

Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal de Minas Gerais

Plano de Tese

Teoria de Detecção de Sinais na Fala e Comunicação

Leonardo Carneiro de Araújo

Orientadores: Prof. Hani Camille Yehia

1 Introdução

Nas últimas décadas, vêm se tornando extremamente popular a utilização de inspirações nas habilidades humanas ou na biologia em aplicações tecnológicas. Em tecnologias da fala não poderia ser diferente. Sistemas de síntese e reconhecimento da fala são em geral construídos com base no conhecimento sobre o processamento lingüístico humano. Em muitos aspectos as abordagens existentes mostram-se bem sucedidas mas ainda bem longe de atingir um desempenho compatível com o humano. Os sistemas que atingem uma melhor performance geralmente fazem uso de grandes bases de dados, experienciam ainda o problema de serem dependentes de locutor e muitas vezes ainda requerem uma longa etapa de treinamento e ajuste de parâmetros. Talvez os problemas em tecnologia da fala possam ser resolvidos não apenas com uma melhor engenharia, mas também com uma colaboração mais intensa entre a tecnologia e a ciência cognitiva, levando a descobertas de princípios fundamentais no processamento de linguagem que podem ser muito úteis em tecnologia da fala.

Os homens possuem uma capacidade surpreendente para lidar com situações adversas no processamento lingüístico, se falharmos na tarefa de percepção da fala temos a capacidade de recuperar a informação criando hipóteses sobre o que foi dito e se encontramos novas situações, temos a capacidade de aprender e adaptar. Esta notável capacidade na percepção e compreensão nos traz motivação e inspiração para projetar novos sistemas de reconhecimento automático de fala. As investigações sobre o processamento da fala não são apenas úteis ao desenvolvimento de sistemas inteligentes de fala, mas também à própria tarefa de entender como é que se dá o processamento humano da fala, para a construção de teorias lingüísticas que podem ser utilizadas para entender a mente humana e para diagnosticar e tratar pessoas com deficiências na fala (afasia).

Neste trabalho, pretende-se investigar a percepção da fala e o processamento lingüístico sob a ótica da teoria estatística de detecção de sinais e teoria da informação. Para consolidar os modelos a serem criados, pretende-se realizar experimentos psicolingüísticos. Procurar-se-á estudar a influência das pistas acústicas na percepção dos estímulos, assim como a influência de elementos lingüísticos, como a sintaxe e o contexto para o aparecimento do significado em sentenças. Uma outra possibilidade para este trabalho seria avaliar qual é a influência da informação visual na comunicação falada e como esta informação pode ser útil no preenchimento de lacunas no processo de criação de significado.

2 Comunicação na Fala

Language is the blood of the
soul into which thoughts run
and out of which they grow.

Oliver Wendell Holmes

O processamento auditivo pode ser dividido em dois ramos principais, sob o ponto de vista de comunicação na fala: a clássica psicofísica da audição e o reconhecimento de pistas acústicas, que serão posteriormente mapeadas em informações linguísticas. O primeiro relaciona-se aos atributos físicos, psicológicos e limitações no processo de transdução auditiva dos estímulos acústicos em disparos neuronais. O segundo está relacionado à identificação, classificação e interrelação de padrões auditivos que criam significado dentro do contexto da experiência comunicativa do ouvinte.

A cognição da fala pode ser definida como o processo de se impor um significado a uma experiência perceptual da fala que até então não possuía significado algum. A psicofísica da audição busca descobrir qual o “poder de resolução” do mecanismo auditivo, para melhor entender como as variáveis dos estímulos acústicos participam na percepção do evento acústico. Algumas características dos estímulos sonoros podem então ser definidas, tais como: pitch, volume, timbre, duração e localização. Todas estas características, com exceção do timbre, são intuitivas e podem, até certo ponto, ser medidas. Não apenas esses atributos, mas também seu efeito sob variação e a inter-relação entre eles podem também, de certa forma, ser medidos. É possível determinar quais são os atributos responsáveis pela criação da qualidade do significante? Até que ponto eles determinam o processo de montagem do significado? Qual o tipo de relação existente entre eles? Todas estas questões são difíceis de ser respondidas. Faz-se necessária uma investigação sobre quais são os aspectos passíveis de criar e mudar o significante para criar novos significados.

A fala, por outro lado, é um sinal multidimensional que suscita uma associação lingüística. Para que exista um código eficiente de comunicação, algum tipo de categorização perceptual absoluta tem que ser feita a partir de seu conteúdo. O processo de audição de um som da fala é um processo perceptivo ainda mais desafiador, pois parece haver uma discrepância entre o estímulo e a experiência do ouvinte com relação a ele. Através do mecanismo cognitivo da fala como um todo, há uma transformação de uma entrada acústica contínua em uma mensagem auditiva discreta composta de palavras, frases e sentenças. Os elementos perceptivos da fala levariam a uma percepção final através de um caminho ainda não bem definido, levando, em

última instância, a uma imagem mental que cria significado e torna a comunicação possível. Esses elementos perceptíveis podem ser pistas acústicas, tais como a localização dos formantes¹, a duração do som, a tonalidade, etc. Essas pistas possuem influência direta na percepção da fala. Por exemplo, a localização dos formantes pode ser utilizada para encontrar quais são as vogais mais prováveis, embora já tenha sido mostrado que a contaminação causada nos sons adjacentes seja ainda mais relevante nesta determinação e categorização da vogal [4].

