AS BASES ACÚSTICA E ARTICULATÓRIA DAS TEORIAS DE PERCEPÇÃO DA FALA

Gustavo NISHIDA[1]

RESUMO: O objetivo deste artigo é apresentar, a partir de uma pesquisa bibliográfica, duas soluções articulatórias para o problema da percepção da fala. Embora ambas sejam soluções articulatórias, mostramos que os primitivos de análise perceptual são de natureza distinta. Enquanto a Teoria Motora da Percepção da Fala (LIBERMAN; MATTINGLY, 1985) propõe uma saída mentalista e modular, a Teoria do Realismo Direto da Percepção da Fala (FOWLER, 1996) decide lidar com unidades reais. Além dessas diferenças, apresentamos como os componentes acústicos e articulatórios tomam lugar nas teorias: para a primeira, como o articulatório é de natureza mental, o sinal acústico é tido como um epifenômeno; por sua vez, a segunda considera o sinal acústico como parte integrante da percepção, pois a sua proposta é antes de tudo multissensorial.

PALAVRAS-CHAVE: Primitivos da percepção da fala. Teoria motora da percepção da fala. Teoria do realismo direto da percepção da fala. Base acústica. Base articulatória.

Introdução

Desde a década de 1950, com os estudos conduzidos por Liberman e seus colaboradores[2] nos Laboratórios *Haskins*,[3] há o chamado *problema* para a percep-

1 Professor do Departamento de Comunicação e Expressão da Universidade Tecnológica Federal do Paraná, Campus Curitiba. *nishida.gustavo@gmail.com*

2 Tais estudos e reflexões estão reunidos em Liberman (1996[1957]).

3 Os Laboratórios *Haskins* são um instituto de pesquisa (privado e sem fins lucrativos) sobre fala, linguagem, leitura e suas bases biológicas. Foi fundado em 1935 e possui sede, desde 1970, em New Haven (Connecticut – EUA). Também apresenta filiação formal com a Universidade de Connecticut e a Universidade de Yale. Para mais informações sobre o instituto, visite www.haskins.yale.edu.

ção da fala. Trata-se do fato de que nesse período, sob as premissas da linguística estruturalista, tais pesquisadores buscavam descobrir quais eram as pistas acústicas que promoviam distinção fonêmica das consoantes da língua inglesa, com o auxílio de um sintetizador de fala chamado *Pattern Playback*[4] (que convertia espectrogramas desenhados à mão em som). O problema reside justamente na descoberta de que os sons da língua inglesa não apresentavam uma relação de um para um entre o sinal acústico e a articulação realizada, i.e., não havia uma única característica acústica que promovesse a distinção de um /p/ de um /t/, uma vez que dentro da própria categoria fonêmica havia uma variação. Em outras palavras, tais achados mostravam que a consoante tinha suas transições alteradas a depender do contexto vocálico em que se encontrava. Para os autores, isso sugeria que a percepção da fala somente era possível se o ouvinte estivesse sendo guiado pela articulação e não pela audição, devido à falta de invariância do sinal acústico.

A proposta de que a percepção é guiada pela articulação e não pela audição é chamada de solução articulatória para o problema da percepção da fala. Ela surge a partir desses achados e promove a proposição de duas teorias para tratar do assunto: a Teoria Motora da Percepção da Fala (LIBERMAN; MATTINGLY, 1985) – doravante TM – e a Teoria do Realismo Direto de Percepção da Fala (FOWLER, 1996) – doravante TRD. Diante da extravagância dos primitivos teóricos (OHALA, 1996), há certa relutância em se aceitar tais hipóteses justamente pelo fato de essas propostas não nos serem intuitivas. Ora, é comum nos perguntarmos: se a fala é transmitida no meio acústico, como e por que percebemos as articulações? É claro, tal extravagância teórica não é amparada apenas com esses primeiros achados dos Laboratórios *Haskins*. Há evidências experimentais que sugerem a adoção de primitivos articulatórios para a análise da percepção da fala, tal como o Efeito *McGurk*[5] (MCDONALD; MCGURK, 1976).

4 Para obter informações sobre o funcionamento do *Pattern Playback* ver a Introdução e o Capítulo 2 de Liberman (1996[1957]).

5 Esse efeito será apresentado nas próximas seções.

Diante disso, temos por objetivo geral apresentar essas duas teorias de percepção da fala (TM e TRD), pois elas apresentam soluções distintas para o problema encontrado na década de 1950. Na seção seguinte, deixaremos clara a maneira como cada uma das teorias considera a contraparte acústica e articulatória da percepção da fala. Para realizar isso, apresentaremos as explicações de cada uma das teorias sobre o Efeito *McGurk*. A nossa proposta é a de que, por mais que o fenômeno de "perceber a fala" seja observacionalmente o mesmo, as teorias tomam os elementos acústicos e articulatórios como distintos, o que as faz incomensuráveis (KUHN, 1975[1962]).

Antes de avançar com o texto, é preciso salientar que nossa reflexão não traz novos dados para o debate. Trata-se de uma pesquisa bibliográfica com o intuito de deixar claro que as teorias sobre a percepção da fala consideram distintos os elementos envolvidos no "circuito comunicativo". Isso se faz necessário por dois motivos: 1) para se ter em conta quais são os elementos de interesse para uma determinada teoria, de modo que os respectivos adeptos consigam conduzir suas pesquisas em sintonia com tais premissas teóricas; e 2) para que se tenha mais discussões acerca da pluralidade teórica na linguística (BORGES NETO, 2004).

Duas teorias de base articulatória para tratar a percepção da fala

A Teoria Motora de Percepção da Fala

As origens da TM datam da década de 1950 com os trabalhos de Liberman e colegas, nos *Laboratórios Haskins*. Nesses trabalhos inaugurais discutiam-se quais seriam os padrões acústicos responsáveis pela percepção de um dado som da fala. Os pesquisadores sugeriram que os padrões acústicos de fala sintetizada (tal como o *locus* das transições de consoantes oclusivas) tinham que ser alterados para que um "percept" fonético invariante fosse percebido

como uma categoria fonêmica em diferentes contextos. Tais observações provenientes desses dados apontavam para o fato de que os objetos da percepção não estariam no sinal acústico. Os objetos da percepção da fala talvez estivessem em processos motores simbólicos. Após aproximadamente 30 anos do levantamento dessa proposta, os autores estabelecem quais são as bases para uma teoria de percepção da fala de base articulatória. A versão mais conhecida da teoria é a de 1985, na qual os autores apresentam a sua versão revisada[6] (LIBERMAN; MATTINGLY, 1985).

