29.4%
restaurante snr 00	-12%	-10.4%	-7.9%	-8.5%	-11.4%	-15.5%	-24.1%	-29.1%	-25.6%
restaurante snr -5	-11.4%	-17.7%	-15.5%	-19%	-24.4%	-27.8%	-33.5%	-39.6%	-21.5%
rua snr 20	40.5%	41.5%	41.8%	37.3%	37.3%	39.2%	20.3%	9.2%	-5.4%
rua snr 15	32.3%	29.1%	26.9%	25.9%	23.7%	19.3%	14.6%	1.3%	-12.3%
rua snr 10	21.5%	19.6%	17.4%	16.8%	10.1%	6%	-1.3%	-11.4%	-15.2%
rua snr 05	6.6%	8.9%	9.8%	4.1%	5.4%	-0.3%	-7.9%	-11.4%	-10.8%
rua snr 00	-0.6%	-0.6%	-1.6%	-4.7%	-4.4%	-8.5%	-14.2%	-18.7%	-10.8%
rua snr -5	0%	0.3%	1.9%	3.5%	2.5%	1.3%	1.3%	-4.7%	-9.8%
metrô snr 20	38.6%	39.9%	40.5%	39.2%	34.8%	31.3%	22.2%	3.2%	-10.1%
metrô snr 15	28.5%	29.4%	35.1%	31.3%	27.8%	21.5%	12.3%	0.9%	-13.3%
metrô snr 10	6%	6%	5.1%	6.6%	5.7%	3.2%	-2.8%	-11.4%	-17.4%
metrô snr 05	-6%	-5.7%	-5.1%	-2.5%	-6%	-5.7%	-4.7%	-7.3%	-8.2%
metrô snr 00	-4.1%	-8.9%	-7.6%	-5.7%	-6%	-3.2%	-5.4%	-10.4%	-4.7%
metrô snr -5	0.6%	3.2%	1.6%	1.3%	0%	-0.9%	-4.7%	-9.2%	-6.3%
trem snr 20	63.9%	58.5%	57.6%	58.2%	57.9%	47.8%	32%	8.5%	-14.9%
trem snr 15	56.3%	56.6%	58.2%	52.8%	48.7%	40.5%	33.2%	11.7%	-15.8%
trem snr 10	44%	39.2%	36.7%	33.9%	28.8%	23.1%	14.2%	-1.9%	-10.4%
trem snr 05	20.6%	21.8%	18%	16.8%	15.5%	9.5%	1.9%	-4.4%	-6.6%
trem snr 00	7.6%	6.6%	4.7%	5.1%	2.5%	-0.3%	-3.5%	-9.5%	-4.4%
trem snr -5	5.1%	5.1%	4.1%	1.9%	2.5%	-1.9%	-1.9%	-4.1%	-2.2%



**Tabela C.8:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de exposição, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	78.9%	79.2%	80.1%	81.1%	81.1%	80.8%	80.4%	80.4%	79.5%
aeroporto snr 20	43.2%	45.4%	44.8%	43.5%	43.5%	45.4%	48.3%	49.2%	50.5%
aeroporto snr 15	30%	32.5%	35.3%	37.2%	38.8%	38.5%	39.1%	40.4%	39.1%
aeroporto snr 10	15.8%	18.9%	19.2%	19.9%	22.1%	21.1%	24%	25.9%	25.2%
aeroporto snr 05	-1.9%	-0.9%	1.9%	4.1%	2.5%	4.1%	5%	5.4%	6.6%
aeroporto snr 00	-4.4%	-0.9%	-2.8%	-2.2%	-0.9%	-0.3%	-2.5%	-1.6%	-1.9%
aeroporto snr -5	-14.8%	-14.8%	-14.8%	-13.6%	-15.1%	-15.5%	-15.5%	-14.5%	-13.6%
balbulcio snr 20	51.1%	51.1%	53.3%	54.3%	54.9%	55.5%	56.5%	54.9%	54.9%
balbulcio snr 15	31.2%	31.2%	33.4%	32.8%	34.1%	35.3%	37.2%	38.2%	38.5%
balbulcio snr 10	14.5%	17%	18%	18.9%	21.8%	22.7%	24.3%	24%	23.3%
balbulcio snr 05	-3.5%	-3.5%	-2.8%	-2.8%	-4.4%	-1.9%	0.3%	0.9%	-1.9%
balbulcio snr 00	-7.9%	-9.5%	-9.8%	-7.6%	-7.9%	-7.6%	-7.6%	-6.9%	-7.6%
balbulcio snr -5	-0.9%	-0.6%	-1.3%	-1.6%	-2.5%	-1.6%	-1.3%	-1.6%	-2.5%
carro snr 20	56.5%	55.2%	56.2%	54.9%	55.2%	55.8%	55.5%	55.8%	55.8%
carro snr 15	41%	42.9%	41.6%	42%	42.3%	42.6%	43.2%	43.2%	43.5%
carro snr 10	19.9%	22.1%	25.6%	27.1%	27.1%	25.2%	24.9%	24.3%	23.7%
carro snr 05	4.7%	5.4%	5.4%	5%	6.3%	5.4%	6.3%	6.6%	6%
carro snr 00	1.6%	2.5%	2.5%	3.2%	2.2%	2.8%	2.8%	2.2%	1.3%
carro snr -5	4.4%	4.4%	5%	4.1%	2.8%	3.5%	3.8%	4.1%	4.7%
exposição snr 20	31.2%	31.9%	34.1%	34.4%	34.7%	35.3%	36.9%	35%	34.7%
exposição snr 15	9.5%	7.9%	12%	13.2%	11.7%	12.6%	13.2%	14.8%	16.1%
exposição snr 10	2.8%	1.6%	2.5%	3.2%	1.3%	1.6%	1.6%	3.8%	3.2%
exposição snr 05	-7.3%	-7.3%	-9.1%	-7.9%	-6.3%	-5%	-5%	-4.4%	-6%
exposição snr 00	-8.8%	-7.3%	-5.7%	-5.7%	-6.3%	-8.5%	-6.6%	-7.6%	-9.1%
exposição snr -5	-6.3%	-7.9%	-8.8%	-9.1%	-8.5%	-6.9%	-6.6%	-8.2%	-6.3%
restaurante snr 20	39.7%	39.1%	38.2%	38.5%	37.9%	38.5%	38.8%	38.2%	38.2%
restaurante snr 15	18.9%	18%	18%	19.6%	20.8%	19.2%	19.9%	21.8%	21.5%
restaurante snr 10	5%	7.6%	8.8%	10.4%	11.4%	11.4%	12.6%	12%	11.7%
restaurante snr 05	-6%	-6.9%	-7.6%	-4.7%	-5.7%	-5.4%	-5%	-5%	-2.2%
restaurante snr 00	-9.1%	-12%	-11.4%	-11%	-10.1%	-9.1%	-9.5%	-11%	-11.7%
restaurante snr -5	-2.5%	-3.8%	-5.4%	-6%	-8.5%	-8.2%	-7.9%	-8.8%	-9.8%
rua snr 20	42%	42.9%	43.5%	46.4%	45.7%	46.7%	46.1%	46.4%	44.5%
rua snr 15	27.1%	27.1%	28.4%	29%	29.3%	29%	30%	31.5%	30.9%
rua snr 10	19.9%	22.1%	23%	21.1%	23%	22.7%	22.7%	24%	23%
rua snr 05	7.3%	6%	5.4%	3.5%	5%	4.7%	6.3%	6.9%	9.1%
rua snr 00	-0.9%	-1.3%	0%	-2.2%	-0.9%	-0.9%	-0.6%	-1.3%	-0.9%
rua snr -5	1.9%	1.6%	1.6%	1.3%	1.9%	1.6%	0.9%	0.3%	0.6%
metrô snr 20	31.5%	32.8%	34.7%	36.6%	36.9%	38.5%	38.5%	39.1%	38.5%
metrô snr 15	24.3%	22.1%	21.8%	21.1%	22.4%	22.7%	24.3%	26.8%	27.8%
metrô snr 10	0.6%	3.8%	3.5%	2.5%	1.3%	5.4%	3.8%	4.4%	3.8%
metrô snr 05	-6.6%	-6.3%	-4.4%	-6%	-4.4%	-6.3%	-6.3%	-6.6%	-6.3%
metrô snr 00	-4.7%	-4.1%	-3.5%	-2.8%	-3.2%	-2.5%	-4.7%	-3.2%	-3.5%
metrô snr -5	1.3%	0.9%	0.3%	0.6%	1.6%	1.3%	1.6%	0.9%	0.3%
trem snr 20	63.4%	63.4%	63.4%	65%	65.3%	64.7%	62.5%	63.4%	63.4%
trem snr 15	53.6%	54.3%	55.5%	54.9%	55.2%	55.5%	58%	57.1%	57.1%
trem snr 10	41.3%	44.2%	43.2%	44.5%	45.1%	44.5%	42.6%	43.2%	43.8%
trem snr 05	17.4%	16.7%	16.4%	18.9%	20.8%	20.8%	21.1%	21.1%	19.6%
trem snr 00	7.3%	6.6%	6.6%	6%	5.4%	6%	6.6%	6.9%	7.3%
trem snr -5	6%	5.7%	5.4%	5.7%	5.7%	5.4%	6.3%	5%	4.7%