O estímulo de entrada é entendido como um fluxo contínuo de sinal acústico (e também movimento facial e gestual quando em uma comunicação face-a-face). Por um outro lado, a fala percebida é composta por uma mensagem auditiva discreta. Existe então uma transformação de um mundo contínuo externo em um mundo discreto interno. “Para identificar qual o papel da estrutura linguística dentro da totalidade da linguagem devemos primeiramente considerar o ato individual da fala e traçar o que ocorre no circuito da fala do indivíduo. Este ato supõe ao menos dois indivíduos, o mínimo elegível para tornar o circuito completo” [1]. Esse processo de transformação entre os mundos discreto e contínuo acontece duas vezes em qualquer processo de comunicação; primeiramente no transmissor, quando criando o sinal acústico da fala e, posteriormente, no receptor, quando percebendo a entrada acústica da fala. Para que a mensagem seja transmitida corretamente, é necessário evitar qualquer erro que possa ocorrer em qualquer uma das três partes do processo: transmissor, meio e receptor. O sucesso da transmissão da mensagem é medido continuamente através da realimentação, em primeira instância, do falante consigo mesmo, corrigindo eventuais erros de locução, por exemplo e, posteriormente, através da realimentação fornecida pelo ouvinte receptor da mensagem. Para uma eficiente transmissão de mensagens estas realimentações devem ser consideradas.

Quais são as pistas que retêm nossa atenção e fazem a conexão entre dois mundos aparentemente díspares? A maior parte dos pesquisadores na área tentam realizar a separação em pistas dinâmicas e pistas estáticas. Foi mostrado que as pistas estáticas são mais efetivas em influenciar a percepção da fala[5]. A localização dos formantes, a distribuição espectral do som e a forma da boca no início de um segmento são exemplos de pistas estáticas. A transmissão de energia entre uma consoante e a vogal seguinte seria um exemplo de pista dinâmica. Embora tenha sido demonstrado que a percepção da fala depende primariamente de outras propriedades dinâmicas de mais alta ordem, as pistas estáticas são as mais utilizadas nas teorias atuais.

¹Formantes são bandas de energia no sinal acústico da fala relacionados à configuração do trato vocal.

3 Lingüística

Em sua teoria sobre a linguagem, Ferdinand de Saussure preocupa-se com a correspondência entre palavras e idéias. “No cérebro (...), onde os fatos da consciência, que chamaremos de conceitos, encontram-se associados com as representações de signos lingüísticos ou imagens acústicas que são usadas para expressá-los.” De acordo com Saussure, o signo lingüístico cria uma ligação associativa no cérebro entre o conceito e a imagem acústica ². Ele discute a dualidade como uma ligação, mudando ao longo do tempo. Saussure enfatiza a linguagem como forma, um sistema de imagens sonoras ligadas à idéias.

Saussure dividia a variável ‘palavra’ em significado e significante, e utilizava ‘símbolo’ para atribuir ao signo lingüístico o que chamamos de significado. O símbolo nunca é completamente arbitrário, ele não é vazio, existe uma ligação entre o significado e o significante. “A ambigüidade desapareceria se as três noções aqui expostas fossem designadas por três nomes, cada uma sugerindo e opondo-se às demais. Proponho manter a palavras signo para designar o todo e substituir o conceito e a imagem sonora respectivamente por significado e significante; os dois últimos termos possuem a vantagem de indicar a oposição que os separa um do outro e dos demais dos quais eles fazem parte. No que concerne ao signo, se estou satisfeito com isso, isto é simplesmente porque eu não conheço palavra alguma para substituí-la, a linguagem ordinária não sugere nenhuma outra.”

“Signos que são completamente arbitrários realizam melhor que outros o ideal do processo semiológico; é por isto que a linguagem, o mais complexo e universal sistema de expressão, é também o mais característico; neste sentido a lingüística pode tornar-se o padrão principal dentre todos os ramos da semiologia, embora a linguagem seja apenas um sistema semiológico em particular.

