

# ALGUNS ASPECTOS DA PERCEPÇÃO DA FALA

Este capítulo apresenta alguns aspectos da percepção da fala relevantes para o estudo em questão. Enfocamos, em especial, a complexidade envolvida na percepção da fala, a percepção de aspectos prosódicos e divergências teóricas quanto aos mecanismos perceptuais. Para desenvolver as ideias centrais deste capítulo, foram utilizados como fontes de referência Bregman (1994), Handel (1993) e Repp (1984), a quem remetemos o leitor para uma visão aprofundada sobre o assunto.

## 4.1 A COMPLEXIDADE DA PERCEPÇÃO

A produção da fala acontece a partir da integração de sons e de gestos. A percepção também é um fenômeno multimodal, pois se vale simultaneamente da audição e da visão. Entretanto, o principal sentido em que se baseia a percepção é, sem dúvida, a audição, um processo que, como veremos, é ativo e sujeito a influências contextuais.

Eventos sonoros são uma das formas básicas que utilizamos para interagir com o ambiente e são responsáveis pela compreensão da realidade espacial e temporal que nos cerca. Ao analisarmos o fenômeno da percepção sonora, nos deparamos com um fato intrigante: não há correspondência direta entre os eventos

sonoros percebidos e sua configuração física. Na história do estudo científico da fala, especialmente após o surgimento de ferramentas de medição e visualização acústica, mostrou-se complexa a tarefa de isolar na onda acústica as unidades perceptuais supostamente nela instanciadas. Como um ouvinte consegue captar da fala de seu interlocutor as informações que este procurou expressar, tornando a linguagem humana uma estratégia de comunicação eficaz?

Uma das principais razões que tornam possível a comunicação é o fato de que a percepção é um processo ativo. Além dos sentidos, que captam do ambiente os estímulos sonoros e visuais, os processos cognitivos desempenham um papel fundamental na percepção da fala e são responsáveis pela decodificação de informações a partir do sinal da fala, notadamente um fenômeno complexo.

A onda acústica é inerentemente ambígua em vista da continuidade e da sobreposição dos eventos sonoros. Como já mencionamos, não há correspondência simples e direta entre partes do sinal acústico e um segmento fonético. A fala envolve uma série de eventos acústicos que acontecem em menos de um segundo e que são unidos não sequencialmente, de forma complexa. Uma das principais fontes de complexidade na fala, responsável pela sobreposição de eventos sonoros, é a coarticulação, fenômeno decorrente da interação na movimentação dos articuladores e, conseqüentemente, entre partes sucessivas do sinal.

Apesar da complexidade do sinal da fala, podem ser identificadas regularidades tanto em termos de padrões de movimento e posicionamento dos articuladores (espaciais), quanto em termos da sequenciação dos segmentos no contínuo da fala (temporais).

Audition is a spatial sense as well as a temporal sense. [...] The ability to localize and thereby distinguish among sound sources is critical to our ability to perceive individual events. [...] time is fused with space to provide the framework for listening and that the rate of change temporally and spatially specifies the events themselves. (HANDEL, 1993, p. 174).

A percepção da fala é sensível a propriedades temporais, como taxa de articulação do enunciado, e fatores prosódicos, como a posição de um som na sentença. Os ouvintes depreendem os eventos do sinal acústico, baseando-se em propriedades da onda em um ponto do tempo e em relações entre partes da onda ao longo do tempo, a fim de gerar eventos perceptuais maximamente coerentes. É necessário, portanto, processar a onda acústica em duas dimensões temporais sobrepostas, que se referem à organização de padrões em curtos períodos de tempo e por longos períodos de tempo. Tal processamento deve ocorrer simultaneamente, uma vez que as propriedades acústicas de um mesmo evento sonoro e

entre diferentes eventos sonoros não são independentes, mas antes se influenciam de forma mútua.