**Tabela C.9:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de restaurante, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	78.2%	77.8%	75%	71.8%	68.4%	64.6%	58.2%	40.8%	17.1%
aeroporto snr 20	49.4%	50%	52.8%	53.8%	45.6%	44%	35.4%	30.7%	9.2%
aeroporto snr 15	40.5%	39.9%	37%	37%	33.9%	29.7%	24.1%	16.1%	-0.6%
aeroporto snr 10	25.6%	29.4%	32.3%	32.9%	29.1%	26.3%	19%	11.4%	-6%
aeroporto snr 05	5.1%	9.8%	13%	13%	11.7%	8.2%	6.6%	0.6%	-9.2%
aeroporto snr 00	-1.3%	4.1%	5.1%	3.5%	2.5%	0.9%	-0.9%	-4.7%	-10.8%
aeroporto snr -5	-15.8%	-13%	-15.5%	-15.8%	-14.6%	-18%	-11.4%	-16.1%	-13.6%
balbulcio snr 20	54.7%	55.4%	55.4%	53.8%	53.2%	51.3%	48.1%	38.9%	6.3%
balbulcio snr 15	36.1%	37.7%	37.3%	38.9%	34.5%	34.2%	27.8%	21.2%	2.8%
balbulcio snr 10	22.5%	19.9%	17.1%	15.2%	13.3%	15.2%	16.1%	7.3%	-6.6%
balbulcio snr 05	1.6%	3.2%	2.5%	5.4%	6%	0.9%	4.4%	-2.5%	-9.2%
balbulcio snr 00	-6.3%	-7.3%	-9.5%	-7.6%	-8.2%	-9.5%	-9.8%	-12%	-15.2%
balbulcio snr -5	-4.1%	-4.4%	-6%	-5.7%	-6.6%	-5.7%	-7%	-7.9%	-12%
carro snr 20	55.4%	56.3%	62.7%	62%	60.1%	53.8%	53.8%	39.9%	20.9%
carro snr 15	45.9%	48.1%	41.5%	44%	43.7%	40.2%	41.1%	31%	7.6%
carro snr 10	27.8%	32%	29.1%	29.7%	27.8%	26.3%	24.4%	20.6%	0.6%
carro snr 05	6.6%	7.3%	8.2%	8.9%	7.9%	9.2%	5.4%	1.6%	-1.6%
carro snr 00	1.6%	1.3%	3.8%	2.8%	1.9%	-0.9%	0%	-0.3%	-4.7%
carro snr -5	6.6%	6.3%	4.7%	4.4%	3.5%	3.5%	3.5%	2.5%	0.6%
exposição snr 20	31.3%	32%	32%	31%	29.4%	26.3%	22.5%	21.2%	3.5%
exposição snr 15	9.5%	8.5%	12.3%	10.1%	9.2%	6.6%	6.3%	10.8%	-9.5%
exposição snr 10	4.7%	5.4%	2.8%	4.4%	2.2%	2.8%	-5.4%	-4.7%	-6.6%
exposição snr 05	-9.2%	-4.7%	-3.8%	-4.7%	-5.1%	-6.3%	-9.2%	-13.3%	-13%
exposição snr 00	-7.6%	-4.1%	-8.2%	-8.2%	-5.4%	-4.7%	-8.5%	-13.6%	-13.6%
exposição snr -5	-8.2%	-7.6%	-5.1%	-2.5%	-4.7%	-4.4%	-7%	-7%	-7.9%
restaurante snr 20	38.9%	42.4%	40.8%	39.2%	43.7%	40.5%	32.6%	25%	5.4%
restaurante snr 15	21.2%	24.7%	22.5%	24.7%	23.1%	22.8%	19.3%	15.2%	1.3%
restaurante snr 10	11.1%	14.2%	12%	13%	10.8%	7.6%	6%	2.8%	-9.2%
restaurante snr 05	-4.7%	-1.9%	-3.2%	-1.3%	0.9%	0.6%	-1.3%	-5.1%	-15.2%
restaurante snr 00	-9.5%	-11.7%	-13%	-13.6%	-11.7%	-9.8%	-11.7%	-7.9%	-16.8%
restaurante snr -5	-6.3%	-14.6%	-15.2%	-14.6%	-17.4%	-17.4%	-18.7%	-20.3%	-20.3%
rua snr 20	39.2%	38.6%	40.2%	41.8%	36.1%	34.2%	29.7%	19.6%	3.8%
rua snr 15	29.4%	26.6%	25.3%	26.6%	23.7%	24.7%	17.7%	17.7%	1.3%
rua snr 10	23.1%	21.2%	21.2%	22.5%	21.2%	16.1%	13%	4.1%	-1.9%
rua snr 05	7.9%	10.4%	11.4%	14.6%	13.6%	10.1%	10.4%	2.8%	-2.5%
rua snr 00	0.3%	2.2%	0.6%	0%	2.2%	1.3%	0.9%	-6.3%	-6.6%
rua snr -5	1.9%	1.9%	0.3%	0%	1.6%	1.6%	-0.6%	-1.3%	-3.2%
metrô snr 20	32.6%	34.8%	32.3%	32.3%	37.3%	31.6%	28.2%	19%	-3.2%
metrô snr 15	20.9%	24.1%	27.2%	25.3%	25.6%	27.5%	24.4%	15.5%	-0.9%
metrô snr 10	1.3%	3.5%	1.9%	7%	5.7%	5.4%	3.5%	0.6%	-10.8%
metrô snr 05	-5.1%	-3.2%	-6%	-5.1%	-3.8%	-4.1%	-6.3%	-6.6%	-13.6%
metrô snr 00	-2.5%	-3.2%	-1.9%	-3.8%	-5.1%	-3.2%	-5.1%	-4.7%	-11.4%
metrô snr -5	0.6%	0%	0.6%	0.9%	1.6%	0.6%	1.3%	-0.6%	-9.2%
trem snr 20	62.3%	63.3%	67.7%	65.8%	62.7%	59.2%	54.1%	46.8%	24.4%
trem snr 15	58.9%	59.5%	62.3%	60.8%	57%	54.1%	52.2%	39.2%	16.8%
trem snr 10	46.5%	47.2%	44%	40.2%	39.2%	35.8%	31.3%	25.6%	11.4%
trem snr 05	21.2%	24.1%	23.1%	22.8%	26.9%	22.5%	20.3%	13.6%	-0.6%
trem snr 00	6%	6.6%	7%	7.9%	9.2%	7.6%	5.1%	2.2%	-5.1%
trem snr -5	7%	5.7%	5.1%	5.1%	3.2%	3.5%	3.2%	-1.3%	-3.8%

**Tabela C.10:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de restaurante, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	77.9%	78.2%	78.2%	77.6%	77.6%	78.5%	79.5%	78.9%	78.9%
aeroporto snr 20	42.6%	43.2%	43.5%	46.1%	46.4%	47.6%	48.3%	48.6%	48.9%
aeroporto snr 15	30.3%	32.5%	32.5%	33.4%	35.6%	35.3%	38.5%	40.1%	40.4%
aeroporto snr 10	16.4%	18.3%	19.6%	20.2%	22.7%	22.7%	20.8%	23%	25.2%
aeroporto snr 05	-3.8%	-0.3%	0.9%	2.5%	3.8%	3.8%	4.7%	5.7%	6%
aeroporto snr 00	-3.2%	-1.6%	-0.9%	-0.6%	-1.9%	0%	0%	0.3%	0.3%
aeroporto snr -5	-13.2%	-12.6%	-14.5%	-14.8%	-13.9%	-13.6%	-15.5%	-14.8%	-14.5%
balbulcio snr 20	51.4%	51.1%	51.4%	52.4%	54.6%	55.2%	55.8%	56.5%	55.2%
balbulcio snr 15	32.5%	33.4%	32.2%	36.3%	34.7%	34.4%	35.6%	37.5%	36.9%
balbulcio snr 10	12.6%	13.9%	17.7%	18.6%	18.3%	19.6%	21.1%	21.5%	21.8%
balbulcio snr 05	-5.4%	-4.7%	-4.1%	-4.1%	-3.8%	-3.8%	-3.8%	-0.9%	0.9%
balbulcio snr 00	-8.2%	-7.9%	-8.5%	-7.9%	-8.2%	-7.9%	-7.6%	-6.9%	-5.7%
balbulcio snr -5	-1.3%	-1.6%	-0.9%	-2.2%	-3.2%	-1.9%	-2.2%	-1.9%	-2.8%
carro snr 20	55.5%	56.5%	55.2%	54.3%	54.3%	54.3%	55.2%	57.4%	56.5%
carro snr 15	42.3%	42.3%	43.2%	43.5%	43.2%	44.2%	44.2%	45.4%	46.1%
carro snr 10	18.9%	21.8%	24%	25.9%	25.6%	25.9%	26.2%	27.1%	28.7%
carro snr 05	2.2%	3.5%	6%	4.7%	5%	7.3%	6.3%	6.6%	6.6%
carro snr 00	1.9%	2.2%	2.2%	2.2%	2.2%	2.2%	1.6%	0.9%	1.9%
carro snr -5	4.4%	4.7%	4.1%	4.7%	5%	5%	5.4%	5.4%	6%
exposição snr 20	28.4%	31.2%	33.1%	32.5%	31.9%	30.9%	30%	30%	29.7%
exposição snr 15	6.9%	7.6%	7.9%	6.9%	8.2%	12.3%	10.1%	9.1%	8.8%
exposição snr 10	2.8%	2.2%	2.5%	3.2%	3.8%	3.5%	3.8%	2.2%	3.2%
exposição snr 05	-7.9%	-9.5%	-12%	-11%	-10.1%	-8.2%	-8.2%	-6.3%	-6.9%
exposição snr 00	-7.3%	-8.8%	-10.1%	-9.8%	-9.1%	-7.9%	-8.8%	-7.6%	-6.6%
exposição snr -5	-5.4%	-7.3%	-7.9%	-6.6%	-6.3%	-4.4%	-7.3%	-6.9%	-8.2%
restaurante snr 20	40.4%	41%	39.4%	38.5%	37.5%	37.2%	37.2%	37.9%	38.5%
restaurante snr 15	17.4%	16.1%	17.4%	16.4%	17.4%	18.9%	20.5%	19.9%	19.6%
restaurante snr 10	2.8%	7.3%	6.6%	9.8%	10.7%	10.4%	11.7%	10.1%	8.8%
restaurante snr 05	-5.7%	-5%	-4.7%	-4.4%	-4.7%	-5.4%	-7.6%	-6%	-5.7%
restaurante snr 00	-8.5%	-9.1%	-10.4%	-9.8%	-8.8%	-9.1%	-10.4%	-9.5%	-8.8%
restaurante snr -5	-2.2%	-3.5%	-3.8%	-4.1%	-3.5%	-2.5%	-5%	-5.7%	-6%
rua snr 20	41.3%	43.8%	45.4%	47.3%	47.9%	47%	45.1%	44.2%	43.2%
rua snr 15	28.1%	26.5%	28.4%	30.6%	30.6%	31.2%	31.2%	30.9%	31.5%
rua snr 10	21.5%	19.9%	21.1%	20.8%	22.1%	23.3%	23.3%	22.7%	24.3%
rua snr 05	8.2%	6.6%	6.9%	7.6%	6.3%	6.6%	6.9%	7.3%	8.2%
rua snr 00	-1.3%	0.3%	-0.6%	1.3%	2.2%	1.3%	0.9%	-0.3%	0.3%
rua snr -5	1.3%	1.9%	2.5%	2.2%	2.2%	2.8%	3.2%	3.5%	1.6%
metrô snr 20	30.3%	30.3%	30.6%	34.7%	32.5%	32.5%	33.1%	33.1%	32.8%
metrô snr 15	20.5%	20.2%	22.4%	22.4%	22.7%	23%	24%	23.7%	21.5%
metrô snr 10	-0.3%	-0.9%	0.9%	2.8%	0.6%	0.3%	0.9%	-0.3%	-1.6%
metrô snr 05	-5.4%	-4.4%	-4.7%	-4.7%	-4.4%	-5.4%	-4.7%	-6.3%	-5.4%
metrô snr 00	-4.4%	-4.1%	-2.5%	-2.5%	-1.9%	-2.2%	-1.9%	-1.9%	-1.3%
metrô snr -5	0.3%	1.3%	0.9%	0.6%	1.6%	1.6%	1.6%	1.6%	1.3%
trem snr 20	64.7%	65%	63.4%	63.7%	64.4%	62.8%	63.1%	63.1%	63.1%
trem snr 15	53.9%	54.3%	53.3%	54.9%	55.8%	55.8%	56.5%	57.1%	58.7%
trem snr 10	40.4%	43.8%	44.5%	46.7%	47%	44.8%	43.8%	45.7%	46.4%
trem snr 05	17%	17.4%	17.4%	18.3%	18.9%	19.2%	21.5%	20.2%	20.2%
trem snr 00	6.3%	7.3%	6.6%	6.3%	6.6%	6.3%	5.4%	5.7%	6.6%
trem snr -5	6%	6%	6.6%	5.7%	6%	6.6%	6.3%	6.3%	6.9%