O símbolo da palavras vem sendo utilizado para designar o signo lingüístico ou, mais especificamente, o que é aqui chamado de significante. Eu particularmente sou fortemente contra a utilização desse termo. Uma característica do símbolo é que ele nunca é completamente arbitrário; ele não é vazio, pois existe uma ligação rudimentar entre significante e significado. O símbolo de justiça,

²“Não é o som material, apenas uma coisa física, mas uma impressão psíquica deste som, sua representação nos dá testemunho de nossos sentidos; tal é uma imagem sensorial e, se a chamamos ‘material’, é apenas nos sentidos, e opondo-se ao outro termo associativo, o conceito, usualmente mais abstrato.”

um par de balanças, não poderia ser substituído por qualquer outro símbolo, tal como um tanque” [1].

É objetivo nosso estudar o processo de comunicação, quer apenas o processo de produção ou de percepção, ou ambos. Em uma primeira vista, é necessário entender quais são os processos realizados durante esta conversão entre esses dois mundos díspares. Isto é necessário para o desenvolvimento de uma teoria da linguagem. Freud possuía a sua própria teoria em seus estudos sobre a afasia “Zur Auffassung der Aphasien”. Freud fala de ‘representações de palavras’ como um conjunto de representações que essencialmente envolvem a imagem sonora da fala. “A representação de palavras é mostrada como um complexo fechado de representações, enquanto a representação de objetos é aberta. A representação das palavras não é ligada à representação de objetos por todos os seus constituintes, mas apenas pela sua imagem sonora. Da associação do objeto, a imagem visual representa o objeto de uma maneira similar a esta em que a imagem sonora representa a palavra. As conexões associativas entre palavras (ao invés da associação com a imagem sonora) são representadas por linhas pontilhadas, as conexões da imagem acústica da palavra com associações de objetos não visuais não estão mostradas” [2].

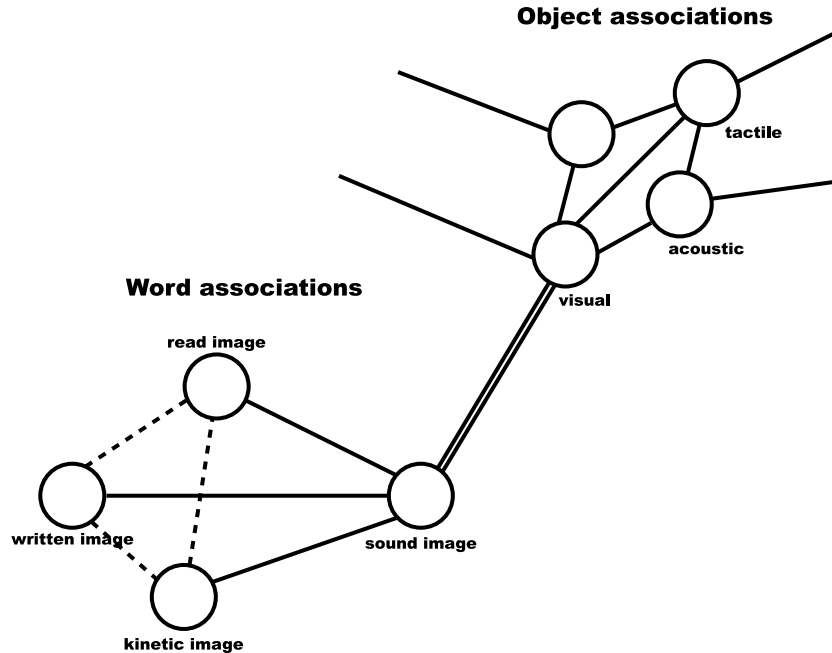


Figura 1: Apresentação das palavras.

4 Teoria de Detecção de Sinais

A Detecção de Sinais está presente em nossa vida cotidiana. Suponha que você esteja em uma cafeteria quando avista alguém que lhe parece familiar. Surge então a dúvida: será que você realmente conhece a pessoa? Existe uma incerteza a cerca da presença ou não do sinal. Será que você deveria se levantar e ir cumprimentar a pessoa correndo o risco de passar por uma situação constrangedora ao perceber que na verdade você não a conhecia? Ou seria melhor fingir não ter visto a pessoa correndo o risco de ofender um amigo? Ambas ações possuem um custo e benefício potenciais. A decisão correta não é clara. Indo além, podemos dizer que a decisão que você tomará pode estar baseada em experiências passadas. Se você acaba de passar por uma situação embaraçosa de acenar para um estranho, é mais provável que em uma próxima ocasião similar você busque evitar este mesmo tipo de situação indesejável.

Suponha que você vá a um radiologista fazer uma tomografia computadorizada (TC). Interpretar a imagem da TC é uma tarefa difícil e requer muita experiência. Nesta análise, existe sempre uma incerteza associada: ou existe um tumor (o sinal está presente) ou não existe (o sinal é ausente). Existem quatro possíveis resultados: um acerto (o tumor está presente e o médico disse ‘existe um tumor’), um erro (o tumor está presente e o médico diz ‘não existe tumor’), um falso alarme (o tumor não está presente e o médico diz ‘existe um tumor’), ou uma rejeição correta (o tumor não está presente e o médico diz ‘não existe tumor’). Existem apenas dois bons resultados sob o ponto de vista de detecção do sinal: acertos e rejeições corretas. Do outro lado existem dois resultados ruins: falsos alarmes e erros. Os dois exemplos mostrados³ são testes de detecção, em que o sujeito do teste deve decidir se o sinal está presente ou não. Para realizar a tarefa, deve-se levar em consideração as condições interrelacionadas em que podem ocorrer as diversas situações.

