

Modelo Dinâmico Neuronal para a Percepção Categorial da Fala

Estela Bicho, Gregor Schoner, Francisco Vaz

Resumo- O fenómeno da percepção categorial tem desempenhado um importante papel na teoria da percepção da fala. Uma das razões provem do facto da percepção categorial se mostrar como o fenómeno básico e fundamental envolvido no interface entre o sinal analógico sensorial (sinal acústico) e a representação discreta e simbólica da linguagem. Outro fenómeno que se manifesta na percepção de sons de fala é a adaptação selectiva. Esta emerge sob a forma de saltos na fronteira categorial. Para reproduzir e explicar a categorização de sons de fala num *continuum* de vozeamento propomos um modelo dinâmico neuronal. Mostramos que este modelo reproduz os padrões típicos observados na percepção categorial, bem como efeitos de histerese e efeitos de adaptação. Comparamos as predições do modelo com os resultados experimentais. Deste estudo concluímos que a percepção categorial pode ser compreendida como o resultado de um processo de competição dentro de uma representação neuronal da informação sensorial.

Abstract- The phenomenon of categorical perception has played an enormous role in the theory of speech perception. One reason is that categorical perception is at the interface between the analog sensory signal and the discrete and symbolic nature of language. Another phenomenon involved in speech perception is selective adaptation. Adaptation effects appear with respect to the location of the categorical boundary along a speech *continuum*. To reproduce and account the categorization of speech sounds within a voicing *continuum* we propose a dynamic neural model. We show that this model is able to reproduce the typical patterns observed in experiments of categorical perception, hysteresis and adaptation effects. We compare the model predictions to experiments with subjects. From this study we conclude that categorical perception may be understood as resulting from competition within a neural representation of sensory information.

I. INTRODUÇÃO

A. Definição do fenómeno e paradigma clássico

O fenómeno da percepção categorial tem desempenhado um importante papel na teoria da percepção da fala. Uma das razões provem do facto da percepção categorial se mostrar como o fenómeno básico e fundamental envolvido no interface entre o sinal analógico sensorial

(sinal acústico) e a representação discreta e simbólica da linguagem.

Falando genericamente, diz-se que a percepção categorial ocorre quando a capacidade de discriminação entre membros de uma mesma categoria é muito fraca, enquanto que a capacidade de discriminação de elementos pertencentes a categorias diferentes é muito boa [1].

Embora a investigação da percepção categorial seja, em princípio, uma área muito vasta permitindo uma certa variedade de métodos, ao longo dos anos tem sido identificada com um paradigma laboratorial particular. O paradigma clássico da percepção categorial envolve duas tarefas, uma de identificação e a outra de discriminação [2]. Na tarefa de identificação de sons, logo após a apresentação do estímulo, o sujeito tem que responder qual o som percebido. Os estímulos que pertencem a um *continuum* físico "dividido" normalmente em duas categorias, representadas inequivocamente pelos estímulos extremos do *continuum*, são apresentados repetidamente numa ordem aleatória para serem classificados numa ou noutra categoria. Com base nos resultados desta tarefa é possível verificar para cada estímulo a percentagem de respostas de uma determinada categoria e definir a função de identificação. Na tarefa de discriminação de sons usa-se tipicamente o paradigma ABX. Neste paradigma para cada tríade de estímulos ABX, em que os dois primeiros estímulos são diferentes, embora possam vir da mesma categoria fonémica, o ouvinte tem que indicar com qual é que o terceiro estímulo é idêntico. Os resultados são em seguida representados numa função de discriminação que mostra a relação entre a posição no *continuum* e a percentagem de respostas correctas para as tríades.

Como referência experimental recente podemos relatar o trabalho de Castro [3]. Neste trabalho foi usado o paradigma experimental clássico da percepção categorial para testar um *continuum* de vozeamento, baseado na fala natural portuguesa, entre as sílabas /ba/ e /pa/, constituído por estímulos que vão desde /pa/ com 0 ms de pré-vozeamento até /pa/ com 70 ms de pré-vozeamento (ou -70ms de VOT), a intervalos de 7 ms (ou -7 ms de VOT). Os resultados obtidos para um dos sujeitos intervenientes nesta experiência são apresentados na figura 1. Os resultados demonstram claramente a existência de percepção categorial para o *continuum* de vozeamento. Como vemos, a discriminação de sons de fala pertencentes a duas categorias diferentes, tais como /ba/ e /pa/, é relativamente fácil, enquanto que a discriminação

entre dois /pa/ de um mesmo orador é quase inexistente. Por outro lado, diferentes elocuições /pa/ são categorizadas identicamente. Mas, apesar disto as fronteiras entre categorias são flexíveis, ajustando-se com factores tais como contexto fonético, ritmo da fala, experiência linguística, entre outros.

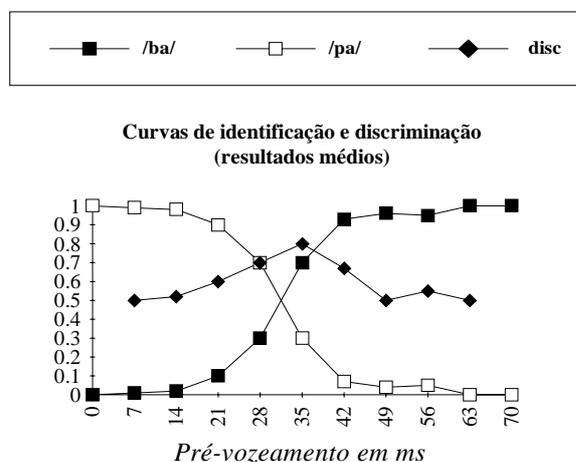


Fig. 1 Curvas de identificação e de discriminação (resultados médios). Em abscissa os estímulos de [pa] com 0 ms de pré-vozeamento a [pa] com 70 ms de pré-vozeamento, a intervalos de 7 ms. Cada ponto da curva de discriminação diz respeito à comparação entre os estímulos imediatamente anterior e subsequente [Castro (1993)].

B. Adaptação Selectiva

Um outro fenómeno envolvido na percepção da fala é a adaptação selectiva. Estes efeitos de adaptação aparecem em relação à localização da fronteira fonémica ao longo do *continuum* de fala. Quando os sujeitos ouvem repetidamente um dos estímulos do *continuum*, que são percebidos como exemplares claros das categorias que representam, o resultado é um salto da fronteira na direcção da categoria perceptual àquela a que o estímulo adaptador pertence [2]. Os efeitos da adaptação selectiva são de origem auditiva e devem-se às relações espectrais entre os estímulos de teste e o adaptador [4].

C. Histerese

Em outras áreas da percepção, a reorganização discreta dos perceptos, à medida que os parâmetros do estímulo são variados continuamente, tem sido relacionada com cooperatividade. Este efeito é caracterizado por regimes multi-estáveis e uma espécie de memória perceptual designada por histerese [5]. Por exemplo, o trabalho em fusão binocular e em percepção visual de movimento revelou efeitos de histerese [6], em que a organização de um percepto depende não só de estímulo corrente, mas também da prévia organização perceptual. Uma questão importante que se levanta é até que ponto a cooperatividade pode desempenhar um papel importante na percepção categorial da fala. Para investigar o efeito de cooperatividade, Castro e seus colaboradores [7], têm

vindo a explorar a existência de histerese na percepção fonémica. O fenómeno de histerese manifesta-se quando numa tarefa de identificação de sons, a série de estímulos consiste numa lista em que o valor do estímulo acústico é apresentado desde um extremo do *continuum* até ao outro, e, se esta série de estímulos é apresentada numa experiência em ordem ascendente e noutra na ordem inversa, então resulta que os pontos no *continuum* onde ocorre uma transição perceptual depende da direcção de apresentação. Em cada caso, a transição perceptual é "atrasada", ou seja a fronteira categorial desloca-se no sentido de apresentação da série, tal que é obtida uma região de sobreposição. Nesta região do *continuum* o percepto formado depende da direcção em que a série é apresentada. A manifestação deste fenómeno vem demonstrar que a emergência de um percepto pode depender não só do estímulo corrente mas também do percepto anterior.

D. A motivação da teoria dinâmica não linear

As evidências experimentais mostram que a formação perceptual de padrões (ou de perceptos) e transições entre perceptos exibem um comportamento dinâmico. Em particular, os conceitos de estabilidade e cooperatividade são essenciais para um bom entendimento do modo como são formados perceptos coerentes. Estabilidade e cooperatividade são também conceitos centrais da teoria dinâmica não linear e da teoria da formação de padrões. Consequentemente, há muitas ideias da dinâmica que poderiam ser desenvolvidas para a percepção. A teoria dinâmica revela-se ainda uma bordagem adequada para lidar com os efeitos da adaptação selectiva. Isto acontece, porque o processo de adaptação selectiva e o processo de organização perceptual propriamente dito ocorrem em níveis diferentes e as suas escalas temporais são muito diferentes, o que permite desacoplar com relativa facilidade estes dois processos dentro da teoria dinâmica.

E. Hipóteses de aplicações do modelo

Numa perspectiva bastante ambiciosa e a longo termo, poderíamos explorar a hipótese de integrar o modelo dinâmico neuronal, que aqui apresentamos, num sistema de reconhecimento automático de fala, ao nível da categorização elementar dos sons de fala.

Um sistema de reconhecimento automático de fala é bastante complexo e deve envolver vários níveis de processamento, desde a segmentação e categorização fonémica (porque a fala pode ser interpretada como uma sequência de fonemas) até à interpretação semântica. A filosofia do modelo que propomos poderia ser usada para implementar o nível de segmentação do sinal de fala numa sequência de fonemas. À partida perspectivamos as seguintes vantagens: O problema da invariância poderia ser parcialmente reduzido, pois as fronteiras perceptuais (ou categoriais) do modelo são flexíveis. Summerfield [8] descobriu que quando o ritmo da fala aumenta, as

diferenças entre as consoantes vozeadas e não-vozeadas diminuem em relação ao VOT, e observou que concomitantemente a localização da fronteira perceptual que marca a distinção entre as consoantes vozeadas e não vozeadas desloca-se na direcção de valores de VOT menores. Adicionalmente, resolve-se também o problema para vários oradores. Isto porque, devido aos saltos nas fronteiras sob efeitos de adaptação o nível de categorização fonémica seria capaz de se “adaptar” a um orador particular. Esta propriedade poderia ser também considerada para sistemas de reconhecimento de orador. Relativamente aos efeitos de contexto, estes poderiam ser também já tratados a este nível como uma espécie de memória perceptual.