**Tabela C.11:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de rua, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	77.8%	76.3%	74.7%	71.5%	64.9%	63%	56.6%	36.1%	17.4%
aeroporto snr 20	45.3%	41.5%	40.8%	38%	34.5%	32.3%	28.2%	12.3%	3.5%
aeroporto snr 15	32.3%	34.2%	29.4%	25.3%	22.8%	19.9%	15.8%	5.1%	-3.8%
aeroporto snr 10	18.7%	21.5%	19.3%	17.4%	12.3%	12.3%	10.1%	1.9%	-12.7%
aeroporto snr 05	3.2%	1.6%	3.8%	4.4%	-2.5%	-5.1%	-9.8%	-11.4%	-17.1%
aeroporto snr 00	-3.5%	-5.4%	-6%	-6.3%	-5.7%	-5.7%	-10.1%	-14.9%	-19.3%
aeroporto snr -5	-18%	-22.2%	-21.8%	-22.2%	-24.1%	-27.5%	-26.6%	-27.8%	-23.4%
balbulcio snr 20	50.9%	50.6%	46.2%	42.1%	37.3%	32%	24.1%	13.9%	3.8%
balbulcio snr 15	33.5%	29.4%	31.3%	27.8%	23.4%	16.5%	16.8%	8.2%	1.6%
balbulcio snr 10	19.6%	19.3%	17.7%	14.6%	13.3%	10.8%	7.9%	-4.7%	-10.8%
balbulcio snr 05	-4.1%	-4.1%	0.6%	-0.3%	-5.1%	-7.3%	-11.4%	-17.4%	-24.4%
balbulcio snr 00	-11.7%	-11.4%	-14.9%	-13.6%	-18.7%	-19.9%	-21.2%	-25%	-30.1%
balbulcio snr -5	-2.8%	-4.4%	-9.5%	-11.1%	-13.6%	-18.4%	-20.3%	-25.9%	-26.3%
carro snr 20	58.2%	58.5%	55.4%	56.3%	55.7%	50.6%	48.7%	34.8%	19.6%
carro snr 15	45.3%	45.9%	43%	44.9%	38%	39.6%	34.2%	25.3%	11.4%
carro snr 10	26.6%	30.7%	28.8%	26.3%	26.6%	25.3%	19%	9.5%	-1.6%
carro snr 05	8.5%	7.6%	5.4%	6.3%	7%	6.3%	1.6%	-2.8%	-4.1%
carro snr 00	1.6%	3.8%	3.8%	3.5%	3.8%	0.9%	1.3%	-0.3%	-3.5%
carro snr -5	6%	5.1%	4.7%	4.7%	3.8%	3.5%	3.5%	2.8%	2.2%
exposição snr 20	29.1%	34.5%	33.5%	33.2%	27.5%	22.8%	16.5%	9.2%	3.5%
exposição snr 15	10.8%	13%	17.7%	17.1%	8.2%	4.1%	6%	2.8%	-7.3%
exposição snr 10	2.5%	2.8%	1.9%	0.9%	1.6%	-1.3%	-2.2%	-7.9%	-7.9%
exposição snr 05	-6.3%	-9.8%	-5.7%	-9.5%	-13.9%	-14.6%	-13%	-8.9%	-16.1%
exposição snr 00	-7.9%	-11.7%	-10.1%	-12.7%	-12%	-19%	-18.7%	-16.8%	-19%
exposição snr -5	-8.2%	-8.9%	-10.8%	-11.7%	-16.5%	-24.1%	-20.6%	-20.6%	-15.2%
restaurante snr 20	38%	37.3%	35.4%	32%	32%	24.4%	18.4%	15.2%	-0.9%
restaurante snr 15	21.5%	21.8%	20.6%	22.5%	17.7%	13.6%	11.7%	1.9%	-9.2%
restaurante snr 10	12.3%	12.3%	11.7%	5.7%	6.3%	1.3%	-3.2%	-8.9%	-18.7%
restaurante snr 05	-4.7%	-2.5%	-7%	-7.3%	-7%	-11.7%	-13.3%	-18.4%	-24.7%
restaurante snr 00	-9.8%	-13.9%	-13.9%	-17.4%	-18%	-16.8%	-17.1%	-27.5%	-33.5%
restaurante snr -5	-9.2%	-12.7%	-18%	-21.2%	-21.5%	-30.4%	-31.6%	-34.2%	-37%
rua snr 20	43.7%	40.8%	40.8%	44.6%	37.7%	29.4%	23.4%	15.8%	3.5%
rua snr 15	27.5%	26.3%	24.4%	24.4%	25%	20.9%	18.7%	13.6%	2.5%
rua snr 10	20.9%	19%	17.4%	16.8%	12.7%	11.4%	8.5%	1.3%	-7.9%
rua snr 05	6.6%	9.2%	11.7%	7.9%	6.6%	2.8%	2.5%	-5.1%	-12.3%
rua snr 00	-2.2%	-4.4%	-5.4%	-7.6%	-9.5%	-9.5%	-15.5%	-16.8%	-21.8%
rua snr -5	0.3%	-0.3%	-3.2%	-1.3%	-3.5%	-3.8%	-6.6%	-5.4%	-10.4%
metrô snr 20	36.4%	34.5%	36.7%	40.8%	40.2%	30.7%	24.4%	18.7%	0.9%
metrô snr 15	23.4%	24.4%	28.8%	29.4%	21.2%	19%	16.5%	13.9%	0.3%
metrô snr 10	5.4%	5.1%	9.5%	10.8%	2.5%	1.3%	-0.6%	-2.8%	-10.4%
metrô snr 05	-6%	-7%	-5.1%	-7.9%	-10.4%	-8.9%	-9.8%	-13.9%	-11.4%
metrô snr 00	-0.6%	-9.8%	-7.3%	-9.2%	-10.8%	-12.3%	-17.1%	-14.2%	-14.6%
metrô snr -5	0.6%	0.9%	-2.2%	-6.6%	-12.3%	-13.6%	-10.8%	-7.9%	-5.7%
trem snr 20	67.7%	65.2%	64.9%	62.7%	62.3%	57%	51.6%	38.9%	21.5%
trem snr 15	59.8%	63%	61.4%	61.1%	59.5%	56%	47.8%	39.9%	16.1%
trem snr 10	46.2%	47.8%	41.1%	42.4%	42.7%	39.6%	33.9%	27.2%	9.5%
trem snr 05	23.7%	24.4%	26.6%	22.8%	24.4%	22.8%	20.3%	10.8%	1.6%
trem snr 00	6%	7.9%	6.3%	8.5%	7.6%	4.4%	5.1%	0%	-6.6%
trem snr -5	5.4%	3.2%	2.8%	0%	1.9%	1.6%	1.3%	0.3%	-5.1%

**Tabela C.12:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de rua, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	77.9%	78.5%	78.5%	80.4%	80.4%	80.4%	79.8%	79.8%	79.8%
aeroporto snr 20	42.3%	42.6%	42.9%	44.8%	43.8%	44.5%	44.2%	44.8%	44.8%
aeroporto snr 15	29.3%	30%	30.3%	29.3%	30.3%	31.9%	32.8%	32.2%	32.8%
aeroporto snr 10	16.1%	14.8%	15.8%	16.4%	18%	18.3%	17.7%	19.6%	19.2%
aeroporto snr 05	-2.2%	-1.3%	-1.3%	-0.6%	0.6%	0.9%	1.6%	2.5%	2.8%
aeroporto snr 00	-3.8%	-3.8%	-3.8%	-3.2%	-2.2%	-3.5%	-1.3%	-2.8%	-2.8%
aeroporto snr -5	-12.9%	-14.5%	-14.2%	-15.8%	-16.4%	-16.7%	-17.7%	-17.4%	-16.1%
balbulcio snr 20	50.5%	51.4%	51.4%	51.7%	51.7%	52.1%	53%	52.4%	52.1%
balbulcio snr 15	30%	32.5%	32.8%	35.3%	35%	34.4%	34.1%	34.1%	34.1%
balbulcio snr 10	12.9%	12.9%	15.5%	16.1%	16.4%	16.7%	18.9%	18.9%	20.2%
balbulcio snr 05	-4.7%	-5.7%	-4.7%	-4.4%	-5.4%	-5.7%	-4.1%	-4.1%	-4.1%
balbulcio snr 00	-8.2%	-8.8%	-7.6%	-8.8%	-9.5%	-9.5%	-10.7%	-11.7%	-11.7%
balbulcio snr -5	-1.3%	-1.6%	-2.2%	-2.5%	-2.5%	-2.5%	-3.2%	-2.8%	-2.2%
carro snr 20	55.2%	55.2%	57.7%	56.2%	56.8%	56.2%	57.7%	57.7%	58%
carro snr 15	41.3%	42%	42%	43.8%	44.5%	43.8%	44.8%	45.7%	43.8%
carro snr 10	18.3%	18.6%	24%	23%	25.2%	26.5%	27.1%	27.4%	25.9%
carro snr 05	3.5%	4.1%	3.8%	5%	4.4%	5%	7.3%	7.3%	8.5%
carro snr 00	2.2%	0.3%	1.3%	1.9%	1.6%	0.6%	0.6%	1.6%	2.2%
carro snr -5	4.4%	4.1%	4.1%	5%	5%	5%	5.7%	6%	6.6%
exposição snr 20	26.8%	29.7%	31.5%	31.9%	34.1%	32.2%	31.2%	30%	30%
exposição snr 15	9.5%	7.6%	9.1%	10.1%	10.1%	8.8%	12%	11.7%	11.7%
exposição snr 10	2.5%	1.9%	1.9%	1.6%	0.9%	0.9%	2.8%	1.3%	2.5%
exposição snr 05	-6.6%	-7.9%	-7.6%	-8.8%	-9.5%	-9.5%	-6%	-5%	-5%
exposição snr 00	-7.3%	-8.2%	-6.9%	-7.9%	-6.9%	-5.7%	-7.3%	-6.3%	-5.7%
exposição snr -5	-4.4%	-6.9%	-7.3%	-8.8%	-7.3%	-7.9%	-8.5%	-9.1%	-8.8%
restaurante snr 20	39.7%	39.4%	39.4%	38.8%	39.1%	39.1%	37.9%	38.8%	38.5%
restaurante snr 15	15.8%	15.1%	16.1%	17%	18%	18.3%	18.6%	20.2%	21.5%
restaurante snr 10	2.8%	7.9%	9.8%	10.1%	12%	12%	11.7%	13.6%	13.6%
restaurante snr 05	-6.3%	-5.7%	-5.7%	-6.3%	-7.3%	-4.1%	-5%	-4.1%	-3.8%
restaurante snr 00	-8.2%	-9.8%	-10.4%	-8.8%	-9.5%	-9.1%	-8.8%	-8.5%	-9.8%
restaurante snr -5	-2.5%	-2.5%	-2.5%	-3.2%	-4.7%	-5.4%	-5.7%	-5.4%	-6.6%
rua snr 20	40.4%	42.6%	42.3%	42.3%	43.2%	44.2%	44.2%	42.9%	42.9%
rua snr 15	25.9%	28.1%	26.8%	27.1%	27.4%	27.8%	27.4%	26.8%	27.1%
rua snr 10	21.1%	21.5%	20.5%	21.1%	20.8%	20.8%	21.5%	21.5%	21.8%
rua snr 05	8.5%	7.9%	10.4%	7.6%	6.6%	6.6%	8.2%	6.6%	6.6%
rua snr 00	-2.2%	-1.3%	-1.6%	-0.9%	-1.3%	-1.9%	-2.8%	-1.6%	-0.9%
rua snr -5	1.3%	1.3%	1.6%	1.9%	1.9%	1.9%	0.9%	0.6%	0.6%
metrô snr 20	31.5%	31.2%	30.6%	35%	35.6%	37.5%	37.2%	38.2%	37.9%
metrô snr 15	20.8%	21.1%	20.2%	21.1%	22.4%	25.2%	23.3%	23.7%	24.6%
metrô snr 10	0.3%	0.6%	1.6%	-0.9%	0%	0%	1.9%	4.4%	5.7%
metrô snr 05	-4.7%	-6.3%	-5%	-6.6%	-6%	-6%	-4.4%	-6%	-5.4%
metrô snr 00	-4.1%	-5.4%	-2.8%	-3.5%	-3.2%	-4.4%	-5.7%	-5%	-2.5%
metrô snr -5	0.3%	1.3%	0.6%	0%	0.9%	0.9%	0.9%	0.6%	0.9%
trem snr 20	64.4%	65%	65%	65.9%	66.2%	68.1%	68.8%	67.5%	67.5%
trem snr 15	54.6%	54.3%	53.3%	54.9%	55.2%	55.5%	57.1%	58.4%	59.3%
trem snr 10	41%	39.4%	42.6%	43.5%	46.1%	46.4%	45.7%	46.4%	46.4%
trem snr 05	17.4%	17.4%	18%	18.6%	19.2%	21.1%	22.4%	23.3%	23%
trem snr 00	6.3%	7.3%	6.9%	6.9%	7.3%	6.3%	6.3%	5.7%	5.4%
trem snr -5	5.4%	6.3%	6.3%	6.6%	6.9%	7.3%	6.6%	6.3%	5.4%

