
Análise espectral e fonética do português do Brasil

Vistas às dificuldades de reconhecimento da fala em portadores de implantes cocleares

Aluno:

Sergio Aguirre

Orientador:

Stephan Paul, Dr. Eng.

19 de janeiro de 2016

1 Introdução

A necessidade de especificar particularidades e dificuldades no reconhecimento de fala em usuários de implante coclear em português do Brasil está vinculada às estratégias e/ou técnicas de processamento de sinais a serem implementadas e otimizadas. Assim, é apresentada uma revisão bibliográfica que busca relacionar o campo de avaliação de reconhecimento de fala com análises objetivas espectro-temporais e dificuldades auditivas oriundas das particularidades dos fonemas do português brasileiro, em utilizadores de implante coclear.

A tarefa está subdividida como demonstra a Figura 1.1.

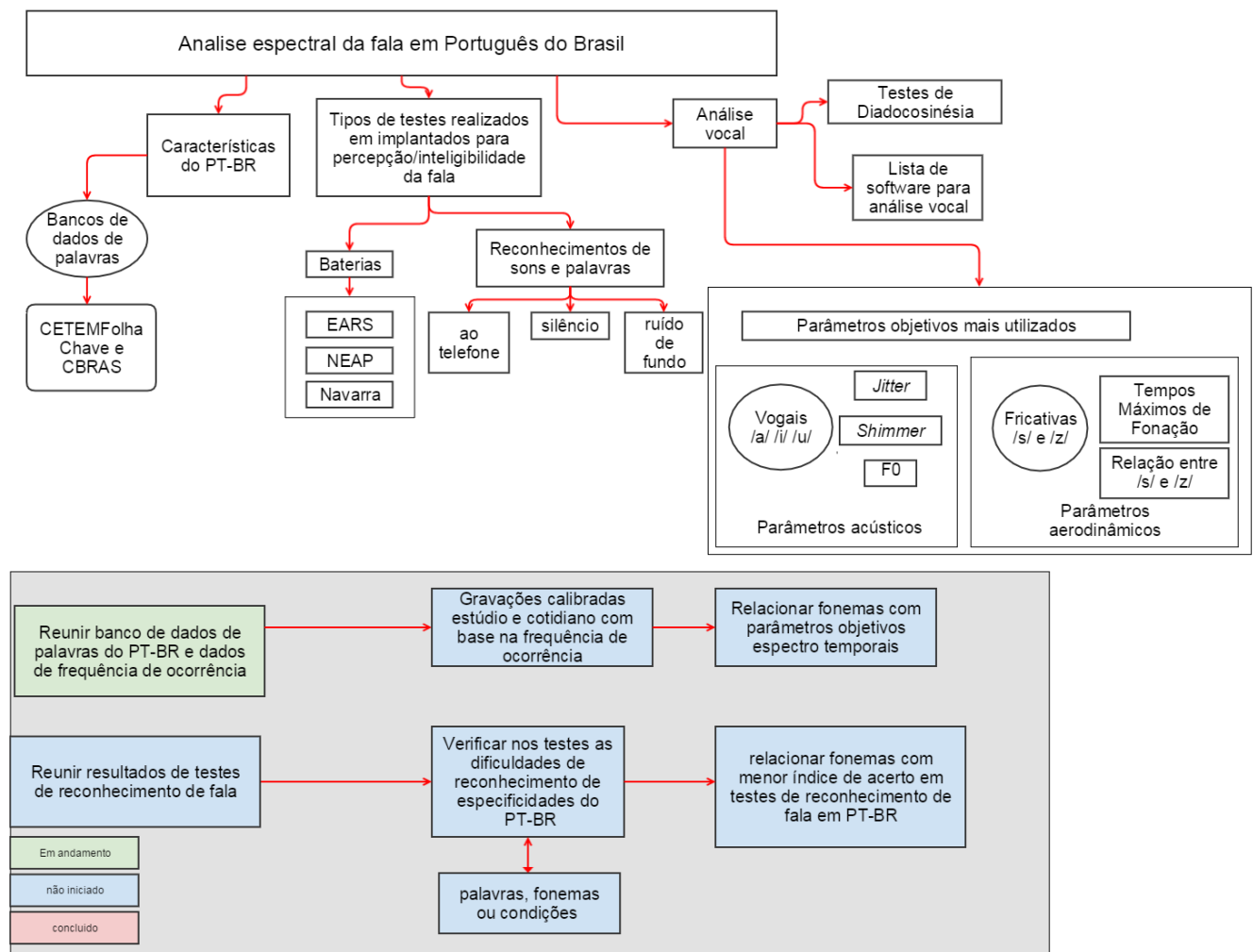


Figura 1.1: Organização inicial da pesquisa bibliográfica de análise espectral e temporal de propriedades do português brasileiro aplicado ao reconhecimento de fala.