The reason that one single cue is not necessary and sufficient for identification is the interactive and context-dependent nature of sound production. The attributes of a sound are not independent. The vibration patterns, rersonances, vibrato and so on are interactive; each aspect of sound production influences another. This influence spreads across a broad time span, and thus each aspect has an imprint on all others. For example, speaking rate and duration influence coarticulation and the harmonic spectrum. (HANDEL, 1993, p. 222)

A audição acontece em níveis diferentes de processamento cognitivo, por vezes utilizando todos ao mesmo tempo, por vezes, apenas um ou dois deles. Ela se baseia nas propriedades físicas dos sons, mas, em nível mais elaborado, se baseia também em propriedades dos sons que não guardam uma correspondência direta com a cadeia sonora (como brilho, *warmness* etc.). Em um nível ainda mais avançado em termos de processamento, a audição de sons equivale à percepção de objetos e está atrelada à associação do evento com significados.

Se não é possível alcançar significados sem passar pelas propriedades acústicas, por outro lado, as propriedades acústicas só são percebidas dentro da estrutura oferecida pelo significado, em uma circularidade que é característica do processamento linguístico. A fala é percebida em termos do significado das cadeias sonoras, uma vez que a relação entre os sons da fala e seu significado é arbitrária.

O objetivo central de Handel (1993) é defender que os princípios subjacentes à percepção de eventos visuais são os mesmos que subjazem eventos auditivos. Retoma então os princípios de organização do campo visual propostos pela Gestalt, estendendo-os à modalidade sonora. Objetos tendem a ser agrupados por: (1) similaridade; (2) proximidade (espacial ou temporal); (3) continuidade (elementos que seguem na mesma direção); (4) movimentação conjunta; (5) simetria e fechamento (elementos que formam objetos simétricos e fechados). Essa proposta é compatível com os modelos teóricos adotados neste trabalho, uma vez que, se a percepção da fala e percepção de eventos visuais apresentam princípios de funcionamento semelhantes, reforça-se a ideia de que a linguagem se vale de processos cognitivos gerais.

## 4.2 PERCEPÇÃO CATEGÓRICA

Ainda que, do ponto de vista da produção, a fala seja essencialmente contínua e apresente um alto grau de sobreposições, a percepção consegue isolar eventos

e relacioná-los a conceitos bem delimitados, discretos. Tal tarefa envolve a segmentação do sinal da fala e a categorização de unidades significativas.

Categorization refers to processes in which different objects and events are placed or sorted into a small number of categories. Objects and events within a single category share a common property, although they differ in other irrelevant properties. [...] Categorizing is a ubiquitous perceptual cognitive process found in all domains and in no way specific to speech. (HANDEL, 1993, p. 266).

A categorização é um processo que intermedeia toda a experiência humana com o mundo e é fundamental à capacidade linguística e à tarefa de percepção da fala. “Perception of speech sounds requires categorization. This is because different instances of the same vocal gesture inevitably vary acoustically and, as a result, a wide range of physically different sounds must nevertheless represent the same phonemic entity.” (LIEBENTHAL et al., 2005, p. 1621).

Há certos eventos cuja percepção não se dá de forma gradual, mas em que, pelo contrário, ocorrem mudanças abruptas de uma classe para outra, ainda que os estímulos ambientais, em si, sejam contínuos. Esse tipo de percepção ficou conhecido como percepção categórica.

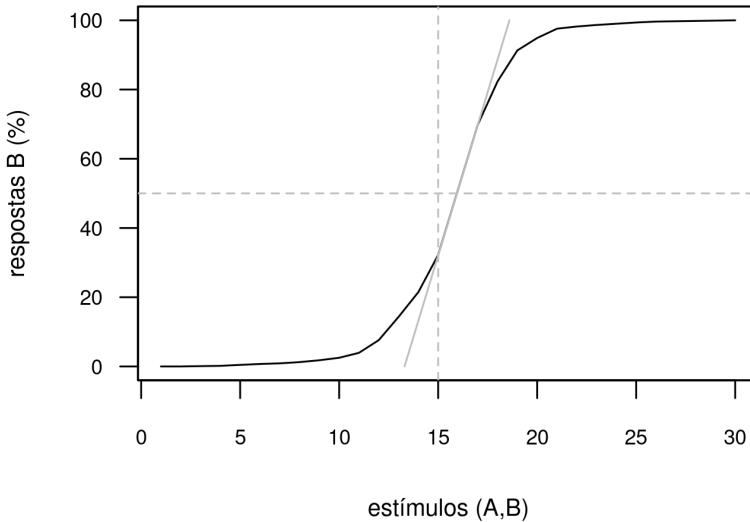
CP [categorical perception] is an important phenomenon in cognitive science because it involves the interplay between humans’ higher-level conceptual systems and their lower-level perceptual systems. [...] Another reason why CP is theoretically important is that [it] offers a potential account for how the apparently symbolic activity of high-level cognition can be grounded in perception and action. (GOLDSTONE; HENDRICKSON, 2009, [p. 1])