A novidade com que é tratada a percepção da fala pela TM se dá basicamente por dois pontos adotados logo de saída. O primeiro deles se refere aos objetos da percepção, que seriam gestos fonéticos pretendidos pelo falante. Tais gestos seriam "representados no cérebro como comandos motores invariáveis que 'chamam' os movimentos dos articuladores em certas configurações linguisticamente significativas" (LIBERMAN; MATTINGLY, 1985, p. 237).

É preciso, contudo, deixar claro aqui que a proposta da TM assume a dicotomização entre os elementos fonéticos e fonológicos da gramática fônica da linguagem. Ou seja, no âmbito fonético, assume-se que os "comandos motores invariáveis" seriam a realidade física subjacente às noções fonéticas tradicionais, i.e., trata-se de um olhar mais minucioso (uma tentativa de trazer mais elementos físicos para as representações) sobre o que estaria por trás da realização fonética (ou implementação) de um som. Por sua vez, no âmbito fonológico,

> [...] os gestos por si só devem ser vistos como grupos de traços, tais como "labial", "oclusivo", "nasal", mas esses traços são atributos dos eventos gestuais, não eventos em si. Perceber uma sentença, então, é perceber um padrão específico de gestos pretendidos. (LIBERMAN; MATTINGLY, 1985, p. 238)[7]

6 Em Nishida (2012) apresento como se dá o processo de "construção" da versão revisada, que só foi possível após alguns marcos teóricos como Chomsky e Halle (1968) e Fodor (1983).

7 No original: "[Phonologically, of course,] the gestures themselves must be viewed as groups of features, such as 'labial', 'stop', 'nasal', but these features are attributes of the gestural events, not events as such. To perceive an utterance, then, is to perceive a specific pattern of intended gestures".

Em outras palavras, pode-se notar que a concepção de gestos pretendidos surge como uma solução para os efeitos de coarticulação que fazem com que os gestos não fossem manifestados diretamente no sinal acústico ou em movimentos articulatórios observáveis ou implementados.

O segundo ponto proposto pela teoria é o referente ao fato de que produção e percepção da fala são tomados como intimamente ligadas. Isso se dá (e só seria possível) porque elas devem compartilhar de um mesmo conjunto de invariantes, i.e., para a proposta, o *link* entre a produção e a percepção não pode ser uma associação aprendida. Para os autores, ela é inatamente especificada, requerendo "apenas desenvolvimento epigenético para trazê-la à tona" (LIBERMAN; MATTINGLY, 1985, p. 238). Esse ponto da teoria, claramente de influência gerativo-transformacional, considera a fala especial e, por isso, como sendo passível de tratamento modular:

> Neste ponto, a percepção dos gestos ocorre em um módulo especializado, importantemente diferente do módulo auditivo, responsável também pela produção de estruturas fonéticas, e parte da grande especialização da linguagem. (LIBERMAN; MATTINGLY, 1985, p. 238)[8]

Em suma, pode-se dizer que a saída dos autores para a falta da invariância acústica dos fonemas de uma dada língua é a de propor símbolos subjacentes à produção que, mesmo sem pistas presentes durante a sua realização física, seriam recuperados por estruturas modulares que fariam "a conversão do sinal acústico para o gesto automaticamente" (LIBERMAN; MATTINGLY, 1985, p. 238). Para entender melhor como se dá essa "conversão" do sinal acústico em gesto, os autores definem módulo como uma "estrutura neural especial, desenhada para tomar vantagem sistemática, mas apenas de uma única relação entre um *display* proximal de um órgão do sentido e alguma propriedade de um

8 No original: "On this claim, perception of the gestures accurs in a specialied mode, different in important ways from auditory mode, responsible also for the production of phonetic structures, and part of the larger specialization for language".

objeto distal" (LIBERMAN; MATTINGLY, 1985, p. 241). Como se pode notar, a existência de um módulo específico para a fala garante que:

> [...] um resultado em todos os casos é que não há, primeiro, uma representação cognitiva do padrão proximal que é modalmente geral, seguido pela tradução de uma propriedade distal particular; ao invés disso, a percepção de uma propriedade distal é imediata, isso significa que o módulo fez todo o trabalho pesado. (LIBERMAN; MATTINGLY, 1985, p. 241)[9]

Como se pode notar, a TM propõe um tratamento diferente para a percepção da fala, uma vez que

> [...] a percepção da fala não é explicada por princípios que se aplicam à percepção de sons em geral, mas deve, ao contrário, ser vista como uma especialização para os gestos fonéticos. Incorporar uma conexão biologicamente baseada entre a percepção e a produção é uma especialização que impede os ouvintes de tomar o sinal como sons ordinários, mas permite que eles usem a relação sistemática, ainda especial, entre sinal e gesto para perceber o gesto. A relação é sistemática porque resulta de uma dependência de fidelidade entre os gestos, movimentos articulatórios, formatos de cavidade oral e sinal acústico. Ela é especial porque ocorre apenas na fala. (LIBERMAN; MATTINGLY, 1985, p. 240-241)[10]

9 No original: "A result in all cases is that there is not, first, a cognitive representation of the proximal pattern that is modality-general, followed by translation to a particular distal property; rather, perception of the distal property is immediate, which is to say that the moule has done all the hard work".

10 No original: "[...] speech perception is not to be explained by principles yhat apply to perception of sounds in general, but must rather be seen as a specialization for phonetic gestures. Incorporating a biologically based link between perception and production, this specialization prevents listeners from hearing the signals as an ordinary sound, but enables them to use the systematic, yet special, relation between signal and gesture to perceive the gesture. The relation is systematic because it results from lawful dependencies among gestures, articulators movements, vocal-tract shapes, and signal. It is special because it occurs only in speech".