A busca por características do português brasileiro, é iniciada através do estudo de *corpus*¹ disponíveis com a classificação e organização por frequência de ocorrência das palavras. A partir da classificação morfológica² tem-se o intuito de realizar um estudo dirigido ao reconhecimento de fala e seus problemas com fonemas mais correntes no português brasileiro.

Paralelamente estuda-se métodos de avaliação de reconhecimento de fala, desde testes individuais até os conjuntos de testes, denominados baterias. Tais baterias são configuradas como um grande estudo multidisciplinar e podem ser adequadas para línguas diferentes das quais são originalmente propostas.

A intenção da pesquisa dos tipos de testes é a identificação de relações entre as dificuldades de reconhecimento de fala (em variados tipos de ambientes) por pessoas com implante coclear e parâmetros objetivos de caráter espectro temporal, especificamente para o português brasileiro, o qual não recebeu uma adaptação de uma bateria de testes até o momento. Ainda pretende-se realizar gravações acústicas de testes que tenham com finalidades distintas, como de diadococinesia³, buscando também relação com parâmetros objetivos.

A partir de tal levantamento de dados pode-se auxiliar na estratégia de processamento de captação e transdução para usuários de implante coclear. Além disso é possível montar a proposta de um teste ou do início de uma proposta para adaptação de uma bateria de testes específicos para utilizadores de implantes cocleares falantes do português brasileiro.

Gravações calibradas serão realizadas com o propósito de auxiliar na fase de implementação de diferentes estratégias de processamento, cruzando informações de nível de pressão sonora (NPS), *loudness*⁴, com características espectrais dos fonemas. Ainda é possível relacionar as características espectrais e temporais com características dos fonemas, *e.g.* modo de articulação, sonoridade, e papel das cavidades (oral ou nasal).

2 Classificações e características do português Brasileiro (PT-BR)

esta seção apresenta

Produção dos sons

Para contextualizar as classificações é importante que a produção da fala seja apresentada, sendo que, segundo Medeiros *et al.* [1], a voz é produzida pelo trato vocal, a partir de um som básico gerado na laringe. A laringe localiza-se no pescoço e é um tubo alongado, no interior do qual ficam as pregas vocais que são formadas por músculos e mucosa, em posição horizontal. A produção da voz se dá quando no processo de expiração (fluxo de ar egressivo)

¹explicar corpus

²estudo da estrutura, da formação e da classificação das palavras

³Segundo [1] é a capacidade que uma pessoa tem em realizar movimentos rápidos alternadamente, na fala a repetição de sílabas é utilizada.

⁴Grandeza associada a sensação básica de volume sonoro [sone].

as cordas vocais aproximam-se e vibram. Vibrações mais rápidas geram sons mais agudos enquanto vibrações menos rápidas produzem sons mais graves.

Sons produzidos com as pregas vocais afastadas parcialmente, segundo Seara *et al.* [?], sem restrições de ar ao longo da laringe são denominados sons não vozeados ou surdos.

O aparelho fonador é apresentado na Figura 2.1

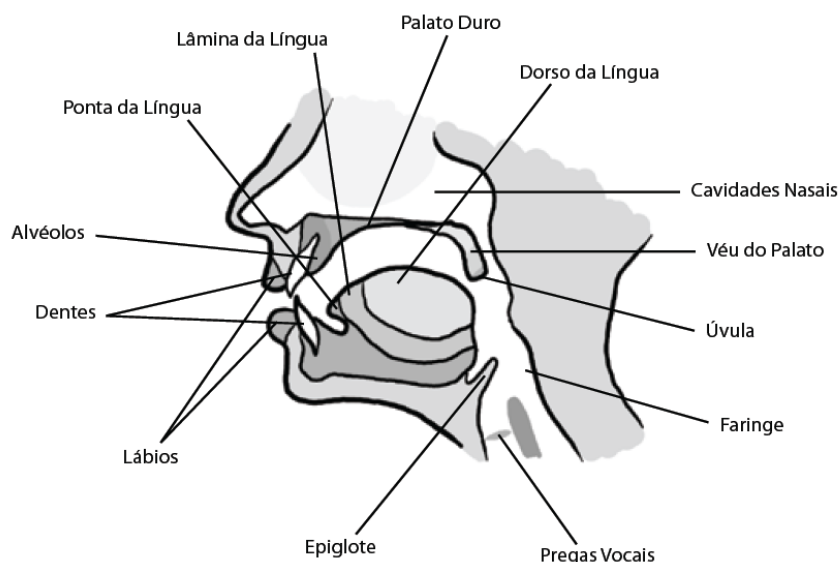


Figura 2.1: Esquema dos órgãos articulatorios do aparelho fonador humano, retirado de Seara *et al.* [?].

Para o estudo de fonemas a partir dos movimentos articulatorios, pode-se separar o estudo entre vogais e consoantes.