**Tabela C.13:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de metrô, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	78.2%	76.6%	74.1%	71.5%	68.4%	61.4%	42.7%	21.2%	-5.1%
aeroporto snr 20	41.1%	37.7%	32.6%	29.7%	25.6%	23.1%	17.7%	0.3%	-14.9%
aeroporto snr 15	31.6%	27.8%	23.4%	21.5%	11.7%	8.9%	-0.3%	-14.6%	-21.2%
aeroporto snr 10	15.2%	18.7%	16.8%	11.1%	5.4%	0.3%	-11.4%	-15.5%	-24.1%
aeroporto snr 05	-0.6%	0.3%	-2.8%	-7.9%	-8.5%	-11.4%	-15.5%	-25.9%	-23.1%
aeroporto snr 00	-5.1%	-6%	-6.3%	-9.5%	-12.3%	-16.8%	-23.4%	-25%	-22.5%
aeroporto snr -5	-14.9%	-18%	-18%	-22.2%	-25%	-25.9%	-22.8%	-19.9%	-7.6%
balbulcio snr 20	47.2%	45.3%	37.7%	26.6%	27.2%	16.5%	8.9%	-6.3%	-18.4%
balbulcio snr 15	29.1%	28.5%	27.2%	23.7%	19.3%	13.9%	7.6%	-3.5%	-19.9%
balbulcio snr 10	16.5%	11.7%	10.8%	10.1%	6.6%	4.1%	-4.4%	-16.5%	-21.5%
balbulcio snr 05	-2.5%	-0.9%	-3.8%	-5.4%	-8.2%	-15.2%	-18%	-21.2%	-26.3%
balbulcio snr 00	-9.2%	-9.5%	-13%	-16.8%	-19.6%	-22.8%	-23.4%	-24.1%	-22.8%
balbulcio snr -5	-3.8%	-7.9%	-10.4%	-13.9%	-19.9%	-21.8%	-27.8%	-32.9%	-24.7%
carro snr 20	55.1%	58.5%	56.6%	50%	46.8%	40.5%	30.1%	15.5%	-8.2%
carro snr 15	42.1%	38.9%	38%	37%	31.6%	25%	19%	1.6%	-16.1%
carro snr 10	24.4%	22.2%	24.4%	18%	15.2%	8.9%	-3.2%	-9.8%	-20.3%
carro snr 05	6%	5.1%	7.9%	8.9%	9.8%	5.1%	-1.3%	-7.9%	-13.3%
carro snr 00	3.5%	2.5%	2.2%	4.1%	5.7%	4.4%	2.5%	-2.8%	-6.3%
carro snr -5	4.7%	6%	6%	6%	5.7%	5.7%	3.2%	-3.5%	-7.6%
exposição snr 20	38.3%	37.7%	33.5%	33.9%	29.7%	26.9%	19.3%	6%	-8.5%
exposição snr 15	11.4%	13.6%	15.8%	13.3%	13%	6.6%	0%	-3.2%	-14.2%
exposição snr 10	-1.3%	-0.3%	0.9%	-2.2%	-6.6%	-11.7%	-7.9%	-11.4%	-15.8%
exposição snr 05	-4.1%	-4.7%	-4.1%	-4.7%	-7.9%	-7.3%	-8.2%	-13%	-16.5%
exposição snr 00	-7.9%	-4.7%	-7.3%	-6.6%	-7.3%	-5.4%	-8.9%	-12.3%	-16.5%
exposição snr -5	-4.4%	-2.5%	-6%	-3.8%	-5.4%	-6.6%	-12%	-7.9%	-11.1%
restaurante snr 20	36.7%	36.7%	32.3%	28.5%	26.9%	22.2%	11.7%	-0.9%	-21.5%
restaurante snr 15	18.7%	25.3%	23.7%	19.3%	15.8%	7.3%	1.3%	-12%	-20.3%
restaurante snr 10	8.5%	7%	3.5%	-1.3%	-6.6%	-8.5%	-11.4%	-16.1%	-22.5%
restaurante snr 05	-5.7%	-6.6%	-8.5%	-12.3%	-12.7%	-19.9%	-17.4%	-26.3%	-26.9%
restaurante snr 00	-12%	-14.6%	-16.5%	-18%	-19%	-20.3%	-26.3%	-27.2%	-15.5%
restaurante snr -5	-9.8%	-14.9%	-19.9%	-24.4%	-25.9%	-30.1%	-34.8%	-36.1%	-30.4%
rua snr 20	45.3%	41.5%	36.7%	36.4%	32.9%	22.8%	14.6%	0.6%	-12.7%
rua snr 15	25%	24.1%	20.6%	21.5%	15.5%	11.7%	6.3%	-7.9%	-19.6%
rua snr 10	19.3%	20.3%	16.1%	14.6%	11.7%	1.9%	-1.9%	-13.3%	-23.7%
rua snr 05	6.3%	6.3%	3.8%	2.8%	1.6%	-4.4%	-9.5%	-16.1%	-22.5%
rua snr 00	-5.1%	-1.9%	-6.3%	-11.7%	-14.9%	-13.3%	-16.5%	-14.6%	-16.5%
rua snr -5	1.3%	0.3%	0.3%	1.3%	0%	-1.6%	-7.6%	-8.2%	-15.8%
metrô snr 20	39.9%	39.6%	43.4%	44%	42.4%	37%	26.6%	12.3%	-5.4%
metrô snr 15	26.9%	29.1%	32.3%	34.2%	25.3%	21.2%	14.6%	6.3%	-6.3%
metrô snr 10	6.6%	8.9%	12.7%	14.9%	9.2%	8.5%	0.3%	-0.3%	-13%
metrô snr 05	-4.4%	-7%	-4.4%	-4.4%	-4.1%	-4.4%	-7.9%	-10.1%	-13.6%
metrô snr 00	-4.7%	-9.5%	-9.2%	-7.9%	-7.6%	-9.5%	-12%	-15.5%	-14.6%
metrô snr -5	-0.3%	1.3%	-0.6%	-1.3%	-2.5%	-5.7%	-9.8%	-12.3%	-15.5%
trem snr 20	64.9%	64.6%	59.2%	58.5%	55.1%	54.4%	37.7%	16.8%	-8.2%
trem snr 15	55.4%	59.2%	61.4%	54.7%	50.3%	39.6%	29.4%	9.8%	-17.7%
trem snr 10	44%	41.8%	38.3%	36.4%	29.7%	27.2%	19%	-2.2%	-15.2%
trem snr 05	19.6%	20.6%	20.6%	16.5%	13.3%	9.2%	3.8%	-5.1%	-19%
trem snr 00	8.5%	5.1%	4.4%	4.1%	2.5%	-0.3%	-5.7%	-10.4%	-15.8%
trem snr -5	6%	4.1%	4.4%	4.1%	4.7%	3.2%	0.6%	-2.5%	-7%