O que deve ficar claro é o tratamento especializado que a TM dá à fala ao se basear em uma abordagem modular. Essa proposta, além do poder explicativo proporcionado, dá conta da questão com que toda teoria de percepção da fala vai se deparar: como dar conta da falta de relação biunívoca entre os padrões acústicos e as categorias fonêmicas? Para a TM só há uma saída possível para lidar com essa questão: "os padrões acústicos proximais devem ser os objetos distais percebidos. Embora a relação entre o gesto e o sinal acústico não seja estreita [...]" (LIBERMAN; MATTINGLY, 1985, p. 238). Para os proponentes da teoria, essa relação só pode ser garantida ao assumir os gestos pretendidos como primitivos da percepção.

Por sua vez, a especialidade da fala se manifesta não só na falta da relação biunívoca entre pistas acústicas e uma categoria fonêmica. Os autores sugerem que é possível propor que o *timing* dos gestos realizados não é o mesmo dos gestos envolvidos em um dado símbolo, i.e., nível simbólico e implementacional não precisam necessariamente estarem estreitamente ligados. Em suas palavras:

> [...] os movimentos dos gestos em decorrência de um único símbolo não são tipicamente simultâneos, e os movimentos decorrentes de símbolos sucessivos sempre se sobrepõem extensivamente. Esta coarticulação significa que a mudança do formato do trato vocal, e, por consequência, o sinal resultante, é influenciado por vários gestos ao mesmo tempo. Portanto, a relação entre gesto e sinal é sistemática de uma maneira que é peculiar à fala. (LIBERMAN; MATTINGLY, 1985, p. 238-239)[11]

Esse olhar diferenciado sobre a fala sugere que um mesmo gesto pode possuir mais de um correspondente acústico. Essa "saída" daria conta, por exemplo, do fato de que as transições das consoantes variassem de acordo

11 No original: "[...] the movements for gestures implied by a single smbol are typically not simultaneous, and the movements implied by sucessive symbols often overlap extensively. This coarticulation means that the changing shape of the vocal tract, and hence the resulting signal, is influenced by several gestures at the same time. Thus, the relation between gesture and signal, though certainly systematic, is systematic in a way that is peculiar to speech".

com o contexto (vogal adjacente à consoante). Em suma, isso garantiria o fato de que percebemos os gestos pretendidos (e não os realizados) como sendo uma mesma consoante.

A Teoria do Realismo Direto de Percepção da Fala

Uma segunda saída para o problema da percepção da fala é a saída articulatória da Teoria do Realismo Direto da Percepção da Fala (TRD) de Fowler (1996). Para sistematizar a proposta da teoria, vamos tomar como referência Fowler (1996). É preciso salientar essa questão por dois motivos: 1) a TRD se encontrava "dispersa" em trabalhos de natureza experimental que apontavam para um tratamento de natureza realista sobre a percepção; 2) é com essa publicação que a autora sistematiza sua proposta em decorrência de uma "confusão" que Ohala (1996) havia feito ao considerar TM e TRD como equivalentes, i.e., tratava-se, segundo Ohala, de teorias extravagantes sobre a percepção da fala pois tomavam como primitivos a articulação e não os sons.[12]

De maneira sintética, pode-se dizer que a TRD considera a percepção da fala a partir de duas premissas: uma relacionada à natureza dos primitivos de percepção da fala; e outra, aos indivíduos. Passemos a cada uma delas separadamente.

Primeiramente, Fowler assume que "os gestos fonológicos são ações públicas[13] do trato vocal que causam estrutura nos sinais acústicos da fala" (1996, p. 1731). Isso significa que, ao contrário da TM, os gestos do trato vocal possuem propriedades invariantes próprias e seriam eles mesmos os

12 É preciso salientar, ainda, que TM e TRD parecem não competir. As disputas se dão mais entre uma abordagem acústica contra uma articulatória. Por este motivo, não apresentaremos aqui contra-argumentações entre TM e TRD pelo simples fato de elas não existirem. Aos interessados em conhecer mais a disputa entre uma abordagem acústica e uma articulatória ver Ohala (1996), Fowler (1996) e Diehl, Lotto e Holt (2004).

13 A autora considera os gestos fonológicos como *ações públicas* pelo fato de eles serem reais.

componentes fonológicos de uma sentença: Uma vez que há propriedades articulatórias invariantes, há "resultados" acústicos especificados que permitem que a percepção das propriedades fonológicas seja direta, i.e., não haveria um módulo específico fazendo o "trabalho pesado" de traduzir os estímulos acústicos em intenções (articulatórias) do falante.

Com relação aos indivíduos, a TRD assume que os usuários de uma língua percebem os gestos porque os sistemas perceptuais têm uma função universal de perceber as causas do mundo real em mídias como a luz, o ar e as superfícies. Em outras palavras, há aqui uma mudança de concepção com relação aos indivíduos, trata-se de seres que percebem o seu nicho de sobrevivência. Isto é, por sermos percebedores,[14] percebemos a fala (e o mundo em geral) de maneira real e direta pois isso promove sobrevivência.

Essas premissas acentuam um ponto de discordância com relação à TM. Trata-se do fato de a TRD não tomar a percepção da fala como especial. Para Fowler (1996), não há motivo para se propor sistemas especializados para tratar de cada sentido animal. Sua abordagem sugere que um sistema motor de percepção da fala funciona a partir de um mesmo princípio que um para percepção visual do sistema de locomoção: os seres percebem os movimentos reais, sejam eles referentes à fala ou à visualização dos movimentos envolvidos na locomoção, porque os seres captam as informações de seu nicho de maneira direta e real.

Uma maneira de compreender a proposta da TRD é considerar que a percepção direta dos gestos reais obedece a funções universais de percepção. Para isso, a autora utiliza a proposta realista e evolucionista dos "sistemas perceptuais" de Gibson (1966). Essa proposta sugere que os seres em geral percebem seu nicho de maneira direta e multissensorialmente para promover a sobrevivência da espécie.

O que se pode notar a partir dessas premissas é que a autora tenta inserir conhecimentos sobre outros sentidos para compreender a percepção da

14 Fowler utiliza a palavra *perceiver*.

fala. Ao fazer isso, Fowler diminui ainda mais as "distâncias" entre os dados de produção e percepção da fala, sintonizando os achados acerca da produção e percepção da fala. O que queremos dizer com isso é que a autora se filia a uma teoria de percepção mais ampla. Da mesma forma que a TM se vê revisada a partir de marcos teóricos (mencionados anteriormente), a TRD fundamenta a proposta de que percebemos os gestos reais do trato vocal a partir das teorias de produção da fala desenvolvidas por Fowler e Saltzman (1993) e de fonologia de Browman e Goldstein (1986, 1992).