II. MODELO

A. Estrutura conceptual

O modelo que propomos consiste em três níveis de representação como é ilustrado na figura 2 e que passamos a descrever.

a) *Nível de VOT*

O nível de entrada (ou de VOT) é o nível de representação mais baixo. Neste, é feita uma representação discreta dos estímulos acústicos através de um conjunto de detectores acústicos equidistantes colocados ao longo do eixo de VOT. Cada um destes detectores acústicos sinaliza a presença de um estímulo acústico com um valor de VOT particular.

b) *Nível auditivo*

O nível intermédio, ou auditivo, é onde ocorrem os efeitos de adaptação. O nível de representação do nível auditivo consiste em três camadas de neurónios diferentes, como é ilustrado na figura 2. Cada neurónio da camada Z, camada de entrada do nível auditivo, é especializado numa parte particular do eixo de VOT e assume-se que há sobreposição. Estes neurónios são excitados proporcionalmente à sua sensibilidade ao valor do estímulo acústico, sinalizado por um dos detectores do nível de VOT. Isto conduz em geral à situação em que mais do que um neurónio da camada Z está activo. Desta maneira é necessário um processo de normalização. Este processo é realizado na camada Y, em que todos os neurónios estão desligados excepto o neurónio correspondente ao maior z_i . Quando dois (ou mais) neurónios da camada Z tem o mesmo valor de excitação e este é o máximo valor então na camada Y dois (ou mais) neurónios ficam ligados. Os neurónios y_i possuem um limiar mínimo de activação porque na ausência do estímulo acústico não queremos que eles fiquem ligados devido a quaisquer pequenas flutuações nos neurónios de entrada.

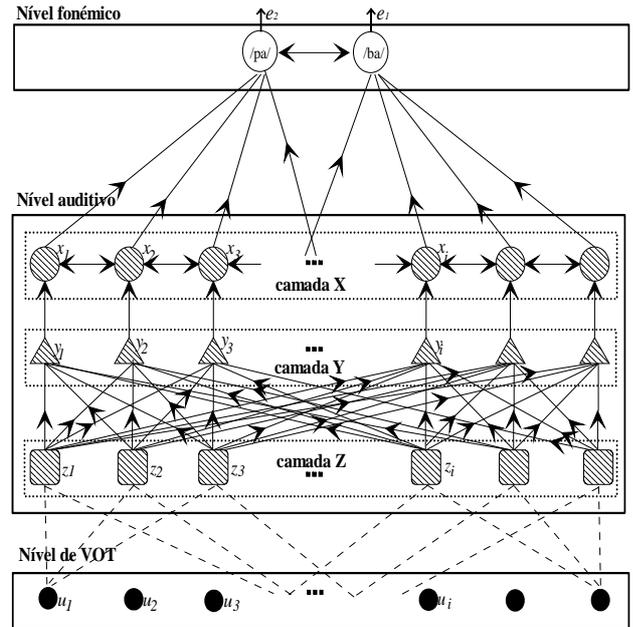


Fig. 2 Níveis de representação para o modelo dinâmico neuronal para a percepção categorial e adaptação selectiva

Consideramos que a saída do nível auditivo define um espaço contínuo (caracterizado por um parâmetro a). Mas, este espaço é amostrado através de um conjunto de detectores equidistantes ao longo do eixo a , a figura 3 ilustra este processo. Os neurónios x_i representam estes detectores e o índice i rotula a_i . $x_i = \pm 1$ representa uma detecção no nível auditivo do valor a_i e $x_i = 0$ representa a ausência de detecção deste valor.

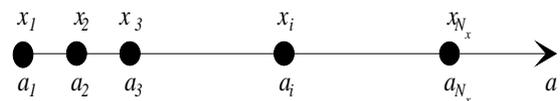


Fig. 3 Amostragem do espaço definido pelo parâmetro a

Os neurónios x_i constituem a última camada do nível auditivo e formam um campo dinâmico de neurónios que competem globalmente. Esta competição conduz à activação dominante de um único neurónio da camada X, este é o neurónio que na situação corrente melhor representa o estímulo acústico. O papel desta camada é particularmente importante na situação em que mais do que um neurónio da camada Y está ligado. Esta situação irá ocorrer em consequência dos efeitos de adaptação.

Como podemos ver, o nível auditivo é basicamente um filtro complicado que na ausência de adaptação iria conduzir a uma relação unívoca entre os valores do estímulo acústico (VOT), sinalizados por u_i e a_i . Na figura 4a) ilustra-se o mapeamento dos valores de VOT em valores de a na ausência de adaptação.

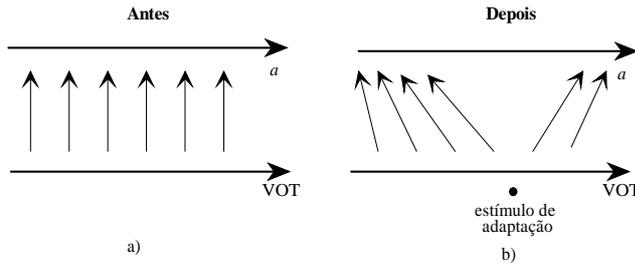


Fig 4 Em a) ilustra-se o mapeamento dos valores de VOT em valores de a na ausência de efeitos de adaptação. E em b) ilustra-se o mesmo mapeamento após a adaptação

A adaptação é introduzida através de uma distorção sistemática deste mapa, tal que, após alguns períodos de detecção de um valor particular a_i , o seu valor de VOT correspondente (estímulo de adaptação) e valores de VOT vizinhos são mapeados em valores de a que estão mais afastados do valor de a_i do que antes da adaptação, ver figura 4.b). Uma maneira simples de realizar isto consiste em reduzir a sensibilidade dos neurónios da camada Z que estão activos, isto é, enfraquecer a força do seu acoplamento ao nível de VOT. Esta variação da sensibilidade é dinâmica se assumirmos que é gradual durante a adaptação e durante a recuperação. Um aspecto importante a ter em mente é que para que os efeitos da adaptação sejam sentidos ao longo das simulações, a recuperação da sensibilidade tem que ser muito mais lenta do que a sua diminuição.

c) *Nível fonémico*

O nível de representação mais elevado é o nível fonémico, onde a saída das operações realizadas no nível auditivo são organizadas em termos de um percepto fonémico. Este nível consiste em dois neurónios dinâmicos que recebem estimulação do nível auditivo. Os neurónios competem entre si e a sua mútua activação exclusiva representa um estado perceptual, que conduz a qualquer uma das duas respostas categoriais.

B. *Formalização matemática*

a) *Nível de VOT*

Os detectores do nível de VOT são simplesmente células do tipo ligadas/desligadas que sinalizam a presença de um valor particular do estímulo acústico:

$$u_i = \begin{cases} 1 & \text{se estímulo} = VOT_i \\ 0 & \text{se estímulo} \neq VOT_i \end{cases} \quad (1)$$

$u_i = 1$ sinaliza a presença de um estímulo acústico com valor VOT_i . Neste procedimento está implícito uma amostragem do eixo de VOT pelas células u_i . O índice i rotula o valor de VOT_i .

b) *Nível auditivo*

Cada neurónio z_i é sensível a uma zona particular do eixo de VOT e não responde equivalentemente a todos os

valores compreendidos dentro do seu campo receptivo. A função de sensibilidade de um neurónio z_i é dada por

$$v_i = \begin{cases} V_{0i} \left(1 - \frac{2(CRF_i - VOT)}{RF} \right) & \text{se } CRF_i - \frac{RF}{2} < VOT < CRF_i \\ V_{0i} \left(1 + \frac{2(CRF_i - VOT)}{RF} \right) & \text{se } CRF_i < VOT < CRF_i + \frac{RF}{2} \\ 0 & \text{caso contrário} \end{cases} \quad (2)$$

Onde,

V_{0i} representa o valor máximo de sensibilidade

RF é o campo receptivo destes neurónios

CRF_i é o centro do campo receptivo do neurónio z_i

Os neurónios z_i são excitados pelas células do nível de VOT do seguinte modo

$$z_i = v_{il} u_l \quad \text{para } i = 1, 2, \dots, N_x \wedge l = 1, 2, \dots, N_x \quad (3)$$

onde v_{il} representa a sensibilidade do neurónio z_i ao valor do estímulo acústico representado por u_l .

Na figura 5 é ilustrado o padrão de excitação, na ausência de adaptação, gerado nos neurónios da camada Z quando a célula u_6 do nível de VOT está ligada.

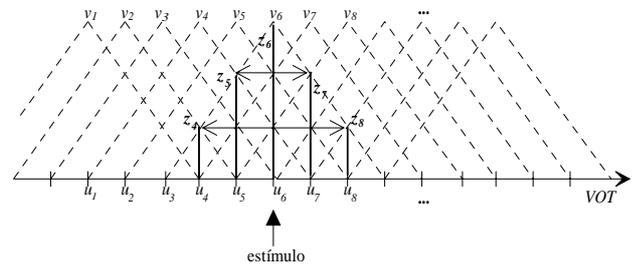


Fig 5 Padrão de excitação, na ausência de adaptação, gerado nos neurónios da camada Z quando a célula u_6 do nível de VOT está ligada.

Os neurónios da camada Y são excitados do seguinte modo:

$$y_i = \begin{cases} 1 & \text{se } z_i \text{ é o máximo valor } \wedge \text{ é superior ao limiar } l_y \\ 0 & \text{caso contrário} \end{cases} \quad (4)$$

onde l_y representa o valor do limiar mínimo de activação destes neurónios. Na figura 5 é ilustrada a função realizada por esta camada.

