

PERCEPÇÃO DE C OU CC OCLUSIVAS POR OUVINTES NATIVOS DE PORTUGUÊS EUROPEU¹

Amália Andrade
Centro de Linguística da Universidade de Lisboa

1. Introdução

Em Português Europeu (PE, doravante), como é sabido, a não manifestação fonética de vogais altas não-acentuadas, em particular do “e mudo”, é frequente, dando lugar a sequências relativamente longas de consoantes. No encadeamento de palavras e clíticos, a não manifestação fonética do “e mudo” pode estender-se a sequências de várias sílabas (5-6), se determinadas condições prosódicas e coarticulatórias (em sentido lato) forem satisfeitas. Um exemplo extremo de sequência de consoantes é a realização que a seguir se transcreve: [pedsksdʃprt̪ar̃ɐ̃] (“pede-se que se despertarem”).²

A ausência de manifestação fonética de uma vogal pode ser acompanhada de modificações do contexto segmental adjacente. A possibilidade de ocorrência de alongamento da consoante precedente, nomeadamente quando a vogal sem realização fonética é um “e mudo”, está atestada em trabalhos anteriores (e.g. Companys Maldonado, 1954; Mateus e Delgado-Martins, 1982; Andrade, 1982; Brandão de Carvalho, 1987).³ Na maioria dos casos, as observações focam a posição final absoluta (/C--#/); alguns trabalhos incidem sobre posições mediais, como Andrade (1982) e sobretudo (1993a,b). Os resultados acústicos de Andrade (1993a,b), por outro lado, são compatíveis com a hipótese de que a ausência de realização fonética do “e mudo” entre duas oclusivas (C₁C₂) pode estar associada a diferenças

acústicas finas em função do preenchimento ou não da posição entre as oclusivas, a nível da representação mental. Quando essa posição é preenchida por um “e mudo”, a explosão de C_1 pode ser acompanhada de um “reforço” da explosão da oclusiva precedente.

Os dois tipos de dados permitem a hipótese de que os falantes de PE tendem a realizar sequências de duas oclusivas resultantes da não realização fonética de um “e mudo” como um ataque complexo, preservando a identidade da oclusiva inicial.⁴ Na secção 3 expõem-se, de forma resumida, os dados acústicos e aerodinâmicos disponíveis e relevantes para o PE.

O trabalho que aqui se apresenta faz parte de um conjunto de experiências com o objectivo de examinar a *importância relativa da duração do intervalo de silêncio*⁵ e da ocorrência ou não de uma explosão de amplitude elevada associada a C_1 , na percepção de uma sequência de duas oclusivas por parte de ouvintes nativos de PE. O presente estudo, ainda em curso, visa determinar *a fronteira a partir da qual, fazendo variar apenas a duração do intervalo de silêncio, os ouvintes deixam de detectar duas oclusivas de pontos de articulação diferentes para passar a ouvir apenas uma.*

2. Alguns aspectos relevantes de estudos sobre a duração do intervalo de silêncio oclusivo realizados para outras línguas

Os estudos acústicos sobre a duração de grupos consonânticos não são muito vastos e baseiam-se num número relativamente reduzido de línguas, como o Sueco (e.g. Lindblom, 1973), o Inglês (e.g. Klatt, 1973; Haggard, 1973), o Francês (e.g. Bartkova e Sorin, 1987; Nishinuma, Duez e Paboudjian, 1991)⁶. Os resultados dos diferentes estudos estão, em geral, em concordância, mostrando que a duração de uma consoante, excepto no caso das laterais, segundo Bartkova e Sorin, (1987), é mais breve quando ocorre em sequência consonântica do que quando ocorre isoladamente, em outros contextos. Haggard (op. cit.), por exemplo, verificou que [t] em “cost” é c. de 21% mais breve do que em “cot”; Klatt (1973) observou que a duração de [s] diminui c. de 28% em sequências de duas consoantes e 33% em grupos de três consoantes.

O “efeito de compressão” varia em função de diversos factores como a posição na palavra e na frase (e.g. Lindblom, 1973; Klatt,

1973; Haggard, 1973; Bartkova e Sorin, 1987), a posição relativamente ao acento, os modos e os pontos de articulação envolvidos, o número de consoantes da sequência, a qualidade da vogal antecedente (e.g. Nishinuma et alii, 1991).⁷ De certo modo na linha de pensamento de Lindblom (1981), é de admitir que a compressão das consoantes em grupo consonântico é regida pela relação entre características prosódicas e "patterns" temporais de organização articulatória. Segundo Klatt, o grau de informação transportada pelas consoantes é também um factor condicionante: a duração é tanto maior quanto maior é o grau de informação; a redundância permite a abreviação.

Diversos estudos perceptivos têm incidido sobre o papel do intervalo de silêncio na detecção por parte dos ouvintes da ocorrência de oclusivas e na sua identificação. Restringimo-nos, aqui, aos que incidem sobre a percepção de oclusivas singulares e sequências de duas oclusivas. Repp (1977) examinou o efeito perceptivo da variação da duração do silêncio em posição intervocálica, com base em estímulos gerados a partir das sequências [ab] ou [ad] e [be] ou [de] e diferentes intervalos de silêncio inseridos entre eles (daí resultando sequências como [abbē] e [abde]). O intervalo de silêncio foi variado entre 15ms e 200 ms. Repp verificou que quando o silêncio era da ordem dos 15 ms os ouvintes passavam a perceber *uma consoante: a consoante em ataque* (C_2 de C_1C_2). Para que as *duas consoantes de pontos de articulação distintos* fossem percebidas, o intervalo de silêncio tinha de atingir os *140 ms*. Um intervalo ainda maior, *200 ms*, era necessário para a percepção de [abbē] como *geminada*. Dorman, Raphael e Liberman (1979) levaram a cabo experiências semelhantes às de Repp, baseadas em estímulos dissilábicos, [bebbe], [begde] e (numa delas) [bedde], em que variaram o silêncio intervocálico. Os estímulos eram sintéticos numa das experiências e naturais noutra experiência. Na primeira verificaram que a percepção de uma oclusiva implicava um intervalo intervocálico de c. 30 ms; para obter 75% de identificação correcta das duas consoantes o intervalo tinha de ser c. 58 ms. Os resultados das experiências baseadas em estímulos sintéticos e naturais são convergentes.⁸

Tendo em consideração os trabalhos de diversos autores sobre os mecanismos auditivos, é lícito pensar que o requisito de um intervalo mínimo de silêncio para a percepção da oclusiva inicial de uma sequência de duas oclusivas (ou simplesmente da oclusiva intervocáli-

ca) é em parte atribuível a uma característica importante das fibras nervosas auditivas, a adaptação (cf. Delgutte, 1980; Greenberg, 1995), e em parte, também, à intervenção de factores a um nível mais central, nomeadamente restrições quanto ao tempo de transmissão e integração da informação.

3. Alguns dados acústicos e aerodinâmicos sobre obstruintes singulares e em sequências consonânticas, existentes para o PE

O alongamento das consoantes em posição final absoluta e a sua menor duração ou compressão no interior de uma sequência consonântica pode ser confirmado em termos qualitativos através dos dados aerodinâmicos das figuras 1 e 2, de Andrade (1982). As figuras sugerem ainda, entre outros aspectos, que a duração consonântica em sílaba não acentuada depende da presença ou ausência de realização da vogal nuclear. Vejamos mais concretamente.

Na figura 1, os traçados de pressão oral (Po) e de fluxo de ar oral (Uo), mostram que, em *posição final*, a duração do período de oclusão de um [t] (gráficos superiores) e de constrição de um [s] (gráficos inferiores) é menos longo quando a consoante é seguida de vogal [ɐ] (“cita”, [sitɐ]) do que quando a consoante não é seguida de segmento vocálico, quer este corresponda a um “e mudo” (“cite”, [sit]) quer a um /u/ (“cito”, [sit^w]).⁹ A duração consonântica apresenta-se mais longa no caso do “e mudo” do que no caso da vogal arredondada: isto é o segmento consonântico final tende a ser um pouco mais longo se, para além de não haver realização vocálica, não há especificação de arredondamento. As curvas de pressão sugerem que, em contexto de [ɐ], a preparação do vozeamento se inicia bastante cedo (vd. pico inicial com descida clara da curva de Po); tal não se verifica nos outros contextos, em que efectivamente não há vozeamento final. As curvas de fluxo de ar reflectem uma diferença significativa entre a coronal+arredondamento e a coronal simples (associada ao “e mudo”): no segundo caso, e não no primeiro, a distensão da oclusiva é seguida de um período de ruído longo e de amplitude elevada.

A figura 2, também de Andrade (1982), inclui traçados de Po e Uo correspondentes a sequências *mediais*, nomeadamente [pɹ], [pɹɪ], [p^wɹ] de “a pesar”, “a pisar”, “a posar” (gráficos superiores) e [k^wpɹ] de “é prático pesar” (gráfico inferior). Observa-se nas curvas superio-

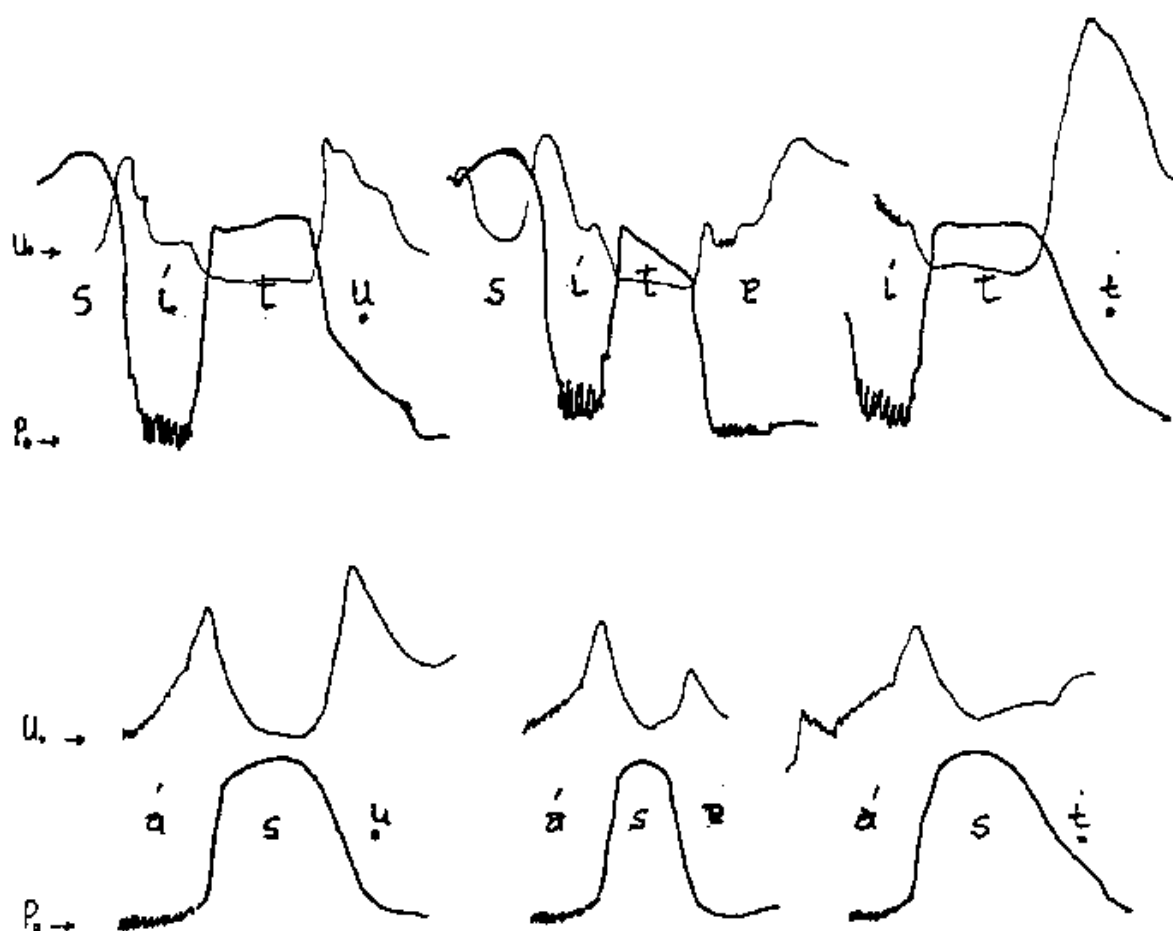


Figura 1 – Traçados de pressão oral (Po) e fluxo de ar oral (Uo) correspondentes a seqüências envolvendo presença/ausência de manifestação fonética de vogal em posição final absoluta.

res que o intervalo oclusivo continua a ser mais longo nas seqüências sem realização de vogal do que quando a realização vocálica tem lugar; mas a diferença manifesta-se de forma muito atenuada. Este facto está em consonância com a observação da ocorrência de redução da duração consonântica em seqüências de consoantes já referida. Os valores de pico de pressão (Po) da oclusiva, por seu lado, são de novo mais elevados no caso das seqüências consonânticas do que no da oclusiva diante de [i].

A seqüência de labial e coronal permite um grau acentuado de sobreposição de gestos articulatórios. O facto da duração total e das durações associáveis à oclusiva e à fricativa serem maiores em [p^wz] do que em [pz] sugere que a especificação de arredondamento impõe uma restrição à co-produção da labial e da coronal.