**Tabela C.14:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de metrô, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	78.5%	79.8%	79.8%	80.1%	80.1%	80.1%	79.8%	81.1%	79.5%
aeroporto snr 20	43.8%	43.2%	42.3%	42.3%	42.9%	44.2%	43.8%	42.3%	41.3%
aeroporto snr 15	30%	33.8%	30.6%	31.2%	32.8%	32.8%	32.5%	32.2%	31.2%
aeroporto snr 10	13.6%	14.5%	16.7%	19.6%	16.1%	15.8%	15.1%	13.9%	14.2%
aeroporto snr 05	-0.6%	-2.8%	-1.6%	-1.6%	-0.6%	0.3%	-0.6%	-0.6%	-0.6%
aeroporto snr 00	-5%	-4.7%	-5.7%	-5%	-4.4%	-6%	-4.7%	-5.4%	-4.7%
aeroporto snr -5	-13.6%	-16.1%	-13.9%	-14.5%	-12%	-13.9%	-12%	-14.8%	-15.8%
balbulcio snr 20	50.2%	49.8%	49.5%	49.8%	47.9%	50.2%	50.2%	49.5%	47.9%
balbulcio snr 15	30.9%	30.9%	30.3%	30.6%	29.7%	29.7%	29%	29.3%	28.7%
balbulcio snr 10	12%	14.2%	15.1%	14.5%	16.1%	18.6%	18.9%	16.7%	17.7%
balbulcio snr 05	-4.4%	-4.1%	-3.5%	-2.8%	-1.3%	-3.8%	-4.7%	-3.8%	-3.5%
balbulcio snr 00	-7.9%	-9.5%	-9.1%	-10.1%	-10.4%	-10.7%	-10.7%	-8.8%	-7.9%
balbulcio snr -5	-1.3%	-0.9%	-1.9%	-1.9%	-1.9%	-1.3%	-0.9%	-2.2%	-2.2%
carro snr 20	55.5%	56.5%	55.2%	55.8%	56.8%	57.4%	55.2%	54.3%	55.2%
carro snr 15	42%	42.6%	43.2%	41.3%	40.4%	39.7%	41.6%	42%	41.6%
carro snr 10	16.7%	19.6%	19.2%	18.9%	20.5%	21.1%	22.7%	23.7%	23.7%
carro snr 05	4.4%	4.4%	5.7%	5.4%	5%	5.4%	5.7%	6.3%	5.4%
carro snr 00	1.9%	1.3%	1.9%	1.6%	2.5%	1.9%	1.9%	2.2%	2.5%
carro snr -5	4.4%	4.7%	4.7%	4.7%	4.4%	4.4%	4.1%	4.4%	4.7%
exposição snr 20	30.6%	30.9%	33.8%	36.6%	36.3%	36%	36.3%	36.9%	36.9%
exposição snr 15	9.5%	10.4%	14.2%	12.9%	12.6%	13.2%	13.9%	12.9%	12%
exposição snr 10	2.8%	0.9%	3.5%	1.9%	0%	0%	-1.6%	-0.3%	-0.9%
exposição snr 05	-6.9%	-7.3%	-6.9%	-6.9%	-7.3%	-5.4%	-5%	-4.7%	-4.1%
exposição snr 00	-7.6%	-6.3%	-8.2%	-8.2%	-7.3%	-8.5%	-7.9%	-8.5%	-9.1%
exposição snr -5	-6.6%	-8.2%	-6.9%	-7.3%	-6.6%	-7.9%	-8.5%	-8.2%	-7.3%
restaurante snr 20	39.1%	38.5%	38.2%	37.5%	37.2%	37.5%	37.9%	37.9%	36.9%
restaurante snr 15	15.8%	17.7%	18%	18.9%	19.2%	18%	17.7%	19.2%	19.2%
restaurante snr 10	3.8%	3.8%	7.3%	7.9%	7.9%	9.1%	8.8%	8.8%	9.5%
restaurante snr 05	-6%	-5.4%	-6%	-6.6%	-5.4%	-5.7%	-6.3%	-4.4%	-6%
restaurante snr 00	-8.2%	-8.5%	-11%	-10.1%	-10.4%	-9.5%	-10.7%	-11.4%	-11.4%
restaurante snr -5	-2.5%	-3.5%	-5.4%	-5%	-6.6%	-6.6%	-6%	-8.5%	-10.7%
rua snr 20	41.6%	41%	42.3%	42%	43.2%	42.6%	42.9%	46.7%	45.4%
rua snr 15	24.6%	27.1%	26.8%	24.6%	23.3%	24.6%	24.6%	26.2%	26.2%
rua snr 10	20.2%	18.6%	19.2%	20.2%	21.5%	21.5%	22.1%	22.4%	19.6%
rua snr 05	7.6%	8.5%	8.2%	7.3%	5%	5.4%	6%	6.6%	6.6%
rua snr 00	-1.6%	-1.3%	-2.8%	-1.9%	-1.3%	-2.5%	-3.2%	-5%	-3.8%
rua snr -5	1.3%	1.3%	1.3%	0.9%	1.9%	1.6%	1.6%	1.3%	0.9%
metrô snr 20	32.5%	32.2%	33.1%	36%	37.2%	36.9%	38.8%	40.1%	41%
metrô snr 15	23.7%	19.6%	20.2%	21.8%	24%	26.2%	28.4%	28.1%	27.4%
metrô snr 10	-0.3%	2.8%	3.5%	4.7%	6.6%	6%	3.8%	4.1%	4.4%
metrô snr 05	-6%	-6.9%	-6%	-7.3%	-8.5%	-7.6%	-8.8%	-7.6%	-4.7%
metrô snr 00	-3.5%	-4.4%	-3.8%	-4.4%	-2.8%	-1.9%	-4.1%	-3.2%	-2.5%
metrô snr -5	0.9%	-0.3%	-0.3%	-0.3%	0.3%	0.3%	0.3%	0.3%	0.3%
trem snr 20	64.7%	64%	65%	64.7%	65.9%	65.9%	66.6%	65.9%	65.6%
trem snr 15	53.9%	53.6%	53.9%	55.8%	55.5%	56.5%	56.5%	56.2%	55.5%
trem snr 10	39.7%	40.1%	41.3%	43.2%	42.3%	42.3%	42.6%	42.3%	44.2%
trem snr 05	18.3%	19.2%	19.2%	20.2%	23%	20.5%	21.5%	21.1%	21.8%
trem snr 00	6.9%	7.3%	6.6%	6.6%	6.9%	6.9%	6.9%	8.8%	9.1%
trem snr -5	6%	6%	6.3%	6%	5.4%	5.4%	5%	5.7%	5%

**Tabela C.15:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de trem, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,1 a 0,9

alfa	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
limpo	79.1%	76.3%	73.1%	70.9%	65.8%	59.8%	0%	29.7%	6.3%
aeroporto snr 20	46.5%	43%	43.7%	44%	36.4%	30.7%	0%	18%	0.6%
aeroporto snr 15	37.3%	38.3%	34.5%	32.6%	26.3%	21.2%	0%	11.7%	0%
aeroporto snr 10	17.1%	23.4%	21.2%	22.5%	16.8%	11.4%	0%	5.1%	-13%
aeroporto snr 05	5.1%	4.4%	6.3%	3.2%	1.9%	3.2%	0%	-8.5%	-12.7%
aeroporto snr 00	-1.6%	-0.6%	-7%	-5.7%	-8.2%	-7.6%	0%	-12.7%	-18.7%
aeroporto snr -5	-18.4%	-21.2%	-19.3%	-21.8%	-28.2%	-33.5%	0%	-28.8%	-25%
balbulcio snr 20	52.5%	54.1%	50.9%	44%	38.6%	25.6%	0%	11.7%	-4.1%
balbulcio snr 15	35.4%	34.2%	32.3%	32%	30.4%	26.3%	0%	7.3%	-4.1%
balbulcio snr 10	18.4%	20.6%	20.3%	19.3%	16.1%	15.2%	0%	3.2%	-14.9%
balbulcio snr 05	-3.2%	-3.5%	-1.3%	-4.1%	-5.1%	-9.2%	0%	-14.6%	-23.7%
balbulcio snr 00	-10.8%	-13%	-15.5%	-19.3%	-19.6%	-23.1%	0%	-25.9%	-25%
balbulcio snr -5	-2.5%	-4.1%	-10.4%	-11.4%	-14.6%	-15.5%	0%	-28.8%	-29.4%
carro snr 20	57%	58.9%	60.1%	57.9%	57%	50.3%	0%	34.8%	11.1%
carro snr 15	47.2%	51.3%	52.2%	53.2%	45.3%	42.4%	0%	31.6%	8.5%
carro snr 10	28.5%	33.9%	36.4%	36.4%	38.3%	37%	0%	19.6%	2.8%
carro snr 05	6.6%	7.3%	8.5%	10.4%	8.5%	5.4%	0%	2.5%	-4.4%
carro snr 00	1.6%	4.1%	2.2%	0.3%	1.9%	-1.6%	0%	-7.6%	-14.9%
carro snr -5	5.7%	6%	3.8%	1.6%	0.9%	1.3%	0%	-1.9%	-6%
exposição snr 20	30.4%	33.2%	34.2%	29.1%	30.1%	30.7%	0%	14.6%	-2.5%
exposição snr 15	12.3%	11.7%	13%	12.7%	11.4%	9.2%	0%	-1.3%	-6.6%
exposição snr 10	-2.2%	-1.6%	0%	-5.1%	-1.9%	-0.9%	0%	-16.8%	-18%
exposição snr 05	-10.1%	-8.2%	-9.8%	-7%	-11.7%	-10.1%	0%	-24.7%	-26.6%
exposição snr 00	-8.2%	-10.1%	-10.4%	-13.6%	-16.1%	-18.4%	0%	-28.8%	-29.1%
exposição snr -5	-7%	-11.4%	-11.4%	-13%	-15.8%	-20.3%	0%	-17.1%	-21.8%
restaurante snr 20	38.3%	37.7%	40.5%	36.4%	33.9%	29.4%	0%	9.8%	-7.3%
restaurante snr 15	20.9%	24.1%	25.9%	20.6%	16.8%	13.9%	0%	5.1%	-15.2%
restaurante snr 10	13.6%	10.8%	8.5%	9.2%	6.3%	3.5%	0%	-9.5%	-17.7%
restaurante snr 05	-4.1%	-1.9%	-6.3%	-10.1%	-9.2%	-14.9%	0%	-18%	-27.5%
restaurante snr 00	-11.4%	-15.2%	-15.5%	-18.7%	-17.1%	-19%	0%	-24.1%	-31%
restaurante snr -5	-7.9%	-10.8%	-14.9%	-24.4%	-29.4%	-30.7%	0%	-34.2%	-34.5%
rua snr 20	43.4%	38.9%	39.9%	39.9%	33.9%	33.2%	0%	19.6%	0.9%
rua snr 15	29.4%	24.4%	24.4%	25.6%	22.5%	19.3%	0%	4.1%	-2.8%
rua snr 10	21.8%	19%	15.2%	12.3%	9.5%	6.6%	0%	1.9%	-6%
rua snr 05	5.4%	14.2%	13.6%	13%	9.2%	8.2%	0%	-0.9%	-9.5%
rua snr 00	-0.3%	-6.3%	-5.4%	-7%	-9.8%	-12.3%	0%	-14.2%	-14.2%
rua snr -5	0.9%	-0.6%	-3.2%	-1.9%	-3.2%	-5.7%	0%	-7%	-11.4%
metrô snr 20	36.4%	32.9%	33.2%	37.3%	38.3%	32.6%	0%	10.4%	-11.4%
metrô snr 15	25%	25.6%	29.7%	26.6%	20.3%	15.5%	0%	7.9%	-13.3%
metrô snr 10	1.6%	8.2%	9.8%	5.7%	5.1%	5.4%	0%	-15.2%	-21.8%
metrô snr 05	-6%	-5.7%	-6%	-10.1%	-11.7%	-11.7%	0%	-13%	-19.6%
metrô snr 00	-1.9%	-4.7%	-8.9%	-8.2%	-13.3%	-19.6%	0%	-26.6%	-18.7%
metrô snr -5	1.3%	-0.6%	-0.9%	-6%	-12.7%	-15.5%	0%	-22.2%	-13.6%
trem snr 20	66.8%	69%	72.2%	71.2%	69.9%	63.6%	0%	48.1%	19.9%
trem snr 15	59.8%	64.9%	65.8%	65.8%	65.2%	59.5%	0%	50.6%	20.9%
trem snr 10	45.9%	47.2%	44.6%	41.8%	44.6%	39.2%	0%	27.8%	14.6%
trem snr 05	23.7%	23.4%	26.9%	24.1%	21.2%	21.8%	0%	15.5%	4.4%
trem snr 00	6.3%	4.7%	4.7%	4.7%	4.1%	4.1%	0%	0.3%	-3.5%
trem snr -5	4.7%	1.3%	0%	-0.3%	-0.6%	-2.8%	0%	-4.4%	-10.1%