O que o leitor deve estar se perguntando a essa altura da exposição da proposta da TRD é como os dados sobre a fala podem ser acomodados em uma teoria geral de percepção.

Primeiramente, Fowler (1996) sugere que tudo que percebemos é decorrente de eventos públicos presentes no ambiente ou nicho de atuação do percebedor. Tais eventos provêm informações que dão pistas para a sobrevivência dos animais, uma vez que, através de investigação empírica, os animais (ou percebedores) coletam tais informações para guiar suas ações.

As informações que guiam a ação dos seres são coletadas pelos sistemas perceptuais independentemente da mídia que veicula as informações sobre a sua fonte de estimulação. Para a proposta, só há sobrevivência se os órgãos do sentido perceberem o que causou os estímulos e não as mídias que "transportam" tais pistas, i.e., "essencialmente para sua sobrevivência, eles percebem os componentes do seu nicho que causaram essa estrutura" (FOWLER, 1996, p. 1732). Esse aspecto da teoria é inovador por salientar o interesse pela multissensorialidade para compreender a percepção da fala.

Ao fim, Fowler embasa sua proposta de que percebemos gestos reais diretamente com o argumento sobre a seleção natural:

> Esses sistemas perceptuais foram moldados pela seleção natural para obedecerem à função de familiarizar os percebedores com os componentes de seus nichos. A percepção auditiva apenas pode ser selecionada para a mesma função. Não há

> vantagem de sobrevivência em ouvir ar estruturado, mas há vantagem, por exemplo, para localizar um animal pesado fora do campo de visão e para detectar de uma maneira, com relação a si mesmo, que se trata de algo pesado. (FOWLER, 1996, p. 1732)[15]

Em suma, a proposta da TRD de se adotarem primitivos perceptuais como gestos reais está em sintonia com uma abordagem dita ecológica (influenciada por Gibson – 1966), pois tenta dar conta da percepção dos sons da fala sem considerá-la especial ou com características exclusivas. A autora tenta relacionar a sua teoria a questões de ordem evolutiva, uma vez que perceber diretamente os gestos e o ambiente promove a sobrevivência dos seres.

O que é o acústico e o articulatório para cada uma das teorias de percepção da fala

Após a apresentação esquemática da TM e da TRD, trataremos nesta seção de como cada uma das teorias considera os aspectos acústicos e articulatórios no processo de percepção da fala. Para realizar essa discussão é importante não perder de vista que o fenômeno de perceber a fala deve ser visto como o mesmo para todas as abordagens. O que muda para cada uma delas é o estatuto que se dá para cada um dos processos envolvidos durante o processo perceptual. Diante disso, podemos dizer que todas as teorias têm que levar em conta que perceber a fala envolve pelo menos três processos físico-fisiológicos.

15 No original: "These perceptual systems were shaped by natural selection to serve the function of acquainting perceivers with components of their niches. Auditory perception can only have been selected for the same function. There is no survival advantage to hearing structured air, but there is an advantage, for example, to locating a large lumbering animal out of view and to detecting which way, in respect to one's self, it is lumbering".

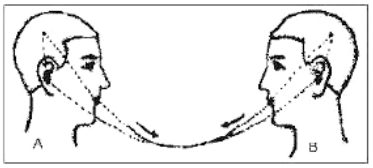

Figura 1: Circuito comunicativo entre A e B proposto por Saussure (1969, p. 19).

Considerando a Figura 1, acima, podemos dizer que, primeiramente, é indispensável que tenhamos um falante (A) que é considerado a fonte de emissão da fala. Esse indivíduo, com o auxílio de seus órgãos e articuladores, faz com que seu aparelho fonador ressoe e produza sons que, em segundo lugar, serão transmitidos pelo ar até os ouvidos de seu interlocutor. Terceiro, também é fundamental levar em conta que para se perceber a fala precisamos de um ouvinte (B)[16] que receba e interprete essas vibrações do ar.

Essas condições são as mesmas para todas as teorias. Em outras palavras, estamos dizendo que o fenômeno de perceber a fala é o mesmo para todas elas, uma vez que é inegável que há o envolvimento da movimentação de articuladores (uma contraparte articulatória) que promovem vibrações das partículas de ar que, por sua vez, são transmitidas (uma contraparte acústica) até os ouvidos de um receptor presente num dado momento comunicativo.[17]

O que queremos apresentar nas linhas a seguir é que esse mesmo fenômeno de perceber a fala tem seus elementos tomados de maneira distinta a depender da teoria que o está considerando. Focaremos principalmente nos papéis assumidos pelas contrapartes acústica e articulatória envolvidas no processo de percepção e produção da fala.

16 Nessa abordagem de natureza psico-física sobre a percepção da fala, é possível falar em ouvinte. Uma abordagem de base realista e, consequentemente, mais ampla é preciso falar em percebedores, já que se trata de seres que percebem o mundo multissensorialmente.

17 É claro que é possível perceber fala através de outros meios, como via telefone, por gravações ou pela interceptação de uma mensagem que não foi endereçada diretamente aos nossos ouvidos (i.e., quando se escuta a conversa de outra pessoa). No entanto, o que não devemos perder de vista é que em todos esses casos alguém estava produzindo fala para alguém ouvir, i.e., há uma contraparte articulatória e uma acústica.