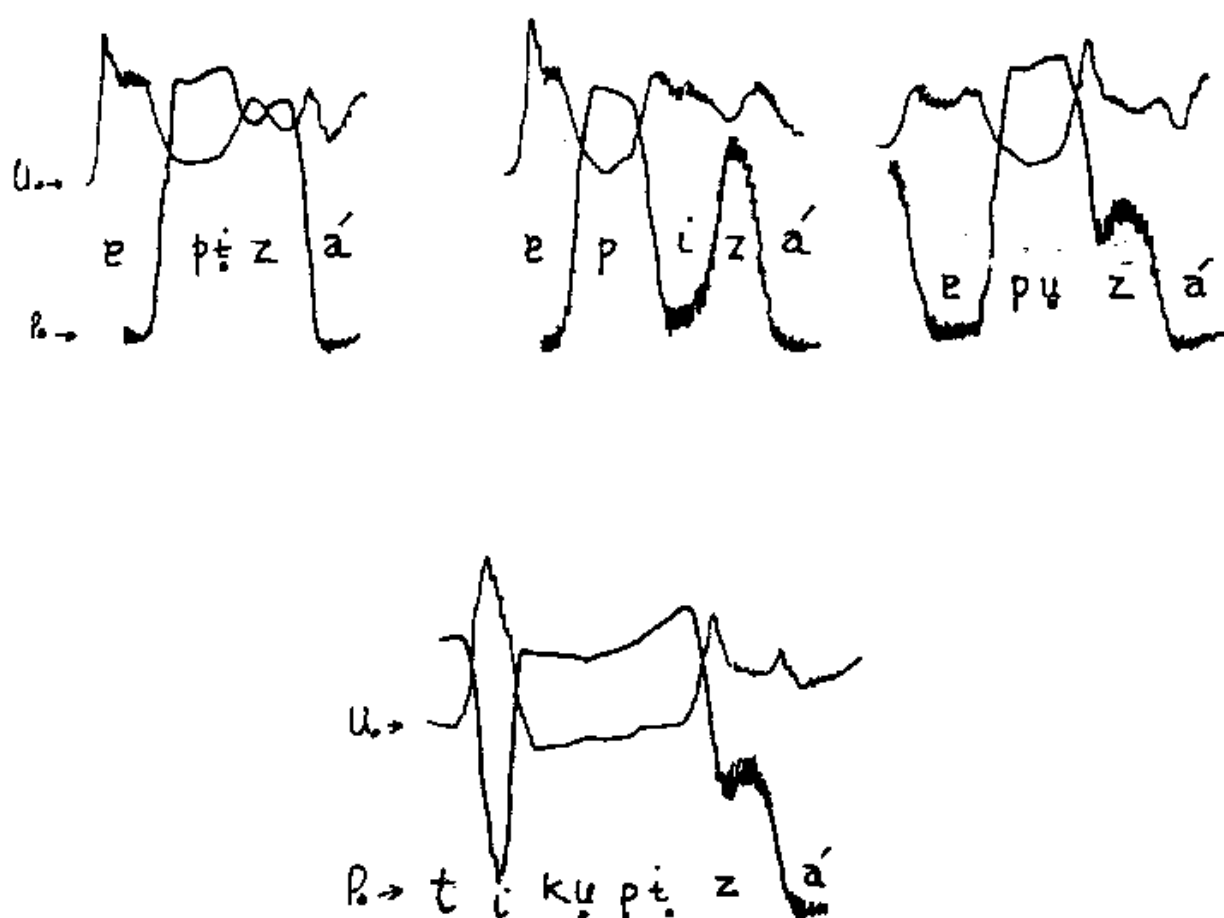


Figura 2 – Traçados de pressão oral (Po) e fluxo de ar oral (Uo) correspondentes a sequências em posição medial envolvendo presença/ausência de manifestação fonética de vogal alta não acentuada.

O gráfico inferior corresponde a uma junção de palavras em que a ausência de realização vocálica em posição final e em sílaba inicial deu lugar à sequência de três consoantes: duas oclusivas, e uma fricativa, associadas aos pontos de articulação velar+labial+coronal. No traçado de Po é possível identificar três segmentos associados às consoantes em questão. É de notar que, no seu total, a duração da sequência corresponde aproximadamente ao *triplo da duração* da oclusiva labial diante de [i], em [piz]. Este gráfico e os restantes sugerem, assim, que na ausência de realização vocálica, a co-produção das consoantes¹⁰ não implica a sua fusão.

Passamos agora a centrar-nos apenas em sequências mediais que envolvem oclusivas surdas. No âmbito de um estudo acústico mais recente sobre o vocalismo átono pré-acentual, analisaram-se as dura-

ções segmentais correspondentes às sequências /tVk/, em que V é ou /i/ ou um “e mudo” (como em “praticaram”, “orticária”, “bibliotecária”, “abotecaram”). Mais especificamente, mediram-se as durações correspondentes aos períodos (a) de oclusão da coronal e da velar, (b) do VOT da coronal ou simplesmente da duração da sua explosão quando esta não é seguida de vozeamento, e (c) do segmento vocálico. A figura 3 representa os valores de duração de segmentos acústicos associados às sequências /tVk/ em posição pré-acentual de palavras

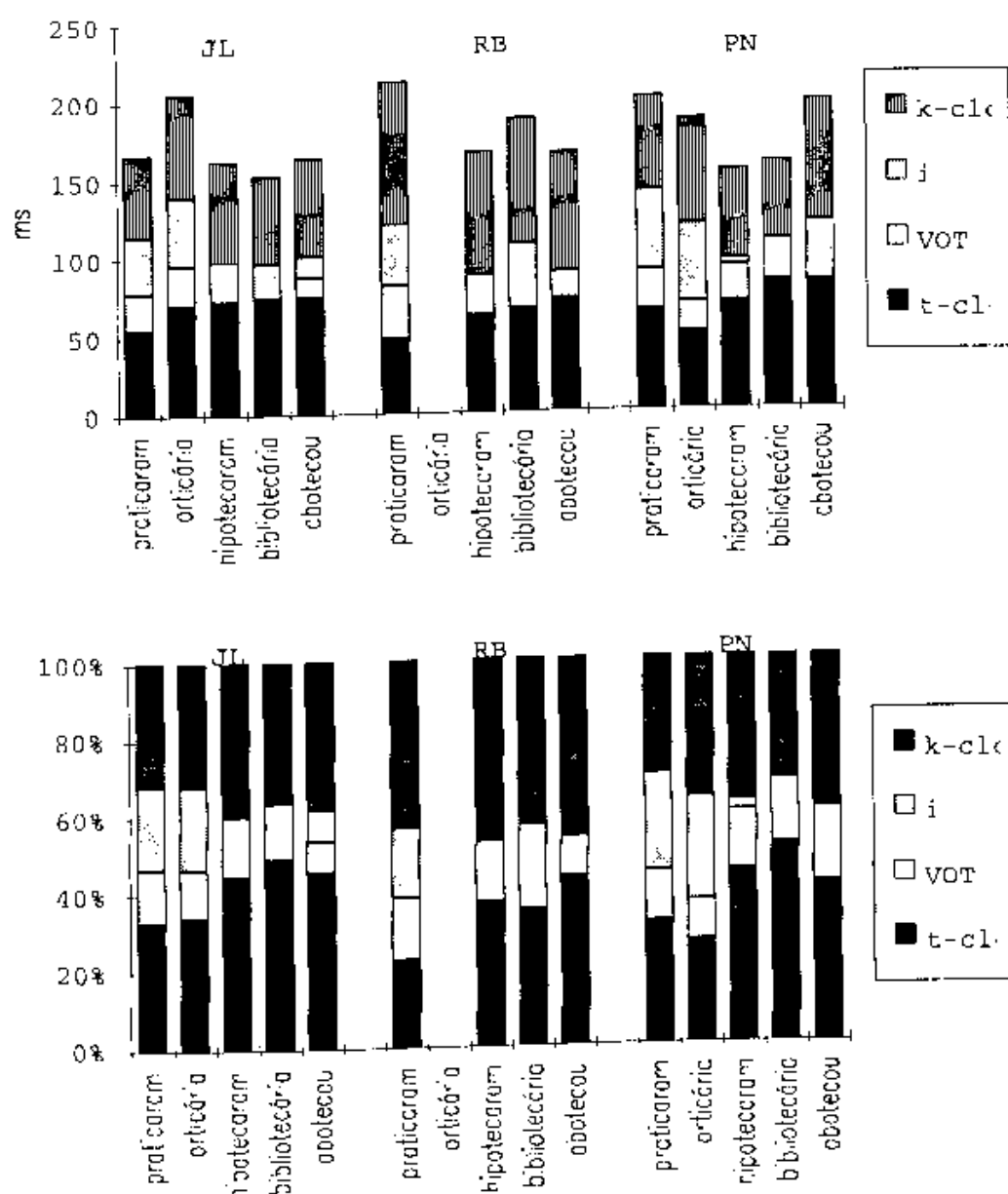


Figura 3 – Duração dos intervalos de oclusão e VOT da coronal, do segmento vocálico e da oclusão velar (/tVk/) das palavras “praticaram”, “orticária”, “hipotecaram”, “bibliotecária”, “abotecou” realizadas em início de frase por três falantes da região de Lisboa. Os valores por informante correspondem a médias de três realizações.

em início de frase (declarativa) realizadas por três falantes.¹¹ O gráfico inferior contém uma normalização dos mesmos valores relativamente à duração total da sequência /tV_k/.

No que respeita ao “e mudo”, apenas se encontraram breves indícios de vozeamento no caso de “hipotecaram” e de “aboteceu”, dos falantes JL e PN, respectivamente. Pode-se observar que, tanto em termos absolutos como em termos relativos, embora a diferença não seja muito grande, a duração da oclusão coronal é sistematicamente menor em contexto de [i], do que na ausência de realização fonética de “e mudo” ou quando este está associado a um segmento de voz muito breve. Do gráfico que segue pode-se inferir, ainda, que a duração da coronal que antecede o “e mudo” é idêntica ou superior à duração da consoante da sílaba acentuada seguinte. A convergência entre estes dados e os dados aerodinâmicos reforça a ideia apenas apontada acima de que na ausência de realização fonética de uma vogal, a *identidade da consoante em ataque é preservada* pelos falantes. Esta hipótese requer, ainda, verificação empírica adequada, como é evidente.

Em Andrade (1993a,b)¹² analisaram-se as durações segmentais (cf. figura 4) e a amplitude da explosão oclusiva em sequências associadas a estruturas silábicas diferentes, do tipo C₁VC₂ e C₁C₂ em que V é um “e mudo”, C₁ é uma oclusiva labial surda e C₂, uma oclusiva coronal, também surda. As sequências em questão ocorrem em posição pré-acental nas palavras “atapetaram”, “adaptaram”, “petardo” e “raptados”, as primeiras duas em início de frase e as duas últimas em fim de frase.¹³ Não se encontraram, nas realizações analisadas, exemplos claros de “ee mudos” vozeados ou não-vozeados. Para alguns dos falantes encontraram-se casos de breves segmentos (8-14 ms) interconsonânticos associados a “pet” passíveis de serem interpretados como vozeados. No entanto, é apenas em 8% desses casos (3/36) que o segmento vozeado é acompanhado de uma ligeira excitação do segundo formante. Tais segmentos estão incluídos, na figura 4, como parte integrante da explosão. Os gráficos desta figura mostram que as durações dos intervalos de oclusão e as suas relações não reflectem diferenças sistemáticas relacionáveis com duas estruturas subjacentes distintas: a verificarem-se essas diferenças, esperar-se-ia uma maior duração da labial no caso de “atapetar” e “petardo” do que em “adaptar”.

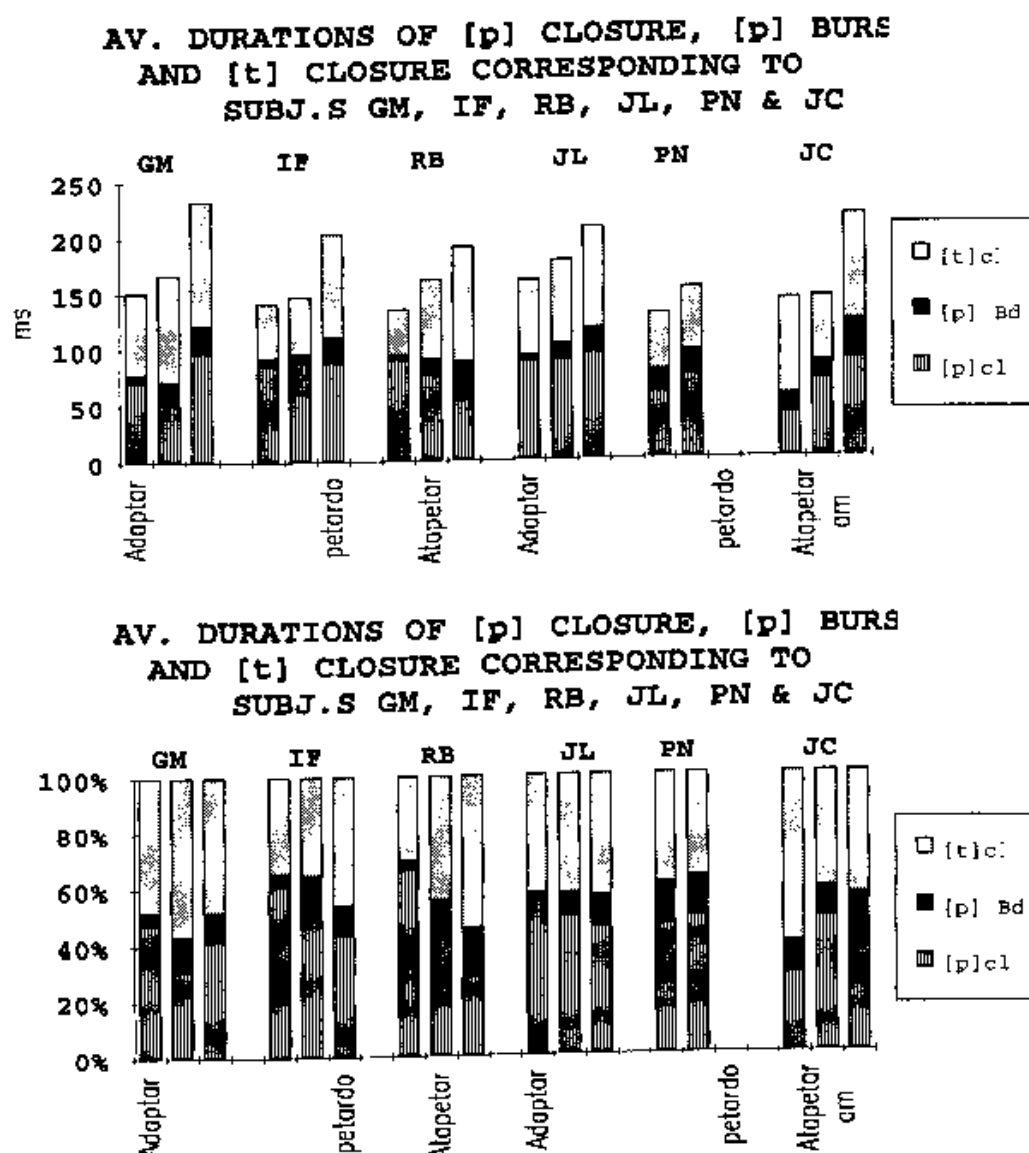


Figura 4 – De Andrade (1993b). Duração dos intervalos de oclusão e explosão da labial e oclusão da coronal (/p(V)t/) das palavras “adaptaram”, “atapetaram”, “petardo” (da esquerda para a direita), realizadas em início de frase por três falantes da região de Lisboa. Os valores por informante correspondem a médias de três realizações. Os seis informantes são da região de Lisboa.

