

De como a representação motora integra a auditiva

Eleonora Cavalcante Albano*

A mesa-redonda que se fecha com este trabalho deu continuidade a uma discussão iniciada em 1991, no *II Encontro Nacional sobre Aquisição da Linguagem* (doravante ENAL), sobre "a idéia de uma sensorimotricidade auditivo-vocal plástica o bastante para subjazer a todas as vocalizações humanas" (Albano, 1992, p. 56). Em 1997, no IV ENAL, o estudo do tema já avançara o suficiente para permitir uma defesa explícita de uma certa comensurabilidade entre a Fonética e a Fonologia pelo uso do "gesto articulatório como unidade fônica abstrata" (Albano, 1999a). São alguns dos desdobramentos dessas questões que hoje ocupam esta mesa.

Ana Luiza Navas, com o trabalho "Estudo experimental sobre a natureza das representações fonológicas", mostrou que, nos trava-línguas usados no estudo experimental dos lapsos, as palatais são tão favorecidas quanto nos primeiros estágios da aquisição de linguagem (Hernandorena, 1999), o que contrasta, mais uma vez, com a sua baixa frequência de ocorrência nos inventários fonológicos das línguas (Maddieson, 1984). Isso levanta a questão de como e por que uma estrutura sensório-motora que é estável e até preferida num nível mais baixo de organização pode, não obstante, tornar-se instável a ponto a ser evitada em níveis mais altos.

Aglael Gama-Rossi, com o trabalho "Relações entre produção e percepção na aquisição da duração da vogal na aquisição da duração no português brasileiro" (doravante PB), mostrou, por outro lado, que, no que toca à implementação do ritmo característico de uma língua, uma motricidade ainda imatura pode contornar as suas próprias limitações para soar competente sem, de fato, reproduzir o padrão maduro.

* UNICAMP – Universidade Estadual de Campinas.

Finalmente, Maria de Jesus Gonçalves, com o trabalho "O papel do *sensorio-auditivo* na codificação de mensagens pictoideográficas em comunicação alternativa" demonstrou o fato – surpreendente – de que uma audição puramente passiva, em razão de déficits motores devidos a paralisia cerebral, desempenha, não obstante, o papel de auxiliar a fixação de padrões visuais na comunicação alternativa. Isso sugere que, mesmo na ausência de prática, a fala pode, de alguma forma, representar-se e fixar-se na memória permanente. A pergunta que emerge é: trata-se, de fato, de uma representação puramente auditiva?

O conjunto desses trabalhos suscita uma preocupação com os limites de uma representação fônica que se assente sobre bases motoras. Pode uma unidade gestual, do tipo advogado pela Teoria da Ação (Kelso et al., 1986) e pela Fonologia Articulatória (dora-vante FAR; Browman e Goldstein, 1986, 1989, 1992), constituir a representação fônica padrão e contemplar não só a produção mas também a percepção da fala? Não seria a neutralidade sensorio-motora das unidades tradicionais, tais como o segmento e o traço fonético, mais apropriada, afinal, para exprimir os níveis mais abstratos do conhecimento fônico?

Este trabalho argumenta que uma representação fônica dinâmica *de base motora* pode ser suficientemente genérica e abstrata para estar no centro da conexão entre as demais representações envolvidas no falar. A chave para isso reside em não reduzir a motricidade fônica ao simples movimento dos articuladores e encará-la como meio de implementar e, portanto, codificar ativamente macro e microrritmos que estão presentes também em todo o sensorio envolvido na fala, bem como em algumas outras atividades quotidianas. Em outras palavras, se o sensorio é também dinâmico, isto é, se é definido por parâmetros que evoluem no tempo, uma representação por meio de oscilações abstratas, orquestradas preferencialmente, mas não obrigatoriamente, no trato vocal, pode ser perfeitamente apta a mediar processos perceptuais de vários tipos.

O argumento divide-se em três partes.

Primeiro, é preciso mostrar que a dinâmica da produção de fala não está localizada neste ou naquele articulador, mas envolve o trato vocal como um todo. A discussão da dupla especificação das consoantes palatais, levantada por Ana Luiza Navas, será retomada a esse respeito.