Por exemplo, no período Estruturalista, o meio acústico era visto como um canal de transmissão de informações. Nesse domínio, era esperado que os invariantes perceptuais fossem encontrados, pois assim o sistema fonológico de uma dada língua conseguiria representar as distinções suficientes e necessárias para a transmissão satisfatória de uma mensagem para um dado destinatário. Tais considerações sobre o conteúdo informativo (e distintivo) das mensagens estão presentes nas páginas iniciais dos *Preliminaries of Speech Analisys* (JAKOBSON; FANT; HALLE, 1952):

> Qualquer distinção mínima carregada pela mensagem coloca o ouvinte diante de uma situação de duas escolhas. Dentro de um determinado idioma cada uma das oposições tem uma propriedade específica que a diferencia de todas as outras. O ouvinte é obrigado a escolher ou entre duas qualidades polares de uma mesma categoria, tais como grave *versus* agudo, compacto *versus* difuso, ou entre a presença e ausência de uma certa qualidade, tais como vozeado *versus* não vozeado, nasalizado *versus* não nasalizado, incisivo *versus* raso (plano). (JAKOBSON; FANT; HALLE, 1952, p. 3)[18]

Ainda mais, a procura por unidades mínimas distintivas (tanto para a produção quanto para a percepção da fala) era justificada pelo fato de os usuários da língua terem acesso exclusivamente ao meio acústico. Diante disso, esse domínio passa a ser o foco dos estudos do período, uma vez que a contraparte acústica do processo de produção da fala em um dado sistema comunicativo passa a ser considerada fundamental (e a única saída) para tratar a percepção.

É importante notar que, ao se olhar exclusivamente para a mensagem, deixa-se de lado o ouvinte envolvido no processo perceptual. Melhor dizendo: o ouvinte possui um papel passivo na percepção da fala. Isto é, considerando

18 No original: "Any minimal distinction carried by the message confronts the listener with a two-choice situation. Within a given language each of these opposition has a specific property which differentiates it from all the others. The listenner is obliged to chose either between two polar qualities of the same category, such as grave vs. acute, compact vs. diffuse, or between the presence and absence of a certain quality, such as voived vs. unvoiced, nasalised vs. non-nasalized, sharpened vs. Non-sharpened (plain)".

que todas as distinções já estão presentes no domínio acústico, o destinatário apenas recebe passivamente a mensagem através desse meio (canal).

Antes de passarmos às considerações articulatórias dos períodos seguintes, é preciso salientar (principalmente ao leitor atento) que, ao nos debruçarmos sobre o domínio articulatório, é possível notar que desde o Estruturalismo já há a utilização do termo gesto, que hoje nos remeteria diretamente (e, por consequência, anacronicamente) a uma abordagem de natureza articulatória sobre a linguagem. Entretanto, nesse período o termo gesto era tratado de outra maneira. E esse é o ponto interessante para a discussão realizada aqui. Em suma, falava-se de gesto muito mais como sinônimo de movimento, que, devido às propostas estruturalistas, não apresentava relevância linguística, i.e., não produzia distinções fonológicas. Isso se deve ao fato de que o movimento dos articuladores estava vinculado muito mais à fonética do que à fonologia, i.e., o gesto estava vinculado ao que se produzia e não ao que se acreditava realizar (TROUBETZKOY, 1970[1939]).

De modo distinto do período estruturalista, o Gerativismo imprime um outro olhar para o sinal acústico e para as articulações. Embora fonética e fonologia apresentassem uma tímida tentativa de tratar aspectos fonéticos como fonológicos na proposta do *Sound Pattern of English* (SPE) (CHOMSKY; HALLE, 1968), o sinal acústico continua tratado como epifenômeno dentro do processo de produção e percepção da fala.

Do ponto de vista da produção da fala, o sinal acústico é o resultado físico de toda a implementação motora dos símbolos que compõem a gramática fonológica das línguas, i.e., trata-se do resultado da realização motora das regras simbólicas envolvidas nas transformações mentais e abstratas que precedem a enunciação. Do ponto de vista da percepção, o sinal acústico será o estímulo proximal que alimentará a mesma gramática fonológica a partir da audição. Em seguida, fazendo uso de transformações, o módulo fonético traduzirá os estímulos proximais para que se recuperem os estímulos distais,

i.e., o módulo fonético fará com que essa parte física (as ondas sonoras) seja convertida em símbolos de base articulatória. Para a TM, tais articulações sequer precisam ser realizadas, uma vez que o módulo fonético consegue recuperar o comando neuromotor (o gesto) pretendido.

Como se pode notar, ao contrário do estruturalismo, o usuário da língua do gerativismo possui papel importante no que concerne à implementação das regras armazenadas (como representações mentais) que geram a cadeia da fala. E estando esse falante em "posse" de um sistema que converte símbolos de base articulatória (tal como a proposta de Chomsky e Halle (1968)) em ondas sonoras, nada mais esperado que o ouvinte (um interlocutor num dado momento comunicativo qualquer), também "equipado" pelo mesmo sistema que converte símbolos em ondas sonoras e ondas sonoras em símbolos, consiga recuperar as articulações pretendidas pelo emissor de uma mensagem.

Em outras palavras, cabe ao interlocutor (o ouvinte) perceber o dado articulatório, pois o sinal acústico é considerado consequência da articulação. Esse aspecto confere ao domínio acústico um lugar secundário no processo de percepção da fala. Em suma, no gerativismo, há um ouvinte parcialmente ativo, porque ele recebe o sinal acústico e o módulo fonético faz todo o trabalho de recuperar os estímulos distais (as articulações pretendidas) através da recepção e tradução dos estímulos proximais (as ondas sonoras) a partir de regras fonológicas. A citação de SPE salienta a preocupação de que a fonologia dê conta tanto da produção quanto da percepção, evidenciando uma tentativa de integrar na fonologia alguns aspectos tradicionalmente tidos como fonéticos:

> Dada a estrutura superficial de uma sentença, as regras fonológicas de uma língua interagem com certas restrições fonéticas universais para derivar todos os fatos determinados gramaticalmente sobre a produção e a percepção desta sentença. (CHOMSKY; HALLE, 1968, p. 293)[19]

19 No original: "Given the surface structure of a sentence, the phnologica rules of the language interact with certain universal phonetic constrains to derive all grammatically determined facts about the production and perception of this sentence".

Seguindo essa lógica, ao pensarmos no domínio articulatório da produção e percepção da fala, a TM gramaticaliza o gesto, porque ele passa a fazer parte da fonologia da língua. Nesse modelo, o gesto (ainda concebido como o movimento pretendido dos articuladores) assume as características estáticas, mentais e simbólicas compatíveis com a abordagem gerativo-transformacional de SPE.