À primeira vista, os resultados dos estudos de 1993 parecem não estar em concordância com os resultados dos estudos anteriores. Com efeito, não se encontram aqui, para a labial, durações correspondentes a cerca de metade da duração total da sequência CC. A maioria dos falantes, no entanto, evidencia uma tendência para produzir a explosão da labial, com uma amplitude significativamente superior em “atapetar” e “petardo” do que em “adaptar”; de forma menos marcada, também a

duração da explosão tende a ser maior nos dois primeiros casos do que nos dois últimos. A figura 5 ilustra estas observações: os círculos representam amplitude, e as barras, duração. Para uma interpretação correcta da figura é necessário explicitar que os valores de amplitude representados correspondem à *diferença* entre o pico de amplitude da explosão obtido por meio de uma janela de 6 ms e a amplitude da vogal seguinte ([a]). Esta última foi obtida centrando-se a janela de análise no terceiro ciclo de vozeamento. Assim, os valores máximos de amplitude da explosão estão representados pelos círculos de valores mais baixos no eixo y. Os resultados aqui descritos parecem reflectir a utilização de estratégias articulatórias distintas na realização dos dois tipos de estruturas: a maior amplitude da explosão poderá, por exemplo, resultar da intervenção combinada de actividade articulatória laríngea e supra-laríngea (Andrade, 1993a, b): um certo grau de glotalização poderá explicar a ocorrência de uma amplitude de explosão relativamente elevada.

Dois dos informantes afastam-se da estratégia que se acaba de descrever. Um deles, pura e simplesmente não diferencia as realizações das sequências consonânticas de “adaptar” e “atapetar” (cf. Andrade, 1993a); por outro lado é de apontar que produz a labial com uma intervalo oclusivo superior a 60% do total de duração de CC; nestes dois casos, e no caso de “petardo” verifica-se uma duração da oclusão um pouco superior aos 50%, mas com amplitude e duração de explosão elevadas. O sexto informante (JC), por seu lado, apresenta uma significativa redução da labial, que chega a realizar como contínua breve, em “adaptar”, mas assegura a realização oclusiva com uma duração da ordem dos 50% de CC ou mais, e uma explosão nítida em “atapetar”, e “petardo”. Os dados sugerem que a labial é realizada por este informante como um elemento em *coda* da sílaba anterior no primeiro caso, e no segundo caso, como um elemento em *ataque*.

O confronto dos resultados dos diferentes estudos sugere o seguinte:

- Na ausência de realização fonética de uma vogal não acentuada, nomeadamente de um “e mudo”, no contexto de duas oclusivas, a identidade da oclusiva precedente é preservada através da sustentação da constricção de modo a que a duração consonântica (oclusão+explosão) seja cerca de 50% ou mais da duração total da sequência CC, e/ou

[p] BURST AMPLITUDE AND DURATION
CORRESPONDING TO SUBJ.S GM, IF, RB, JL
PN & JC

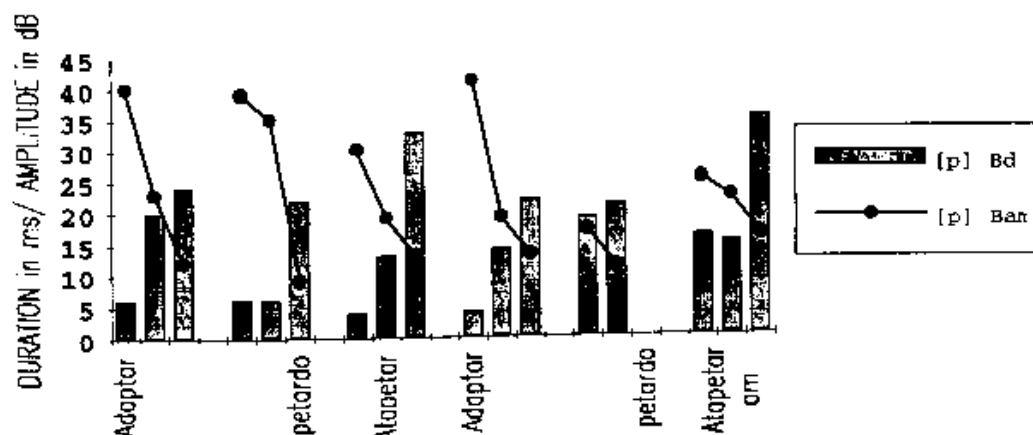


Figura 5 – De Andrade (1993b). Duração e amplitude da explosão da labial (/p(V)t/) das palavras “adaptaram”, “atapetaram”, “petardo”, “bibliotecária”, “abotcou” realizadas em início de frase por três falantes da região de Lisboa. Os valores por informante correspondem a médias de três realizações. Os seis informantes são da região de Lisboa.

através da criação de uma pressão elevada atrás da constrição, de modo a gerar uma explosão de amplitude significativa. As durações consonânticas mínimas parecem ser da ordem dos 40% (c. de 30% se apenas a oclusão é tomada em consideração).

- O facto da relação temporal no interior da sequência CC ser melhor preservada no caso da sequência [tk] do que em [pt] pode ser em parte, pelo menos, resultante da natureza dos articuladores envolvidos: a articulação labial permite maior co-articulação do que a coronal.

- A duração consonântica tende a ser um pouco menor quando a realização vocálica tem lugar do que na sua ausência;¹⁴ as diferenças acentuam-se em posição final absoluta (e possivelmente em posição pré-pausa).

- Como foi apontado na Introdução, a oclusiva inicial de uma sequência de duas oclusivas (pelo menos resultante de não realização fonética de vogal), não perdendo a sua identidade, integra um ataque complexo.

- Quanto à oclusiva labial de “adaptaram” vs “atapetaram” e “petardo”, o seu estatuto é menos claro: mesmo quando há diferenciação das características da labial, a identidade tende a ser preservada pela maioria dos falantes (cf. duração da oclusão), o que permite pen-

sar que tende também a integrar o ataque seguinte; no entanto, como os dados de JC sugerem, os falantes podem ter uma “solução” diferente.¹⁵ Permitimo-nos especular que talvez esta estratégia de JC no que respeita à labial de “adaptaram” seja idêntica à que deu lugar historicamente à glidização das oclusivas latinas em coda.

4. Estudo perceptivo

4.1. Aspectos metodológicos

Apresentam-se em 4.2 os resultados parcelares de dois testes de identificação: teste 1 (resultados de 9 informantes) e teste 2 (resultados de três informantes apenas).

Os estímulos dos dois testes foram construídos por meio do programa Speech Station (da Sensimetrics), a partir de realizações naturais, produzidas por um dos falantes utilizados em estudos anteriores (e.g. Andrade, 1993a, 1994). O estímulo referência do teste 1 é uma realização de “atapetaram” (cf. figura 6) e o do teste 2, uma realização de “adaptaram”. Trata-se de palavras iniciais de frases declara-

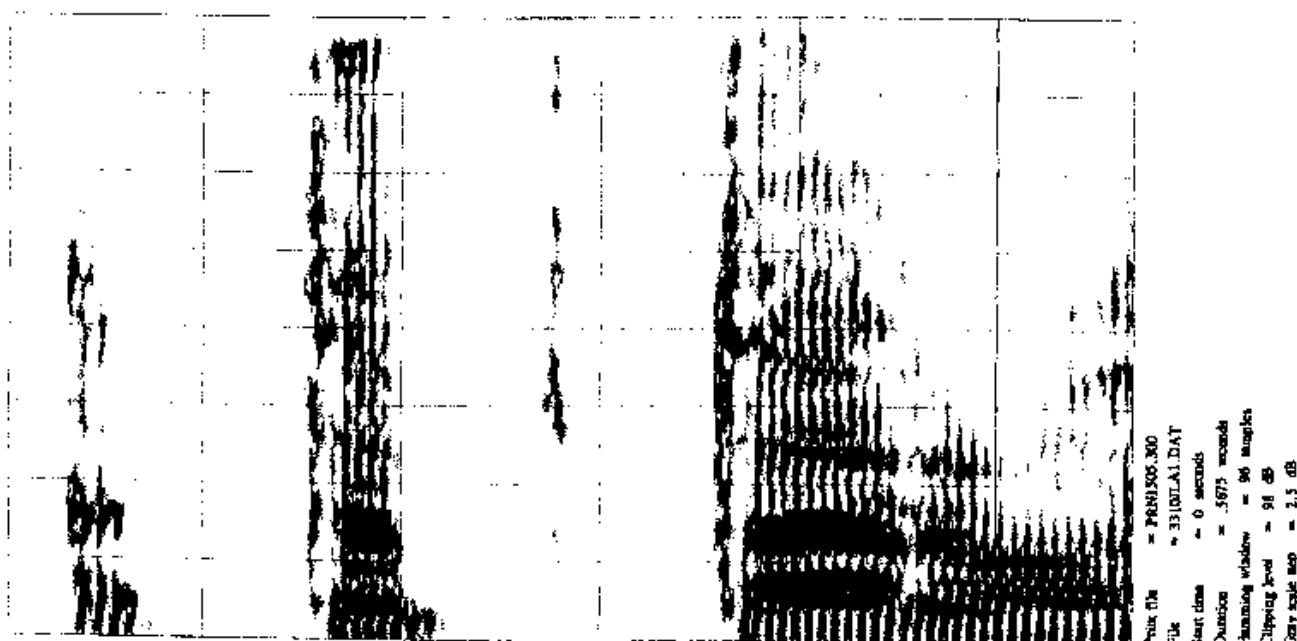


Figura 6 – Spectrograma da realização natural de JL de “atapetaram” que constitui o estímulo de referência do teste 1.

tivas produzidas com um débito considerado normal pelo falante. Nestes itens naturais, a duração da sequência [pt], e da coronal inicial é de 158ms e 50 ms em “atapetaram”, e 142 ms e 38 ms em “adaptaram”, respectivamente. A amplitude de pico da explosão da coronal da sílaba acentuada é 62 dB em “atapetaram” e 57 dB em “adaptaram”.¹⁶

Na construção dos estímulos, extraiu-se a explosão labial e manipulou-se a duração do intervalo correspondente à sequência labial+coronal.

A fim de minimizar o eventual efeito da utilização de uma palavra e de um logátomo e, ainda, a dispersão da atenção devido ao facto de a variação do intervalo oclusivo estar inserido no interior de uma sequência relativamente longa,¹⁷ *dirigiu-se a atenção do informante* para a região do sinal onde se localiza a especificação da variável, pedindo de forma explícita que dissesse se ouvia “(ata)petaram” ou “(ata)taram”. Com esse objectivo, a apresentação de cada estímulo auditivo, no decorrer do teste, foi sempre precedida pela apresentação num écran de computador da representação ortográfica das duas opções sob a seguinte forma: “1- petaram” e “2 – taram”. O informante era avisado da ocorrência eminente do estímulo através da mudança de cor das representações no écran.

Recorda-se, quanto à segunda opção (“taram”), o que foi dito sobre os resultados de outros autores que estudaram anteriormente o papel da duração do intervalo de silêncio na percepção de uma ou duas consoantes de pontos de articulação distintos: quando apenas consoante era ouvida, dominava sempre a consoante em ataque (C_2 de C_1C_2). A definição da tarefa foi concretizada depois de verificado em teste piloto que com a redução da duração, a consoante dominante era efectivamente a segunda (C_2 de C_1C_2).

No teste 1, utilizaram-se 18 estímulos, incluindo o estímulo natural. A duração do intervalo oclusivo foi variada por intervalos de 10 ms, de modo a constituir um quase-contínuo com os valores extremos de 50 ms (cf. figura 7) e 210 ms. A diferença entre a duração do intervalo de oclusão da coronal precedente (“arapetaram”; 90 ms) e a duração do intervalo correspondente à sequência [pt] varia entre -40 ms e +210 ms.

No teste 2 utilizaram-se 19 estímulos. Neste caso a duração do intervalo de silêncio tem os valores extremos de 35ms e 210 ms e varia

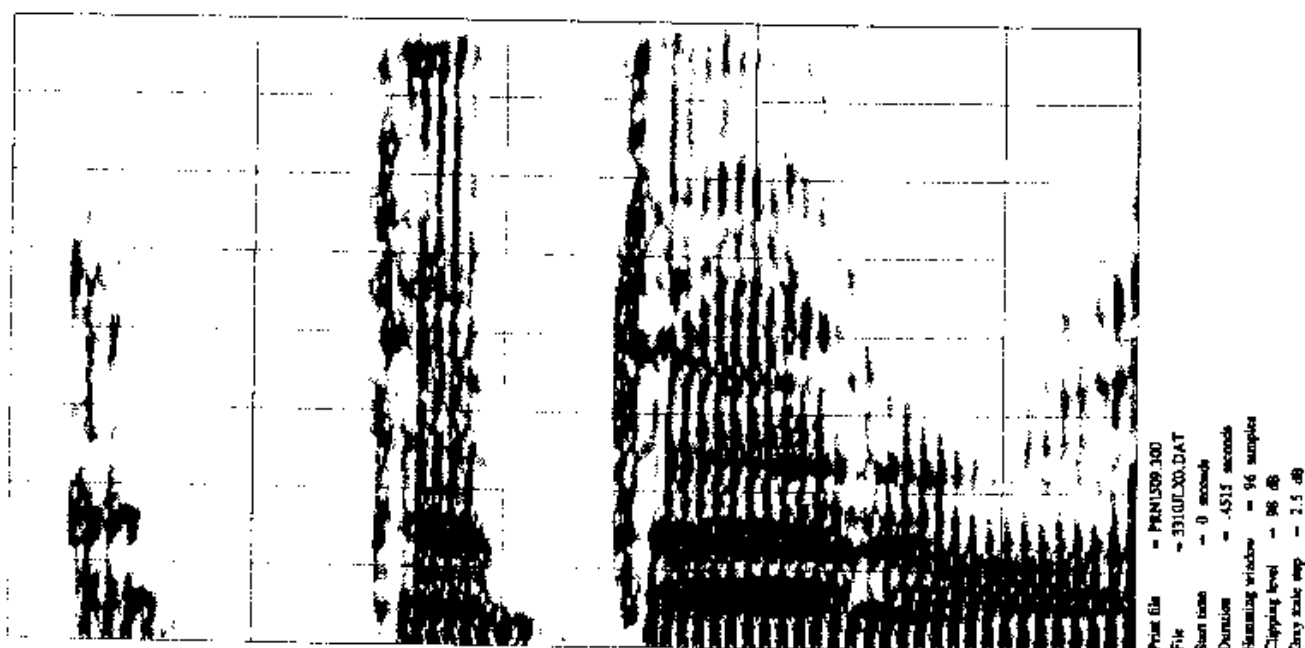


Figura 7 – Espectrograma de estímulo do teste 1: tem intervalo de silêncio editado, sem explosão e com duração de 50ms.

por etapas de 10 ms, excepto na extremidade inferior do contínuo; entre os estímulos com intervalos de silêncio iguais a 50 ms e 35 ms não há mais estímulos.