Considerando eventos sonoros dispostos em um contínuo, a percepção categórica consiste em estabelecer uma maior proximidade entre eventos pertencentes a uma mesma categoria e uma maior separação entre eventos pertencentes a categorias distintas. Em termos funcionais, trata-se de uma forma de interpretar a realidade quando é desnecessário processar detalhes finos, privilegiando o estabelecimento de grandes contrastes.

A percepção categórica pode ser testada por meio de experimentos de identificação, em que participantes são apresentados a estímulos interpolados entre dois sinais não ambíguos e devem classificá-los ou nomeá-los numa tarefa de escolha forçada. A partir do ajuste logístico dos resultados do teste, é traçada, para cada participante, uma curva psicométrica, que modela a probabilidade de resposta como função do parâmetro físico de interpolação dos estímulos. As curvas psicométricas consistem em gráficos que mostram a porcentagem de respostas conferidas a uma classe *versus* os estímulos apresentados aos participantes,

como mostra esquematicamente a FIG. 8, a seguir. Tal curva reflete a dinâmica de categorização dos estímulos. No caso da percepção categórica, espera-se uma mudança brusca na identificação em função dos estímulos apresentados, como se pode ver, sobretudo, entre os estímulos 11 e 19 da FIG. 8.

**FIGURA 8** – Exemplo de curva psicométrica obtida em experimento de identificação



A curva psicométrica é definida pelos parâmetros seguintes:

- a) localização do intercepto ou limiar, região onde a curva psicométrica transformada atinge o valor de 50% no eixo vertical, que corresponderia a uma resposta ao acaso. O intercepto indica o ponto no contínuo dos estímulos onde ocorre a transição de percepção entre as categorias.
- b) a inclinação no ponto do intercepto da curva transformada. A inclinação se relaciona diretamente a quão “categórica” é a percepção ou quão abrupta é a transição entre as categorias.

Na percepção da fala, operamos tanto com a percepção categórica quanto com a percepção gradual, em que os eventos são percebidos de forma contínua. Além disso, é notável que a percepção categórica não se restrinja à linguagem ou à dimensão sonora: pode acontecer na percepção de eventos sonoros e visuais linguísticos e não linguísticos.

Sabe-se que se, de um lado, a percepção categórica pode ser adquirida por meio de aprendizagem e, por outro, está presente no desenvolvimento humano desde os primeiros meses. Ao que tudo indica, consiste em um fenômeno em certa medida inato, em certa medida, adquirido. Sabe-se também que ela não é exclusiva à espécie: alguns animais exibem padrões de respostas a estímulos compatíveis

com a percepção categórica observada em humanos. Ainda se conhece pouco sobre como a percepção categórica acontece em termos cerebrais.

### 4.3 PERCEPÇÃO E ASPECTOS PROSÓDICOS

Aspectos prosódicos exercem grande influência na percepção da fala. Uma das razões para isso é a natureza inerentemente temporal da audição. “The temporal nature of the auditory world lends a special importance to time and the timing of sounds. Time provides a framework for auditory events.” (HANDEL, 1993, p. 173). Eventos sonoros, em geral, são identificados com base na forma com que os sons se iniciam e terminam, na organização temporal de sons que se repetem e na regularidade do intervalo de ocorrência entre sons.