**Tabela C.16:** Taxa de acertos de palavras de um sistema treinado com locuções limpas adaptado utilizando modelos fonéticos corrompidos com ruído de trem, testado com locuções limpas e ruidosas e com  $\alpha$  variando de 0,01 a 0,09

alfa	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
limpo	77.9%	77.9%	78.2%	78.2%	78.9%	78.5%	80.1%	80.1%	79.8%
aeroporto snr 20	41.6%	42.3%	42.6%	46.1%	46.7%	45.4%	45.1%	45.7%	47%
aeroporto snr 15	29.3%	29%	31.5%	29.7%	31.5%	33.1%	34.7%	36.9%	36.6%
aeroporto snr 10	17%	16.7%	16.1%	15.8%	16.7%	17.4%	17.7%	16.1%	16.4%
aeroporto snr 05	-1.9%	-2.2%	-0.9%	0.3%	0.6%	1.6%	3.2%	3.8%	5%
aeroporto snr 00	-2.8%	-4.1%	-4.4%	-4.7%	-2.5%	-2.8%	-1.3%	-0.6%	-0.9%
aeroporto snr -5	-12.3%	-13.6%	-13.9%	-13.6%	-15.5%	-16.1%	-15.5%	-18.3%	-18%
balbulcio snr 20	50.2%	52.1%	52.7%	53%	51.7%	50.8%	50.8%	51.4%	52.1%
balbulcio snr 15	32.5%	32.5%	31.5%	36.3%	36.3%	35%	35%	35%	34.1%
balbulcio snr 10	12.9%	14.5%	15.5%	16.4%	18.6%	17.7%	18.6%	17.7%	19.2%
balbulcio snr 05	-4.1%	-4.4%	-3.5%	-3.5%	-3.2%	-3.5%	-4.4%	-3.2%	-2.8%
balbulcio snr 00	-8.8%	-8.8%	-10.1%	-9.1%	-9.5%	-10.1%	-11%	-12%	-10.1%
balbulcio snr -5	-1.3%	-1.3%	-2.8%	-2.5%	-2.8%	-2.5%	-2.5%	-2.8%	-3.2%
carro snr 20	55.2%	55.8%	55.8%	54.9%	55.2%	56.2%	56.2%	55.8%	57.4%
carro snr 15	42%	42.9%	43.5%	43.8%	44.8%	45.1%	46.1%	47%	49.2%
carro snr 10	18.9%	18%	22.7%	22.4%	22.4%	25.9%	26.2%	26.8%	27.4%
carro snr 05	2.8%	2.8%	5.4%	5%	5.7%	7.3%	8.5%	8.8%	7.9%
carro snr 00	2.2%	1.9%	0.3%	1.9%	3.5%	2.8%	2.2%	1.3%	1.6%
carro snr -5	3.5%	3.8%	4.4%	4.1%	4.4%	4.7%	4.4%	5%	5%
exposição snr 20	28.1%	32.2%	33.4%	32.5%	30.9%	31.5%	30%	31.5%	31.2%
exposição snr 15	6.9%	6.6%	6.6%	7.9%	8.2%	7.3%	11.7%	11.4%	12.9%
exposição snr 10	1.9%	1.3%	1.6%	0.6%	0.3%	-0.6%	0%	-0.3%	-2.2%
exposição snr 05	-7.6%	-8.8%	-12%	-12%	-12%	-10.4%	-9.1%	-9.8%	-9.5%
exposição snr 00	-7.3%	-9.1%	-6.6%	-9.5%	-7.6%	-8.8%	-8.5%	-7.3%	-7.6%
exposição snr -5	-4.7%	-7.6%	-7.6%	-6.6%	-6.3%	-7.9%	-6.6%	-8.2%	-8.2%
restaurante snr 20	39.7%	39.4%	40.1%	38.2%	39.1%	39.1%	39.1%	38.8%	38.5%
restaurante snr 15	17.4%	16.4%	16.1%	16.1%	17.4%	18.3%	19.9%	19.6%	20.2%
restaurante snr 10	2.8%	7.9%	8.5%	8.8%	9.8%	10.7%	12%	12%	12.6%
restaurante snr 05	-5.7%	-4.7%	-5.4%	-6%	-6%	-6%	-5.4%	-4.1%	-4.1%
restaurante snr 00	-7.9%	-8.8%	-9.1%	-11.4%	-8.8%	-9.8%	-9.1%	-9.5%	-8.5%
restaurante snr -5	-2.2%	-2.8%	-3.2%	-3.2%	-4.1%	-4.7%	-4.7%	-5%	-6.3%
rua snr 20	39.4%	43.2%	45.1%	45.4%	45.1%	44.8%	46.1%	42.9%	42.9%
rua snr 15	29.3%	28.4%	27.1%	28.4%	29.3%	30%	29%	29.3%	29.7%
rua snr 10	20.5%	20.2%	19.2%	20.5%	20.8%	21.5%	21.5%	20.8%	21.8%
rua snr 05	9.1%	8.2%	7.9%	7.6%	7.9%	7.3%	7.9%	7.6%	7.6%
rua snr 00	-1.9%	-1.3%	-1.9%	-0.9%	-0.9%	-0.6%	-0.9%	-1.6%	-0.9%
rua snr -5	1.3%	1.3%	1.9%	1.9%	1.9%	1.9%	1.9%	1.6%	1.3%
metrô snr 20	29.3%	29.7%	29.3%	30.9%	30.6%	31.5%	34.4%	34.4%	35.3%
metrô snr 15	21.1%	20.5%	19.9%	20.5%	22.7%	23.7%	24.6%	25.6%	24.3%
metrô snr 10	0.3%	0%	0%	-1.6%	-0.6%	0.3%	0.3%	1.6%	1.9%
metrô snr 05	-5%	-5.7%	-5%	-6%	-6.3%	-7.6%	-6.3%	-6%	-6%
metrô snr 00	-6%	-6%	-4.1%	-1.3%	-1.6%	-1.9%	-3.2%	-2.8%	-2.5%
metrô snr -5	0.3%	-0.3%	0%	0.6%	0.9%	0.6%	1.3%	1.6%	1.6%
trem snr 20	64.4%	65%	67.5%	68.1%	68.1%	67.8%	68.1%	67.8%	66.9%
trem snr 15	54.3%	52.7%	54.3%	54.9%	56.2%	56.5%	56.5%	57.1%	59.3%
trem snr 10	41.6%	42%	42.6%	43.5%	42.9%	44.5%	44.2%	46.7%	46.1%
trem snr 05	17.4%	18.6%	18.9%	17.4%	20.2%	20.5%	21.1%	23%	23.7%
trem snr 00	6.9%	6.6%	6.6%	6%	5.7%	5.7%	5.7%	6%	6.3%
trem snr -5	5.7%	5.7%	5.7%	6%	5.4%	5.4%	4.7%	5%	5%

# Anexo D

## Artigo resultante da tese

Artigo apresentado no *International Workshop on Telecommunications (IWT)* evento biênio patrocinado pelo Instituto Nacional de Telecomunicações (Inatel), realizado entre os dias 3 a 6 de maio de 2011 no Rio de Janeiro, Brasil.

# Multi-style Training Analysis for Robust Speech Recognition

Thales Antonio Fernandes Valerio  
Instituto Nacional de Telecomunicações - Inatel  
P.O. Box 05 - 37540-000  
Santa Rita do Sapucaí - MG - Brazil  
tafvalerio@gmail.com

Carlos Alberto Ynoguti  
Instituto Nacional de Telecomunicações - Inatel  
P.O. Box 05 - 37540-000  
Santa Rita do Sapucaí - MG - Brazil  
ynoguti@inatel.br

**Abstract**—The objective of this work is to investigate the degradation introduced by noise in a continuous ASR (Automatic speech recognition) and decide whether or not it is gainful to use an extract of this noisy environment to train an ASR system's parameters.

**Index Terms**—Automatic speech recognition, Robust speech recognition.

## I. INTRODUCTION

Despite decades of research on noise robustness, leading researchers in the field have called on a serious effort to improve recognizer performance in noise [1]. The main reason for poor accuracy in noise is a mismatch between the original conditions of the data used to train the system and the actual noisy environment it is tested in.

Noise robustness methods can be classified by how they address this problem. Standard approaches are:

- Front-end Compensation;
- Inherently Robust Front-end;
- Model-based Compensation.

In front-end compensation, noise is explicitly removed from the observed speech to better match the clean models of speech. These methods provide an estimation of the clean speech parameterization in order to reduce the mismatch between training (clean) and recognition (noisy) conditions. This way, the clean version of the speech is recognized using models trained under clean conditions. In this category, we can find the following methods: parameter mapping [2][3][5], spectral subtraction [6], statistical enhancement [4], compensation based on clean speech models [7][8][9][10].

In the Inherently Robust Front-end approach, the goal is to seek for speech features that are immune to noise. A common assumption in the speech parameterization methods is that the speech is considered to be independent of the noise. Among the various methods that use this approach we cite: application of liftering windows [11], methods based on auditory models [12][13][14][15], mel-scaled cepstrum [14][16], discriminative parameterizations [17][18], slow variation removal [16][19][20][21][22] and inclusion of time derivatives of parameters [23][24].

Finally, in model-based compensation, acoustic model parameters can be adapted to reflect the effects of noise. This

technique allows to adapt the HMM's (Hidden Markov Model) emission matrix to account for the degree to which the noise will affect its mean and variance elements. However improvement in results typically come with a significant computational cost [25]. Among the methods that use this approach, we can cite: HMM decomposition [26][27], state dependent Wiener filtering [28], statistical adaptation of HMMs [16] and contamination of the training database [29].

A schematic view of these process in an ASR system is shown in Figure 1.

This work can be classified in the third approach, more precisely in the contamination of the training database method. Several ideas can be tested: one could train a single system with all noise types and SNRs. Alternatively, one would train a specific ASR for each noise type and level and switch among all of these systems according to the incoming speech. Some questions arise from these approaches:

- Is it possible to have a single system that can deal with all types of noise in all possible SNR scenarios? In this case, how would be the performance of such system?
- Besides the noise type, the actual SNR of the incoming speech is also an important parameter to take into account. How to deal with this problem?
- In another scenario, we could have a specific ASR for a given noise type. The question to be answered here is whether if the gain worth the effort.
- For the second scenario, what would happen if we choose the wrong noise type (and therefore the wrong system) for a given situation?

The aim of this work is to search for answers for the first three questions. The last one is left for a future work. The rest of this paper is organized as follows: in Section II the experimental apparatus used for the tests is described. Section III shows the tests that were performed, together with the results and analysis. Finally, Section IV brings the final conclusions for this work.

## II. EXPERIMENTAL SETUP

In order to answer the questions a series of experiments were performed. They will be described and analyzed later in this work. Before entering in detail about the results, it is

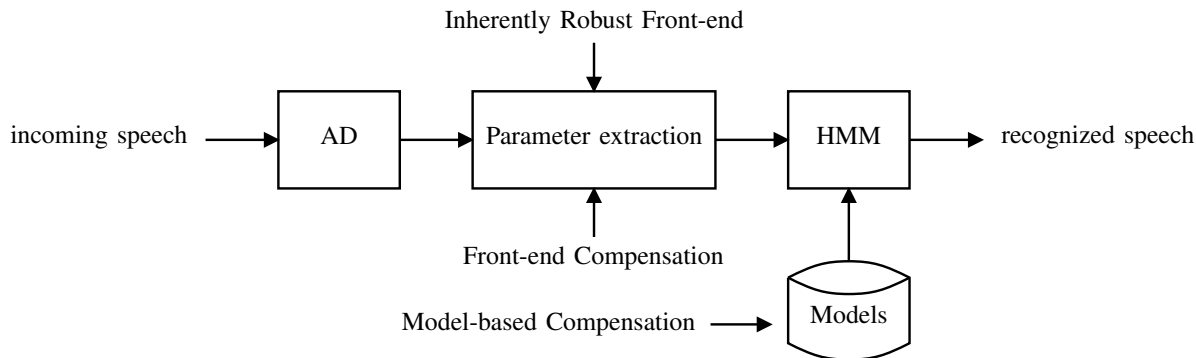


Fig. 1. Different approaches for speech recognition and their point of application in an ASR system.

necessary to describe the experimental setup used to perform the tests, and this section is dedicated to this task.