O véu do palato apresentado na Figura 2.1 é o articulador responsável pelas características fonéticas de vogais, nasais ou orais.

Classificações quanto à produção:

- | | |
|--|--|
| <ul style="list-style-type: none"> • Das vogais <ul style="list-style-type: none"> – Altura: alta, média alta, médias baixa e baixa – Língua: anterior, posterior e central – Lábios: arredondados e não arredondados | <ul style="list-style-type: none"> • Das consoantes <ul style="list-style-type: none"> – Vozeamento: – Ponto de articulação: – Modo de articulação: |
|--|--|

2.1 Caracterização morfológica

A classificação morfológica dá-se na forma das classes gramaticais (substantivo, verbo, advérbio, pronome, numeral, preposição, conjunção, interjeição, artigo e adjetivo). Ainda é possível tratar a classificação quanto ao número de sílabas, ao acento e ao padrão silábico. Tal classificação é importante, no tocante a frequência de ocorrência, para a compreensão das características da fala comum no PT-BR,



2.2 Classificações dos fonemas

Segundo a *International Phonetic Association* [2], o Alfabeto Fonético Internacional (AFI) ou (IPA)¹ tem como objetivo desenvolver o estudo da fonética e as aplicações práticas dessa ciência. Assim a partir da sua criação em 1886 a associação desenvolve uma representação escrita dos sons, primariamente baseado no alfabeto romano e com adição de letras e símbolos de outras variedades. É composto atualmente² por 107 letras, 52 diacríticos, e quatro marcas de prosódia.

A Figura 2.2 apresentada como exemplo as classificações de fonemas consoantes baseado no (IPA)

Símbolos fonéticos do IPA para consoantes (pulmônicas)

	Bilabial	Labiodental	Dental	Alveolar	Palatoalveolar	Retroflexa	Palatal	Velar	Uvular	Faringal	Glotal
Oclusiva	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Vibrante				r					ʀ		
Tepe		ʋ		ɾ		ɽ					
Fricativa	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Fricativa lateral				ɬ ɮ							
Aproximante		ʋ		ɹ		ɻ	j	ɰ			
Aproximante lateral				l		ɭ	ʎ	ʟ			

Figura 2.2: Esquema dos órgãos articulatórios do aparelho fonador humano, retirado de Schwindt [3].

2.3 Estudo do *corpus* Brasil

Segundo Sardinha [4] para linguística o *Corpus* (conjunto de documentos) é a coleta e a exploração de dados linguísticos coletados com dado critério, com o propósito de pesquisa de uma língua ou de uma variedade linguística.

O *corpus* analisado tem um bilhão de palavras, trata-se do *Corpus Brasileiro*³ do Grupo de Estudos de Linguística de Corpus GELC, que está sediado no Centro de Pesquisas, Recursos e Informação de Linguagem (CEPRIL), Programa de Pós-Graduação em Linguística Aplicada (LAEL) da PUCSP. Disponibilizado *on-line* gratuitamente e apresenta a seguinte estatística quanto à frequência de ocorrência:

conseguir quantificação por morfologia

¹Sigla em inglês de *International Phonetic Alphabet*.

²Segundo Schwindt [3] a ultima revisão é datada de 2005.

³*Corpus Brasileiro* anotado, v.2.3, versão de 6 de agosto de 2014.

2.4 Características do PT-BR

Nesta seção são apresentadas as características referentes ao PT-BR encontradas em trabalhos na área de linguística e

3 Testes de percepção e inteligibilidade da fala

Testes de percepção e de inteligibilidade da fala são aplicados em pacientes com implantes cocleares afim de mensurar a qualidade auditiva do implantado. Existem baterias com vários testes e métricas definidas em testes internacionais que podem ser adaptados para outras línguas, diferentes da de origem do teste.

Para o português falado no Brasil em geral propõe-se testes comparativos entre ouvintes normais (ON), pacientes com implantes cocleares (IC) e pacientes com aparelho de amplificação sonora individual (AASI), utilizando conjunto de palavras a partir de critérios de escolha dos pesquisadores.

Enumerar testes e baterias discriminando-as

- Testes de sustentação de fonemas
implantes cocleares (IC) *vs* aparelho de amplificação sonora individual (AASI) *vs* ouvintes normais (ON) []
 - listas de reconhecimento de vocábulos monossílabos e dissílabos
 - listas de reconhecimento de sílabas sem sentido
 - listas de reconhecimento de sentenças
 - lista de palavras
- Processamento digital
 - Extensão máxima de frequência vocal (EMFV)
 - Frequência fundamental vocal F_0
 - Variabilidade de F_0
 - Extensão dinâmica
 - Índice de fonação suave (SPI)
 - Intensidade objetiva vocal falada (média leitura voz alta falada de 70 dB NPS)
 - Variabilidade da amplitude

Baterias de testes...