Cada estímulo foi apresentado 10 vezes. Cada um dos testes inclui 10 estímulos extra no início e no fim da lista que não foram tidos em consideração na análise. O conjunto total de itens foi dividido em 8 subconjuntos. Estes foram apresentados como pequenos testes relativamente independentes, para evitar a fadiga dos ouvintes.

A montagem dos testes foi realizada com auxílio do programa MTAPE, concebido conjuntamente por R. Berkovitz e pela experimentadora, e realizado pelo primeiro.¹⁸

Os ouvintes responderam, carregando numa de duas teclas do teclado de um computador PC. A apresentação de cada estímulo auditivo foi primeiro anunciada no écran e só depois efectivada. Após a apresentação do estímulo auditivo, os informantes tinham 4 segundos para responder. O registo automático da resposta foi complementado com informação do tempo de resposta. Esta informação é um pouco grosseira, indo apenas até às décimas de segundo, devido às limitações inerentes ao próprio PC.

Os informantes pertencem, na sua maioria à faixa etária dos 20-30 anos, à exceção de um, pertencente à geração dos 40-45 anos, a qual participara já em outras experiências realizadas por nós. A proveniência geográfica dos ouvintes é diversa, mas predominantemente centro-sul. Todos os informantes têm formação universitária, nomeadamente em linguística, excepto um (SA).

4.2. Os resultados e discussão

As figuras 8 e 9 a seguir, correspondem ao teste 1, e contêm as médias das respostas dos 9 informantes analisados. A fronteira perceptiva representada pelo cruzamento das curvas de identificação ao nível dos 50% incide na duração do intervalo de silêncio de 90 ms. Os 75% de identificação de oclusiva singular e de sequência C_1C_2 correspondem, respectivamente, às durações de 70 ms e 100 ms. Durações inferiores a 70 ms (60 e 50 ms) parecem ser preferidas pelos ouvintes em associação com a consoante singular; por sua vez, o aumento da duração para além dos 100 ms também resulta em percentagens de identificação de [pt] mais elevadas. Estas são francamente boas quando o intervalo oclusivo é de 120 ms ou mais.

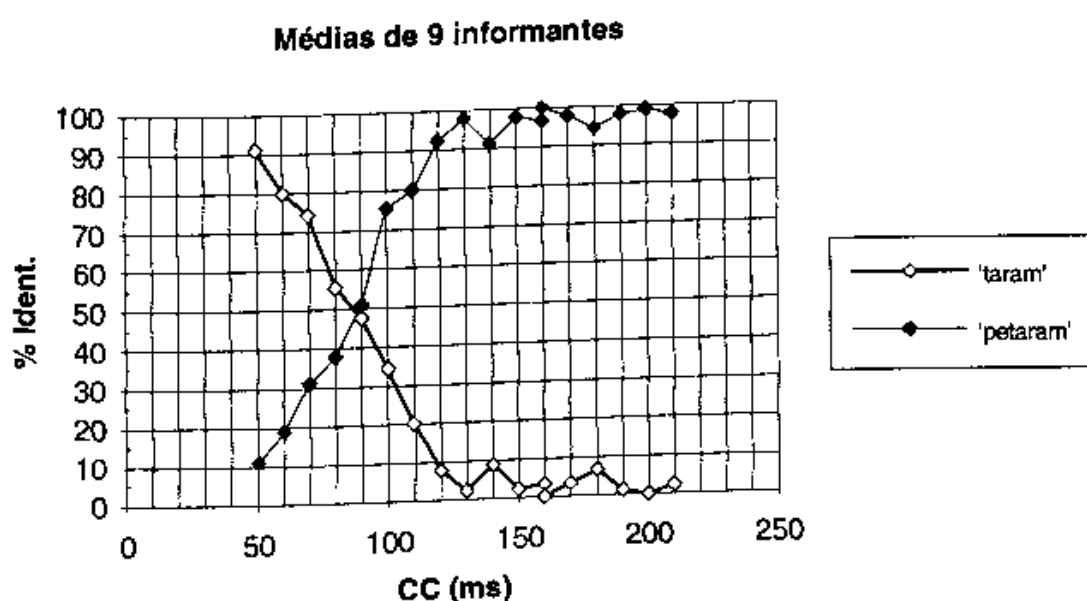


Figura 8 – Médias das respostas de nove informantes: identificação de “taram” ou “petaram”, em função da duração do intervalo de silêncio intervocálico.

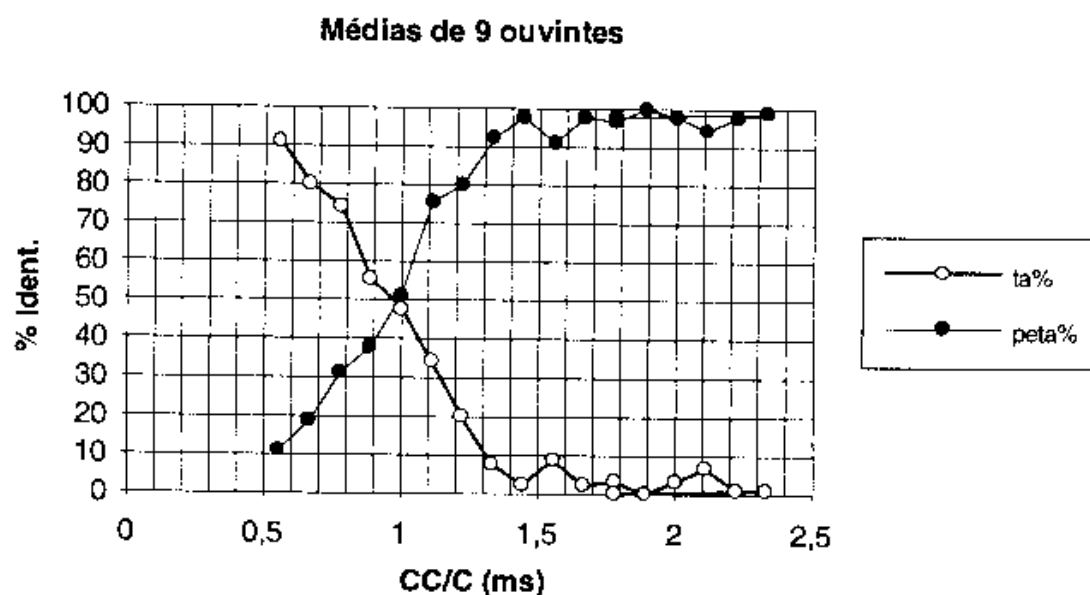


Figura 9 – Médias das respostas de nove informantes: identificação de “taram” ou “petaram”, em função da razão entre a duração do intervalo de silêncio intervocálico e a duração da consoante coronal anterior (“atapetaram”).

Se tivermos em conta a relação temporal inter-consoantes, na palavra, podemos verificar que a fronteira definida pelas médias das respostas dos ouvintes coincide, curiosamente, com a duração da consoante singular anterior, conforme se pode observar na figura 9. As figuras 10a-10i permitem verificar que a gama de variação inter-indi-

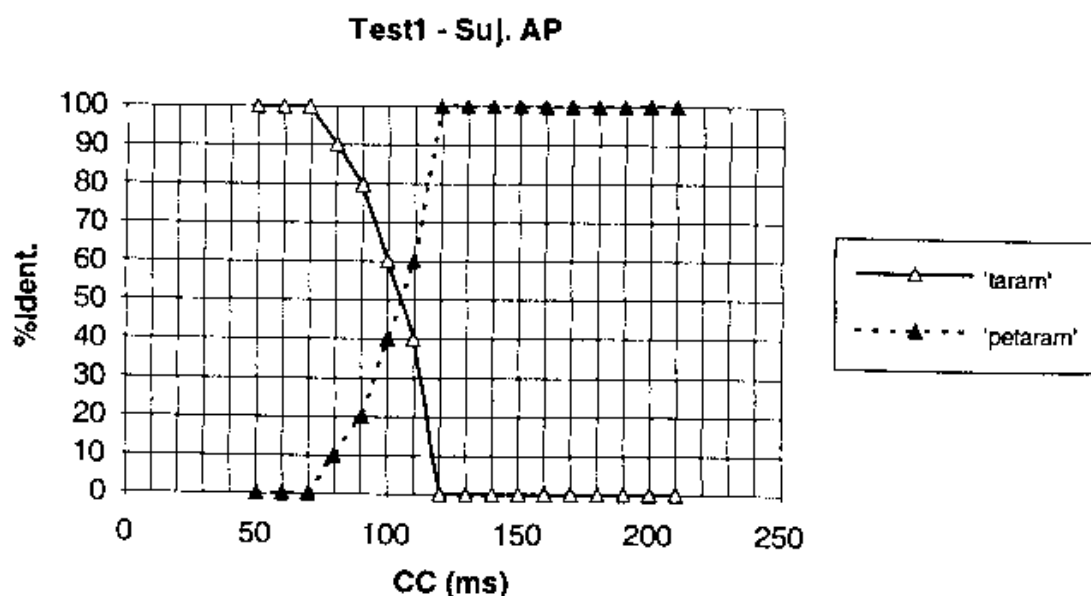


Figura 10a – Respostas de AP: identificação de “taram” ou “petaram”, em função da duração do intervalo de silêncio intervocálico.

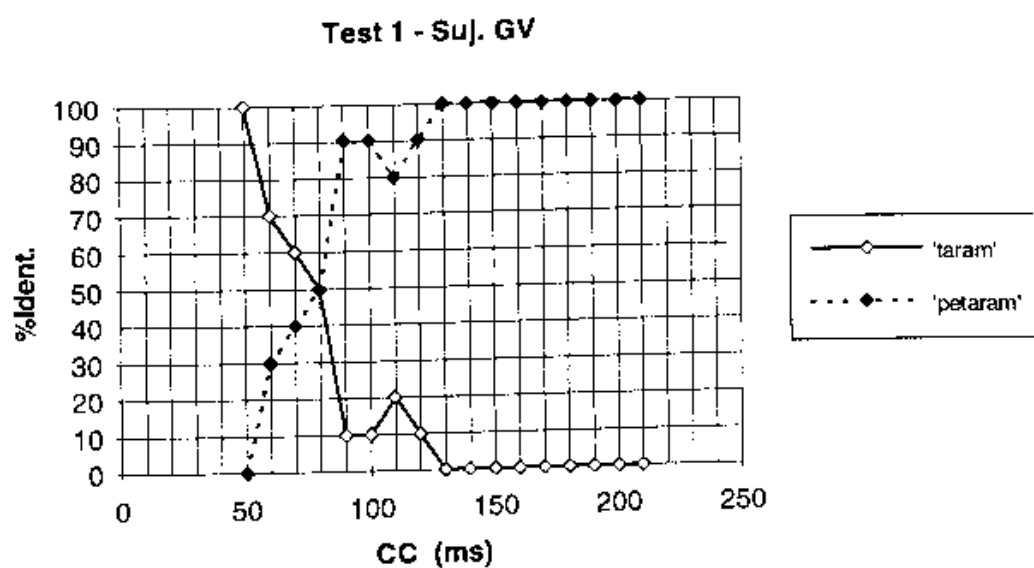


Figura 10b – Respostas de GV: identificação de “taram” ou “petaram”, em função da duração do intervalo de silêncio intervocálico.

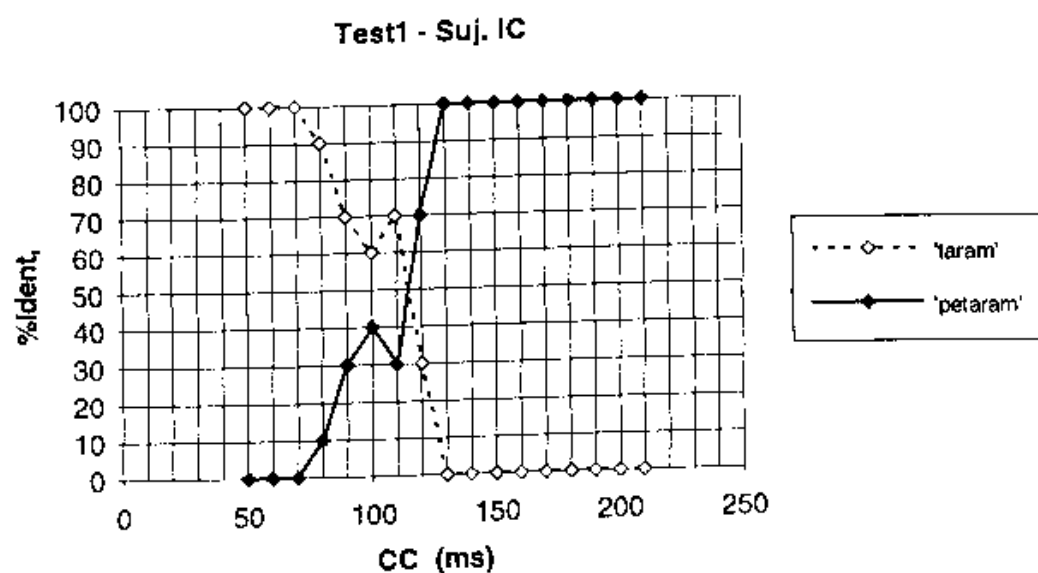


Figura 10c – Respostas de IC: identificação de “taram” ou “petaram”, em função da duração do intervalo de silêncio intervocálico.