#### A. Speech recognition engine

The engine used for the experiments is a phoneme based, continuous density HMM developed in [30]. Each phone consists of a 3 state HMM, with the allowed transitions among states as shown in Figure 2.

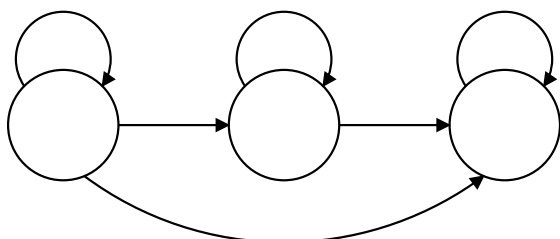


Fig. 2. Phone model used in the ASR system.

The acoustic parameters were the 12 MFCC, together with their first and second derivatives, leading to feature vectors of dimension 36. For each state, a mixture of 10 multidimensional gaussian distributions with diagonal covariance matrix was used.

The phonetic transcription of each vocabulary word was performed by 36 context independent phone models, and a bigram language model was also used to improve the recognition performance [30].

#### B. Database

The speech corpus were comprised of 40 adult speaker, 20 men and 20 women [31]. Each of them recorded 40 utterances in Brazilian Portuguese. The material that was recorded came from phonetically balanced sentences suggested in [32], and have a total of 694 words. Therefore, the problem at hand can be considered as speaker independent, continuous speech recognition with a medium size vocabulary.

All audio files were recorded in a low noise environment, at 8kHz sample rate and 16-bit coded.

To generate the noise corrupted versions of these recordings, the noises from Aurora database [29] were used. These are airport, babble, car, exhibition, restaurant, street, subway and

train. Six versions of each utterance were created by electronically adding these noises at the following SNRs: 20, 15, 10, 5, 0 and -5 dB.

### III. RESULTS

In this section, the tests and the results that support our conclusions are described. Some comments and conclusions are shown as well.

#### A. Baseline system

The baseline proposed here is a system trained with clean utterances and also tested with clean utterances. The result for this scenario was 82.90% of words correctly recognized considering substitution and deletion errors. However including errors due to insertion this total decline to 75%. For this reason this system has a 75% word accuracy. These and next results were calculated using the *sclite* tool, provided by NIST [33].

#### B. System trained with clean speech and tested with noisy speech

The second test were performed with the same system (trained with clean speech), but tested against a database corrupted with noise.

The results, summarized in Table I show a dramatic performance drop, a result that is expected due to the great mismatch between the noise conditions in the training and testing material.

TABLE I  
WORD ACCURACY RESULTING FROM A SYSTEM TRAINED WITH CLEAN UTTERANCES AND TESTED AGAINST THE CONTAMINATED CORPUS.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
Airport	41.4%	31.6%	18.2%	0.4%	-4.10%	-12.5%
Babble	48.2%	31.5%	14.5%	-5.2%	-7.6%	-0.9%
Car	55.7%	41.3%	22.3%	2.7%	2.5%	3.8%
Exhibition	31.1%	10.2%	-0.7%	-7.5%	-6.3%	-2.8%
Restaurant	38.1%	19.5%	6.9%	-5.4%	-9.4%	-0.4%
Street	41.8%	23.8%	19.2%	7.0%	-3.1%	1.7%
Subway	27.7%	17.6%	-0.4%	-6.5%	-3.1%	0.8%
Train	64.8%	55.7%	37.3%	18.1%	5.3%	5.3%

### C. System trained with all noise types and SNRs

From the previous results it is clear that the next step is to train a ASR system using a corrupted training material. The first approach was to train the system with all noise types and SNRs. The underlying hypothesis is that a system that is exposed to all possible situations should be more robust and therefore have better performance. Table II shows the results for this tests.

TABLE II

WORD ACCURACY RESULTING FROM A SYSTEM TRAINED WITH AUDIOS CONTAMINATED WITH ALL NOISES AND ALL NOISE LEVELS (SNR) AND TESTED AGAINST THE CONTAMINATED CORPUS

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
Airport	52.8%	49.7%	42.9%	22.6%	7.1%	-11.2%
Babble	49.4%	44.0%	33.1%	16.5%	-1.1%	-0.9%
Car	52.1%	61.5%	53.9%	21.1%	-1.0%	3.2%
Exhibition	61.5%	56.2%	33.1%	14.7%	3.3%	2.2%
Restaurant	51.3%	52.0%	41.7%	17.7%	2.6%	-1.8%
Street	43.5%	35.5%	31.9%	21.9%	-0.3%	3.1%
Subway	54.8%	50.6%	35.8%	16.7%	4.7%	0.8%
Train	58.7%	66.9%	63.5%	55.9%	25.6%	-1.3%

It can be seen that, in general, the results are better than in previous case, but it not true for all situations. For example the test with the car noise with SNR = 20 dB presented a worst result when compared with the previous test, an unexpected result.

One possible explanation for this result is that utterances with low SNR have lost the speech information. Therefore, the system is trained mainly with noise and not with the actual speech signal. In this way, the speech recognition system becomes a noise recognition system.

To test this hypothesis, another test set was performed, now using only utterances with “high” SNRs. The results of these tests are shown in the sequel.

### D. System trained with all noises, but only with SNR = 20 dB and SNR = 15dB

As observed in the previous section, training an ASR system with all SNRs doesn’t lead to a good performance maybe because in this case the system is starting to model the noise instead of the speech.

To verify this point, we trained a system with all noise types, but only with SNR = 15dB and SNR = 20 dB. The results of these tests are shown in Table III

The results os these tests show a big improvement when compared to the previous ones. Therefore, the hypothesis that when training an ASR system with high noise levels lead to a poor performance because the system starts to model only the noise is verified.

An interesting point is that the performance is better even for heavily degraded signals (lower SNRs), corroborating the above hypothesis.

TABLE III

WORD ACCURACY FOR A SYSTEM TRAINED WITH ALL NOISE TYPES BUT ONLY NOISE LEVELS OF 15 dB AND 20 dB AND TESTED AGAINST THE CONTAMINATED CORPUS.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
Airport	75.3%	69.4%	53.9%	24.0%	4.2%	-3.3%
Babble	73.8%	62.0%	41.7%	14.5%	0.5%	4.6%
Car	68.5%	68.2%	48.5%	15.2%	-0.8%	1.9%
Exhibition	73.4%	59.4%	31.6%	5.1%	-0.4%	1.6%
Restaurant	68.5%	63.1%	38.7%	14.2%	-1.7%	0.5%
Street	64.4%	51.2%	46.0%	31.6%	5.4%	1.7%
Subway	69.2%	64.8%	40.0%	8.1%	0.6%	1.7%
Train	75.6%	76.1%	64.4%	44.6%	13.5%	-0.4%

### E. Systems trained and tested with the same noise type

The final question to be answered is whether if is there any gain by training a system to be used for a specific noise type. Intuitively, this approach should lead to a better performance due to a better acoustic matching between the training and testing conditions.

To verify this for each noise used on this paper a HMM was trained using as its training database utterances corrupted with each of these noise and a SNR = 20 dB. On Table IV each line represents values when the testing utterances have been corrupted with a determined noise and SNR inserted on a HMM trained with the same noise.

TABLE IV

WORD ACCURACY RESULTING FROM SYSTEMS TRAINED AND TESTED WITH THE SAME NOISE TYPE. AS BEFORE, ONLY UTTERANCES WITH SNR = 15 dB AND SNR = 20 dB WERE USED TO TRAIN THE SYSTEM.

	SNR20	SNR15	SNR10	SNR5	SNR0	SNR-5
Airport	77.5%	73.6%	56.3%	28.8%	4.4%	-1.1%
Babble	70.1%	68.7%	56.1%	15.9%	2.5%	2.6%
Car	71.8%	72.4%	56.3%	15.5%	0.6%	3.0%
Exhibition	75.8%	70.3%	48.8%	14.5%	-4.1%	-0.3%
Restaurant	70.0%	63.5%	49.6%	21.0%	1.0%	-0.2%
Street	63.0%	53.6%	39.1%	24.3%	3.1%	-0.2%
Subway	75.3%	71.1%	48.1%	13.6%	1.4%	3.1%
Train	70.2%	75.9%	68.6%	52.4%	14.4%	-4.00%

The results show that, in general, the performance of these systems is better when compared to a system trained with all noises. However, in some cases, the performance is worst. Therefore, we can conclude that a system trained with all noise types have a better cost benefit.

## IV. CONCLUSIONS

In this work, a study about the adequacy of the training material to build a robust ASR were carried out.

The first tested hypothesis was that to better cope with noise, we should train an ASR system with different noises and SNRs. The test result showed that only part of this hypothesis is true: as the signal to noise ratio falls, the noise mode starts to dominate, the noisy speech means move towards the noise

and the variances shrink. Eventually, the noise dominates and there is little or no information left in the signal. The net effect is that the probability distributions trained on clean data become very poor estimates of the data distributions observed in the noisy environment and recognition error rates increase rapidly.

This result led to a new set of tests, where the system was trained with all noise types, but only with utterances with “high” SNR, that is, SNR = 20 dB and 15 dB. This strategy led to a great improvement in the system’s performance, corroborating the above conclusion.

The last question to be answered was if there is any gain by training a system to be used for a specific noise type. Intuitively, this approach should lead to a better performance due to a better acoustic matching between the training and testing conditions. However, the test results showed that there is a little gain in most of cases, and the performance is even worst in some cases.

Therefore, the best strategy seems to be to train a single ASR system with all noise types but only with high SNRs.