A camada de saída do nível auditivo consiste num campo dinâmico não-linear de neurónios, identificados por x_i , que competem globalmente. A dinâmica do seu comportamento é descrita por um sistema de equações diferença, que na forma matricial é dada por:

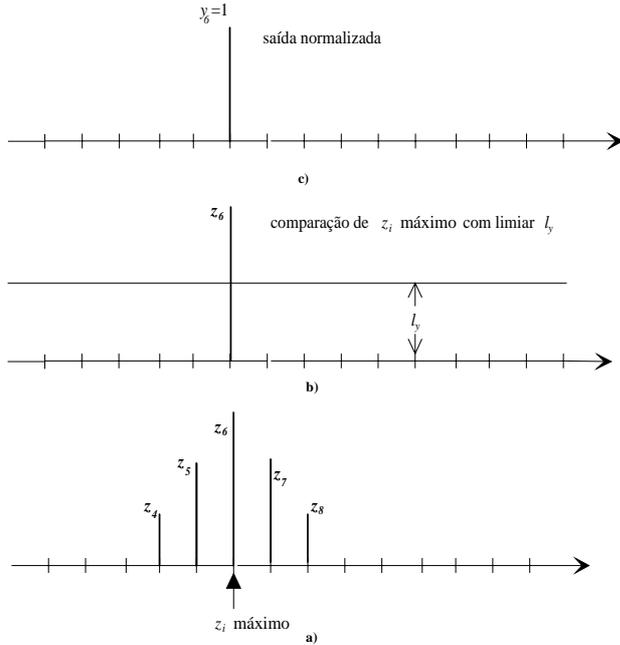


Fig 6 a) Seleção do neurónio z_i com a máxima excitação. b) Comparação de z_i máximo com o limiar l_y . Ilustração da função realizada pela camada Y. c) saída da camada Y do nível auditivo

$$\mathbf{x}(n+1) = \mathbf{x}(n) + \mathbf{h}(\mathbf{y}, \mathbf{x}(n)) \quad [-1, 1] \text{ mod } 2 \quad (5)$$

$\mathbf{x}(n+1)$ e $\mathbf{x}(n)$ são vectores coluna, de dimensão N_x que representam o estado dos neurónios da camada X nos instantes $n+1$ e n respectivamente; $\mathbf{h}(\mathbf{y}, \mathbf{x}(n))$ é o campo de vectores total e \mathbf{y} é um vector coluna que representa a entrada da camada X. \mathbf{y} e $\mathbf{h}(\mathbf{y}, \mathbf{x}(n))$ são também vectores de dimensão N_x . Nós restringimos cada variável x_i ao espaço de fase circular $[-1, 1] \text{ mod } 2$ de modo a ter soluções limitadas. Por conseguinte, apenas pontos fixos no intervalo $[-1, 1]$ são importantes.

O campo de vectores $\mathbf{h}(\mathbf{y}, \mathbf{x}(n))$ foi definido de modo a capturar as especificações do problema através de forças componentes que definem atratores ou repulsores do sistema dinâmico. No projecto de tais forças componentes fizemos uso da linguagem da teoria qualitativa de sistemas dinâmicos [9, 10, 11]. As contribuições das várias forças componentes para o campo de vectores total são tratadas aditivamente. As contribuições individuais são definidas de modo a gerar o diagrama de fase desejado, caracterizado por soluções invariantes e respectivas estabilidades, na ausência de todas as outras contribuições [12].

O campo de vectores $\mathbf{h}(\mathbf{y}, \mathbf{x}(n))$ é a soma de três forças componentes:

$$\mathbf{h}(\mathbf{y}, \mathbf{x}(n)) = \mathbf{h}_{\text{info}}(\mathbf{y}, \mathbf{x}(n)) + \mathbf{h}_{\text{comp}}(\mathbf{x}(n)) + \mathbf{h}_{\text{stoch}}(n) \quad (6)$$

onde $\mathbf{h}_{\text{info}}(\mathbf{y}, \mathbf{x}(n))$ representa a informação que transita da camada Y para a camada X, dentro do nível auditivo. $\mathbf{h}_{\text{comp}}(\mathbf{x}(n))$ é a componente que determina a competição

entre os neurónios da camada X. Finalmente, $\mathbf{h}_{\text{stoch}}(\mathbf{x}(n))$ representa a natureza estocástica dos neurónios x_i . Para o neurónio i estas várias componentes são dadas por:

$$h_{\text{info},i}(y_i, x_i(n)) = \varepsilon(y_i)(x_i(n) - x_i^3(n)) \quad (7)$$

$$h_{\text{comp},i}(\mathbf{x}(n)) = \gamma_x \sum_{j=1(j \neq i)}^{N_x} x_j^2(n)(x_i^3(n) - x_i(n)) \quad (8)$$

$$h_{\text{stoc},i}(n) = \sqrt{Q_x} \xi_x(n) \quad (9)$$

onde $\varepsilon(y_i)$ é a função de entrada do neurónio x_i e tem a forma

$$\varepsilon(y_i) = y_i - l_e \quad (10)$$

em que l_e tem que ser escolhido de modo a que tenhamos $\varepsilon(y_i) > 0$ se y_i estiver ligado e $\varepsilon(y_i) < 0$ se y_i estiver desligado. γ_x é o parâmetro do modelo que determina a força da componente de competição. Q_x é a magnitude da natureza estocástica dos neurónios da camada X. E, por fim $\xi_x(n)$ representa ruído branco gaussiano. A introdução desta componente permite a fuga a pontos fixos instáveis.

Os neurónios x_i que tem uma função de entrada negativa, $\varepsilon(y_i) < 0$, ficam inibidos e os que possuem uma entrada positiva ficam activados. Mas no final, devido ao processo de competição apenas um destes neurónios vai ficar ligado. São dois os factores que determinam qual dos neurónios, que têm uma função de entrada positiva, vai ganhar a competição. O primeiro é a excitação inicial dos neurónios, $x_i(0)$, e o segundo é a sua natureza estocástica.

Os fenómenos de adaptação vão ocorrer na camada Z. Como já vimos uma forma de lidar com a adaptação consiste em enfraquecer a sensibilidade de todos os neurónios z_i que têm o estímulo de entrada dentro do seu campo receptivo. Isto é conseguido fazendo o parâmetro V_{ok} dinâmico e diferente para cada neurónio z_i . A dinâmica de V_{ok} pode muito simplesmente ser dada pela seguinte equação de diferença:

$$V_{ok}(n+1) = V_{ok}(n) - \alpha'_{\text{adap}}(V_{ok}(n) - V_{0\text{min}}) - \alpha'_{\text{recup}}(V_{ok}(n) - V_{0\text{max}}) \quad (11)$$

onde

α'_{adap} representa a taxa efectiva de decaimento de V_{ok} .

α'_{recup} é a taxa efectiva de recuperação de V_{ok} .

α'_{adap} e α'_{recup} são dados pelo produto de duas componentes:

$$\alpha'_{\text{adap}} = z_i^2 \alpha_{\text{adap}} \quad (12)$$

$$\alpha'_{\text{recup}} = (1 - z_i^2) \alpha_{\text{recup}} \quad (13)$$

em que α_{adap} e α_{recup} são parâmetros que determinam a taxa de decaimento e recuperação de V_{ok} respectivamente. Como vemos, para a adaptação o valor de z_i também é importante, porque um dado adaptador não provoca a

mesma adaptação em todos os neurónios z_i . Os que sofrem os maiores efeitos de adaptação são aqueles que num dado momento são mais sensíveis ao estímulo adaptador e por conseguinte os que ficam mais excitados.

Na figura 7 ilustramos o perfil das funções de sensibilidade após a adaptação com um estímulo de valor VOT_4 , o padrão de excitação gerado nos neurónios da camada Z em resultado dos efeitos de adaptação e no topo da figura podemos observar que o neurónio da camada Y que agora fica ligado é o y_8 . É interessante comparar o padrão de excitação gerado nos neurónios da camada Z na ausência e na presença de feitos de adaptação.

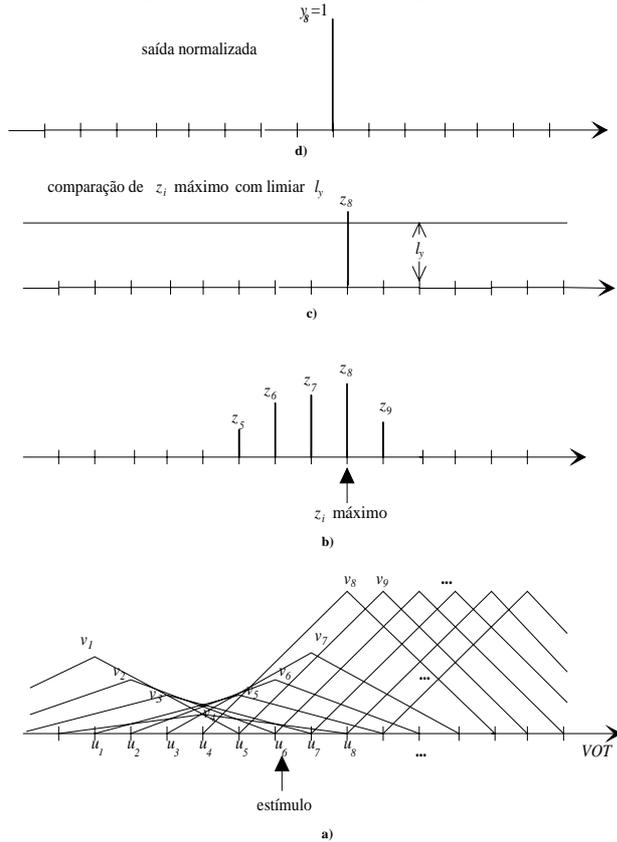


Fig 7 a) Funções de sensibilidade após a adaptação com VOT_4 . b) padrão de excitação gerado sob os efeitos de adaptação nos neurónios da camada Z quando a célula u_6 do nível de VOT está ligada e saída da camada Y. c) Comparação de z_i máximo com limiar l_y . d) Saída da camada Y.

c) *Nível fonémico*

A saída de cada neurónio é identificada pela variável perceptual e_i , onde i é um índice que rotula fonemas. Um código que achamos conveniente no projecto da dinâmica do sistema é o seguinte: quando o sistema relaxa para um ponto fixo de e_i , perto de +1 ou -1 então o estado perceptivo é o fonema indexado por i (neurónio i ligado) e quando o sistema relaxa para um ponto fixo de e_i perto de zero então o fonema correspondente não é "percebido" (neurónio i desligado). Este comportamento é modelado por equações dinâmicas para as variáveis perceptuais que têm a seguinte forma:

$$e_i(n+1) = e_i(n) + g_i(\mathbf{e}(n), \mathbf{x}), \quad e_i \in [-1, 1] \text{ mod } 2 \quad i = 1, 2 \quad (14)$$

onde $\mathbf{e}(n)$ exprime a dependência do campo de vectores total g_i em relação aos valores das variáveis perceptuais no instante n , e \mathbf{x} é o vector de saída do nível auditivo. Este sistema de equações dinâmicas começa com uma condição inicial $\mathbf{e}(0)$, que nas simulações é escolhida de modo a representar a história perceptual prévia. Nós restringimos e_i ao espaço de fase circular $[-1, 1] \text{ mod } 2$ de modo a ter soluções limitadas. Como consequência, apenas pontos fixos no intervalo $[-1, 1]$ são importantes.