Segundo, é preciso mostrar que também na fala natural e não apenas na fala sintética ocorrem situações em que um mesmo sinal acústico é interpretado como pertencente a duas categorias distin-

tas, a depender do contexto adjacente. Uma representação auditiva estática, que não contemple fronteiras móveis de categorias, simplesmente não pode dar conta desse fato. Já uma representação auditiva dinâmica prevê a ambigüidade perceptual de um gesto articulatório conforme a força relativa dos demais gestos do contexto.

Finalmente, é preciso mostrar que a interpretação neuromotora da distinção entre gestos vocálicos e consonantais, corroborada pelo trabalho de Aglael Gama-Rossi, está na base de uma hierarquia rítmica capaz de responder pelas correlações entre o sensorio e o motor na fala. Essa hierarquia, apesar de refletir sobretudo a motricidade fônica, é passível de resgate por via alternativa porque envolve oscilações de baixa frequência que, em casos de severo comprometimento motor como aqueles estudados por Maria de Jesus Gonçalves, podem ser transferidas a outros setores da experiência.

1 **Contra a especificação pontual, rígida do gesto articulatório**

Como mostrou D'Angelis (1998), o sucesso da FAR em lidar com as relações entre os vários modos de articulação não é muito maior do que o dos modelos fônicos estáticos. Em Albano (1999b), acresci a essa insatisfação uma outra, com o tratamento das vogais no modelo, e sugeri que processos que envolvem alterações de modo, tais como a vocalização de consoantes, pudessem, talvez, ser abordados pela ótica proposta por Sproat e Fujimura (1993) para o tratamento da lateral velarizada do inglês. Esses autores decompõem, com base em evidências de raios X, a lateral "escura" em dois gestos: um da ponta da língua, entendido como consonantal, e outro, do corpo da língua (no caso, o dorso), entendido como vocálico. A gradiência do efeito auditivo de velarização em diferentes posições da palavra e do grupo prosódico é, então, facilmente racionalizada pela sincronização relativa dos dois gestos.

Vou sugerir, agora, que definir gestos via articuladores é uma aproximação grosseira que, embora tendo servido à FAR para fins de implementação do modelo via síntese articulatória, não tem sentido do ponto-de-vista teórico. A Teoria Acústica da Produção da Fala (Fant, 1960) mostrou como certas configurações distintas do trato vocal geram sinais acústicos idênticos e, assim, racionalizou definitivamente o princípio da "equivalência motora", há muito conhecido na área clínica (Rodrigues, 1989, p. 15-28) e hoje documentado também em termos experimentais (Gay et al., 1981).

Ora, para conviver com a equivalência motora, uma teoria da dinâmica articulatória tem de admitir complementação e/ou compensação entre áreas diferentes do trato. O exemplo clássico é a vogal [u], em que as constrictões labial e velar tendem a se compensar mutuamente conforme o contexto (Perkell et al., 1993). Também a vogal [a], ao menos em línguas em que não contrasta com [æ], [ɑ] ou [ɒ], envolve compensação entre a raiz e o dorso da língua, de forma a apresentar, conforme o contexto, variantes mais ou menos posteriores.

Para dar conta desse tipo de fenômeno, um modelo fônico dinâmico deve especificar constrictões por região do trato – uma ou mais – e não por articulador. Usar o articulador dominante para denominar tais regiões é uma mera conveniência terminológica e não implica a sua participação obrigatória na obtenção de configurações articulatórias eficazes. Em suma, um tal modelo precisa resgatar explicitamente o elo previsto pela Teoria Acústica entre os aspectos articulatórios e acústicos do som de fala.

Um exemplo extremo, da área clínica, ajudará a ilustrar esse ponto. Os espectrogramas da Figura 1 mostram duas pronúncias da sílaba [da] por um foneticista: a normal, com a constrictão na região dental que constitui o padrão do português e de outras línguas românicas, e a que poderíamos chamar “compensatória extrema”, usada na reabilitação de pacientes glossotomizados, com constrictão pós-dental obtida por forte retração do lábio inferior de modo a formar uma câmara de ressonância frontal com os dentes superiores:

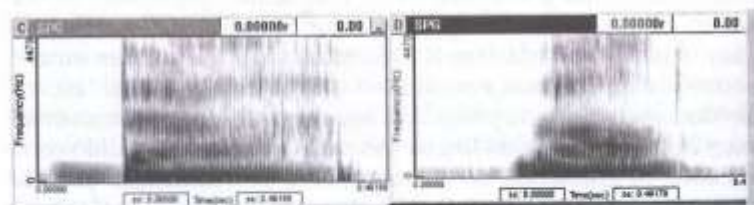


Figura 1. A sílaba [da], produzida com a lâmina da língua em (c) e com o lábio inferior retraído em (d).