A partir da proposta da Fonologia Articulatória (FAR) (BROWMAN; GOLDSTEIN, 1992), poderíamos conjeturar que o sinal acústico não seria importante para o desenvolvimento da teoria. Do ponto de vista da sua proposta de base dinâmica, a FAR tinha como objetivo tratar dos aspectos puramente articulatórios da produção da fala – ou da ação envolvida na sua produção (GOLDSTEIN; FOWLER, 2003). No entanto, ao considerarmos as propostas da TRD, conseguimos notar que o domínio acústico desempenha papel fundamental na efetivação do processo comunicativo, i.e., é possível notar que tanto o domínio acústico quanto o articulatório devem ser tomados como fundamentais para a produção e percepção da fala.

Contrariamente às abordagens estritamente auditivas dos estruturalistas e simbólicas dos gerativistas, a dinâmica sugere que, embora o primitivo acional e perceptual sejam de base articulatória e real, o sinal acústico possui uma posição de suma importância para a transmissão das informações acerca dos movimentos realizados pelos articuladores no interior do trato vocal. As ondas sonoras são fundamentais para que os estímulos sejam interceptados pelos usuários da língua a partir da estimulação dos órgãos do sentido que compõem o sistema (multi)perceptual humano.

Por exemplo, um indivíduo tem a tarefa de articular um determinado som da sua língua. Ele o faz através da imagem mental que tem estocada e que é relativa ao conjunto dos articuladores que têm de ser acionados para o cumprimento dessa tarefa. Aliado a isso, o falante sabe qual é o tempo de ativação de cada articulador, a maneira como cada articulador se coordena com os demais e o movimento que esse articulador tem de descrever (e.g.,

ele pode se encostar totalmente noutro para realizar uma plosiva ou apenas se aproximar maximamente dele para produzir fricção suficiente para a realização de uma fricativa). Toda essa complexa organização das tarefas envolvidas produz o sinal acústico que "carrega" ao receptor da mensagem todas essas informações que têm de ser interceptadas por ele, para que ele consiga "perceber" o que foi articulado.

Como se pode notar, acústico e articulatório são fundamentais para a percepção da fala no arcabouço da TRD. Ao contrário de se falar em decodificação ou tradução das informações articulatórias presentes no sinal acústico, é preciso tratar a percepção como um processo ativo de interceptação dos estímulos disponíveis nas mídias informacionais que carregam as características dos movimentos realizados pelos usuários das línguas. É essa característica direta e realista que possibilita a incorporação da variável tempo nas representações fonológicas que, por sua vez, é percebida pelos usuários de uma determinada língua.

Mais detalhadamente, a TRD propõe que o indivíduo que "percebe" o sinal acústico tem de depreender dele relações temporais entre os movimentos dos articuladores. Tais variações de relações temporais fazem com que se percebam, por exemplo, as diferenças entre um /l/ de início de sílaba ou palavra como *light* e um de coda silábica ou de final de palavra como *dark*. Sproat e Fujimura (1993) não trabalharam com análises perceptuais sobre a natureza ora *light* e ora *dark* do /l/ da língua inglesa. A questão de seu trabalho não era se havia variantes *dark* ou *light*, e, sim, se só havia essas variantes. Se pensarmos categoricamente e se assumirmos essa hipótese bigestual de organização das laterais do inglês, é indiscutível que as diferentes relações temporais são veiculadas pelo sinal acústico e, consequentemente, percebidas. Para se corroborar essa hipótese ainda é preciso verificar se os usuários da língua percebem as diferenças gradientes que se distribuem no *continuum* entre uma categoria e outra.

Por fim, falta ainda mencionar que a proposta da TRD de dar importância de mesmo peso para os níveis acústico e articulatório só é viável se os envolvidos na percepção da fala forem tidos como usuários ativos da língua. Ao contrário da percepção passiva do estruturalismo e parcialmente ativa do gerativismo, o percebedor da TRD se desloca em seu nicho em busca de informações nas mais diferentes mídias (ar, tato ou luz refletida) para interceptar as fontes dos estímulos que alimentam os sistemas perceptuais. Ou seja, como os usuários são ativos, eles buscam outras fontes de estimulação na ausência de informações em uma das mídias. Essa é a base de um sistema multimodal de percepção que o faz totalmente distinto da abordagem abstrata, simbólica e estática da TM. Um exemplo disso é quando tentamos conversar em um ambiente com um som muito alto. Diante dessa situação, imediatamente (e ativamente) buscamos informações visuais com relação aos movimentos dos articuladores de nosso interlocutor. Isso somente é possível se considerarmos a percepção como um sistema complexo que tem suas modalidades perceptuais ajustadas (e integradas) em decorrência das alterações (in)disponibilizadas pelo ambiente.

Um fenômeno com duas interpretações

Uma maneira de entender como as teorias de percepção da fala consideram os elementos acústicos e articulatórios envolvidos na sua produção e recepção é tomar um fenômeno comentado por cada uma das abordagens e mostrar como elas lidam com esses componentes de maneira distinta. Antes de prosseguirmos com a apresentação e considerações acerca do Efeito McGurk (MCDONALD; MCGURK, 1976), é preciso deixar claro que essa disputa entre TM e TRD não existe – pelo menos publicamente (DASCAL, 2006). Nosso objetivo é deixar claro como cada uma das teorias considera tais elementos envolvidos na percepção da fala. Voltamos a dizer: tais considera-

ções se fazem necessárias para que os estudos perceptuais sejam teoricamente embasados, uma vez que, sendo teoricamente orientados, pode haver coleta maior de dados em benefício de uma teoria ou a testagem dela. Diante disso, passaremos agora ao Efeito McGurk (doravante EM) para depois argumentar como cada uma das teorias lida com ele.

McDonald e McGurk (1976) publicam os resultados de um experimento que inaugura um olhar multimodal sobre a percepção da fala. Nesse estudo, os autores solicitavam que os participantes, num primeiro momento, ouvissem amostras de sílabas e identificassem o que estava sendo ouvido. Num segundo momento, os indivíduos eram solicitados a assistirem a um vídeo com o rosto de uma pessoa produzindo tais sílabas. A novidade dessa segunda etapa é que a imagem estava dublada por outras sílabas.