Existem diversas formulações para a Teoria de Detecção de Sinais (SDT, Signal Detection Theory) e que servem a diferentes propósitos. Ela pode, por exemplo, ser utilizada como uma técnica de análise de dados. “A descrição mais simples sobre a performance do observador em uma tarefa de detecção, tal como a probabilidade de fornecer respostas corretas ou o número de vezes que um sinal é detectado quando apresentado, constituem uma visão simplificada e imperfeita do que está acontecendo. Uma análise baseada em Modelos de Detecção de Sinais fornece uma medida melhor de tais coisas como a facilidade com que o observador detecta o sinal”[13]. A SDT pode também ser

³Ambos exemplos citados não foram originalmente propostos pelo presente autor.

utilizada na interpretação de diferenças entre diversas condições experimentais. A descrição teórica de dados pode levar em consideração os diferentes mapeamentos feitos devido às diferentes condições experimentais utilizadas. A SDT fornece uma organização para coleções de muitos dados e sugere como as mudanças procedimentais manifestam-se na performance. Uma terceira aplicação para a SDT, que será o foco neste trabalho, ocorre quando ela “pode ser tomada como um modelo psicológico, isto é, uma descrição do processo atual que um observador vivencia quando está tomando uma decisão. (...) Ela fornece uma explicação plausível de como as informações podem ser combinadas para se tomar uma decisão. (...) Como qualquer descrição de um processo psicológico, os modelos de detecção de sinais são na melhor das hipóteses idealizações. (...) No entanto, qualquer dúvida sobre a teoria de detecção de sinais como uma descrição psicológica não reduz o seu valor como uma maneira para se medir performance e organização de dados.” [13]

5 Detectabilidade do Sinal

A teoria da detectabilidade do sinal lida com o seguinte problema: julgar em uma observação o que é sinal e o que é ruído. No artigo “The Theory of Signal Detectability” de Petersom, Birdsall e Fox [8], é apresentado o problema em que a um observador são dadas amostras de uma tensão variando no tempo, durante um determinado intervalo de observação, e a ele cabe decidir se o que é observado é um ruído ou uma combinação de sinal e ruído. Para responder a esta questão, qual método deve o observador utilizar, e qual seria o receptor que é uma realização desse método? A Teoria da Detecção de Sinais fornece maneiras para quantificar a capacidade de se discernir entre sinal e ruído, partindo do ponto em que praticamente toda inferência e tomada de decisão são realizadas na presença de incertezas. O conceito subjacente é similar à razão sinal-ruído quando é conveniente separar eventos importantes do ruído de fundo.

A tarefa de detecção está sempre susceptível a uma incerteza criada pela presença de ruídos, que podem ser distinguidos em dois tipos com relação à sua origem: ruído interno e ruído externo. O ruído externo refere-se à qualquer sinal presente que interfira no sinal almejado. No exemplo da tomografia computadorizada, podemos ter fatores ruidosos que são parte do processo fotográfico ou algo presente nos tecidos da pessoa que venham a assemelhar-se com um tumor. Podemos tentar ficar livres desses ruídos externos, mas provavelmente é impossível eliminá-los completamente. O ruído interno refere-se às características ruidosas do processo de detecção para determinar a probabilidade de existir um sinal ou não. No exemplo da tomo-

grafia computadorizada, o ruído interno é associado ao julgamento do médico sobre a existência ou não de um tumor. A resposta interna é inerentemente ruidosa.

Toda tarefa de detecção é basicamente uma questão do tipo SIM ou NÃO, significando que ‘o sinal está presente’ para a resposta SIM e ‘o sinal não está presente’ para a resposta NÃO. O observador experimenta dois tipos de tentativas, para as quais ele responderá SIM ou NÃO: *tentativas de ruído* e *tentativas de sinal* (também chamadas *tentativas de sinal mais ruído*). O primeiro tipo representa as tentativas em que existe sinal. O outro tipo representa as tentativas em que o sinal está presente.

Sob uma primeira vista, parece fácil avaliar o desempenho de experimentos deste tipo. Poderíamos medir quão boa foi a performance do observador simplesmente contando o número de vezes em que o sinal é detectado. Este tipo de evento é chamado *acerto* (*hit*): dizer SIM a um sinal. A proporção de acertos, a *taxa de acertos* (*hit rate*), h , é a razão dada entre a frequência de acertos e a frequência de tentativas com sinal.

$$h = \frac{\text{Número de respostas SIM a sinais}}{\text{Número de julgamentos com sinais}}. \quad (1)$$

Gostaríamos então de maximizar h para melhorar o desempenho do observador.