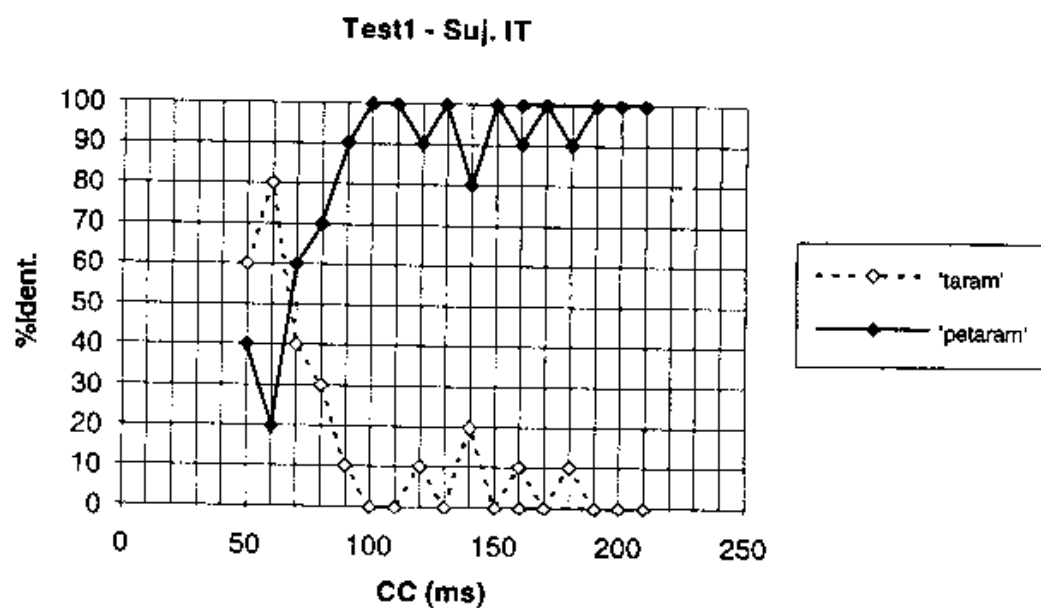


Figura 10d – Respostas de IT: identificação de “taram” ou “petaram”, em função da duração do intervalo de silêncio intervocálico.

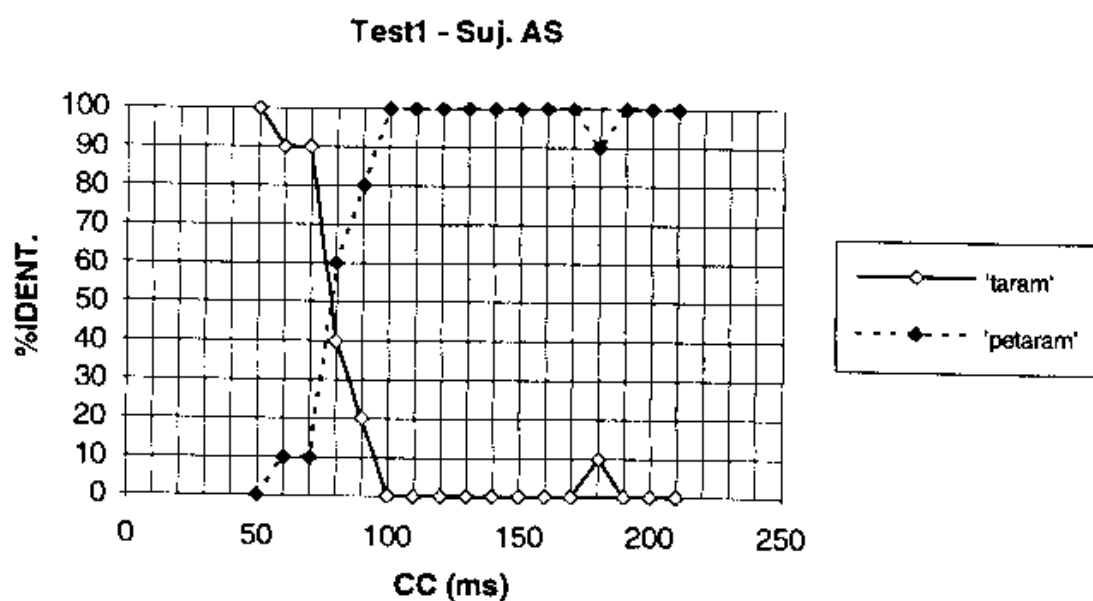


Figura 10e – Respostas de AS: identificação de “taram” ou “petaram”, em função da duração do intervalo de silêncio intervocálico.

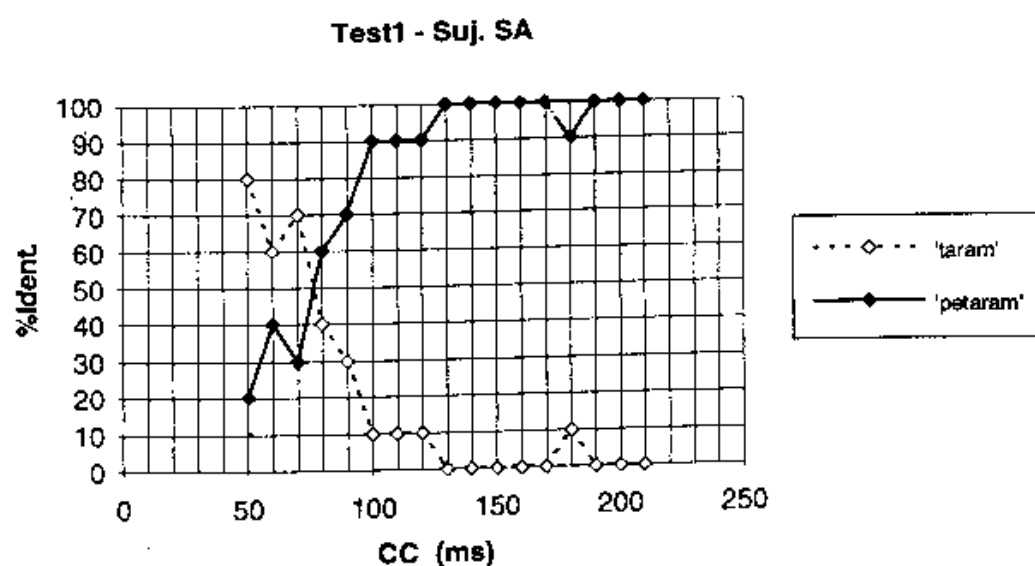


Figura 10f – Respostas de SA: identificação de “taram” ou “petaram”, em função da duração do intervalo de silêncio intervocálico.

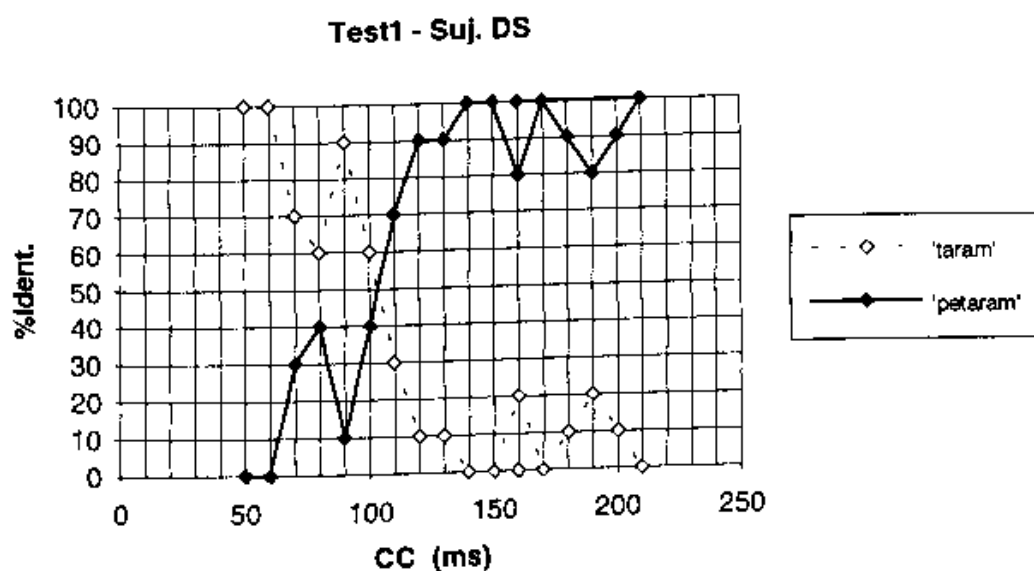


Figura 10g – Respostas de DS: identificação de “taram” ou “petaram”, em função da duração do intervalo de silêncio intervocálico.

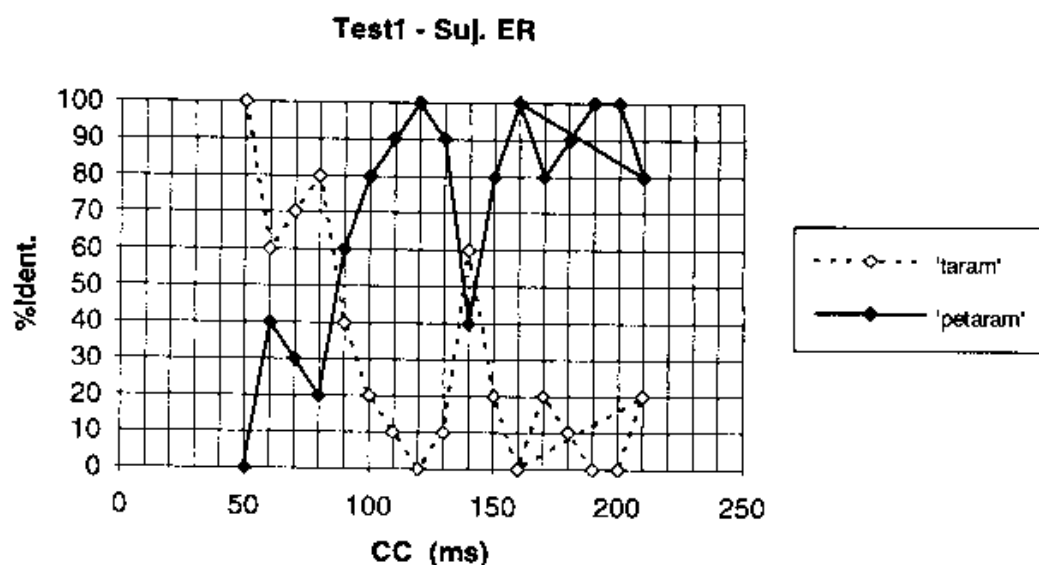


Figura 10h – Respostas de ER: identificação de “taram” ou “petaram”, em função da duração do intervalo de silêncio intervocálico.

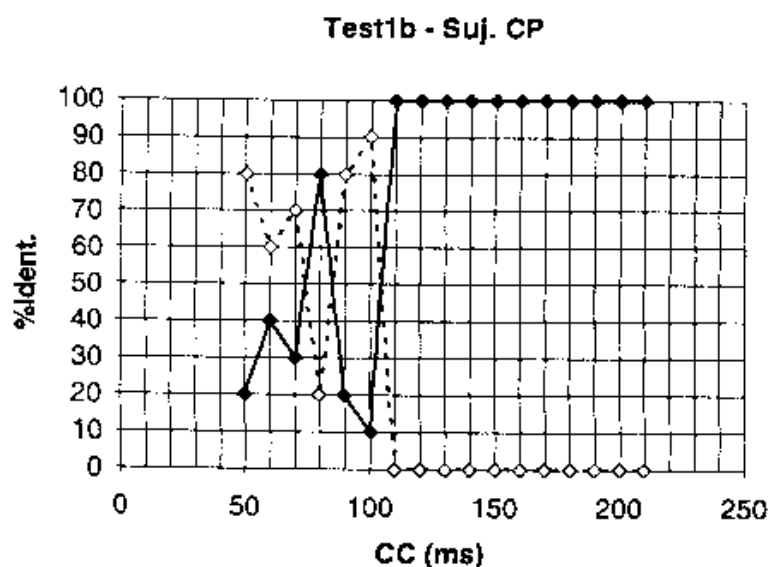


Figura 10i- Respostas de CP: identificação de “taram” ou “petaram”, em função da duração do intervalo de silêncio intervocálico.

dual quanto à localização da fronteira é de 70-120 ms, o que está aliás de acordo com o que foi apontado acima com base nos valores e identificação médios superiores da ordem dos 75% (ou mais) e indica que os valores médios são significativos para o parâmetro e a população em causa.

Os resultados obtidos não convergem com qualquer dos valores de fronteira entre C e CC (de pontos de articulação distintos) referidos na secção 2; ficando aproximadamente a meio dos 140 ms de Repp (1977) e dos 58 ms de Dorman, Raphael e Liberman (1979). Não se nos afigura uma explicação evidente para estas diferenças. É de notar, no entanto, que os valores de fronteira daqueles dois trabalhos correspondem a valores próprios de grupo consonântico e de consoante singular, para o falante que utilizámos (nas condições de estilo e débito em que foi recolhido o material de fala). Assim, uma hipótese a considerar é a de que os ouvintes estabeleceram a fronteira em coerência com as características temporais do enunciado. Uma hipótese alternativa é a de que pura e simplesmente, por razões a determinar, mas possivelmente devido a factores de natureza auditiva e outros a um nível mais central (cf. secção 2), os informantes não ouvem a labial quando o intervalo de silêncio é inferior a c. de 90 ms. Note-se que o material utilizado como estímulos neste estudo não é idêntico ao dos outros trabalhos.

As curvas de resposta individuais reflectem uma tendência bastante generalizada para os valores de percentagem mais elevados ocorrerem associados à sequência consonântica. A informante CP (cf. figura 10i) manifesta esta tendência de forma muito acentuada: evidenciando um comportamento errático quando as durações são inferiores a 120s, em contraste com uma grande estabilidade na identificação do grupo consonântico, quando a duração é igual ou superior a 120 ms. Esta informante voltará a ser objecto de comentário, um pouco mais adiante.