O campo de vectores $g_i(\mathbf{e}(n), \mathbf{x})$ é a soma de três forças componentes:

$$g_i(\mathbf{e}(n), \mathbf{x}) = g_{\text{info},i}(e_i(n), \mathbf{x}) + g_{\text{comp}}(\mathbf{e}(n)) + g_{\text{stoch},i}(n), \quad i = 1, 2 \quad (15)$$

onde:

$g_{\text{info},i}(e_i(n), \mathbf{x})$ representa a contribuição da informação auditiva na dinâmica do nível fonémico e tem a forma:

$$g_{\text{info},i}(e_i(n), \mathbf{x}) = \beta_i(\mathbf{x})(e_i^3(n) - e_i(n)), \quad i = 1, 2 \quad (16)$$

onde $\beta_i(\mathbf{x})$ é a função de entrada do neurónio fonémico i e define o seu grau de activação em relação ao valor corrente detectado no nível auditivo, i.e do parâmetro a . A função de acoplamento entre os níveis auditivo e fonémico tem a seguinte forma

$$\beta_i(\mathbf{x}) = \sum_{k=1}^{N_x} w_{ik} |x_k| - t_e \quad \text{para } i = 1, 2 \quad (17)$$

Nesta equação

t_e é um limiar que representa o potencial em repouso dos neurónios do nível fonémico. O valor deste limiar determina o tempo de relaxação do sistema para o estado em que os dois neurónios fonémicos ficam desligados, o que deve acontecer na ausência de estímulo acústico.

w_{ik} é a intensidade da ligação entre o neurónio e_i do nível fonémico e o neurónio x_k do nível auditivo. O seu valor define a sensibilidade do neurónio fonémico e_i ao valor a_k .

A função de sensibilidade dos neurónios fonémicos aos valores do parâmetro auditivo a é definida do seguinte modo:

$$w = \begin{cases} w_0 + \frac{w_0 - h}{a_{\text{sup}} - a_{c1}} (a - a_{\text{sup}}) & \text{se } i = 1 \\ w_0 - \frac{w_0 - h}{a_{c2} - a_{\text{inf}}} (a - a_{\text{inf}}) & \text{se } i = 2 \end{cases} \quad (18)$$

onde a_{sup} e a_{inf} representam os limites superior e inferior do espaço definido por a respectivamente. a_{c1} e a_{c2} são os pontos onde vão ocorrer as fronteiras de estabilidade ao longo do *continuum* e o parâmetro h , como iremos ver na próxima secção, é o responsável pelo posicionamento das fronteiras de estabilidade ao longo do *continuum*. Este ser

escolhido de modo a que estas ocorram em a_{c1} e a_{c2} . $g_{comp,i}(\mathbf{e}(n))$ representa a dinâmica intrínseca que consiste na competição entre os dois neurónios. Esta componente tem a forma:

$$g_{comp,i}(\mathbf{e}(n)) = \gamma_o \gamma_j e_j^2(n) (e_i^3(n) - e_i(n)), \quad i, j = 1, 2 \wedge i \neq j \quad (19)$$

onde γ_o representa a magnitude da dinâmica de competição e $\gamma_j > 0$ reflecte o grau de competição entre neurónios. As ligações entre os neurónios são puramente inibitórias. Por exemplo, isto significa que a activação do neurónio fonémico 1 (i.e. /ba/), pode inibir a activação do neurónio fonémico 2 (i.e. /pa/).

Nós admitimos uma natureza estocástica para a saída dos neurónios que é modelada por uma força estocástica gaussiana:

$$g_{stoch,i}(n) = \sqrt{Q} \xi_i(n), \quad i = 1, 2 \quad (20)$$

onde $\xi_i(n)$ é ruído branco gaussiano que não está correlacionado entre passos subsequentes e $Q > 0$ é a força do ruído. A inclusão desta força estocástica permite explicar a variabilidade no desempenho perceptual, transições espontâneas entre perceptos e fugas de perceptos instáveis, i.e., pontos fixos instáveis.

C. Estudo analítico

A seguir apresentamos os resultados dos estudos analíticos realizados para o campo dinâmico não linear de neurónios que constitui a camada X, para a dinâmica da adaptação selectiva que ocorre na camada Z do nível auditivo e para o nível fonémico. Nestes estudos aplicámos os métodos *standard* da teoria de sistemas dinâmicos. Primeiro determinámos os pontos fixos e em seguida avaliámos a sua estabilidade em relação aos vários parâmetros das equações.

a) Estudo analítico para a camada X do nível auditivo

A componente determinística para o campo de vectores total é dada pela soma das componentes de informação e de competição. Por conseguinte temos

$$h_{dt,i}(y_i, x_i(n)) = h_{info,i}(y_i, x_i(n)) + h_{comp,i}(\mathbf{x}(n)) = \left(-\alpha(y_i) + \gamma_x \sum_{j=1, j \neq i}^{N_x} x_j^2(n) \right) \left(x_i^3(n) - x_i(n) \right) [-1, 1] \text{ mod } 2 \text{ para } i = 1, \dots, N_x \wedge i \neq j \quad (21)$$

Esta equação define pontos fixos em $\mathbf{x}_{fp}^{(0)} = (0, 0, \dots, 0)_{N_x}$ (todos os elementos são nulos), $\mathbf{x}_{fp}^{(k)} = (0, \dots, 0, \pm 1, 0, \dots, 0)_{N_x}$ (todos os elementos são nulos excepto o k ésimo que é ± 1) e $\mathbf{x}_{fp}^{(k)}$ (qualquer combinação de 0's e 1's e diferentes dos dois tipos anteriores). Segundo a teoria linear de estabilidade temos que:

- Uma condição necessária e suficiente para que $\mathbf{x}_{fp}^{(0)} = (0, 0, \dots, 0)_{N_x}$ seja um estado estável é $-2 < \varepsilon(y_i) < 0 \quad \forall i = 1, 2, \dots, N_x \quad (22)$

Esta condição de estabilidade é satisfeita quando os neurónios da camada X não recebem qualquer estimulação dos neurónios na camada Y, situação em que temos $y_i = 0 \quad \forall i = 1, \dots, N_x$. Isto acontece na ausência de estímulo acústico.

- Uma condição necessária para que $\mathbf{x}_{fp}^{(k)} = (0, \dots, 0, \pm 1, 0, \dots, 0)_{N_x}$, detecção com sucesso do valor a_k , seja um estado estável é

$$\gamma_x - 2 < \alpha(y_i) < \gamma_x \quad \forall i = 1, 2, \dots, N_x \wedge i \neq k \wedge 0 < \alpha(y_k) < 1 \quad (23)$$

- Uma condição suficiente para que $\mathbf{x}_{fp}^{(k)}$ (vector com qualquer combinação de 0's e 1's e diferentes dos dois tipos anteriores) sejam pontos fixos instáveis é

$$\varepsilon(y_i) < \gamma \quad (24)$$

b) Estudo analítico para a dinâmica da adaptação selectiva

A equação que descreve a dinâmica dos fenómenos de adaptação que ocorrem na camada Z do nível auditivo é definida pela expressão (11). Quando o neurónio z_i contém o estímulo acústico dentro do seu campo receptivo a dinâmica do seu parâmetro V_{oi} reduz-se a

$$V_{oi}(n+1) = V_{oi}(n) - \alpha_{adap} z_i^2 (V_{oi}(n) - V_{o\ min}) \quad (25)$$

Esta equação tem como ponto fixo $V_{oi} = V_{o\ min}$ e este é estável se $0 < \alpha_{adap} < 2$. Nesta situação a sensibilidade do neurónio z_i enfraquece. Quando pelo contrário, o estímulo acústico não cai dentro do seu campo receptivo a dinâmica do parâmetro V_{oi} é descrita por

$$V_{oi}(n+1) = V_{oi}(n) - \alpha_{recup} (1 - z_i^2) (V_{oi}(n) - V_{o\ max}) \quad (26)$$

Esta equação tem como ponto fixo $V_{oi} = V_{o\ max}$ e este é um atractor se $0 < \alpha_{recup} < 2$. Agora, assiste-se a uma recuperação da sensibilidade do neurónio z_i aos estímulos que caem dentro do seu campo receptivo.

c) Estudo analítico para o nível fonémico

A componente determinística do campo de vectores total é a soma das componentes de informação e competição. Desta maneira temos

$$g_{dt,i}(\mathbf{x}, \mathbf{e}(n)) = g_{info,i}(\mathbf{x}, \mathbf{e}(n)) + g_{comp,i}(\mathbf{e}(n)) = \left(\beta_i(\mathbf{x}) + \gamma_o \gamma_j e_j^2(n) \right) \left(e_i^3(n) - e_i(n) \right) [-1, 1] \text{ mod } 2 \text{ para } i = 1, 2 \wedge i \neq j \quad (27)$$

Esta equação define pontos fixos em $\mathbf{e}_{fp} = [0 \ 0]^T, [0 \ \pm 1]^T, [\pm 1 \ 0]^T, [\pm 1 \ \pm 1]^T$ para todos os valores de a_k detectados no nível auditivo e
$$\left[\pm \sqrt{\frac{-\beta_2(\mathbf{x}_{fp}^{(k)})}{\gamma_0 \gamma_{21}}} \quad \pm \sqrt{\frac{-\beta_1(\mathbf{x}_{fp}^{(k)})}{\gamma_0 \gamma_{12}}} \right]^T \text{ se } a_k \in]a_{c1}, a_{c2}[.$$

Segundo a teoria linear de estabilidade temos que:

- Uma condição necessária para $\mathbf{e}_{fp} = [0 \ 0]^T$ ser um atractor é

$$0 < t_e < 2 \tag{28}$$

- Uma condição necessária para $\mathbf{e}_{fp} = [0 \ \pm 1]^T$ ser um possível estado estável é

$$a_{c1} > \frac{(h-t_e-\gamma_0\gamma_{12})a_{\max} + (1+t_e-h)a_{\min} + (w_0-t_e-1)a_{c2}}{w_0-t_e-\gamma_0\gamma_{12}}$$

$$a_{c1} > \frac{(h-t_e-\gamma_0\gamma_{12}-2)a_{\max} + (t_e-h)a_{\min} + (w_0-t_e)a_{c2}}{2+w_0-t_e-\gamma_0\gamma_{12}} \tag{29}$$

- Uma condição necessária para $\mathbf{e}_{fp} = [\pm 1 \ 0]^T$ ser um possível estado estável é

$$a_{c1} > \frac{(h-t_e-1)a_{\max} + (\gamma_0\gamma_{21}+t_e-h)a_{\min} + (w_0-t_e-\gamma_0\gamma_{21})a_{c2}}{h-t_e-1}$$

$$a_{c1} > \frac{(h-t_e)a_{\max} + (\gamma_0\gamma_{21}+t_e-h-2)a_{\min} + (2+w_0-t_e-\gamma_0\gamma_{21})a_{c2}}{w_0-t_e} \tag{30}$$

- Uma condição suficiente para que $\mathbf{e}_{fp} = [\pm 1 \ \pm 1]^T$ sejam repulsores é

$$a_{c1} > \frac{(h-t_e-\gamma_0\gamma_{12}-1)a_{\max} + (1+\gamma_0\gamma_{21}+t_e-h)a_{\min} + (w_0-t_e-\gamma_0\gamma_{21}-h)a_{c2}}{w_0-t_e-\gamma_0\gamma_{12}-1}$$

$$a_{c1} > \frac{(h-t_e-\gamma_0\gamma_{12})a_{\max} + (\gamma_0\gamma_{21}+t_e-h)a_{\min} + (w_0-t_e-\gamma_0\gamma_{21})a_{c2}}{w_0-t_e-\gamma_0\gamma_{12}} \tag{31}$$

- Os pontos fixos $\left[\pm \sqrt{\frac{-\beta_2(\mathbf{x}_{fp}^{(k)})}{\gamma_0 \gamma_{21}}} \quad \pm \sqrt{\frac{-\beta_1(\mathbf{x}_{fp}^{(k)})}{\gamma_0 \gamma_{12}}} \right]^T$

existem apenas se

$$a_k \in \left] \frac{(t_e-h)a_{\max} + (w_0-t_e)a_{c1}}{w_0-h}, \frac{(t_e-h)a_{\min} + (w_0-t_e)a_{c2}}{w_0-h} \right[$$

e no nosso espaço de fase são repulsores.

D. Parâmetros do modelo

a) Estabelecimento dos parâmetros para o nível fonémico

Os parâmetros do modelo a ser fixados são: $a_{inf}, a_{sup}, a_{c1}, a_{c2}, \gamma_0, \gamma_{12}, \gamma_{21}, w_0, t_e, h, Q$ e N_{ite} . Os critérios que se seguem são úteis para a escolha de um conjunto básico de valores:

- O VOT deve ser transposto para o parâmetro a do modelo de modo a que num extremo da escala o sistema é monoestável, no outro extremo é também monoestável e no meio o sistema é biestável, ou seja
 - regimes monoestáveis para $a \in [a_{inf}, a_{c1}]$ e $a \in [a_{c2}, a_{sup}]$ e
 - regime biestável para $a \in [a_{c1}, a_{c2}]$
- A escolha deve ser tal que implique de algum modo uma simetria. Podemos escolher por exemplo, $a_{inf} = -1, a_{sup} = 1, a_{c1} = -0.5, a_{c2} = 0.5$ e $\gamma_{12} = \gamma_{21} = 1$. A simetria imposta por esta escolha de valores, para estes parâmetros, torna a escolha dos outros parâmetros mais elegante.

- Por facilidade de implementação nós usamos dinâmica discreta no tempo. Um preço a pagar por esta facilidade é que a escala de tempo não é fixa. Isto acontece porque o tempo correspondente a uma iteração, no computador, é arbitrário. Deste modo, as constantes de relaxação no modelo não têm uma base fisiológica. A constante de tempo de relaxação, " τ_{rel} ", pretendida para cada atractor deve ser relativamente pequena. Escolhemos $\tau_{rel} \approx 5$ porque este valor implica uma relação razoável entre o tempo de relaxação e o tempo de uma iteração no computador.

- Devem ocorrer transições estocásticas, isto é, queremos que ocorram transições perceptuais espontâneas devido a flutuações nas variáveis perceptuais. Isto é importante porque os estímulos ambíguos conduzem à biestabilidade e nós queremos modelar este comportamento.

- Do critério anterior determinamos que o desvio padrão (SD), ou variabilidade, permitida para as variáveis nos estados estacionários deve ser aproximadamente 0.1, mas pode variar um pouco. Este valor foi escolhido empiricamente. Para $SD \approx 0.1$, as transições estocásticas nas simulações aparecem com uma probabilidade adequada.

Após escolhidos os parâmetros $a, a_{c1}, a_{c2}, \gamma_{12}, \gamma_{21}$, o tempo de relaxação e a variabilidade em regime estacionário para as variáveis perceptuais, todos os outros parâmetros do modelo podem ser estimados. Neste processo, tomamos como base alguns argumentos que fundamentam as acções realizadas.

Argumentos:

1. Condições para regimes estáveis e instáveis:

A condição necessária para que $[0 \pm 1]^T$ seja um estado estável é dada por (29). Substituindo os parâmetro já estabelecidos nesta equação podemos reescreve-la da seguinte forma

$$a_{c1} > \frac{0.5w_0 - 2.5t_e + 2h - \gamma_0 - 1.5}{w_0 - t_e - \gamma_0}$$

$$a_{c1} > \frac{0.5w_0 - 2.5t_e + 2h - \gamma_0 - 2}{2 + w_0 - t_e - \gamma_0} \quad (32)$$

A condição necessária para que $[\pm 1 \ 0]^T$ seja um estado estável é dada por (30). Substituindo os parâmetro já estabelecidos nesta equação obtemos

$$a_{c1} > \frac{0.5w_0 - 2.5t_e + 2h - 1.5\gamma_0 - 1}{h - t_e - 1}$$

$$a_{c1} > \frac{0.5w_0 - 2.5t_e + 2h - 1.5\gamma_0 + 3}{w_0 - t_e} \quad (33)$$

A condição necessária para $[0 \ 0]^T$ ser um estado estável é dada por (6.22).

A condição suficiente para que $[\pm 1 \ \pm 1]^T$ seja um estado instável toma agora a seguinte forma

$$a_{c1} > \frac{0.5w_0 - 2.5t_e + 1.5h - 1.5\gamma_0 - 2}{w_0 - t_e - \gamma_0 - 1}$$

$$a_{c1} > \frac{0.5w_0 - 2.5t_e + 2h - 2.5\gamma_0}{w_0 - t_e - \gamma_0} \quad (34)$$

2. Tempo de relaxação

O tempo de relaxação para um ponto fixo é dado por

$$\tau_{rel} \approx \frac{1}{|\lambda_A|} \quad (35)$$

onde $|\lambda_A|$ representa o maior valor próprio da matriz Jacobiana \mathbf{A} , do campo de vectores. Como escolhemos $\tau_{rel} \approx 5$ vem que $|\lambda_A| \approx 0.2$. Isto implica que o valor próprio correspondente associado à equação do mapa dinâmico é

$$\lambda = 1 - |\lambda_A| \approx 0.8 \quad (36)$$

3. Relaxação para $[0 \ 0]^T$

Neste ponto fixo os valores próprios de $\mathbf{I+A}$, onde \mathbf{A} é a matriz Jacobiana da equação (15) (campo de vectores para o nível fonémico) são dados por

$$\lambda_i = 1 - t_e \quad \text{para } i = 1, 2 \quad (37)$$

Agora como para este estado pretendemos $\tau_{rel} \approx 5$ das equações (35) e (36) tiramos que

$$\lambda_{max} = 0.8 \quad (38)$$

Como $\lambda_1 = \lambda_2 = \lambda_{max}$ da equação (37) obtemos

$$t_e = 0.2 \quad (39)$$

que satisfaz a condição (28).

4. $[\pm 1, 0]^T$ deve ser estável para $a_k > a_{c1}$

Neste ponto os valores próprios de $\mathbf{I+A}$ são

$$\lambda_1 = 1 - 2\beta_1(a_k) \quad (40)$$

$$\lambda_2 = 1 - 2\beta_2(a_k) - \gamma_0 \quad (41)$$

Em $a_k = a_{c1}$ temos um ponto de bifurcação pois para este valor do parâmetro auditivo o estado $[\pm 1, 0]^T$ torna-se instável. Se escolhermos para esta instabilidade o tipo fonte (*source*), então devemos fazer com que em $a_k = a_{c1}$ λ_1 e λ_2 atravessem simultaneamente o valor 1. Isto implica que

$$\beta_1(a_{c1}) = 0 \quad \wedge \quad \beta_2(a_{c1}) = \gamma_0 \quad (42)$$

De seguida substituindo (17) em (42) obtemos as seguintes relações

$$h = t_e \quad (43)$$

$$\frac{2}{3}w_0 - \frac{2}{3}h = \gamma_0 \quad (44)$$

.A relaxação mais rápida deve ocorrer nos extremos do espaço definido pela camada X do nível auditivo. Para o estado $[\pm 1, 0]^T$ a relaxação mais rápida ocorre para $a_k = a_{sup} = 1$. para este valor do parâmetro auditivo o ponto fixo $[\pm 1, 0]^T$ é o único estado estável. Se escolhermos para a relaxação mais rápida $\tau_{rel} = 5$, de (35), (36) e (40) temos que

$$w_0 = t_e + 0.1 \quad (45)$$

Finalmente os parâmetros h , w_0 e t_e podem ser determinados. Das equações (39) e (43) vem que $h = 0.2$, em seguida substituindo t_e em (45) obtemos $w_0 = 0.3$ e por fim substituindo w_0 e t_e em (44) tira-se γ_0 que pode ser sintonizado para o valor $\gamma_0 = 0.1$.