A abordagem acima é sem dúvida ingênua. Para aumentar h , o observador poderia simplesmente responder SIM a todos os julgamentos, levando assim h ao seu valor máximo (um) mas, por outro lado, o observador estaria julgando que um sinal está presente toda vez em que ele não está, o que é chamado de *falso alarme*. Obviamente não queremos um observador que nos forneça muitos falsos alarmes. O que queremos realmente é maximizar a taxa de acertos e minimizar a taxa de falsos alarmes. A *taxa de falsos alarmes*, f , é definida como a razão entre frequência de falsos alarmes e a frequência de tentativas com sinal.

$$f = \frac{\text{Número de falsos alarmes}}{\text{Número de julgamentos com ruído}}. \quad (2)$$

Estas duas verossimilhanças descritas acima são convencionalmente utilizadas na Teoria de Detecção de Sinais. As outras duas são a *taxa de erros* e a *taxa de rejeições corretas*, mas estas são redundantes, pois são dadas por aquelas:

$$\text{taxa de erros} = 1 - h, \quad (3)$$

$$\text{taxa de rejeições corretas} = 1 - f. \quad (4)$$

As quatro verossimilhanças descritas acima são apresentadas abaixo na forma de uma tabela, que mostra uma visão completa das verossimilhanças. Existem dois tipos de julgamentos, ruído ou sinal, e há dois tipos de resposta, SIM ou NÃO. Como cada tipo de resposta pode ocorrer para cada tipo de estímulo, existem quatro possíveis resultados que são identificados pelos seus nomes.

Tipo de julgamento		Resposta	
		NÃO	SIM
	Ruído	Rejeição Correta	Falso Alarme
Sinal	Erro	Acerto	

Uma outra terminologia também é utilizada para designar as idéias expressas aqui. Acertos e falsos alarmes são chamados *positivos verdadeiros* e *positivos falsos*, e rejeições corretas e erros são chamados *negativos verdadeiros* e *negativos falsos*. Na literatura de epidemiologia, a *sensitividade* de um teste é a sua taxa de acertos, e a *especificidade* é a taxa de rejeições corretas.

Uma questão central na visão estatística do processo de detecção é a utilização de conceitos estatísticos chamados ‘função de verossimilhança’ e ‘curvas ROC’ (na qual ROC vem do inglês *Receiver Operating Characteristic*, característica de operação do receptor). A verossimilhança funciona no sentido oposto da probabilidade. Se a probabilidade permite prever resultados desconhecidos baseando-se em parâmetros conhecidos, então a verossimilhança permite determinar os parâmetros desconhecidos baseando-se em resultados conhecidos. De forma matemática, dado o evento B , utilizamos a probabilidade condicional $P(A|B)$ para inferir sobre o evento A . O problema inverso é: dado A , utilizaremos a função de verossimilhança $L(B|A)$ para inferir sobre B .

A inferência é formalizada pelo teorema de Bayes:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \quad (5)$$

A função de verossimilhança é tida como a função de densidade de probabilidade (pdf) sendo uma função do seu segundo argumento, mantendo fixo o primeiro argumento, então

$$b \longrightarrow P(A|B = b), \quad (6)$$

ou qualquer outra função proporcional a esta. A função de verossimilhança de B é a classe equivalente de funções

$$L(b|A) = \alpha P(A|B = b), \quad (7)$$

para qualquer $\alpha > 0$. O valor numérico de $L(b|A)$ é imaterial, tudo o que importa são as razões da forma

$$\frac{L(b_2|A)}{L(b_1|A)}, \quad (8)$$

já que existem invariantes sob a constante de proporcionalidade.

Quando a função de densidade de probabilidade é vista como uma função de parâmetros desconhecidos, ela é chamada de ‘função de verossimilhança’. Toda informação fica retida nos dados observados e então na PDF relacionada. Devemos esperar então uma acurácia na estimação que depende fortemente da sua PDF. Não devemos esperar obter uma boa estimativa para os parâmetros se a PDF depender pouco deles. Espera-se observar um bom desempenho na estimação dos parâmetros o quão mais a PDF for influenciada por eles. Intuitivamente, a ‘agudez’ da função de verossimilhança determina a precisão em que podemos estimar os parâmetros desconhecidos.

As curvas ROC são gráficos da sensibilidade contra $1 - \textit{especificidade}$. Esta pode também ser determinada pelo gráfico de verdadeiros positivos (TP) contra a fração de falsos positivos (FP). As curvas ROC foram desenvolvidas nos anos 1940 como um subproduto da pesquisa em extração de informação de sinais de rádio contaminados por ruídos. Após o ataque a Pearl Harbor, a pesquisa foi motivada pelo desejo de determinar como os radares americanos ‘operadores receptores’ não foram capazes de detectar os aviões japoneses. Na década de 50, tornaram-se úteis em psicofísica, para estudar a detecção humana de sinais. Ficou claro que eram bastante úteis no campo de tomada de decisões médicas e, mais recentemente, mostraram ser úteis na avaliação de performance de técnicas de aprendizado de máquina. Eles também são muito utilizados em epidemiologia e pesquisas médicas.