Por exemplo, o experimento consistia de duas etapas: uma auditiva e outra visual. Na auditiva, os participantes ouviam apenas estímulos sonoros tais como as sílabas [baba]. Sistematicamente, os participantes as identificavam como [baba]. O mesmo ocorria para [dada] e [gaga]. No entanto, na fase visual, os informantes experienciavam uma versão dublada: o vídeo mostrava uma pessoa articulando [baba], mas com o áudio de [gaga]. Ao presenciarem essa versão dublada do estímulo visual, os informantes diziam ter percebido [dada]. Na Tabela 1, abaixo, é possível verificar uma súmula do experimento.

Tabela 1: Súmula dos resultados do Efeito McGurk

Estímulos dublados **Áudio: [baba]** **Vídeo: [gaga]**	
RESPOSTAS	
Estímulo acústico	***Estímulos visual e acústico (dublado)***
[baba]	[dada]

Como se pode notar, o EM se mostra um problema para teorias de percepção da fala que apenas lidam com o sinal acústico, uma vez que, se a percepção fosse unissensorial (ou de exclusividade auditiva), perceber-se-ia apenas o sinal acústico, i.e., ao presenciar a imagem dublada, os indivíduos identificariam as sílabas pronunciadas sem "interferência" dos aspectos visuais.

Para a TM (LIBERMAN; MATTINGLY, 1985), o EM é uma prova cabal para a existência de um processamento modular e especial para a fala. Ao considerarem que há divergência entre sinal físico e real dos domínios acústico e visual-articulatório, a saída é considerar que a fonte da percepção (ou o estímulo distal) esteja na mente do usuário da língua. Em outras palavras, se os elementos físicos relevantes para a percepção não estão no sinal acústico e tampouco no visual, só podem estar na mente.

Figura 2: A mente é o domínio de interesse para a TM.

Considerando a Figura 2, podemos sugerir que, em vez de se procurarem os dados relevantes para a percepção da fala no domínio acústico (como se pretendia no Estruturalismo), a TM está interessada em buscar os elementos simbólicos (mentais) subjacentes à percepção da fala. Isto é, o EM seria o caso que comprova a falta de relação biunívoca entre o sinal acústico e a articulação realizados (implementados) e o nível simbólico da linguagem.

Em contrapartida, a TRD (FOWLER, 1996) busca interpretações a partir dos elementos reais da fala. Para essa teoria, os indivíduos percebem [da] nos dados dublados por que deve haver alguns elementos compartilhados multissensorialmente entre os estímulos experienciados. A explicação endossada pela autora é a mesma exposta por McDonald e McGurk (1976): percebe-se [d] nos estímulos dublados pois deve haver elementos acústicos

compartilhados entre [b] e [d] (pois os estímulos auditivos são [b]) e elementos visuais compartilhados entre [g] e [d] (pois os estímulos visuais são [g]); diante disso, [d] seria o candidato que melhor equaciona o "desencontro" entre as informações acústicas e visuais.

Ainda mais, o dado mais interessante para a abordagem realista de Fowler está na pequena parcela dos informantes do experimento que acabaram percebendo [bagba]. Essa resposta sugere que os indivíduos estão percebendo os estímulos auditivos e visuais (articulatórios) ao mesmo tempo. Trata-se de um ponto importante para a adoção de um olhar multissensorial sobre a percepção da fala.

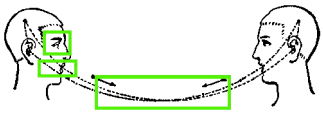

Figura 3: Os domínios de interesse para a TRD são a audição e visão (articulação).[20]

Como se pode notar, a TRD busca suas interpretações com as informações reais e disponíveis aos usuários da língua. Para essa teoria, ao contrário da abordagem mentalista da TM, interessam os domínios acústicos e visuais (articulatórios). Trata-se de uma abordagem realista que busca a integração de todos os sentidos como um sistema multissensorial (GIBSON, 1966), i.e., há uma "expansão" dos domínios de interesse da teoria: enquanto para a TM os elementos acústicos e articulatórios são epifenômenos devido ao seu ponto de partida nitidamente mentalista, a TRD contempla em suas explicações os elementos reais disponíveis a mais de um domínio sensorial.

20 Aqui apenas dissemos que os domínios de interesse para a TRD são a audição, visão e a articulação. Apenas lidamos com esses domínios como uma estratégia didática para que consigamos comparar as teorias de alguma forma. Frisamos isso pelo fato de a proposta da TRD estar em sintonia com a de Gibson (1966). Nela, todos os sentidos são interessantes, pois não se trata de perceber apenas a fala; trata de se perceber o mundo.

Considerações finais

O objetivo deste trabalho é o de trazer à luz algumas teorias de percepção da fala. Como foi possível notar em seu desenvolvimento, as teorias cotejadas aqui poderiam, em princípio, ser tidas como soluções articulatórias para o problema da falta de relação biunívoca entre a articulação e o sinal acústico dos sons. Contudo, um olhar minucioso sobre as propostas teóricas faz com que notemos que as teorias tratam da percepção de maneira distinta, de modo que as contrapartes acústicas e articulatórias de se perceber a fala ganham lugares distintos para cada uma delas. Vale lembrarmos o olhar estritamente direcionado para os fenômenos acústicos do período estruturalista, que desconsiderava o receptor de uma mensagem como responsável pela sua decodificação.

Em contrapartida, a abordagem de base gerativo-transformacional da TM propõe um ouvinte parcialmente ativo que tem por objetivo converter a mensagem acústica em símbolos de base articulatória a partir de regras fonológicas que fazem essa tradução. Aqui, o sinal acústico estimula o sistema perceptual dos ouvintes para que se percebam as articulações pretendidas pelo falante da língua. Ou seja, o domínio acústico é um epifenômeno. O importante para essa teoria é a abstração articulatória pretendida no momento em que se produz a fala.

Por sua vez, há o olhar da TRD que busca os invariantes articulatórios reais que são transmitidos pelo sinal acústico. Nessa perspectiva, os usuários da língua são considerados seres ativos no processo de percepção da fala, pois os percebedores interceptam o sinal acústico em busca de informações acerca do seu ambiente de sobrevivência. Em outras palavras, tanto os aspectos acústicos quanto os articulatórios são importantes para a TRD. Ainda mais, por influência de Gibson (1966), todos os sentidos passam a ser relevantes para a percepção dos seres. Trata-se de um olhar multimodal (e multissen-

sorial) sobre a percepção, i.e., a percepção da fala é entendida de maneira não especial,[21] mas sim como uma capacidade ecológica que integra todos os sistemas perceptuais disponíveis nos seres (ROSENBLUM, 2005).