5. Ruído

A adição de ruído ao mapa dinâmico não altera muito a dinâmica do sistema. Quando há um ponto fixo estável o ruído provoca uma dispersão dos valores das variáveis perceptuais em torno do ponto fixo. A largura da distribuição dos valores das variáveis perceptuais, quando a grandeza do ruído não é muito elevada, pode ser caracterizada pelo desvio padrão que é dado por:

$$SD = \sqrt{\frac{Q\tau_{rel}}{2}} \quad (46)$$

Por convenção \sqrt{Q} designa-se por pré-factor do termo de ruído (força estocástica) e ϵ é o desvio padrão do ruído. Tal como já foi mencionado, a variabilidade permitida para as variáveis perceptuais nos regimes estacionários é aproximadamente de 0.1., $SD \approx 0.1$, conseqüentemente obtemos a partir de (46) $Q \approx 0.004$. Deve-se salientar que a escolha destes valores não é muito rigorosa sendo possível variá-los dentro de certos limites.

6. Número de iterações

Para os valores escolhidos o tempo de relaxação dos estados mais estáveis é de aproximadamente 5 e por norma o número de iterações é escolhido a partir de

$$N_{ite} > 10\tau_{rel} \quad (47)$$

O conjunto de parâmetros $w_0 = 0.3$, $\gamma_0 = 0.1$, $t_e = 0.2$, $h = 0.2$, $\gamma_{12} = \gamma_{21} = 1$ vai ser usado como um conjunto de referência e os parâmetros Q e N_{ite} vão ser variados. Estas variações permitem investigar o papel do nível de ruído e das escalas temporais no desempenho perceptual.

b) Estabelecimento dos parâmetros para o nível auditivo

Um critério muito importante a ter em consideração ao estimar um conjunto de valores para os vários parâmetros que ocorrem no nível auditivo é que este nível deve relaxar mais rapidamente que o nível fonémico.

b1) Determinação dos parâmetros para o campo dinâmico de neurónios

Os parâmetros a ser fixados são N_x , l_ϵ , γ_x , Q_x e N_{itex} . A seguir enumeramos alguns critérios que ajudam a escolher um conjunto básico de valores.

- Foi feita uma amostragem do espaço definido pelo parâmetro a , desde o limite inferior $a_{inf} = -1$ até $a_{sup} = 1$ a intervalos regulares de $\Delta a = 0.2$. Por conseguinte temos 11 neurónios no campo dinâmico, isto é $N_x = 11$.

- O tempo de relaxação para cada um dos atractores deve ser inferior ao tempo de relaxação do nível fonémico. No capítulo anterior fixámos que o menor tempo de relaxação para o nível fonémico era $\tau_{rel, fonémico} \approx 5$, pelo que deveremos ter $\tau_{rel, x} < 5$.

- O desvio padrão (SD), ou variabilidade, permitido para as saídas dos neurónios nos estados estacionários deve ser $SD \leq 0.01$. Este critério é importante porque as saídas dos neurónios x_i são directamente a entrada dos detectores fonémicos.

Até agora escolhemos os parâmetros N_x , a variabilidade em regime estacionário para as saídas dos neurónios x_i e impusemos limites ao tempo de relaxação. A seguir os restantes parâmetros do campo podem ser estimados. Nesta tarefa, tomamos alguns argumentos que fundamentam os vários passos.

Argumentos:

1. Condições para regimes estáveis:

A condição necessária para que $\mathbf{x}_{fp}^{(0)} = (0, 0, \dots, 0)_{N_x}$ seja um estado estável é dada por (22). Isto deve acontecer quando $y_i = 0 \quad \forall i = 1, \dots, N_x$. Substituindo (10) em (22) obtem-se que o parâmetro l_ϵ deve obedecer a

$$0 < l_\epsilon < 2 \quad (48)$$

A condição necessária para que $\mathbf{x}_{fp}^{(k)} = (0, \dots, 0, \pm 1, 0, \dots, 0)_{N_x}$, detecção com sucesso do valor a_k , seja um estado estável é dada por (23). Esta condição deve ser satisfeita quando na camada Y existe pelo menos um neurónio y_i ligado. Se substituirmos (10) em (23) para esta situação podemos reescrever esta última equação da seguinte forma

$$-\gamma_x < l_\epsilon < 2 - \gamma_x \quad \wedge \quad 0 < l_\epsilon < 1 \quad (49)$$

2. Tempo de relaxação

O tempo de relaxação para um ponto fixo é dado por

$$\tau_{rel} \approx \frac{1}{|\lambda_A|} \quad (50)$$

onde λ_A representa o maior valor próprio da matriz Jacobiana, \mathbf{A} , do campo de vectores (6).

Os valores próprios da matriz Jacobiana do campo de vectores estão relacionados com os valores próprios da equação do mapa dinâmico pela seguinte equação

$$\lambda_i = 1 - \lambda_{A,i} \quad (51)$$

3. Relaxação para $\mathbf{x}_{fp}^{(0)} = (0, 0, \dots, 0)_{N_x}$

Neste ponto fixo os valores próprios da equação do mapa dinâmico (5) são

$$\lambda_i = 1 - l_\epsilon \quad \forall i = 1, \dots, N_x \quad (52)$$

De (51) e (52) tiramos que

$$\lambda_{A,i} = l_\epsilon \quad \forall i = 1, \dots, N_x \quad (53)$$

Substituindo (53) em (50) obtemos que o tempo de relaxação para o estado em que todos os neurónios x_i estão desligados é dado por

$$\tau_{rel} = \frac{1}{l_\epsilon} \quad (54)$$

Para este estado podemos escolher, por exemplo, um tempo de relaxação de 2, o que pela equação (54) implica que $l_\epsilon = 0.5$. Este valor para o parâmetro l_ϵ satisfaz a condição de estabilidade (48).

4. relaxação para $\mathbf{x}_{fp}^{(k)} = (0, \dots, 0, \pm 1, 0, \dots, 0)_{N_x}$

Para este ponto fixo os valores próprios da equação vectorial do mapa dinâmico são dados por

$$\lambda_i = \begin{cases} 1 - (l_\varepsilon + \gamma_x) & \text{se } y_i = 0 \\ 1 - 2(1 - l_\varepsilon) & \text{se } y_i = 1 \end{cases} \quad \forall i = 1, \dots, N_x \quad (55)$$

Agora, a partir das equações (51) e (55) obtemos a expressão para os valores próprios da matriz Jacobiana do campo de vectores dado por (6.6)

$$\lambda_{A,i} = \begin{cases} l_\varepsilon + \gamma_x & \text{se } y_i = 0 \\ 2(1 - l_\varepsilon) & \text{se } y_i = 1 \end{cases} \quad \forall i = 1, \dots, N_x \quad (56)$$

No ponto anterior escolhemos $l_\varepsilon = 0.5$, se agora escolhermos $\gamma_x = 0.9$ as condições (47) e (49) são satisfeitas. Introduzindo estes valores em (56) obtemos que o maior valor próprio toma o valor 1.4 o que determina um tempo de relaxação que é aproximadamente 0.9.

5. Ruído

As flutuações estocásticas nas saídas dos neurónios x_i em regime estacionário devem ser suficientemente pequenas para que o desempenho do nível fonémico não seja afectado. Por exemplo, na situação em que não há estímulo acústico o estado estável do campo é aquele em que todos os neurónios tem as suas saídas a zero, mas devido à adição de ruído observam-se flutuações em torno de zero; se estas flutuações forem suficientemente elevadas podem activar o nível fonémico e induzir à detecção de um fonema. Um dos critérios que adoptamos estabelece que a máxima variabilidade permitida em regime estacionário é 0.01. Deste critério e da equação (46) tiramos então que

$$Q_x \leq \frac{2 \times SD_{\max}^2}{\tau_{\text{rel}}} \quad (57)$$

fazendo agora $SD_{\max} = 0.01$ e $\tau_{\text{rel}} = 2$ (maior tempo de relaxação) em (57) obtemos que $Q_{x,\max} \leq 0.0001$. Nas simulações fixamos $Q_x = 0.0001$.

6. Número de iterações

O estado que demora mais a relaxar tem um tempo de relaxação de 2, por conseguinte de (47) deduzimos que o número de iterações a escolher deve ser superior a 20. Desta maneira podemos escolher, por exemplo, $N_{\text{iter}} = 30$.

b3) Estabelecimento do limiar de activação dos neurónios y_i

O limiar de activação dos neurónios y_i é determinado pelo parâmetro l_y . O critério de escolha de um valor para

este parâmetro não é muito rigorosa. As unicas condições a satisfazer são que, o valor de l_y deve ser suficientemente grande para que na ausência de estimulação acústica quaisquer flutuações nos neurónios de entrada não dê origem à detecção de um estímulo acústico, por outro lado, deve ser suficientemente pequeno para que sob os efeitos de adaptação e perante um estímulo acústico na entrada os neurónios y_i excitem os neurónios x_i . Esta ultima condição é necessária porque não queremos que o sistema fique 'surdo', pois nas experiências com sujeitos observa-se que por mais fortes que sejam os efeitos de adaptação estes ouvem sempre alguma coisa. Para o presente propósito fixamos $l_y = 0.1$.

b4) Determinação dos parâmetros para a dinâmica da adaptação selectiva

Perante um estímulo acústico a camada X do nível auditivo relaxa para o estado $\mathbf{x}_{fp}^{(k)} = (0, \dots, 0, \pm 1, 0, \dots, 0)_{N_x}$ com uma constante de relaxação que é de aproximadamente 0.9. E, todos os neurónios z_i que têm o estímulo acústico dentro do seu campo receptivo sofrem efeitos de adaptação. Estes consistem na diminuição do parâmetro V_0 na direcção de $V_{0\text{min}}$ como é descrito pela equação (25), com uma constante de relaxação que é dada por

$$\tau_{\text{adap}} \approx \frac{1}{\alpha_{\text{adap}}} \quad (58)$$

Nas experiências de adaptação selectiva com sujeitos verifica-se que a apresentação de um estímulo acústico uma única vez não produz quaisquer efeitos observáveis de adaptação. Desta maneira, para que os resultados do modelo coincidam com os resultados observados na experiência com sujeitos, devemos adoptar como critério

$$\tau_{\text{adap}} > 0.9 \quad (59)$$

Podemos escolher por exemplo $\tau_{\text{adap}} \approx 3 \times 0.9 = 2.7$, pelo que de (58) podemos escolher $\alpha_{\text{adap}} = 0.35$.