A preocupação com a detectabilidade de sinais e as suas terminologias surgiram com o problema dos radares referido acima. Um tratamento formal para a teoria iniciou-se nos anos 50 pelos matemáticos e engenheiros na Universidade de Michigan, Peterson, Birdsall e Fox [9], e em Harvard e no Instituto de Tecnologia de Massachusetts (MIT) por Van Meter e Middleton. Em 1966, John A. Swets e David M. Green criticaram os métodos tradicionais em Psicofísica pelas suas incapacidades de discriminar entre a real sensibilidade dos sujeitos e suas (potenciais) tendências de respostas. Munson e Karlin [6] conduziram experimentos psicofísicos que levaram ao reconhecimento da importância da teoria da detecção para o observador humano, sugerindo uma teoria similar para o processo humano de detecção.

Como mostrado por Green [3], a Teoria da Detecção de Sinais é uma combinação de duas distintas estruturas teóricas: (i) teoria estatística de decisão, como desenvolvida inicialmente por Wald [12], Neyman e Pearson

[7]; (ii) teoria do observador ideal iniciada por Siegert[11]. A teoria da decisão é utilizada como uma instância para se testar uma hipótese estatística a fim de realizar a tarefa de detecção. Isto torna possível lidar efetivamente com o problema psicofísico de controlar e medir os critérios para discernir a existência de sinais que são empregados pelo observador. A teoria do observador ideal cria uma ligação entre a performance da detecção atingida pelo observador real e a performance da detecção matematicamente ideal.

6 Teoria da Informação

A Teoria da Informação é uma teoria matemática que se inicia com determinadas premissas que definem os aspectos da comunicação e, a partir daí, atinge-se várias conclusões lógicas. A Teoria da Informação define entropia e explica o seu papel na recuperação de informação em mensagens que foram corrompidas por ruído. Esta formulação em Teoria da Informação foi desenvolvida de forma abstrata, baseada em certas estruturas algébricas, permitindo assim a sua aplicação em diferentes áreas.

A linguagem natural possui uma estrutura bem complexa e bem diferente das linguagens artificiais. Apesar das diferenças estruturais, a linguagem natural realiza várias funções da mesma forma que outros códigos e, em princípio, admite o mesmo tratamento teórico. O código da linguagem natural é a sua gramática, que compreende a fonologia, fonotaxe, morfologia e sintaxe, estas formam as regras que regem a formação de sentenças possíveis. Estes aspectos da linguística são bem conhecidos e bem modelados matematicamente. Para o cálculo da entropia é mais conveniente tratar a gramática como uma cadeia de Markov, o que infelizmente ignora alguns aspectos importantes da gramática. Uma vez calculada a entropia, podemos utilizá-la para estabelecer um limite na probabilidade de erro de decodificação para um decodificador ótimo utilizado em conjunto com um classificador acústico, caracterizado pela sua matriz de confusão e representado por um canal ruidoso.

A Teoria da Comunicação desenvolvida por Claude Shannon não aborda os aspectos do sentido semântico de uma mensagem e nem os efeitos desta sobre o sujeito receptor. A noção de *informação* dada por Wilbur Schramm é bem confortável ao nosso conhecimento de mundo. Segundo ele, *informação* é simplesmente qualquer coisa que tenha importância ou que faça diferença. Para Shannon, no entanto, existe uma definição técnica para *informação* que não relaciona *informação* com a idéia de significado. Segundo ele, “os aspectos semânticos da comunicação são irrelevantes para os aspectos de engenharia”[10]. Para Shannon, *informação* refere-se à capacidade de se

reduzir a incerteza. O conceito de entropia foi tomado por Shannon da segunda lei da termodinâmica, que estabelece que a entropia de um sistema isolado tende a aumentar com o tempo, aproximando a um valor máximo de equilíbrio, ou seja, o universo caminha de um estado organizado em direção ao caos, movendo-se de uma condição previsível para uma incerta. A quantidade de informação contida em uma mensagem é capacidade que esta possui de combater a entropia, tornando-se menos imprevisível.

O modelo proposto por Shannon e Weaver considera a comunicação como um sistema em que o fluxo de informação se dá em apenas um sentido: do transmissor para o receptor. Sabemos no entanto que a comunicação humana é estabelecida através de um sistema em que existe um *feedback*, a resposta obtida, quer seja oral, facial ou até mesmo o silêncio, constituindo um elemento que é utilizado pelo transmissor da mensagem para adaptar a sua maneira de transmitir determinada informação. Norbert Wiener preocupou-se com a tentativa humana de controlar a entropia através de *feedback* (realimentação), fazendo uma comparação como o que acontece na comunicação entre máquinas. A realimentação é uma maneira de introduzir aprendizado no sistema, algo que não foi contemplado pelo modelo de Shannon e Weaver.