Em suma, numa abordagem dinâmica, o gesto assume *status* central no modelo. O gesto passa a ser também o movimento real do articulador, não apenas o pretendido, tal como queria a TM.

Por fim, consideramos que saber essas diferenças acerca das teorias de percepção da fala faz com que os fonólogos interessados em percepção comecem a tratá-la dentro dos respectivos arcabouços teóricos, pois, como se pode ver neste trabalho, o primitivo de análise fonológica assumido pelas teorias parece estar em sintonia com os primitivos de percepção da fala propostos pelas respectivas abordagens teóricas.

NISHIDA, Gustavo. Acoustic and articulatory basis of speech perception theories. **Revista do Gel**, v. 11, n. 1, p. 142-167, 2014.

ABSTRACT: *The main issue of this paper is to shed some light on two articulatory solutions to the speech perception problem by comparing different aproaches available in the literature. Although both are articulatory solutions, we argue that the perceptual primitives are ontologically different. The Motor Theory of Speech Perception (LIBERMAN; MATTINGLY, 1985) proposes a mentalistic and modular approach. The Direct Realism Theory of Speech Perception (FOWLER, 1996) considers real units as primitives. Besides these differences, we also present the way that acoustic and articulatory components take place in both theories: in the Motor Theory, the acoustic counterpart is taken as an epiphenomenon, since the theory assumes a mental approach to the articulatory basis; on the other hand, Direct Realism Theory integrates acoustics as part of perception, because of its multisensory approach.*

21 O fato de a proposta da TRD não ser *especial* é uma oposição à abordagem da TM que previa mecanismos específicos (como o módulo fonético) para tratar da fala enquanto um "código especial". Para aprofundar essa questão, consultem Fowler e Rosenblum (1991a) para críticas à especialidade e modularidade da fala; Fowler e Rosenblum (1991b) para problematizar o efeito duplex; e Rosenblum (2005) para salientar os achados neurofisiológicos que embasam a proposta multimodal da percepção.

KEYWORDS: *Speech perception primitives. Motor theory of speech perception. Direct realism theory of speech perception. Acoustic basis. Articulatory basis.*

Referências

BORGES NETO, José. **Ensaios de filosofia da linguística**. São Paulo: Parábola, 2004.

BROWMAN, C.; GOLDSTEIN, L. Towards an articulatory phonology. **Phonology Yearbook**, v. 3, p. 219-252, 1986.

_____. Articulatory Phonology: an overview. **Phonetica**, v. 49, p. 155-180, 1992.

CHOMSKY, N.; HALLE, M. **The Sound Pattern of English**. New York: Harper & Row, 1968.

DASCAL, M. **Interpretação e compreensão**. São Leopoldo: Editora da Unisinos, 2006.

DIEHL, R.; LOTTO, A.; HOLT, L. Speech perception. **Annual review of Psychology**, v. 55, p. 149-179, 2004.

FODOR, J. A. **Modularity of Mind**: An Essay on Faculty Psychology. Cambridge, Mass.: MIT Press , 1983.

FOWLER, C. A.; ROSENBLUM, L. D. Perception of the phonetic gesture. In: MATTINGLY, I. G.; STUDDERT-KENNEDY, M. (Eds.). **Modularity and the Motor Theory**. Hillsdale: Lawrence Earlbaum, 1991a. p. 33-59.

_____. Duplex perception: A comparison monosyllables and slamming doors. **Journal of Experimental Psychology: Human Perception and Performance.** v. 16, p. 742-754, 1991b.

FOWLER, C. Listeners do hear sounds, not tongues. **The Journal of the Acoustical Society of America**, v. 99, n. 3, p. 1730-1741, 1996.

FOWLER, C. A.; SALTZMAN, E. Coordination and Coarticulation in Speech Production. **Language and Speech**, v. 36, p. 171-195, 1993.

GIBSON, J. J. **The senses considered as perceptual systems**. Boston: Houghton Miffin, 1966.

GOLDSTEIN, L.; FOWLER, C. A. Articulatory phonology: a phonology for public language use. In: SCHILLER, N. O.; MEYER, A.S. (Eds.). **Phonetics and phonology in language comprehension and production**. Berlin/Boston: Mouton de Gruyter, 2003. p. 159-207.

JAKOBSON, R.; FANT, G.; HALLE, M. **Preliminaries to speech analysis**. Cambridge: MIT Press, 1952.

KUHN, T. S. **A estrutura das revoluções científicas**. São Paulo: Perspectiva, 1975[1962].

LIBERMAN, A. Some results of research on speech perception. In: _____. **Speech**: a special code. Cambridge, MA: MIT Press, 1996[1957].

LIBERMAN, A. M.; MATTINGLY, I. G. The motor theory of speech perception revised. **Cognition**, v. 21, p. 1-36, 1985.

MCGURK, H.; MACDONALD, J. Hearing Lips and Seeing Voices. **Nature**, v. 264, p. 746-48, 1976.

NISHIDA, G. **Sobre teorias de percepção da fala**. Tese (Doutorado) – Universidade Federal do Paraná, Curitiba, 2012.

OHALA, J. Speech perception is hearing sounds, not tongues. **Journal of the Acoustical Society of America**, v. 99, p. 1718-1725, 1996.

ROSENBLUM, L. D. The primacy of multimodal speech perception. In: PISONI, D.; REMEZ, R. (Eds). **The handbook of speech perception**. Malden: Blackwell, 2005. p. 51-78.

SAUSSURE, Ferdinand de. **Curso de linguística geral**. (Org. Charles Balley e Albert Sechehaye). São Paulo: Cultrix, 1969.

SPROAT, R.; FUJIMURA, O. Allophonic variation in English /l/ and its implications for phonetic implementation. **Journal of Phonetics**, n. 21, p. 291-311, 1993.

TRUBETZKOY, N. **Principes de Phonologie** [Grundzuge der Phonologie]. Paris: Klincksieck, 1970[1939].