Note-se que o movimento descendente de F_2 na transição para a vogal está presente em ambos os casos. O desconforto da pronúncia “compensatória extrema” manifesta-se numa barra de voz muito curta durante a oclusão e em algumas descontinuidades espectrais durante a vogal estacionária. Vale notar que essas distorções não atingem a identidade do [d], embora afetem a inteligibilidade geral.

Uma pergunta logo se impõe: se não é o articulador que específica a constrictão, como obter, através de simples menção a regiões relativamente amplas, o elenco de pontos de articulação atestado nas línguas do mundo?

O caso das fricativas palatais, abordado por Ana Luíza Navas, indica a direção da resposta. Os clínicos sabem que, no ensino das fricativas palatais a portadores de distúrbios articulatórios, costuma-se exagerar um gesto de protrusão labial que ajuda a obter uma constrictão adequada. O que muitos não sabem é que esse gesto não é obrigatório e que pode ser substituído pela formação de um canal longo, largo e côncavo, para trás dos dentes. Esse é, aliás, o padrão de produção das fricativas palatais do inglês (Heffner, 1950).

Uma análise dinâmica, mas não pontual, dirá, a partir dessas observações, que uma fricativa palatal é a realização simultânea de uma vogal e de uma constrictão crítica – isto é, geradora de turbulência – na região coronal.¹ Se essa vogal é do tipo [i] ou do tipo [y] é uma questão que depende da língua, do dialeto ou mesmo do contexto. A vantagem de [y] parece não ser apenas visual: o gesto de protrusão labial leva a língua a assumir uma forma ligeiramente convexa, que também forma um canal largo e longo na região coronal. É possível que assim se obtenha o efeito acústico de [ʃ, ʒ] sem muito esforço.

Fica claro, agora, que o predomínio de [ʃ, ʒ] sobre [s, z] nos lapsos, assim como na aquisição da linguagem, decorre da presença, implícita ou explícita, do gesto vocálico. Muito provavelmente, os erros categóricos, em que [ʃ, ʒ] efetivamente substituem [s, z], resultam da atração desse gesto para onde havia apenas uma constrictão crítica na região coronal. Trata-se de uma atração semelhante à que ocorre nos lapsos com as líquidas de grupos consonantais. Já os erros gradientes, que produzem algo entre [s, z] e [ʃ, ʒ], podem resultar da desorganização do próprio gesto constrictivo, que se desacelera e forma um canal mais largo do que o exigido por [s, z]. Neste caso, não há propriamente um gesto vocálico; o que há é uma queda na frequência da oscilação que gera a constrictão, formando uma passagem de ar com efeitos acústicos semelhantes.

Analogamente, o falante imaturo, como demonstrou Gama Rossi (1999), tem preferência por gestos mais lentos, os quais, em termos dinâmicos, são modelados por oscilações de baixa frequên-

¹ Hernandorena (1999) defende uma perspectiva semelhante no quadro da fonologia auto-segmental.

cia. Já os sistemas consonantais das línguas tendem a evitá-los, assim como tendem a evitar a implementação de dois gestos distintos na mesma região do trato, como é o caso das verdadeiras palatais.

2 Fronteiras categóricas móveis e reconstrução de gestos subjacentes

Um testemunho veemente do acerto da análise bigestual das palatais reside no caso da africacão de [t, d] diante de [i] no PB. No IV ENAL (Albano, 1999a, p. 152), demonstrei que a maior ou menor concentração de energia na região média do espectro do estouro daquelas oclusivas indica um gradiente de palatalização que nem sempre é audível. Restou a pergunta: o que leva um ponto desse gradiente a cruzar a fronteira entre as categorias dental/alveolar e palatal e, assim, ganhar os nossos ouvidos?