3.1 Conjunto de palavras e gravações

A partir da análise quantitativa do *corpus* da Seção 2.3 pode-se avaliar quantitativamente os fonemas com maior frequência de ocorrência, para adaptação de listas de palavras e realizar gravações calibradas (com referência em nível absoluto de pressão) para testes de reconhecimento de palavras em pessoas com implantes cocleares. Pode-se ainda comparar os resultados entre as técnicas de processamento utilizadas nos implantes, relacionando com o NPS e *loudness*.

4 Análise vocal

Esta seção visa reunir conceitos relativos à transdução, análise e processamento do sinal de voz para o processamento digital, observando relações com fonemas do PT-BR.

Para a área de processamento vocal (*speech processing*) há diferentes segmentos que se complementam, assim uma pequena revisão é apresentada, com intuito de reunir conceitos importantes para compreensão das dificuldades inerentes ao português falado no Brasil para que haja um aprimoramento no processamento específico para a língua.

4.1 Processamento vocal para identificação de patologias

4.2 Processamento vocal de reconhecimento de voz

Os avanços dos procedimentos de reconhecimento de sinais analógicos de voz por dispositivos eletrônicos através de processamento digital podem ser empregados na qualificação das estratégias de processamento para o implante coclear.

Segundo Fulop *et al.* [5], em dispositivos eletrônicos de reconhecimento de voz a transdução do sinal contínuo (voz falada) para uma quantidade muito pequena de bits equivalente a um vocabulário extenso associado àquele tipo de sinal com as equivalentes predições baseadas em sintaxe, morfologia e frequência de ocorrência, são derivadas de técnicas de processamento digital de sinais.

4.2.1 Análises espectro temporal

Segundo Fulop *et al.* [5], a compreensão do discurso e da variedade de sons em linguagem pode ser realizada a partir da análise de espectro.

Para Fulop *et al.* [5], existe uma relação bem definida entre o trato vocal e as posições dos picos no espectro da fala, por exemplo, um de F1 (primeira "formante" ou ressonância) varia diretamente com a altura da língua, enquanto F2 varia com a localização da língua (anterior, posterior e central) e posicionamento dos lábios (arredondamento).

4.2.2 STFT

4.2.3 Cepstrum

Cepstral Mean Subtraction (CMS)

4.2.4 Wavlet

4.3 Processamento vocal de implante coclear

4.3.1 Estratégias de extração de características

F0 F1 F2

4.3.2 Recrutamento

4.3.2.1 Compressão banda larga e multi banda

4.3.2.2 Deslocamento de frequência

4.3.3 Efeito do ruído de fundo no reconhecimento

Para Kasturi [?] o processamento vocal com técnicas de redução de ruído para implantes cocleares que melhorem a percepção da fala são baseados em implementações de novos algoritmos que tem um desafio dado a limitação do processador do implante e da complexidade do processamento.

4.3.4 Métodos de melhoramento

4.3.4.1 *Minimum Mean Square Error* (MMSE)

4.3.4.2 *Minimum Mean-Square Error Log-Spectral* (Log-MMSE)

4.3.4.3 *Statistical-Model-based Spectral Subtraction* (SMSS)

4.3.4.4 *Multi-band Spectral Subtraction* (MBSS)

4.3.4.5 Melhoramentos adaptativos

4.4 relações entre espectro e IPA

4.4.1 Características espectrais do português brasileiro PT-BR

Referências Bibliográficas

- [1] Juliana B L Medeiros, Luciane P Spinelli, Maria G Brito, and Videres Petronila M. Alternativa de ação preventiva com educadores do projeto Paz e Cidadania nas escolas e no bairro. Anais do congresso Brasileiro de Extensão Universitária, 2002.
(Citado nas páginas)
- [2] International Phonetic Association. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. A Regents publication. Cambridge University Press, 1999. ISBN 9780521637510. URL https://books.google.com.br/books?id=33BSkFV_8PEC.
(Citado nas páginas)
- [3] L.C. Schwindt. *Manual de linguística*. Editora Vozes, 2014. ISBN 9788532648679. URL https://books.google.com.br/books?id=SGI_CgAAQBAJ.
(Citado nas páginas)
- [4] T.B. Sardinha. *Lingüística de corpus*. Manole, 2004. ISBN 9788520416761. URL <https://books.google.com.br/books?id=i8uJXgeok48C>.
(Citado nas páginas)
- [5] Sean A. Fulop, Kelly Fitz, and Douglas O'Shaughnessy. Signal Processing in Speech and Hearing Technology. *Acoustics Today*, 7(3):8+, July 2011. URL http://acousticstoday.org/issues/2011AT/Jul2011/9E5769E405C5818AC847C94FC61D2E2E/ATCODK_7_3.pdf.
(Citado nas páginas)