7 Plano de Trabalho

Segue abaixo uma breve descrição dos itens que constituem o plano de trabalho:

1. Revisão bibliográfica

Inicialmente deve ser realizada uma revisão bibliográfica contemplando os assunto deste projeto de doutorado. Esta revisão consistirá em buscar e analisar livros e periódicos nacionais e internacionais das áreas que este projeto abrange.

2. Estudos mais aprofundados em Lingüística, Psicolingüística, Teoria da Informação e Detecção de Sinais

Estes estudos serão realizados através de disciplinas a serem cursadas no curso de pós-graduação em Engenharia Elétrica, no curso de pós-graduação em Estatística e no curso de pós-graduação em Lingüística. Além disso, muito do conhecimento necessário não contemplado por estes cursos deverão ser adquirido pelo aluno através de estudos autônomos.

3. Elaboração de experimentos psicolingüísticos

Parte fundamental deste projeto será a elaboração de experimentos de forma a poder corroborar a aplicação de teorias matemáticas sobre

o foco da comunicação falada. Para a aplicação desses experimentos será necessário softwares apropriados, como o DMDX, Praat, dentre outros, e quando necessário deverão ser desenvolvidas aplicações em MatLab, C, C++, Java, etc. Um primeiro e importante experimento a ser realizado está escrito ao final desta secção.

4. **Análise dos resultados**

Após realizados os experimentos, deveremos analisar os seus resultados, selecionar as amostras válidas, detectar qual foi o problema ocorrido com aquelas classificadas como não-válidas, e avaliar estatisticamente os resultados, para tanto realizando medidas estatísticas, testes de hipótese, etc.

5. **Utilização dos resultados obtidos tendo por base a Teoria da Informação e Teoria de Detecção de Sinais**

Outra etapa muito importante do projeto será analisar os dados obtidos sob a ótica da Teoria da Informação e da Teoria de Detecção de Sinais. Nesta etapa será possível mostrar como essas teorias podem ajudar a entender o processamento e percepção da fala, explicando diversos resultados observados. Além disto esta etapa será útil para a construção de modelos que expliquem a percepção humana da fala e que sejam posteriormente utilizados para aperfeiçoar sistemas de compreensão da fala.

6. **Conclusões**

Por fim serão elaboradas as conclusões obtidas ao longo do trabalho e como resultado final será elaborada a tese de doutorado.

Um primeiro experimento, que será muito importante para o curso do projeto, tem como objetivo criar um espaço fonético onde poderemos representar a distância entre fones e fonemas de uma língua. Este espaço será importante para criar uma métrica de forma a estabelecer relações entre fones e fonemas. Além disso, a criação desse espaço será importante para a posterior utilização de teorias estatísticas e matemáticas. Ele poderá ser construído a partir de amostras reais de sinais acústicos da fala ou através de sinais sintéticos onde a suas qualidades poderão ser manipuladas, sendo assim possível estudar a relevância de determinadas pistas acústicas na categorização de um som como um fone ou outro. Para construção deste espaço é necessário estabelecer as distâncias subjetivas entre os sons. Neste processo pretende-se utilizar vários indivíduos como sujeitos que classificarão os sons em termos de suas distâncias relativas. Serão apresentados grupos de 3 sons a cada sujeitos, dentre cada grupo ele deverá classificar qual é o

par mais similar e o par mais dissimilar. Com base nos resultados de vários indivíduos para o grupo de fones da língua podemos utilizar técnicas como o MDS (multidimensional scaling) para criar o espaço de fones ou fonemas. Para o português brasileiro existem 39 fonemas, segundo Albano e Barbosa, ou seja, existem mais de 39 fones. Isto implica na existência de pelo menos 9139 combinações de fones em grupo de 3, um número muito alto para se realizar experimentos psicofísicos com seres humanos. Num primeiro estudo, seria interessante restringir o número de fones a serem analisados. Seria também importante utilizar uma plataforma on-line para os experimentos, para tornar mais fácil de se arrebatar um número maior de sujeitos para o mesmo. Espera-se que os fones que constituem um mesmo fonema possuam em geral uma relativa proximidade contrastando com a relativa distância entre fones de fonemas diferentes. Com relação ao vozeamento, lugar e modo de articulação, esperamos que essas características tenha uma ordem de importância na categorização do som igual à que foram citadas aqui, ou seja, uma fricativa alveolar vozeada estará mais próxima de uma fricativa alveolar não-vozeada do que a uma fricativa labiodental vozeada; esta, por conseguinte, estará mais próxima do que uma plosiva alveolar vozeada. Espera-se ainda mostrar quais as influências nas variações das pistas na categorização dos fones. Com estas informações, torna-se mais factível a aplicação da teoria da informação e da teoria de detecção de sinais na comunicação falada. Podemos então estudar de forma quantitativa os diversos fenômenos estudados em psicolinguística, como por exemplo a influência do contexto na percepção de elementos da fala. Por exemplo, quando um segmento fonético é substituído por um ruído, muitas vezes ele é interpretado como o segmento correto, graças à informação do contexto. Veja alguns casos de frases onde um fone foi substituído por um ruído assinalado por “*”:

- No futebol, João chutou a *ola.
- O professor pegou o aluno passando *ola.
- O cabelo dela é crespo como uma *ola.