Para os neurónios da camada Z que não têm o estímulo acústico corrente dentro do seu campo receptivo assiste-se a uma recuperação da sensibilidade aos estímulos que pertencem ao seu campo receptivo. O fenómeno de recuperação traduz-se segundo a equação (26) num aumento do valor do parâmetro V_0 na direcção de $V_{0\text{max}}$ com uma constante de tempo que é dada por

$$\tau_{\text{recup}} \approx \frac{1}{\alpha_{\text{recup}}} \quad (60)$$

Um aspecto importante a ter em mente é que para os efeitos de adaptação selectiva sejam sentidos ao longo das simulações, a recuperação de V_0 tem que ser muito mais lenta que a sua diminuição. Por conseguinte, devemos ter a seguinte condição

$$\tau_{\text{recup}} \gg \tau_{\text{adap}} \quad (61)$$

Se escolhermos, por exemplo $\alpha_{\text{recup}}=0.004$, de (60) obtemos $\tau_{\text{recup}} \approx 200$ que satisfaz a condição (61).

Para finalizar a escolha de um conjunto de valores para os parâmetros do nível auditivo falta ainda fixar os valores dos parâmetros $V_{0\text{max}}$ e $V_{0\text{min}}$. Para estes parâmetros podemos escolher os valores 1 e 0.15 respectivamente. A escolha de $V_{0\text{min}}=0.15 > l_y$ evita que o sistema fique 'surdo' a um estímulo acústico por mais fortes que sejam os efeitos de adaptação.

III. ENSAIOS EXPERIMENTAIS

A. Avaliação do modelo em relação às propriedades experimentais observadas para a percepção categorial

Para as simulações construímos um pequeno simulador em matlab. A ideia que nós temos em mente é usar este simulador como os psicólogos usam os seus sujeitos quando efectuem experiências da percepção categorial.

A.1 Simulação das tarefas de identificação e predição da discriminação

Método

As tarefas envolvidas na simulação da experiência psicológica de identificação são:

i) Preparar uma lista pseudo-aleatória de valores para o parâmetro a (representando a lista experimental de estímulos acústicos). Pseudo-aleatória significa que a ordem é aparentemente aleatória, mas cada valor diferente do parâmetro aparece o mesmo número de vezes. O uso de um número de vezes igual para cada valor do parâmetro a evita fenómenos de polarização. Para o presente propósito usámos onze valores diferentes para o parâmetro a que vão desde um extremo do regime ao outro. Cada um destes valores é repetido dez vezes.

ii) Corre-se o simulador sobre esta lista de estímulos. Em cada caso a condição inicial para o estímulo seguinte é a condição final da execução anterior (história perceptual).

iii) Para cada valor do parâmetro a toma-se nota do estado final. A condição lógica que classifica se o estado final é o reconhecimento do fonema 1 ou fonema 2 baseia-se na regra do máximo já descrita. Todos os resultados são contados.

iv) Dos resultados obtidos faz-se a curva de identificação, que mostra o número de vezes que cada estímulo foi reconhecido como fonema 1 (/ba/) ou fonema 2 (/pa/).

Esta experiência foi feita para vários valores dos parâmetros N_{ite} e Q .

As curvas de discriminação são preditas a partir dos resultados da tarefa de identificação usando a teoria de que a discriminação é puramente baseada na categorização, isto é, assumimos o pressuposto extremo de que os sujeitos só conseguem discriminar sons de fala a que atribuem categorias fonémicas diferentes [13].

Resultados e discussão

Um dos resultados das simulações da tarefa de identificação e curvas de discriminação preditas para dois passos é apresentado graficamente na figura 8a). Na figura 8b) é apresentado graficamente resultados análogos obtidos em experiências com um dos sujeitos. Na apresentação dos resultados cada ponto na curva de discriminação é relativo à comparação entre o estímulo imediatamente antes e depois dele.

Nesta figura observa-se que há uma partição bem definida dos estímulos em duas categorias distintas, com a fronteira categorial (ou fonémica), zona em que as duas categorias são equiprováveis, muito perto de $a = 0$ que corresponde ao ponto médio do *continuum*. Note-se a excelente concordância com os resultados obtidos com o sujeito, e em particular a localização da fronteira fonémica que para este ocorre muito perto de 35 ms de pré-vozeamento, ponto médio do *continuum* tal como acontece nas simulações.

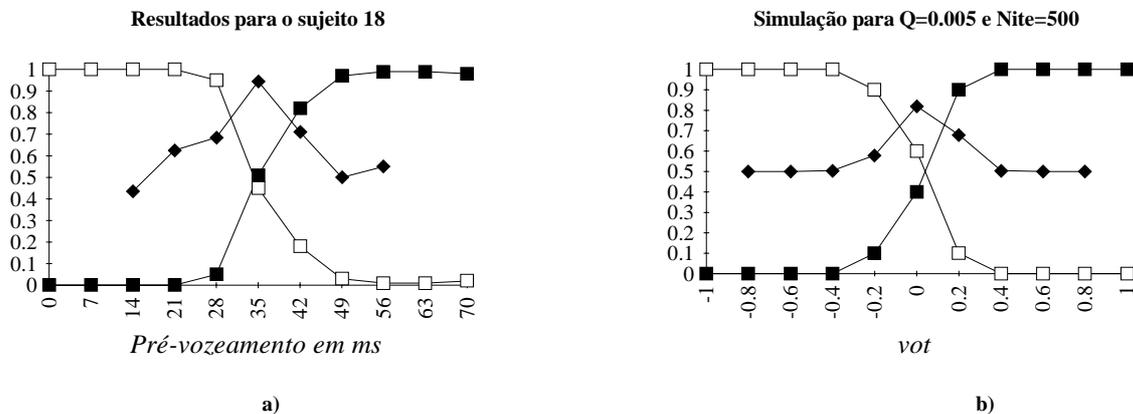


Fig. 8 Curvas de identificação e de discriminação. Em a) Resultados para o sujeito 18. Em b) resultados obtidos por simulação

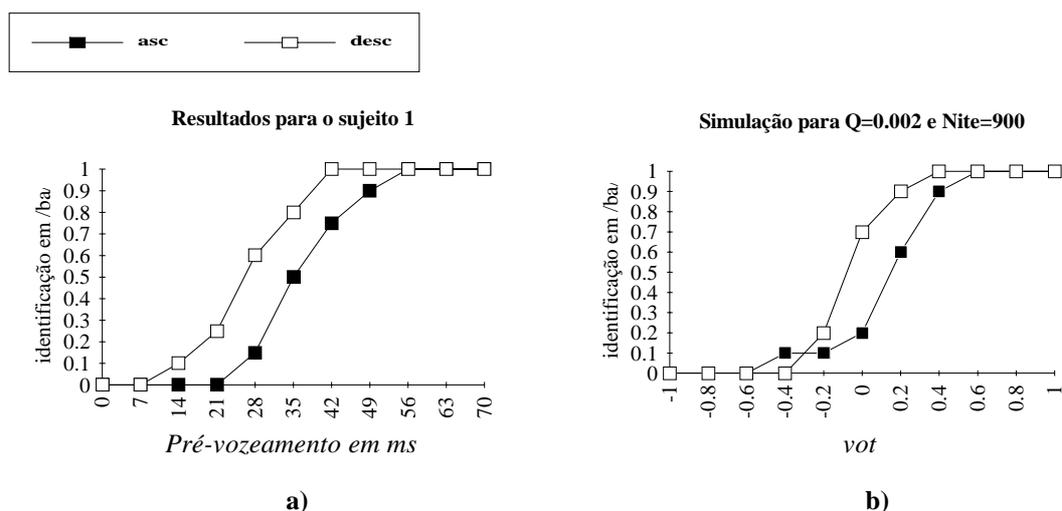


Fig. 9 Padrões de histerese. Em a) padrão de histerese para o sujeito 1. Em b) padrão de histerese obtido por simulação

Das curvas de discriminação emerge que estas exibem valores de 0.5 para estímulos dentro de cada regime monoestável (onde apenas é possível um percepto), e tem picos dentro do regime biestável, onde os dois perceptos são possíveis. Isto está de acordo com o facto de que com diferenças acústicas similares, a discriminação de sons entre membros da mesma categoria fonémica é quase nula (nível do acaso) enquanto que a discriminação é relativamente mais perspicaz entre estímulos que residem em lados opostos da fronteira fonémica. É o que acontece, por exemplo, para a comparação entre -0.2 e 0.2 no modelo e que equivale à comparação entre -28ms e -42ms de VOT nas experiências com os sujeitos.

A.2 Simulações de histerese

Método

As experiências de histerese são uma variante simples da experiência clássica da percepção categorial. A principal diferença é que usamos não uma lista pseudo-aleatória de estímulos mas sim uma lista ordenada. São consideradas duas séries de estímulos, uma ascendente que consiste numa série de estímulos que vão desde o extremo inferior até ao extremo superior do *continuum*, e outra ascendente que é a mesma série mas ordenada de forma inversa. Cada uma destas séries é constituída por onze valores diferentes para o parâmetro *vot*, e cada série é repetida dez vezes.

Resultados e discussão

Na figura 9 são apresentados os resultados experimentais obtidos para um dos sujeitos intervenientes nestas experiências, e é ilustrado um dos resultados da simulação. A figura 9b) mostra que, como era esperado, a apresentação das séries de estímulos induzem o fenómeno de histerese pois o ponto em que é observado uma transição de categoria é diferente para a ordem ascendente e para a ordem descendente, em cada caso a transição perceptual é 'atrasada' em cada série tal que é obtida uma região de sobreposição. Esta região é chamada a magnitude da histerese. Em termos da dinâmica do modelo

isto pode ser compreendido do seguinte modo: À medida que uma série é apresentada sequencialmente no tempo desde um extremo a outro, acontece que o percepto inicial persiste no regime biestável até que o parâmetro *a* atinge a fronteira mais distante deste regime onde o percepto perde a estabilidade. Se a série de estímulos é apresentada na ordem inversa acontece o oposto. Consequentemente os pontos na série em que ocorre a transição perceptual depende da direcção em que a série é apresentada.

B. Avaliação do modelo em relação aos efeitos da adaptação

Os efeitos de adaptação na percepção de sons de fala aparecem em relação à localização da fronteira categorial ao longo do *continuum* de estímulos. Para estudar a adaptação no modelo realizámos uma simulação da experiência de contraste, que como iremos ver produz efeitos idênticos ao paradigma da adaptação selectiva.