Podemos, agora, à luz das considerações acima, responder: são indícios da presença de um [i] ou [y] que possam ser interpretados como concomitantes à formação de turbulência na região coronal. Vejamos, agora, como essa resposta ilumina o comportamento de falantes que não costumam produzir a africada diante de [i], mas que não conseguem evitá-la em palavras como *lingüística*, *místico*, *dístico*, etc.

A Figura 2 exibe espectrogramas das palavras *vítima* e *vitima* pronunciadas por um falante desse tipo e postas na mesma escala de tempo. Note-se que os estouros de ambos os [t] são muito semelhantes, apresentando aproximadamente a mesma duração e a mesma frequência de ressonância. Observe-se, por outro lado, que, em *vitima* (janela a), o ruído excita intensamente o F_3 da vogal antes de dar lugar a uma breve excitação periódica de todos os formantes. Já em *vítima* (janela b), a intensidade de F_3 não é particularmente alta no ruído e cresce ao longo da vogal.

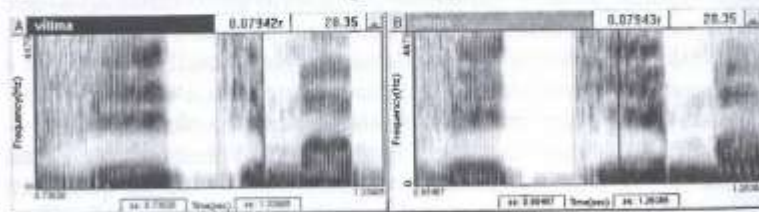


Figura 2. Oclusivas acusticamente semelhantes com efeitos auditivos diferentes em *vítima* e *vitima*.

É interessante observar que o [t] da janela (a) tende a ser ouvido como uma africada, ao passo que o da janela (b) é sempre ouvido como uma oclusiva. O que está em jogo é, evidentemente, a proporção da vogal que se sobrepõe à oclusão. Em *vítima*, ela é significativa; em *vitima*, ela é irrisória. O primeiro [t] tende a ser reconstruído como palatalizado, isto é, concomitante a um [i], porque o deslizeamento da vogal sobre a consoante é muito grande para parecer accidental.

Um achado antigo de Stevens (1985, p. 246-248) traz um interessante complemento a essa interpretação. Trabalhando com estímulos sintéticos, ele manipulou a amplitude de um ruído fixo, de 2-5 kHz, relativamente à amplitude do F_3 da vogal seguinte e descobriu que há uma fronteira categórica entre [s] e [ʃ] em torno da diferença zero. Quando o F_3 da vogal supera a intensidade do ruído na mesma faixa de frequência, os estímulos sintéticos são ouvidos como [s]. Quando é o ruído que supera a intensidade do F_3 da vogal, os mesmos são ouvidos como [ʃ]. Stevens (1989) afirma que esses efeitos decorrem da capacidade do sistema auditivo de detectar discontinuidades em parâmetros espectrais e arrola outros casos em que fronteiras categóricas emergem de uma diferença maior ou igual a 3 db em alguma faixa de frequência do espectro.

Parece claro, agora, a que se presta a diferença de 3 db no caso de [s, ʃ]. Ela permite decidir sobre a presença ou ausência do gesto vocálico característico da palatal durante a produção do ruído fricativo.

3 Significar, cadenciar, organizar: ritmos da fala, ritmos da ação

O caráter ainda precário da interlocução entre a literatura sobre dinâmica articulatória e a crescente literatura sobre a sincronização dos gestos corporais à fala (Mc Neill, 1992) obscurece alguns pontos que sugerem que uma representação fônica dinâmica orquestra ritmos que não são alheios a outras atividades voluntárias (para o esboço de um novo panorama, ver Rizzolatti e Arbib, 1998).

Gama-Rossi (1999) demonstrou que o ritmo da fala não se organiza de cima para baixo ou de baixo para cima, mas do meio para cima e daí para baixo. São as oscilações do nível silábico, que correspondem aos ciclos da mandíbula, que a criança domina mais cedo e traz para a tarefa de cadenciar a fala com os acentos da língua adulta, que são aproximados, inicialmente, pela duração rela-

tiva das vogais. É essa criança prosodicamente competente que vai se tornando segmentalmente mais precisa à medida que encaixa, com maior ou menor sucesso, os gestos mais rápidos e pontuais correspondentes às consoantes na hierarquia que organiza as batidas silábicas em torno dos acentos.