Nestes exemplos o sujeito nem notaria a presença do ruído e perceberia a frase perfeitamente.

8 Disciplinas da pós-graduação já cursadas

Durante a graduação, o mestrado e após o mestrado cursei algumas disciplinas de pós-graduação. A lista destas segue abaixo:

1. Geometria Computacional (Pós-graduação - Eng. Elétrica)
2. Processos Estocásticos (Pós-graduação - Eng. Elétrica)
3. Elementos Finitos (Pós-graduação - Eng. Elétrica)
4. Acústica (Pós-graduação - Eng. Mecânica)
5. Identificação de Sistemas (Pós-graduação - Eng. Elétrica)
6. Redes Neurais (Pós-graduação - Eng. Elétrica)
7. Lógica Fuzzy (Pós-graduação - Eng. Elétrica)
8. Seminários (Pós-graduação - Eng. Elétrica)
9. Processamento de Sinais (Pós-graduação - Eng. Elétrica)
10. Detecção e Análise de Imagens (Pós-graduação - Eng. Elétrica)
11. Psicolinguística (Pós-graduação - Letras)

9 Disciplinas de pós-graduação a serem cursadas

1. Ciência e Tecnologia da Fala (Pós-graduação - Eng. Elétrica)
2. Detecção e Estimacão de Sinais (Pós-graduação - Eng. Elétrica)
3. Probabilidade Avançada (Doutorado - Estatística)
4. Inferência Avançada (Doutorado - Estatística)
5. Análise Multivariada (Doutorado - Estatística)
6. Linguística (Pós-graduação - Letras)
7. Semântica (Pós-graduação - Letras)

10 Cronograma

Cursar Disciplinas Revisão de Literatura	12 meses
Definição da Metodologia Utilização e Adaptação da Teoria Estatística Elaboração de Experimentos	9 meses
Preparação para Qualificação	3 meses
Estudo Aprofundado Análise dos Resultados Elaborações e Elucubrações sobre a Aplicação da Teoria	12 meses
Preparação para a Defesa de Tese	12 meses

11 Conclusão

Ao longo deste trabalho de doutorado pretende-se investigar qualitativamente e, até onde for possível, quantitativamente a aplicação da Teoria da Informação e da Teoria de Detecção de Sinais ao processo de comunicação da fala. Pretende-se aplicar experimentos com sujeitos de forma a corroborar hipóteses ou levantar novas. O curso em que se seguirá ainda não está completamente delineado, o que se pretende é que, ao longo do percurso, estas nuances tornem-se mais claras.

Referências

- [1] Ferdinand de Saussure. *Course in General Linguistics*. 1959.
- [2] Sigmund Freud. *Aphasia*. 1893.
- [3] David M. Green. Psychoacoustics and detection theory. *Journal of the Acoustical Society of America*, 32(10):1189–1203, 1960.
- [4] J. J. Jenkins, W. Strange, and T. R. Edman. Identification of vowels in “vowelless” syllables. *Perception and Psychophysics*, 34(5):441–450, 1983.
- [5] D. W. Massaro. *International Encyclopedia of the Social and Behavioral Sciences*, chapter Speech perception. Elsevier Science Ltd, 2001.
- [6] W. A. Munson and J. E. Karlin. The measurement of the human channel transmission characteristics. *Journal of the Acoustical Society of America*, 26:542–553, 1956.

- [7] J. Neyman and E. Pearson. The testing of statistical hypotheses in relation to probabilities. *Proc. Cambridge Phil. Soc.*, 29:492–510, 1933.
- [8] W. Peterson, T. Birdsall, and W. Fox. The theory of signal detectability. *IEEE Transactions on Information Theory*, 4(4):171–212, 1954.
- [9] W. H. Peterson, T. O. Birdsall, and W. C. Fox. The theory of signal detectability. *IEEE Transactions on Information Theory*, 4(4):171–212, 1954.
- [10] Claude Shannon and Warren Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.
- [11] J. F. Siegert. *Threshold Signals*. McGraw-Hill, New York, 1950.
- [12] Abraham Wald. Contributions to the theory of statistical estimation and testing hypotheses. *Annals of Mathematical Statistics*, 10:299–326, 1939.
- [13] Thomas D. Wickens. *Elementary Signal Detection Theory*. 2002.