O ritmo tem grande importância na percepção da fala (cf. DOOLING, 1974), assim como a entoação (cf. VAISSIERE, 2005). Sabe-se também que os acentos lexicais ao longo de uma sentença funcionam como facilitadores no processamento de expressões linguísticas. Cutler (1976), a partir de testes de tempo de reação, conclui que o processo de compreensão de sentenças é influenciado pela previsibilidade dos acentos seguintes, por meio de sua relação com a estrutura melódica da língua.

Cutler (2005) retoma experimentos sobre a percepção do acento e aponta que as pistas acústicas mais consensualmente aceitas como relevantes são F0 e duração: uma maior duração é sempre associada à sílaba acentuada; ademais, a percepção do acento é influenciada pela trajetória de F0 e pelo ajuste temporal de F0 dentro da sílaba, uma vez que, em línguas com acento lexical, as sílabas acentuadas funcionam como potenciais âncoras para a proeminência no nível da sentença. Não há acordo quanto à influência da intensidade, que, para alguns, seria relevante para a percepção do acento somente de forma indireta, como ênfase espectral: sílabas acentuadas apresentam mais energia nas regiões espectrais de maior frequência.

### 4.4 PERSPECTIVAS TEÓRICAS

Até a década de 70, os estudos sobre a dimensão sonora da linguagem se voltaram prioritariamente para a produção da fala. A percepção era geralmente avaliada em pesquisas ligadas à Psicologia e Psicoacústica, áreas a que devemos grande parte dos conhecimentos básicos sobre processamento e armazenagem de estímulos em geral, entre eles, os sonoros. Naquele período, uma das razões para a separação dos estudos de percepção e produção da fala talvez envolvesse mais

uma consequência teórica que uma divisão de saberes: o paradigma modularista e inatista adotado pela Gramática Gerativa tem como corolário a normalização dos estímulos na percepção da fala, que consiste na ideia de que a percepção linguística implica no mapeamento de eventos fonéticos variáveis em elementos abstratos invariáveis, na estrutura fonológica, em geral supondo a aplicação reversa das regras envolvidas na produção da fala. Nesse processo, seriam filtradas e descartadas as informações redundantes e presumidamente irrelevantes para o reconhecimento lexical, tanto linguísticas – como alofonias –, quanto extralinguísticas –, como características do locutor, do meio de comunicação e ambiente etc. Com isso, surgiram propostas de explicar a percepção da fala a partir da produção.

A ideia de que padrões acústicos levam a inferências sobre a articulação ganhou corpo com o desenvolvimento da *Motor Theory of Speech Perception* (LIBERMAN et al., 1957), segundo a qual a forma com que articulamos padrões sonoros influencia a maneira com que percebemos esses mesmos padrões na fala de outras pessoas. Para essa teoria, a percepção categórica seria exclusiva à fala – como já mencionamos anteriormente, há fortes evidências de que não é – e sua existência se deveria à produção: no contínuo da fala, há pontos de descontinuidade, como as consoantes oclusivas, que são produzidas por manobras articulatórias distintas (envolvendo, por exemplo, lábios, no caso das labiais, e dorso da língua, no caso das velares).

Especialmente na década de 80, ganharam desenvolvimento modelos psicológicos de memória e categorização que propõem o armazenamento detalhado de informações (e.g. teoria dos múltiplos traços de memória) e modelos linguísticos que rompem com a concepção de um léxico formado por elementos abstratos e discretos (e.g. teorias baseadas no uso e em exemplares). A hipótese da normalização começou a ser questionada, com base em evidências experimentais de que a repetição exerce um papel inegável na memória linguística e de que detalhes considerados irrelevantes são armazenados pelos falantes. Detalhes fonéticos exercem influência em níveis superiores do processamento linguístico (HAWKINS, 2003).

Desde então, esforços vêm sendo feitos a fim desenvolver modelos que considerem que o armazenamento e processamento da linguagem envolvem redundância e detalhamento; que o uso da linguagem é determinante para a percepção e para a produção da fala.