É importante frisar que os achados de Gama Rossi corroboram independentemente a distinção acima estabelecida entre gestos vocálicos e consonantais. E, além disso, na medida em que apontam para o seu fundamento neuromotor, contribuem para reforçá-la. Tem-se, assim, um quadro em que os gestos vocálicos e consonantais são intrinsecamente distintos em termos dinâmicos e podem se combinar seja em seqüência, para formar sílabas, seja simultaneamente, para formar segmentos complexos.

Em manuscrito inédito, Plínio Barbosa (em preparação) propõe que esses três fluxos – o dos acentos, o das sílabas e o dos segmentos – sejam modelados através de osciladores acoplados, isto é, que se influenciam uns aos outros em maior ou menor grau. A força dessa influência pode ser avaliada através da interpretação das durações relativas dentro de um modelo matemático do acoplamento. O reexame, à luz desse modelo, de dados publicados na literatura fonética sobre a duração e o ritmo de diversas línguas mostra que, embora essas variem quanto à dominância das batidas acentuais, nenhuma as dispensa inteiramente (ou seja, não existe "ritmo silábico" puro, assim como não existe "ritmo acentual" puro).

Lembrando, agora, o fato, bastante conhecido, de que os acentos da fala são muitas vezes sincronizados com os focos da atenção visual e da ação, fica clara a vantagem de um modelo fônico dinâmico que dê um tratamento unificado à prosódia e aos segmentos (para maiores detalhes, ver Albano et al., 1998; Barbosa e Madureira, 1999; Albano, no prelo). Um tal modelo permite que os ritmos de outras atividades, se modelados também dinamicamente, sejam acoplados e ajustados aos da fala. Evidencia-se aí não só o que a fala tem em comum com a ação em geral mas também o que há de especial na ação vocal, que, no nível segmental, produz informação distintiva numa taxa relativamente rápida.

Para terminar, façamos o exercício de especular sobre as razões da vantagem da condição "falante" no estudo de Maria de Jesus Gonçalves. Até que ponto o paraplégico cerebral com um trato vocal muito comprometido teria condições de incorporar uma representação fônica dinâmica tal como a aqui discutida?

A resposta, a esta altura óbvia, é: ele tem algum acesso ativo aos macrorritmos acentuais e silábicos, que correspondem a oscilações de frequência muito baixa, compatíveis com outros setores da ação, seja essa, nos casos menos severos, do próprio sujeito ou, nos casos mais severos, do interlocutor. Uma tal estrutura rítmica oferece à percepção um suporte hierárquico em que a informação segmental, excepcionalmente codificada em termos puramente auditivos devido à deficiência motora, encontra pontos estáveis de ancoragem.

Se essa hipótese é verdadeira, pode-se prever que uma fala muito fluente e rápida, que não fizesse, através da sua cadência, contato com as dificuldades motoras do sujeito, não apresentaria a mesma vantagem na facilitação do aprendizado da comunicação alternativa. Desacelerar cadenciando, acentuando é uma velha sabedoria clínica que a visão dinâmica da produção de fala vem, finalmente, racionalizar.

Por último, e não menos importante, quero manifestar a minha satisfação pelos caminhos que a equipe do Laboratório de Fonética Acústica e Psicolinguística Experimental (LAFAPE) da UNICAMP, com o auxílio de interlocutores críticos como Maria de Jesus Gonçalves, vem conseguindo abrir desde o último ENAL. Ali enunciei a nossa aspiração de "vislumbrar melhor como estudar a interação entre os macrorritmos da prosódia e os microrritmos dos gestos segmentais" (Albano, 1999a, p. 161). Os trabalhos apresentados nesta mesa testemunham o quanto já avançamos nessa direção.