Os tempos de resposta encontrados para os diferentes informantes reflectem até certo ponto a tendência que se acaba de apontar: com efeito, para quatro dos ouvintes, os tempos de resposta também tendem a ser relativamente mais baixos no caso dos estímulos que obtêm taxas de identificação como CC mais elevadas (cf. figuras 11 e 12). Para outros quatro ouvintes, os tempos de resposta mais baixos não reflectem esta preferência, acompanhando simplesmente os estímulos sentidos como menos ambíguos, independentemente de serem ou não identificados como CC (cf. figura 13). Não manifesta nenhuma destas tendências a informante ER. Esta informante, aliás afasta-se dos restantes ouvintes, apresentando duas fronteiras entre C e CC: uma convergente com a dos outros informantes e outra na região de 140 ms.

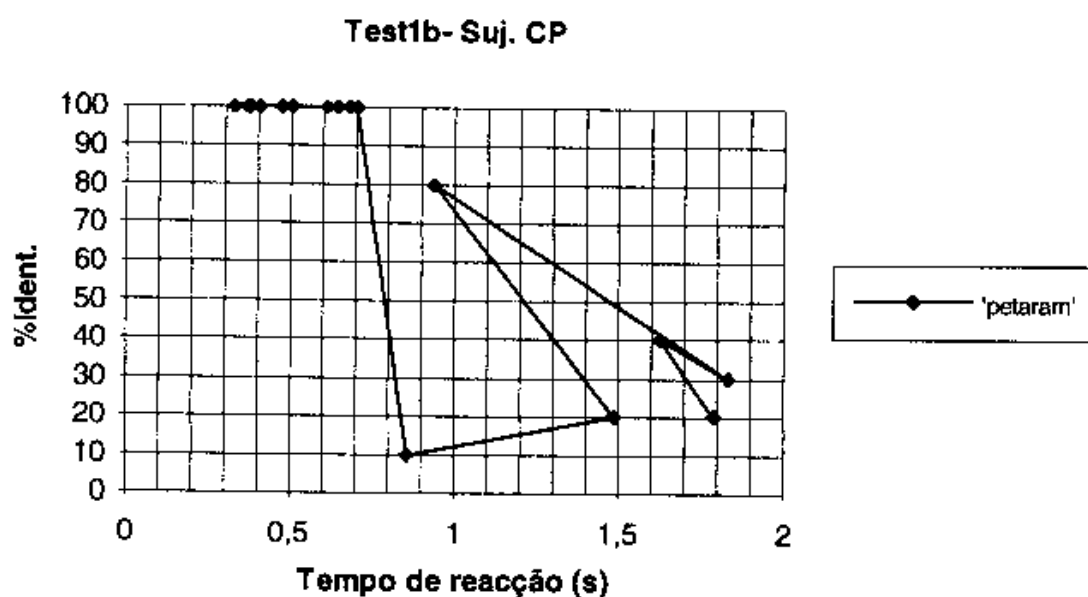


Figura 11- Tempo de reacção das respostas de CP (até à décima de segundo) ao test1.

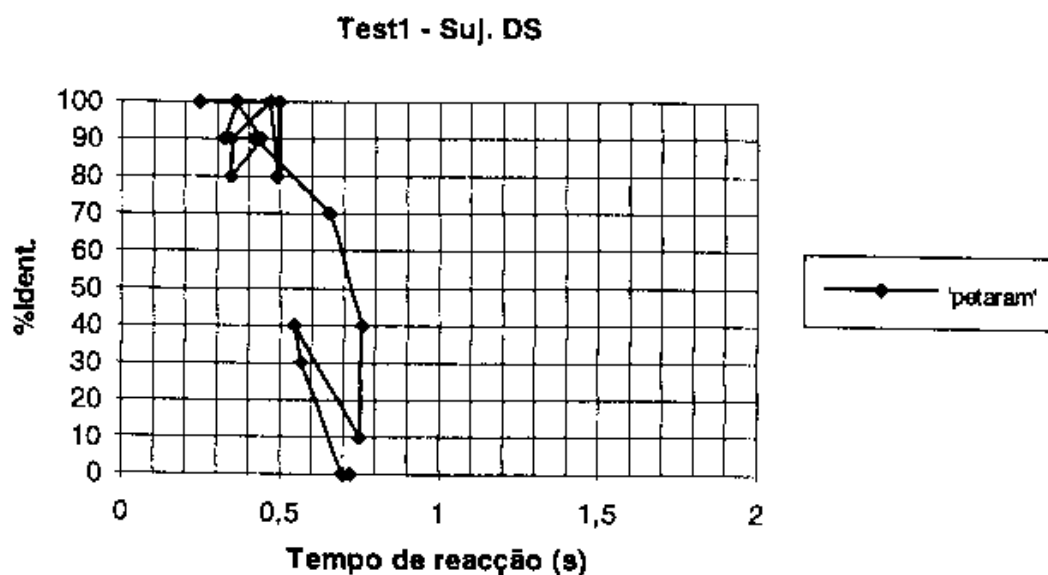


Figura 12- Tempo de reacção das respostas de DS (até à décima de segundo) ao test1. Para esta informante, observam-se tempos de reacção mais baixos em associação com as percentagens de identificação mais elevadas.

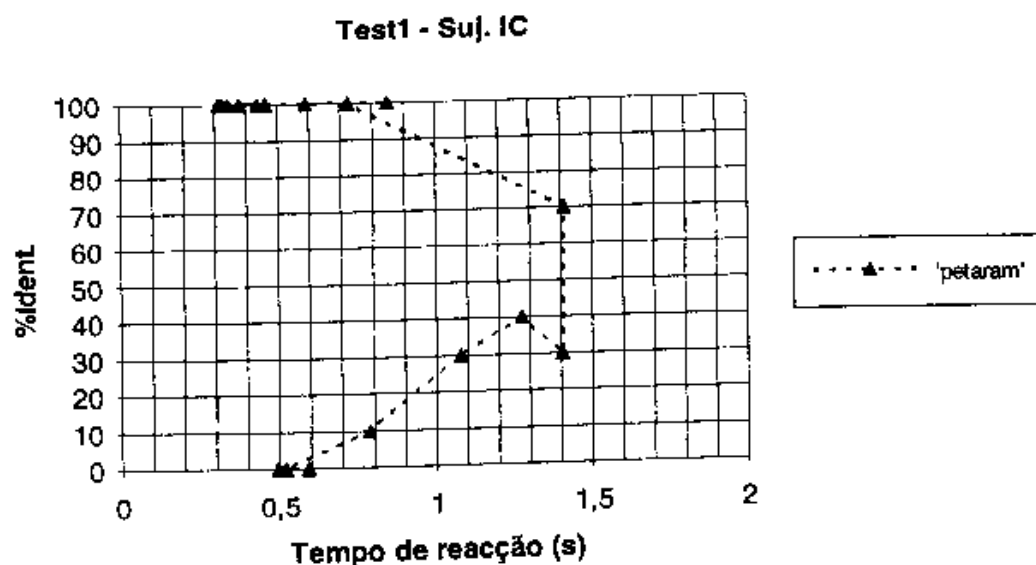


Figura 13- Tempo de reacção das respostas de IC (até à décima de segundo) ao test1. Para esta informante, os tempos mais breves estão claramente relacionados com as respostas extremas de recusa e de melhor aceitação.

A observação de maior variabilidade associada à identificação de uma consoante singular (coronal) do que à identificação da sequência de labial+coronal é compatível com a interpretação de que essa variabilidade resulta da co-ocorrência de propriedades em conflito: é provável que, com os intervalos oclusivos utilizados, e em particular na região de fronteira, *os ouvintes possam detectar a labialidade para além da coronalidade, e que o processamento dessa informação seja dificultado pela sua excessiva proximidade no tempo* (em caso extremo de proximidade, a labialidade não será de todo ouvida). Diferenças inter-individuais na resolução do conflito contribuirão, decerto para a variabilidade dos resultados. Converte neste sentido a observação que se obteve informalmente de que os ouvintes podem detectar a labialidade independentemente da duração e admitir a sua *localização na sílaba antecedente*: como [ɛptɛtár] (e não propriamente [ɛtɛtár]). Note-se que os estímulos que deram lugar a este tipo de respostas têm uma C_i mais longa do que CC. Esta observação baseia-se em dois ouvintes a quem a tarefa de identificação foi apresentada de um modo diferente daquele utilizado nos testes aqui descritos.¹⁹

É altura de se voltar a CP. A figura 14 inclui os resultados obtidos para esta informante em duas sessões distintas realizadas em seguida: test1-b (correspondente também a 10i, acima) é a repetição de

test1-a. A configuração das curvas de resposta permite pensar que houve uma *aprendizagem* da primeira para a segunda sessão. Essa aprendizagem contribuiu para otimizar a identificação das sequências de consoantes quando a duração inter-marcas de labialidade e coronalidade é suficientemente longa; no entanto, a confusão mantém-se se a duração do intervalo oclusivo é menor do que $C_1C_2/C_i = 1.2$ (C_i é a consoante de referência, neste caso a coronal inicial).²⁰

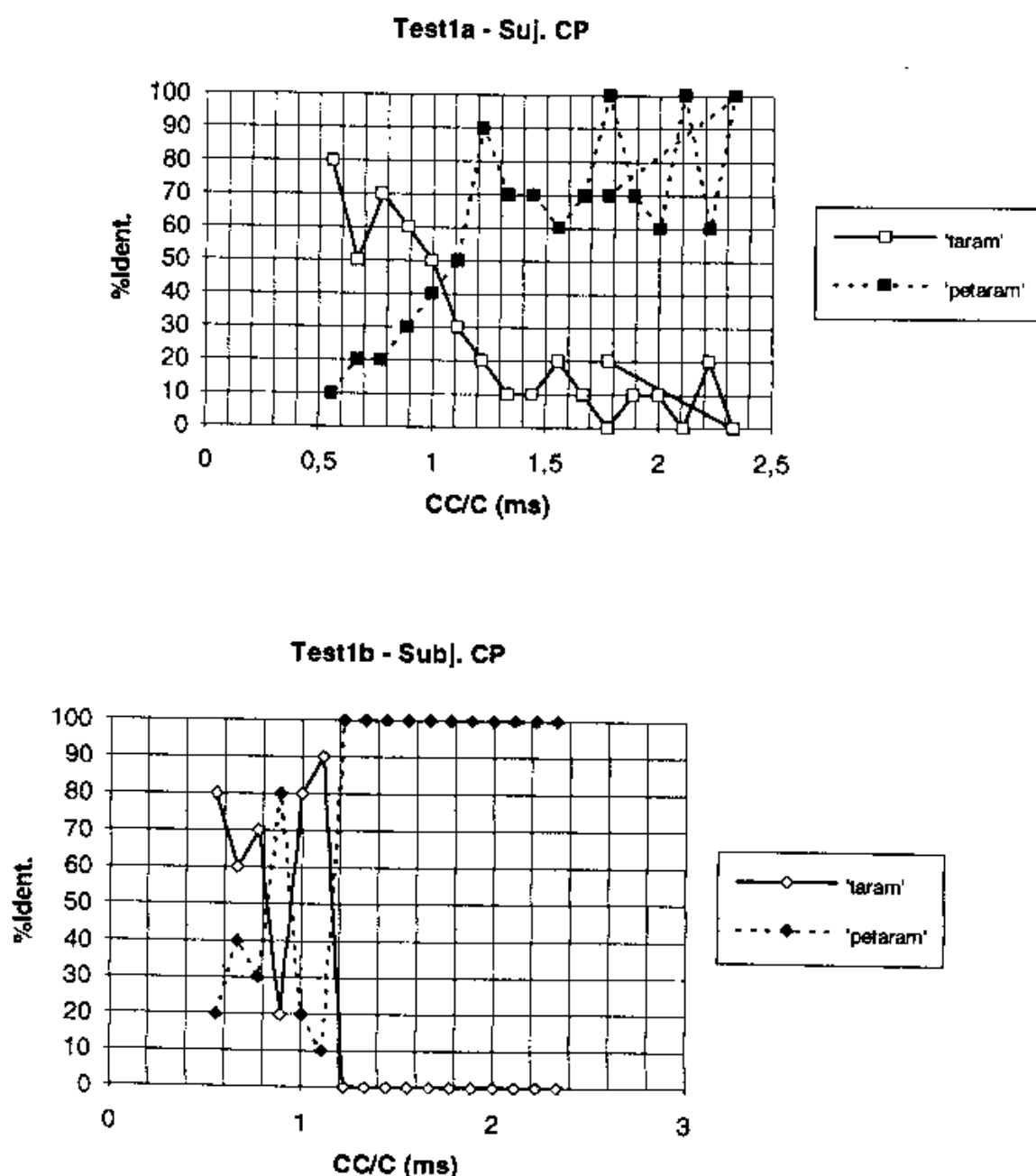


Figura 14 – Respostas de CP: identificação de “taram” ou “petaram”, em função da razão CC/C , em duas sessões distintas.

A verificação deste efeito de aprendizagem indica que a duração do segmento oclusivo desempenha um papel *redundante* na percepção das consoantes, para os ouvintes nativos do PE, ou pelo menos para alguns deles. O facto de, em PE, *não haver distinções segmentais de quantidade* está certamente na base desse estatuto. Estas considerações remetem-nos para o trabalho recente de Rochet e Rochet (1995) sobre a percepção do contraste entre consoantes geminadas e não geminadas por parte de ouvintes nativos de um sistema linguístico que possui esse contraste, o Italiano, e ouvintes nativos de um sistema que não possui o contraste, o Inglês. Em Italiano, a diferença de duração consonântica é acompanhada de uma diferença de duração da vogal antecedente: antes de geminada a vogal tende a ser mais breve do que antes de consoante simples. Tendo isso em conta, os autores examinaram a importância relativa, que os dois grupos linguísticos, atribuem à duração do intervalo oclusivo e à duração da vogal precedente, na identificação de uma geminada ou de uma consoante simples. Os resultados de Rochet e Rochet mostram que as diferenças a nível de sistema condicionam os ouvintes quanto às propriedades em que se baseiam preferencialmente para distinguir as duas classes consonânticas. Enquanto os ouvintes italianos utilizam a duração consonântica e parecem não atender à duração vocálica (será esta a propriedade redundante), os ouvintes anglófonos atendem apenas à duração vocálica (o que está também de acordo com factos fonéticos do Inglês).²¹

Tal como foi apontado anteriormente, os dados recolhidos com o teste 2 (até à data da comunicação) dizem respeito a três informantes: AP, IC e GV. Os resultados em questão permitem, por um lado, verificar a consistência do comportamento dos informantes. A consistência é visível nas próprias curvas de tempos de resposta associadas aos dois testes. Ilustra esta afirmação a figura 15, que representa os resultados de IC.