Simulação da experiência de contraste

Tipicamente, no paradigma da adaptação selectiva os sujeitos são expostos a uma longa sequência de adaptação, que consiste no estímulo adaptador repetido várias vezes. No fim deste período, é apresentada uma série aleatória de estímulos para serem classificados. O resultado é um salto na fronteira fonémica na direcção do estímulo adaptador.

A experiência de contraste consiste num simples procedimento em que a longa sequência de adaptação é substituída por um único estímulo, designado por estímulo de contexto, seguido por um único estímulo de teste. Para os estímulos de contexto, tomam-se estímulos que são considerados pelos sujeitos exemplares claros das categorias que representam, enquanto que os estímulos de teste são seleccionados na vizinhança da fronteira categorial para que tenham um carácter bastante ambíguo. Neste tipo de experiência os sujeitos revelam uma grande tendência para identificarem o item de teste na categoria oposta ao do estímulo de contexto. Um resultado que é

semelhante ao efeito típico da adaptação selectiva. Para Diehl e seus colaboradores [14] esta semelhança sugere que os efeitos de adaptação e contraste podem ser produzidos pelos mesmos processos. Estes autores apresentam alguns estudos que apoiam esta hipótese.

Método

Experiência A

Para a simulação da experiência de contraste foram realizados os seguintes passos [14]:

i) Para estímulos de contexto seleccionaram-se dois estímulos com valores de VOT -0.4 e 0.4. O estímulo -0.4 é principalmente identificado como fonema 2, enquanto que 0.4 é essencialmente classificado como fonema 1.

ii) Escolheu-se para estímulo de teste um valor junto à fronteira categorial, i.e. o valor 0, que possui um carácter muito ambíguo. Teóricamente este estímulo tem uma probabilidade de 50 % de ser classificado em qualquer uma das duas categorias.

iii) Sem os efeitos de adaptação activos, cada combinação dos estímulos de contexto e de teste foi apresentada 10 vezes ao modelo para serem classificados. As condições iniciais

para os neurónios do nível fonémico foram variadas aleatoriamente ao longo de todo o intervalo.

iv) Repetiu-se o passo anterior mas agora com a dinâmica da adaptação activa no modelo.

v) Registaram-se os resultados obtidos.

Experiência B

Repetiu-se essencialmente a experiência A mas agora para todos os restantes estímulos do *continuum*, e o passo iii) foi substituído pelo paradigma aleatório da percepção categorial estando incorporados no modelo efeitos de adaptação.

Resultados e discussão

Os resultados obtidos na experiência A são mostrados na tabela 1. E na figura 10 são apresentados graficamente os resultados da experiência B.

Estímulo de contexto		
$VOT \rightarrow$	-0.4	0.4
Sem adaptação	60%	60%
Com adaptação	100%	0%

Tabela 1 Percentagem de vezes que o estímulo 0 é identificado como fonema 1, para dois estímulos de contexto e para o modelo sem e com efeitos de adaptação. Simulações para $Q=0.005$ e $N_{ite}=300$

Cada valor na tabela 1 representa a percentagem de vezes que o estímulo de teste, $VOT=0$, é identificado na categoria fonémica 1, para cada um dos estímulos de contexto e para o modelo sem e com efeitos de adaptação. Nesta tabela podemos ver que o estímulo de teste, na situação em que o sistema não exhibe efeitos de adaptação, foi classificado 60% das vezes na classe fonémica 1, independentemente do estímulo de contexto. Teóricamente este estímulo de teste tem uma probabilidade de 50% de ser classificado em qualquer uma das duas categorias, acontece porém, que devido ao facto dos tempos de computação serem elevados, este estímulo foi apresentado ao modelo apenas 10 vezes, das quais 6 vezes foi classificado na categoria fonémica 1 e as restantes 4 vezes na categoria fonémica 2. Como vemos temos uma amostragem pequena que se traduz num desvio aparentemente grande dos 50%. Quando o sistema incorpora efeitos de adaptação este padrão de resultados é completamente alterado, o que revela que o efeito do estímulo de contexto na adaptação é muito significativo. Quando o sistema tem a sua dinâmica de adaptação activa e o estímulo de contexto é $VOT=-0.4$, observa-se que o estímulo de teste é sempre classificado na categoria fonémica 1, independentemente das condições iniciais para o nível fonémico. Este resultado é o reflexo de um salto na fronteira fonémica na direcção do estímulo de contexto, que é o responsável pelos efeitos de adaptação. Analogamente, quando o estímulo de contexto é $VOT=0.4$, o estímulo de teste passa a ser inequivocamente “percebido” como fonema 2, em consequência dos efeitos de adaptação que provocam uma mudança na posição da fronteira fonémica na direcção do extremo superior do *continuum*, i.e. na direcção do estímulo de contexto.

Estes resultados, obtidos por simulação, estão em perfeita concordância com os padrões de comportamento observados em experiências com sujeitos e relatados em [14].

Vejamos agora os resultados da experiência de contraste para os restantes estímulos do *continuum* que são ilustrados graficamente na figura 10.

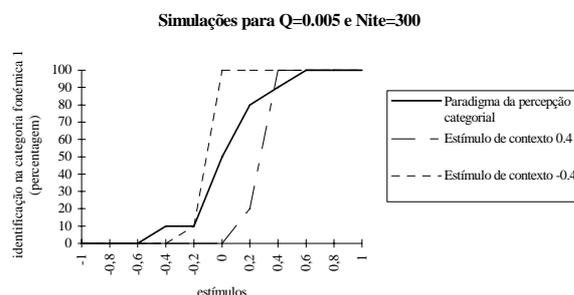


Fig.10 Curvas de identificação na categoria fonémica 1, obtidas por simulação da experiência de contraste para todos os estímulos ao longo do *continuum* e por simulação do paradigma da percepção categorial.

Conforme podemos observar na figura 10, para o paradigma aleatório da percepção categorial o modelo

exibe o padrão clássico esperado para este fenómeno. Nas experiências de contraste, o modelo revela uma elevada tendência para identificar os estímulos de teste na categoria oposta ao do estímulo de contexto. Em particular, observa-se que o estímulo de contexto $VOT=-0.4$, produz um deslocamento na fronteira na direcção do extremo menos vozeado do *continuum*, enquanto que o estímulo de contexto $VOT=0.4$, produz o efeito contrário. Um resultado que é igual ao efeito típico da adaptação selectiva.

VI. CONCLUSÃO

Neste artigo apresentámos um modelo dinâmico neuronal para a categorização de sons de fala num *continuum* de vozeamento. Mostrámos que no paradigma clássico de identificação emerge o padrão típico da percepção categorial. Quando os estímulos são apresentados em séries ascendentes ou descendentes resulta histerese. Esta é uma forma de memória perceptual que não foi construída explicitamente. Ocorrem saltos na fronteira categorial devido a efeitos de adaptação, que são de natureza inteiramente auditiva. No modelo é também modelada a situação em que não há nada para perceber, i.e. a ausência de um estado perceptivo, o que ocorre na ausência de qualquer estímulo acústico. Comparámos as predições do modelo com os resultados experimentais. Deste estudo podemos concluir, que a percepção categorial da fala pode ser compreendida como o resultado de um processo de competição dentro de uma representação neuronal da informação sensorial. Nesta perspectiva a discretização resultante do fenómeno da percepção categorial é o resultado da formação de estados dinamicamente estáveis. Estes estados não são imunes aos processos perceptuais subjacentes, pois dependem da história recente do sistema.

AGRADECIMENTOS

Agradecemos à Doutora São Luís Castro, da fac. de Psicologia e Ciências da Educação, a colaboração dispensada tanto na parte teórica como experimental do fenómeno que aqui estudamos.

Este trabalho foi realizado com o apoio da bolsa de mestrado da JNICT no âmbito do programa CIÊNCIA: BM/2780/92-IA.

REFERÊNCIAS

- [1] Jusczyk (1986), review article by Peter Jusczyk from Handbook of Perception and Human Performance (Boff, K., Kaufman, L. & Thomas, J(eds), 1986).
- [2] Repp B. H. (1984), "Categorical Perception: issues, methods, findings", in N.J. Lass (Ed). Speech and Language: Advances in Basic Research and Practice, vol. 10, p. 243-335.
- [3] Castro S.L. (1993), "Percepção categorial num continuum de vozeamento em fala natural portuguesa", in EPLP'93 1º Encontro de Processamento da Língua Portuguesa Escrita e Falada, 15-18.
- [4] Sawusch J. R., Jusczyk P. (1981), "Adaptation and contrast in the perception of voicing", Journal of Experimental Psychology: Human Perception and Performance 2, p.408-421.
- [5] Williams D., Phillips G., Sekuler R. (1986), "Hysteresis in the perception of motion direction as evidence for neural cooperativity", Nature 324: 253-255, 1986.
- [6] Chang J. J., Julesz B.(1984), "Cooperative phenomena in apparent movement: perception of random-dot cinematograms", Vision Research 24:1781-1788, 1984.
- [7] Castro S.L., Barbosa M.F., Bicho E., Schöner G. (1994), "Cooperativity in the categorical perception of speech: experiment and modelling", European Society of cognitive psychology, Lisboa 1994.
- [8] Summerfield A.Q.(1975), "Aerodynamic versus mechanics in the control of voicing onset in consonant-vowel syllables", in Speech Perception (no.4), Department of Psychology, Queens University of Belfast, 1975.
- [9] Braun M. (1978), "Differential equations and their applications", Springer Verlag, New York, 1978.
- [10] Perko L. (1991), "Differential Equations and Dynamical Systems", Springer Verlag, Berlin.
- [11] Stein D.(1989), "Lectures in the sciences of complexity", Ed Daniel Stein, Department of physics, University of Arizona.
- [12] Schöner G., Dose M.(1992), "A dynamical system approach to task level system integration used to plan and control autonomous vehicle motion", Robotics and Autonomous Systems 10(1992),p253-267-Elsevier.
- [13] Liberman A., Harris K., Hoffman H., Griffith B. (1957), "The discrimination of speech sounds within and across phoneme boundaries", Journal of Experimental Psychology, Vol. 54, No. 5, 1957
- [14] Diehl R. L., Kluender K.R., Parker E.M.(1985), "Are selective adaptation and contrast effects really distinct?", Journal of Experimental Psychology: Human Perception and Performance 11, p.209-220.