Referências bibliográficas

- ALBANO, E. C. *O gesto e suas bordas: para uma fonologia acústico-articulatória do português brasileiro*. Campinas: Mercado de Letras, (no prelo)
- . O gesto articulatório como unidade fônica abstrata: indícios da fala infantil e evidências da fala adulta. In: LAMPRECHT, R. (org.). *Aquisição de linguagem: questões e análises*. Porto Alegre: EDIPUCRS, 1999a, p. 139-165.
- . O português brasileiro e as controvérsias da fonética atual: pelo aperfeiçoamento da fonologia articulatória. *D.E.L.T.A.*, v. 15, n. especial, p. 23-51, 1999b.
- . Esboço de modelo neuropsicologicamente orientado para a aquisição da Fonologia. *Anais do II Encontro Nacional sobre Aquisição de Linguagem*, Porto Alegre, CEAAL, PUCRS, p. 51-75, 1992.

—— et al. A interface fonética-fonologia e a interação prosódia-segmentos. *Estudos Linguísticos XXVII: Anais do XLV Seminário do Grupo de Estudos Linguísticos do Estado de São Paulo (GEL)*. São José do Rio Preto: UNESP-IBILCE, 1998, p. 135-143.

BARBOSA, P. A. Timing in Brazilian Portuguese revisited: new light under a dynamical perspective. IEL-UNICAMP. (manuscrito inédito; em preparação)

——; MADUREIRA, S. *Toward a hierarchical model of rhythm production: evidence from phrase stress domains in Brazilian Portuguese*. Proceedings of the XIVth International Congress of Phonetic Sciences. San Francisco, EUA, 1999, v. 1, p. 297-300.

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

——. Articulatory gestures as phonological units. *Phonology Yearbook 1989*: v. 6, p. 201-251.

——. Articulatory phonology: an overview. *Phonetica*, 49, p. 155-180, 1992.

D'ANGELIS, W. *Traços de modo e modos de traçar geometrias: línguas Macro-Jê e teoria fonológica*. Tese de doutorado inédita. LAFAPE-IEL-UNICAMP, 1998.

FANT, G. *Acoustic theory of speech production*. Haia: Mouton, 1960.

GAMA-ROSSI, A. Relações entre produção e percepção na aquisição da duração da vogal no português brasileiro. Neste volume.

——. *Relações entre desenvolvimento lingüístico e neuromotor: a aquisição da duração no português brasileiro*. Tese de doutorado inédita. LAFAPE-IEL-UNICAMP, 1999.

GAY, T.; LINDBLOM, B.; LUBKER, J. Production of bite block vowels: acoustic equivalence by selective compensation. *Journal of the Acoustical Society of America*, 69, p. 802-810, 1981.

GONÇALVES, M. J. O papel do *sensorio-auditivo* na codificação de mensagens pictoideográficas em comunicação alternativa. Neste volume.

HEFFNER, R. M. *General phonetics*. Madison: University of Wisconsin Press, 1950.

HERNANDORENA, C. L. Aquisição da fonologia e implicações teóricas: um estudo sobre as soantes palatais. In: LAMPRECHT, R. (org.) *Aquisição de língua: questões e análises*. Porto Alegre: EDIPUCRS, 1999, p. 81-99.

KELSO, J.; SALTZMAN, E.; TULLER, B. The dynamical perspective on speech production: data and theory. *Journal of Phonetics*, 14, p. 29-59, 1986.

MADDIESON, I. *Patterns of sound*. Cambridge: Cambridge University Press, 1984.

McNEILL, D. *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago Press, 1992.

NAVAS, A. L. Estudo experimental sobre a natureza das representações fonológicas. Neste volume.

PERKELL, J. et al. Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot "motor equivalence" study. *Journal of the Acoustical Society of America*, v. 93, n. 5, p. 2948-2961, 1993.

RIZZOLATTI, G.; ARBIB, M. A. Language within our grasp. *Trends in Neurosciences*, 21, p. 188-194, 1998.

RODRIGUES, N. *Neurolingüística dos distúrbios da fala*. São Paulo: Cortez/Educ, 1989.

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

STEVENS, K. Evidence of the role of acoustic boundaries in the perception of speech sounds. In: FROMKIN, V. (org.). *Phonetic Linguistics: essays in honor of Peter Ladefoged*. New York: Academic Press, 1985, p. 243-255.

——. On the quantal nature of speech. *Journal of Phonetics*, 17, p. 3-45, 1989.