No que respeita à distinção entre C e CC, os resultados mostram que AP e IC têm um comportamento semelhante, e que GV diverge das duas:

- as primeiras não alteram significativamente a posição da fronteira, *em termos absolutos* (cf. figuras 16-18): observa-se um *abaixamento muito pequeno*, 10 ms precisamente, do test1 para o test2, nestas duas informantes (de 105 ms para 95 ms para AP e de 115 ms

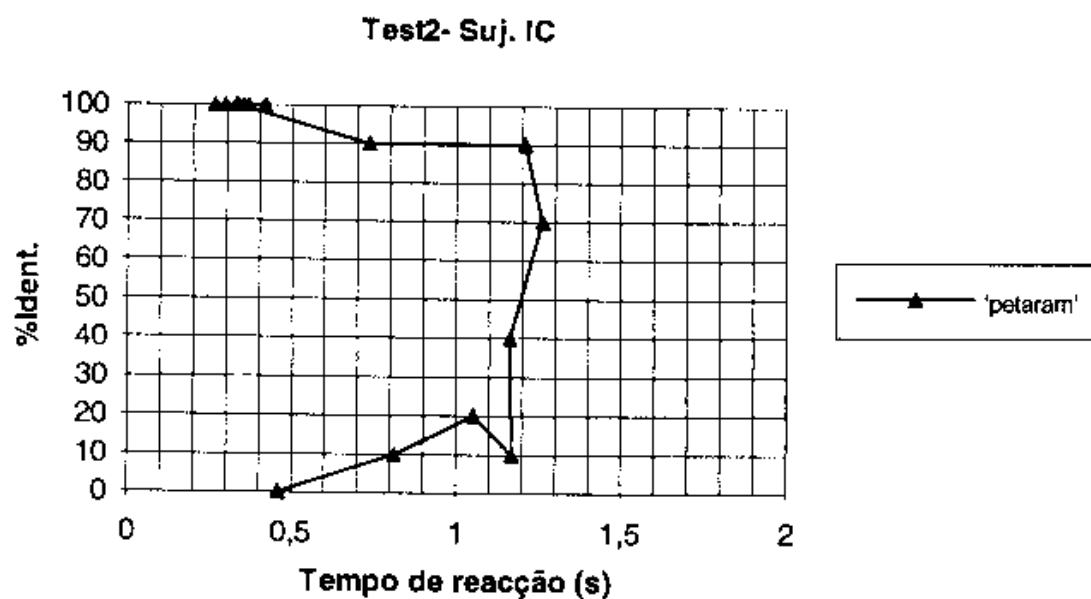


Figura 15 – Respostas de IC ao test2 (baseado em “adaptaram”): tempo de resposta

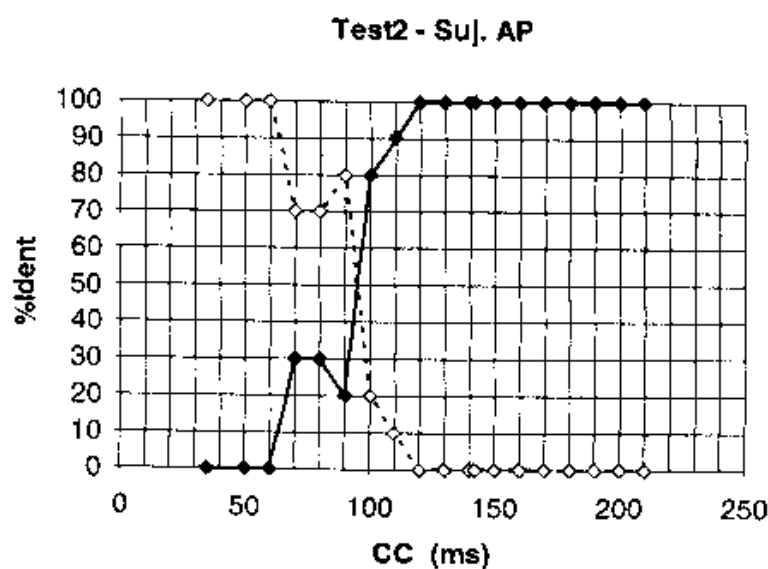


Figura 15 – Respostas de AP ao test2 (baseado em “adaptaram”): identificação de “taram” ou “petaram”, em função da duração de CC.

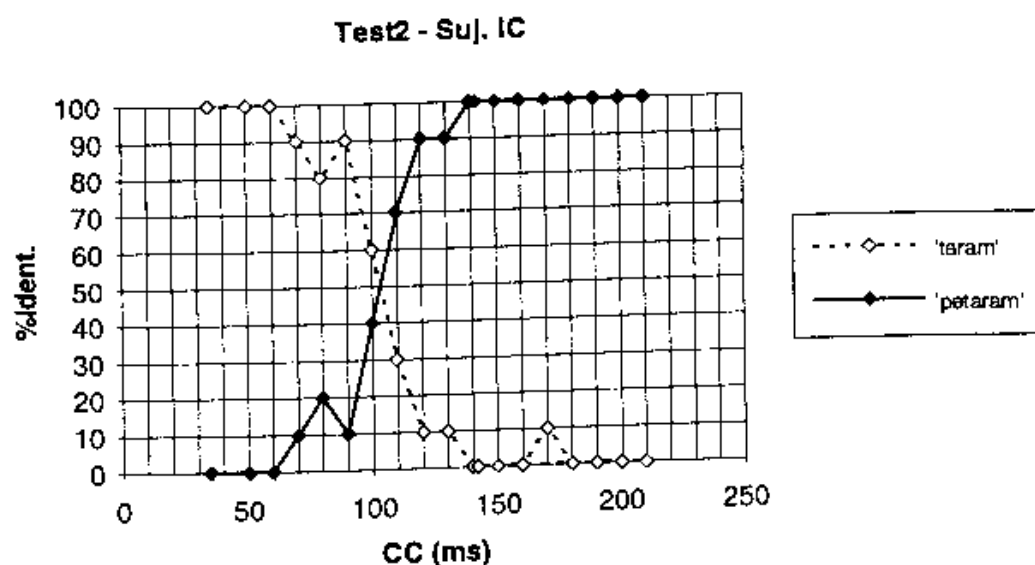


Figura 16 – Respostas de IC ao test2 (baseado em “adaptaram”): identificação de “taram” ou “petaram”, em função da duração de CC.

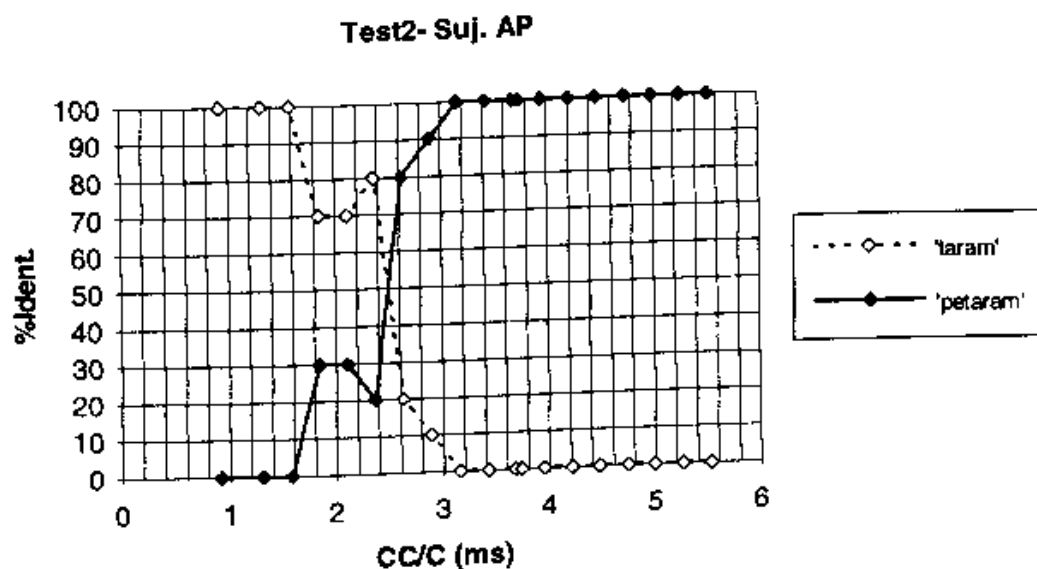


Figura 17 – Respostas de AP ao test2 (baseado em “adaptaram”), em função de C_1C_2/C_i .

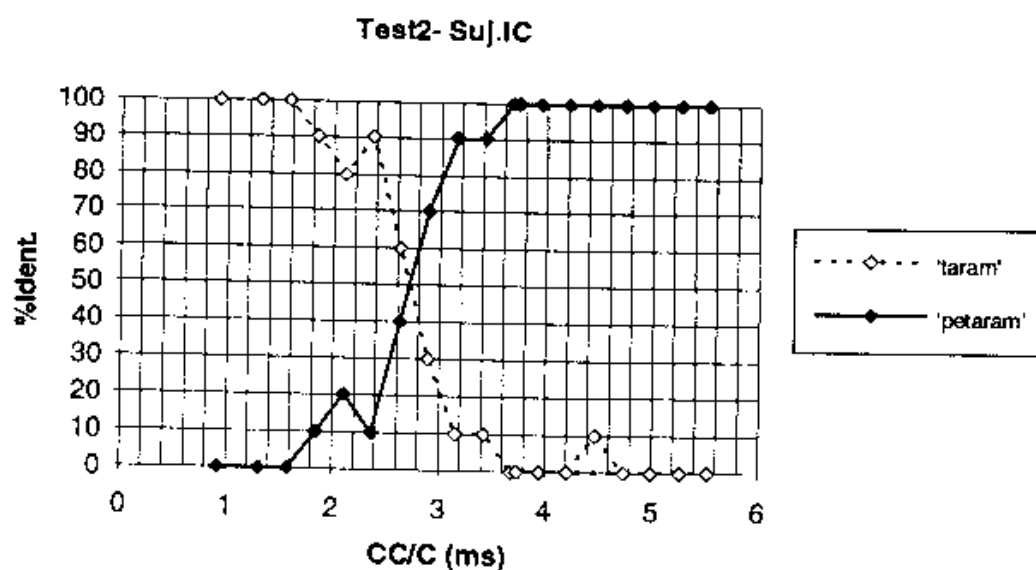


Figura 18 – Respostas de IC ao test2 (baseado em “adaptaram”), em função de CC/C.

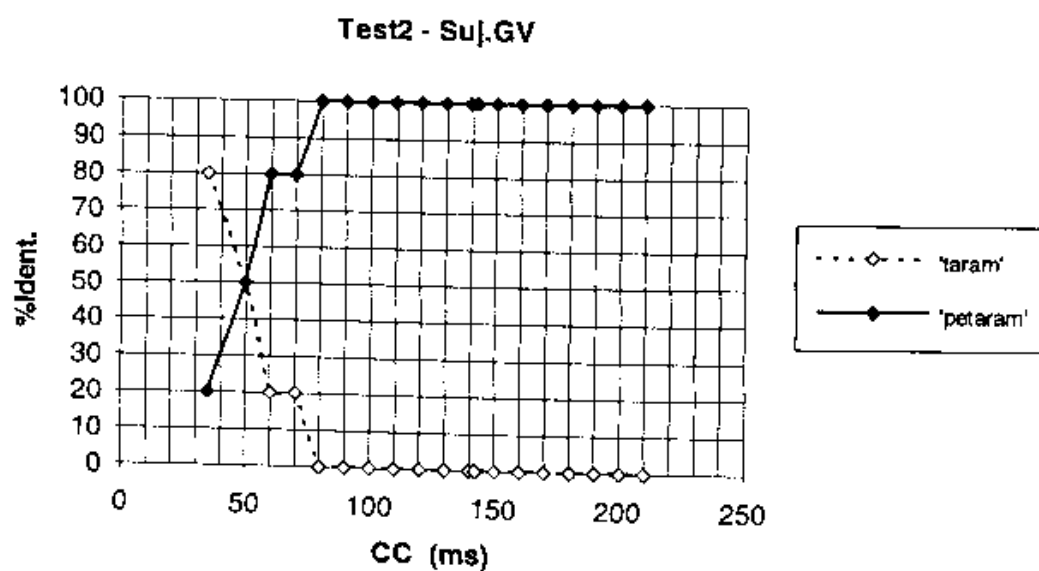


Figura 19 – Respostas de GV ao test2 (baseado em “adaptaram”): identificação de “taram” ou “petaram”, em função da duração de CC.