#### REFERENCES

- [1] S. Furui. 50 years of progress in speech recognition technology: Where we are, and where we should go? From a poor dog to a super cat. Keynote Presentation, ICASSP 2007.
- [2] C. Mokbel and G. Chollet. Speech recognition in adverse environments: speech enhancement and spectral transformations. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -91,1991.
- [3] K. Ohkura and M. Sugiyama. Speech recognition in a noisy environment using a noise reduction neural network and a codebook mapping technique. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -91,1991.
- [4] Y. Ephraim. Statistical model based speech enhancement systems. Proc. IEEE, 80:1526?1555, October 1992.
- [5] F. Seide and A. Mertins. Non-linear regression based feature extraction for connected-word recognition in noise. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -94,1994.
- [6] S.F. Boll. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. on Acoustics, Speech, and Signal Processing, 27:113?120, 1979.
- [7] A. Acero. Acoustical and Environmental Robustness in Automatic Speech Recognition, 1993, Boston, MA, Kluwer Academic Publishers
- [8] R. Stern, B. Raj and P. J. Moreno. Compensation for environmental degradation in automatic speech recognition. Proceedings of ESCA-NATO Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels, 1997.
- [9] P.J Moreno and B. Eberman. A new algorithm for robust speech recognition: the delta vector Taylor series approach. Proceedings of EuroSpeech-97,1997.
- [10] P. J. Moreno, P.J, B. Raj and R. Stern. Data-driven environmental compensation for speech recognition: a unified approach. Speech Communication, 24, 4,1998,267-288.
- [11] J. C. Junqua and H. Wakita. A comparative study of cepstral lifters and distance measures for all pole models of speech in noise. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -89,1989.
- [12] H. Hermanski Hanson, B.A. and H. Wakita. Low dimensional representation of vowels based on all-pole modeling in the psychophysical domain. Speech Communication, 4, (1-3), 1985,181-188.
- [13] O. Ghitza. Auditory models and human performance in tasks related to speech coding and speech recognition. IEEE Trans, on Speech and Audio Processing, 2,1,1994,115-132.
- [14] C.R. Jankowski, J. Hoang-Doan and R. P. Lippmann. A comparison of signal processing front ends for automatic word recognition. IEEE Trans, on Speech and Audio. Processing, 3, 4,1995, 286-293.
- [15] J. C. Junqua and J. P. Haton. Robustness in automatic speech recognition, Kluwer Academic Publishers, 1996.
- [16] P. J. Moreno. Speech Recognition in Noisy Environments, PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania, 1996.
- [17] M. J. Hunt, S. M. Richardson, D. C. Bateman and A. Piau. An investigation of PLP and IMELDA acoustic representations and their potential for combination. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -91,1991.
- [18] A. Torre, A. M. Peinado, A. J. Rubio and P. Garcia. Discriminative feature extraction for speech recognition in noise. Proceedings of EuroSpeech-97, Rhodes, 1997.
- [19] C. Mokbel, J. Monn and D. Jouvet. On-line adaptation of a speech recognizer to variations in telephone line conditions. Proceedings of EuroSpeech-93,1993.
- [20] H. Hermanski, N. Morgan and H. G. Hirsch. Recognition of speech in additive and convolutional noise based on RASTA spectral processing. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -94, 1994.
- [21] P.J. Moreno, R. Stern. Source of degradation of speech recognition in telephone network. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -94,1994.
- [22] A. Anastasakos, E. Kubala, J. Makhoul and R. Schwartz. Adaptation to new microphones using tied-mixture normalization. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -94,1994.
- [23] S. Furui. Speaker-independent isolated word recognition using dynamic features of speech spectrum. IEEE Trans, on Acoustic, Speech and Signal Processing, 34,1986,52-59.
- [24] J. Hernando and C. Nadeu. Speech recognition in noisy car environment based on OSALPC representation and robust similarity measuring techniques. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -94,1994.
- [25] M. Gales and S. Young . The Application of Hidden Markov Models in Speech Recognition, Foundations and Trends® in Signal Processing Vol. 1, No. 3, 2008.
- [26] M. F. J. Gales and S. J. Young. HMM recognition in noise using parallel model combination. Proceedings of EuroSpeech-93,1993.
- [27] M. F. J. Gales. ‘Nice’ model-based compensation schemes for robust speech recognition. Proceedings of ESCA-NATO Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels, 1997.
- [28] S. V. Vaseghi and B. P. Milner. Noise compensation methods for hidden Markov model speech recognition in adverse environments. IEEE Trans, on Speech and Audio Processing, 5,1,1997,11-21.
- [29] H. G. Hirsch and D. Pearce. The Aurora experimental framework for the evaluation of speech recognition systems under noisy conditions. In Proc. ASR-2000, pages 181?188,September 2000.
- [30] C. A. Ynoguti, F. Violaro. Um sistema de reconhecimento de fala contínua baseado em modelos de Markov contínuos. SBrT 2000 - XVIII Simpósio Brasileiro de Telecomunicações, 3 a 6 de Setembro, 2000, Gramado-RS.
- [31] C. A. Ynoguti,. Reconhecimento de fala contínua usando modelos ocultos de Markov. Tese de Doutorado, UNICAMP, Campinas, 1999.
- [32] 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):23-41, 1992.
- [33] www.nist.gov/speech (Accessed in February, 2007).
- [34] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification. Wiley-Interscience, 2nd edition, 2001.
- [35] M.J.F. Gales. Model-Based Techniques for Noise Robust Speech Recognition. PhD thesis, Cambridge University, 1995.
- [36] L. R. Rabiner and R. W. Schafer. Introduction to Digital Speech Processing, Foundations and Trends in Signal Processing, vol 1, no 12, pp 1194, 2007.

# Referências Bibliográficas

- [1] R.P. Lippmann. *Speech recognition by machines and humans*. Speech Communication, 22(1-15), 1997.
- [2] D. Yu, L. Deng, X. He, and A. Acero. *Use of incrementally regulated discriminative margins in MCE training for speech recognition*. In Proc. ICSLP, 2006.
- [3] S. Furui. *50 years of progress in speech recognition technology: Where we are, and where we should go? From a poor dog to a super cat*. Keynote Presentation, ICASSP 2007.
- [4] M. Gales and S. Young . *The Application of Hidden Markov Models in Speech Recognition*, Foundations and Trends in Signal Processing Vol. 1, No. 3 (2007) 195-304.
- [5] J. C. Junqua and H. Wakita. *A comparative study of cepstral lifters and distance measures for all pole models of speech in noise*. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -89,1989.
- [6] H. Hermanski Hanson, B.A. and H. Wakita. *Low dimensional representation of vowels based on all-pole modeling in the psychophysical domain*. Speech Communication, 4, (1-3), 1985,181-188.
- [7] O. Ghitza. *Auditory models and human performance in tasks related to speech coding and speech recognition*. IEEE Trans, on Speech and Audio Processing, 2,1,1994,115-132.
- [8] C.R. Jankowski, J. Hoang-Doan and R. P. Lippmann. *A comparison of signal processing front ends for automatic word recognition*. IEEE Trans, on Speech and Audio. Processing, 3, 4,1995, 286-293.
- [9] J. C. Junqua and J. P. Haton. *Robustness in automatic speech recognition*, Kluwer Academic Publishers, 1996.
- [10] P. J. Moreno. *Speech Recognition in Noisy Environments*, PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania, 1996.

- [11] M. J. Hunt, S. M. Richardson, D. C. Bateman and A. Piau. *An investigation of PLP and IMELDA acoustic representations and their potential for combination*. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -91,1991.
- [12] A. Torre, A. M. Peinado, A. J. Rubio and P. Garcia. *Discriminative feature extraction for speech recognition in noise*. Proceedings of EuroSpeech-97, Rhodes, 1997.
- [13] C. Mokbel, J. Monn and D. Jouvét. *On-line adaptation of a speech recognizer to variations in telephone line conditions*. Proceedings of EuroSpeech-93,1993.
- [14] H. Hermanski, N. Morgan and H. G. Hirsch. *Recognition of speech in additive and convolutional noise based on RASTA spectral processing*. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -94, 1994.
- [15] P.J. Moreno, R. Stern. *Source of degradation of speech recognition in telephone network*. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -94,1994.
- [16] A. Anastasakos, E. Kubala, J. Makhoul and R. Schwartz. *Adaptation to new microphones using tied-mixture normalization*. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -94,1994.
- [17] S. Furui. *Speaker-independent isolated word recognition using dynamic features of speech spectrum*. IEEE Trans, on Acoustic, Speech and Signal Processing, 34,1986,52-59.
- [18] J. Hernando and C. Nadeu. *Speech recognition in noisy car environment based on OSALPC representation and robust similarity measuring techniques*. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -94,1994.
- [19] C. Mokbel and G. Chollet. *Speech recognition in adverse environments: speech enhancement and spectral transformations*. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -91,1991.
- [20] K. Ohkura and M. Sugiyama. *Speech recognition in a noisy environment using a noise reduction neural network and a codebook mapping technique*. Proceedings of Acoustics, Speech, and Signal Processing (ICASSP) -91,1991.
- [21] F. Seide and A. Mertins. *Non-linear regression based feature extraction for connected-word recognition in noise*. Proceedings of Acoustics,Speech, and Signal Processing (ICASSP) -94,1994.



- [22] S.F. Boll. *Suppression of acoustic noise in speech using spectral subtraction*. IEEE Trans. on Acoustics, Speech, and Signal Processing, 27:113–120, 1979.
- [23] Y. Ephraim. *Statistical model based speech enhancement systems*. Proc. IEEE, 80:1526–1555, October 1992.
- [24] A. Acero. *Acoustical and Environmental Robustness in Automatic Speech Recognition*, 1993, Boston, MA, Kluwer Academic Publishers
- [25] R. Stern, B. Raj and P. J. Moreno. *Compensation for environmental degradation in automatic speech recognition*. Proceedings of ESCA-NATO Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels, 1997.
- [26] P.J Moreno and B. Eberman. *A new algorithm for robust speech recognition: the delta vector Taylor series approach*. Proceedings of EuroSpeech-97, 1997.
- [27] P. J. Moreno, P.J, B. Raj and R. Stern. *Data-driven environmental compensation for speech recognition: a unified approach*. Speech Communication, 24, 4, 1998, 267-288.
- [28] M. F. J. Gales and S. J. Young. *HMM recognition in noise using parallel model combination*. Proceedings of EuroSpeech-93, 1993.
- [29] M. F. J. Gales. *“NICE” model-based compensation schemes for robust speech recognition*. Proceedings of ESCA-NATO Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels, 1997.
- [30] S. V. Vaseghi and B. P. Milner. *Noise compensation methods for hidden Markov model speech recognition in adverse environments*. IEEE Trans, on Speech and Audio Processing, 5, 1, 1997, 11-21.
- [31] X.D. Huang, A. Acero, and H.W. Hon. *Spoken Language Processing*. Prentice Hall, 2001.
- [32] M.J.F. Gales. *Model-Based Techniques for Noise Robust Speech Recognition*. PhD thesis, Cambridge University, 1995.
- [33] J. C. Junqua. *The Lombard reflex and its role on human listeners and automatic speech recognizers*. In Proc. JASA, pages 510–524, January 1993.
- [34] D. A. Reynolds, *Channel robust speaker verification via feature mapping Acoustics, Speech, and Signal Processing*, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on 6-10 April 2003.

- [35] C. A. Ynoguti. Reconhecimento de fala contínua usando modelos ocultos de Markov. Tese de Doutorado, UNICAMP, Campinas, 1999.
- [36] Alcain, 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):23-41, 1992.
- [37] H. G. Hirsch and D. Pearce. *The AURORA experimental framework for the performance evaluation of speech recognition systems under noise conditions*. ISCAITRW ASR2000, Automatic Speech Recognition: Challenges for the Next Millenium, 2000.
- [38] C. A. Ynoguti, F. Violaro. Um sistema de reconhecimento de fala contínua baseado em modelos de Markov contínuos. SBrT 2000 – XVIII Simpósio Brasileiro de Telecomunicações, 3 a 6 de Setembro, 2000, Gramado-RS.
- [39] H. Ney. *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.
- [40] *Speech Recognition Scoring Toolkit (SCTK) Version 2.4*. Disponível em: <<http://www.itl.nist.gov/iad/mig//tools/>>. (Acesso em: 10 outubro 2011).
- [41] C. Thompson. *Smarter Than You Think: What Is I.B.M.'s Watson?*. *The New York Times*, New York 16 junho 2010. Disponível em: <<http://www.nytimes.com/2010/06/20/magazine/20Computer-t.html?pagewanted=all>>. (Acesso em: 10 outubro 2011).
- [42] *DeepQA Project: FAQ – IBM Corporation*. Disponível em: <<http://www.research.ibm.com/deepqa/faq.shtml>>. (Acesso em: 10 outubro 2011).