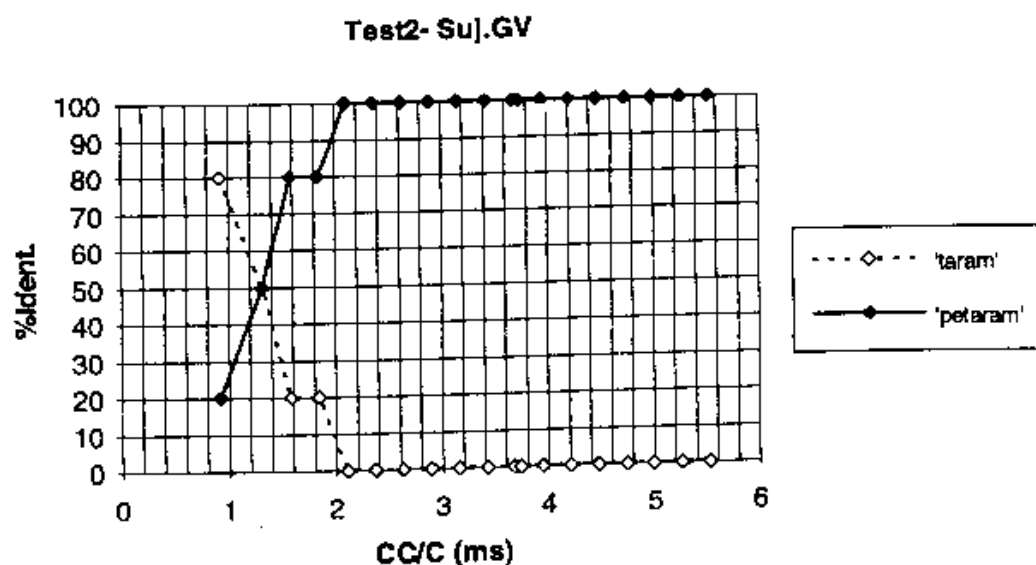


Figura 20 – Respostas de GV ao test2 (baseado em “adaptaram”), em função de CC/C.

para 105 ms em IC); neste caso, o valor de C_1C_2/C_i eleva-se de c. 1.3 para 2.5 no caso de AP, e de 1.4 para 2.7 no caso de IC;

- GV (cf. figuras 19-20), por seu lado, introduz uma alteração na posição da fronteira de 30 ms, estabelecendo-a nos 50 ms; no seu caso, o valor de C_1C_2/C_i mantém-se mais próximo nos dois contextos: 0.9 e 1.3 respectivamente em “atapetaram” e “adaptaram”.

Os dados dos dois testes não permitem determinar quais os factores envolvidos nas diferenças encontradas entre AP e IC, por um lado, e GV, por outro. No entanto, podem-se avançar algumas hipóteses. O comportamento de AP e IC parece reflectir apenas uma pequena influência contextual e ter como referência as durações típicas (em PE) de oclusivas surdas produzidas em sequência, sem fusão, nos “ataques complexos” como se referiu anteriormente em 3. O seu critério aproxima-se de valores absolutos.

Quanto a GV, é oportuno relembrar o que foi dito na Introdução sobre Dorman, Raphael e Liberman (1979). Estes autores verificaram que, ao nível de percentagem de identificação de 75%, a duração mínima necessária para os sujeitos ouvirem duas consoantes era 58 ms; ora precisamente, GV tem 80% de respostas como “pet” quando o intervalo de silêncio é 60 ms, e considera os estímulos ambíguos quando a duração baixa para 50 ms (id.= 50%).²² Os resultados mostram que GV se apoia mais em factores contextuais do que AP e IC. É possível

que os juízos de natureza mais relacional de GV sejam condicionados pela duração de C₁.

Uma hipótese a considerar é a de que, enquanto AP e IC se baseiam num *critério de congruência*, a informante GV coloca a fronteira junto ao *limite de detecção da labialidade*, o qual muito provavelmente depende de condições contextuais. Um dos factores que poderá eventualmente pesar na determinação desse limiar é a amplitude da explosão de C₂: a redução da amplitude poderá implicar que o domínio (mascaramento?) de C₂ sobre C₁ seja menor, mantendo-se tudo o resto igual. Como se viu em 4.1, a amplitude da explosão da coronal é inferior em “adaptaram” do que em “atapetaram” 5 dB, uma diferença que se sabe ser detectável.²³

O trabalho que acabamos de apresentar requer a exploração de uma série de questões. Uma das mais fascinantes entre elas está directamente relacionada com as hipóteses acima enunciadas: a interacção entre factores de ordem mais central e de factores ao nível acústico-auditivo na percepção da fala.

Notas

- ¹ O presente trabalho foi realizado no âmbito do Projecto “Estudo experimental de processos de lenição vocálica do Português Europeu”, PCSH/C/LIN/353/92, financiado pela JNICT.
- ² É interessante apontar que, segundo Bannert (1995), o Sueco pode apresentar sequências de 8 consoantes, “at least in the canonical form (...) across word boundaries, e-g- *skålmskr skratt* (roguish laughter)”. Na fala corrente, no entanto, algumas das consoantes não são realizadas foneticamente.
- ³ A maioria dos trabalhos envolveu análise da duração da consoante precedente à vogal alta sem realização fonética com base em espectrogramas. Os resultados quantitativos da análise de Companys Maldonado, Brandão de Carvalho, Mateus e Delgado Martins não estão disponíveis. Andrade (1982) recolheu também dados de natureza aerodinâmica correspondentes a sequências finais e mediais.
- ⁴ Faz-se a distinção entre realização silábica a nível fonético e estrutura silábica, a nível das representações linguísticas. Os dados de Andrade (1993a,b) correspondem a material lido com naturalidade e um débito normal, obedecendo ao requisito de inteligibilidade. É provável que material produzido num estilo muito informal, sem qualquer requisito de inteligibilidade, dê lugar a resultados significativamente diferentes, pelo menos para certos falantes.
- ⁵ O “silêncio” tem de ser entendido em termos relativos: pode haver alguma energia durante o período de oclusão; esta é, no entanto, muito baixa, a menos que a oclusiva seja uma “fricativizada” (cf. sobre esta questão, para o PE, C. Viana, 1984).

- ⁶ Nishinuma et alii (1991) referem para o Árabe, T. Benkirane (1982).
- ⁷ O factor ponto de articulação implica uma variação menor do que o modo de articulação. Certos efeitos temporais podem ser específicos de determinadas fronteiras de constituintes ou domínios prosódicos.
- ⁸ Verificou-se uma diferença de 15 ms quanto ao intervalo de silêncio necessário para a percepção de uma oclusiva
- ⁹ Os dados aerodinâmicos de Andrade (1982) foram obtidos com base nas realizações da experimentadora. O material em questão foi recolhido no Laboratório de Fonética da Universidade de Leeds, durante uma estadia parcialmente financiada pelo British Council. A recolha do material experimental foi possível graças ao apoio de Eric Brearly.
- ¹⁰ Os trabalhos de electropalatografia de Hardcastle e Roach (1979) e Recasens, Fontdevila e Pallarès (1993) mostram que as oclusivas labiais, dento-alveolares e velares do Inglês e Catalão são co-produzidas consideravelmente em sequências heterorgânicas (sem conflito articulatorio). Embora não seja evidente que a organização co-articulatoria de grupos consonânticos do PE seja idêntica à dessas línguas, é mais do que provável que a sobreposição de gestos também tenha lugar, em algum grau, na realização das sequências de duas oclusivas do PE. Este é um aspecto que se pretende investigar futuramente.
- ¹¹ Isabel Mascarenhas teve uma participação importante na recolha destes dados, enquanto membro da equipa de trabalho do projecto referido na nota 1.
- ¹² Andrade (1993 b) é uma extensão de Andrade (1993 a) do ponto de vista do número de falantes utilizados (incide sobre 6 falantes em vez de 4) e do ponto de vista dos aspectos analisados. Nos dois trabalhos utilizaram-se espectros instantâneos e espectrogramas para analisar as durações de (V₁)C₁(V₂)C₂, em que V₂ corresponde a um "e mudo" ou a um /i/.
- ¹³ As frases são declarativas, produzidas com uma entoação semelhante.
- ¹⁴ Na verdade este ponto requer verificação com dados para uma mesma vogal (e.g. presença/ausência de realização vocálica de "e mudo").
- ¹⁵ Note-se que a diferenciação pode, efectivamente reflectir representações distintas; mas também é possível que os resultados sejam condicionados por factores de outra natureza; como a posição na palavra (cf. Andrade, 1993a), a frequência de ocorrência da palavra ou a representação ortográfica.
- ¹⁶ O controlo das características dos segmentos adjacentes é feito mediante outros testes.
- ¹⁷ Ao incluir várias sílabas está-se a introduzir um número potencial de variáveis considerável.
- ¹⁸ R. Berkovitz é membro da Sensimetrics Co.. A construção deste instrumento de trabalho foi financiado pela JNICT, no âmbito do projecto de que faz parte este trabalho (cf. nota 1).
- ¹⁹ A questão será aprofundada futuramente.
- ²⁰ Um efeito de aprendizagem pode também ter tido lugar com a informante AS, a qual, devido a problemas técnicos, repetiu grande parte do teste, mas em dias distintos. Neste caso, porém, não foi obtido um registo adequado das respostas iniciais, não sendo possível examinar o efeito da repetição.

- 21 Os ouvintes ingleses constituem ainda dois grupos em função do modo como utilizam essa informação.
- 22 Note-se que Dorman et alii (op. cit.) utilizaram oclusivas vozeadas do Inglês, e que estas tendem a aproximar-se das oclusivas não-vozeadas do Português quanto às características de vozeamento e de amplitude da explosão (cf. Viana, 1984).
- 23 Está em curso um novo teste perceptivo em que as características da explosão de C2 são controladas.

5. Referências

- ANDRADE, A. (1982) – "Reduction of unstressed vowels in Portuguese", comunicação apresentada em Seminários do Dept. de Linguística e Fonética da Univ. de Leeds, Dezembro de 1982.
- ANDRADE, A. (1987) – *Um estudo experimental das vogais anteriores e recuadas em Português. Implicações para a teoria dos traços distintivos*. Diss. para Progressão na Carreira de Investigação, INIC, Lisboa, 1989.
- ANDRADE, A. (1993a) – "Estudo experimental de sequências de oclusivas em Português Europeu", *Actas do IX Encontro Nacional da APL, Coimbra, Setembro de 1993*, Colibri, Lisboa, 1994, pp.1-15.
- ANDRADE, A. (1993b) – "An acoustical study of consonantal clusters in European Portuguese", comunicação apresentada em Seminários do Departamento de Linguística da Univ. de Estocolmo, Dezembro de 1993.
- BANNERT, R. (1995) – "Swedish consonant clusters in spontaneous speech: preliminary acoustic-phonetic observations", *Proceedings ICPhs 95, Estocolmo*, pp.616-619.
- BARTKOVA, K. & C. SORIN (1987) – "A model of segmental duration for speech synthesis in French", *Speech Communication*, 6, 3, pp.245-260.
- BENKIRANE, T. (1982) – *Etude phonétique et fonctionnelle de la syllabe en arabe marocain*, Diss. de Doutoramento, Univ. de Provence, Aix-en-Provence.
- BRANDÃO de CARVALHO, J. (1987) – *Le Portugais dans la dynamique ibero-romane: trois exemples de la derive syllabique dans les phénomènes de divergence dialectale*, I, II, Diss. de Doutoramento, Univ. de Paris III.
- BLUMSTEIN, S. & K. STEVENS (1981) – "Perceptual invariance and onset spectra for stop consonants in different vowel environments", *JASA*, 67, pp.648-662.
- BAYLEY, P. & Q. SUMMERFIELD (1978) – "Some observations on the perception of s+stop clusters", *Haskins Lab.s Status Report on Speech Research*, SR53, pp.25-60.

- COMPANYS MALDONADO, M. (1959) – "Quelques remarques sur le phonétisme français et le phonétisme portugais", Actas do IX Congresso Int. de Ling. Românica, Março-Abril 1959, Lisboa, *Boletim de Filologia*, 18, Lisboa, 1961, pp.115-122.
- DORMAN, L. RAPHAEL & A. LIBERMAN (1979) – "Some experiments on the sound of silence in phonetic perception", *JASA*, 65, pp.1518-1532.
- DORMAN, L. RAPHAEL & D. ISENBERG (1979) – "Acoustic cues for a fricative-affricate contrast in word-final position", *Haskins Lab.s Status Report on Speech Research*, SR50, pp.217-229.
- FUJIMURA, O. & J. LOVINS (1978) – Syllables as concatenative phonetic units. In A. Bell and J. Hooper, orgs., *Syllables and Segments*, North Holland, Amsterdam, pp. 107-120.
- HAGGARD, M. (1973) – "Abbreviation of consonants in English pre- and post-vocalic clusters", *J. Phonetics*, 1, pp.9-24.
- HARDCASTLE, W. & P. ROACH, (1979) – "An instrumental investigation of coarticulation in stop consonant sequences" in H. Hollien and P. Hollien, orgs., *Current Issues in the Phonetic Sciences*, John Benjamins, Amsterdam, pp.531-540.
- HAZAN, V. & S. ROSEN (1991) – "Individual variability in the perception of cues to place contrasts in initial stops", *Perception and Psychophysics*, 49, 2, pp.187-200.
- KLATT, D. (1973) – "Durational characteristics of prestressed consonant clusters in English", *Res. Lab. of Electronics QPR 108*. MIT, Cambridge, MA, pp.253-260.
- LACERDA, A. e G. HAMMARSTRÖM (1952) – "Transcrição fonética do Português normal", *Revista do laboratório de fonética experimental da Fac. de Letras da Univ. de Coimbra*, 1, 1952, pp. 119-135; separata, Coimbra, 1953.
- LINDBLOM, B. (1973) – Some temporal regularities of spoken Swedish. In G. Fant and M. Tatham, orgs., *Auditory analysis and perception of speech*, Academic Press, New York, pp.387-396.
- LINDBLOM, B. B. LYBERG & K. HOLMGREN (1981) – *Durational patterns of Swedish phonology: do they reflect short-term memory processes?*, Indiana Univ. Linguistics Club, Bloomington.
- MATEUS, M.H e M.R. DELGADO MARTINS (1982) – "Contribuição para o estudo das vogais átonas [ə] e [u] no Português Europeu", *Biblos*, 18, pp.111-125.
- NISHINUMA, Y., D. DUEZ & C. PABOUDJIAN (1991) – "Automatic classification of consonant clusters in French", *Speech Communication*, 10, pp.395-403.

- RECASENS, D., J. FONTDEVILA, M. PALLARÈS & A SOLANAS (1993) – “An electropalatographic study of stop consonant clusters”, *Speech communication*, 12, 4, pp.335-356
- REPP, B. (1977) – “Perceptual integration and differentiation of spectral information across intervocalic stop closure intervals”, *Haskins Lab.s Status Report on Speech Research*, SR50, pp.115-122.
- ROCHET, B. & A. ROCHET (1995) – “The perception of the single-geminate consonant contrast by native speakers of Italian and anglophones”, *Proceedings ICPHs 95, Estocolmo*, pp.616-619.
- VIANA, C. (1984) – *Etude de deux aspects du consonantisme du Portugais: fricatisation et dévoisement*, Diss. de Doutoramento de 3º Ciclo, Univ. de Estrashurgo.