

UNIVERSIDADE DE LISBOA
FACULDADE DE LETRAS



Percepção dos Tons em Mandarim por Falantes Nativos do Português Europeu

Yingying Peng

dissertação orientada pela Prof.^a Doutora Sónia Marise de Campos Frota, especialmente elaborada para a obtenção do grau de Mestre em Linguística

MESTRADO EM LINGUÍSTICA

2021

Agradecimentos

Agradeço, em primeiro lugar, à minha orientadora, Professora Doutora Sónia Frota, pela paciência, dedicação e apoio durante os últimos três anos, por me ensinar e ajudar sempre.

À Doutora Marisa Cruz, pela disponibilidade, por me ajudar sempre assim que preciso e estar sempre disponível para me ajudar e dar sugestões.

À minha família, especialmente à minha irmã e ao meu pai, obrigada pela compreensão, sugestões e muitos apoios à distância. Obrigada por verem sempre o melhor em mim e nunca me deixarem desistir.

À Inês, Vanessa e ao Bruno, obrigada por serem os meus amigos, obrigada pelas companhias, pelas brincadeiras, pelos jantares, por me ensinarem a fazer bolos. Obrigada pelas conversas, pelos apoios, pelos momentos partilhados e por me fazerem sentir em casa.

À Cláudia Braga e à sua família, por ser a minha primeira amiga portuguesa, pelas brincadeiras, pelo jantar de Natal, pela companhia e generosidade. Obrigada por me aceitar, ajudar e apoiar constantemente sem pedir nada de volta. À Cláudia, obrigada pela amizade, pelas piadas e por partilhar sempre uma visão semelhante à minha.

Ao João Dias, pelas palavras de encorajamento, sugestões e motivações. Obrigada por me acompanhar nos momentos difíceis, por acreditar em mim, ouvir, compreender e dar-me uma visão positiva.

Às minhas amigas Di Yang, Xin Xie, Sihan Wu, Jiayuan Guo e Siyu Luo pelas companhias, pelas forças, conversas, pelos jantares e por todos os momentos partilhados.

Por fim, agradeço a todos os informantes, pelo tempo dedicado e pela paciência dispensada.

Índice

Resumo da Dissertação.....	IV
Abstract.....	V
Lista de Figuras.....	VI
Lista de Tabelas.....	VII
1. Introdução.....	1
2. Enquadramento teórico	4
2.1. Os tons em Mandarim.....	4
2.2. Percepção dos tons do Mandarim.....	6
2.3. Factores que influenciam a percepção dos tons.....	8
2.3.1. A experiência de L1	9
2.3.2. Processamento do tom e do segmento.....	11
2.3.3. Assimilação dos tons.....	13
2.3.4. Pistas fonéticas: falantes de línguas tonais e falantes de línguas não tonais	14
2.3.5. Influência da tarefa e outras condições metodológicas.....	15
2.3.6. Efeito da Ordem.....	16
2.4. Identificação de Tom2 e Tom3: um caso especial	17
2.5. Características Entoacionais do Português Europeu.....	21
2.5.1 Descrição entoacional das frases declarativas	22
2.5.2 Descrição entoacional das frases interrogativas	23
2.5.3. Descrição entoacional de outros tipos frásicos.....	24
2.5.4. Percepção de entoação dos tipos frásicos	25
3. Metodologia.....	26
3.1. Participantes.....	27
3.2. Materiais	27
3.3. Tarefa de percepção.....	29
3.4. Tratamento dos dados.....	30
4. Resultados.....	30
4.1. Análise das respostas: discriminação dos pares de tons diferentes.....	31
4.1.1. Interação entre Tom e ordem	33
4.1.2. Interação entre Segmento e Tom	35
4.1.3. Interação entre Tom, Ordem e Segmento	36
4.2. Análise do tempo de reação: discriminação de pares de tons diferentes.....	38

4.2.1. Interação entre Tom e Ordem.....	40
4.3. Efeitos de Segmento e Tom	41
4.3.1. Segmento e Tom: análise das respostas.....	41
4.3.2. Segmento e Tom: análise do Tempo de Reação.....	42
5. Discussão	44
6. Conclusão.....	48
7. Referências.....	51

Resumo da Dissertação

O presente estudo visa investigar a percepção dos tons do Chinês Mandarim por falantes nativos do Português, determinar os fatores que condicionam a percepção dos tons, nomeadamente a natureza dos contrastes tonais (tipo de tom), a posição do tom (ordem) e a natureza do segmento em que o tom ocorre, e explorar a interação entre variação segmental e variação tonal na percepção dos tons.

Para atingir esse objectivo, foi realizada uma tarefa de percepção em que os participantes avaliaram o grau de semelhança/diferença perceptiva dos contrastes tonais e segmentais.

Os resultados mostraram que os falantes nativos do Português consideram que os pares que partilham características fonéticas parecidas, como é o caso dos pares T1T2 e T4T2, são mais difíceis de discriminar do que os restantes pares de tons. É possível que os falantes recorram mais à altura de F0 na discriminação dos tons, dado os pares com *onset* e *offset* semelhantes (e.g, T1T2, T2T1, T4T2 e T1T4) serem mais difíceis de discriminar.

A posição do tom revelou-se um fator não significativo. Já quanto à presença de contraste segmental e contraste tonal, os resultados confirmaram que os participantes são mais sensíveis ao contraste segmental que ao contraste tonal. Para além disso, o efeito do tom também foi significativo, pois pares com segmentos diferentes e tons diferentes, ou segmentos iguais e tons diferentes, foram percebidos como sendo mais diferentes que, respectivamente, pares com segmentos diferentes e tons iguais ou pares com segmentos iguais e tons iguais. Por outras palavras, os falantes do Português não são “tone deaf”. Diferenças verificadas quanto ao tempo de reação nas condições tom igual e tom diferente apontam no mesmo sentido.

Foi observado ainda que os falantes nativos de Português têm mais dificuldades na discriminação de tons quando a vogal é ‘A’ do que com a vogal é ‘I’. Esse resultado não era esperado e poderá dever-se ao facto de a produção de ‘I’ com os vários tons ser mais estável do que ‘A’ em Mandarim.

Palavra-chave: percepção, tom, Mandarim, Português Europeu

Abstract

The present study aims to investigate the perception of Mandarin tones by Portuguese native speakers, establish the factors that affect tone perception, namely the type of tone contrast, tone order, and the type of segment that bears the tone, as well as explore the interaction between segment variation and tone variation in tone perception.

To this end, a perception task was performed in which participants were asked to assess the degree of similarity/difference of tone and segment contrasts.

The results have shown that tone pairs sharing similar phonetic properties, like T1T2 and T4T2, are the most difficult to discriminate. This suggests that listeners use F0 height as the main perceptual cue, since pairs with similar F0 *onset* and *offset* (e.g., T1T2, T2T1, T4T2 and T1T4) are considered more similar than the others.

Tone order was found not to be a significant factor for tone perception. By contrast, the difference between segmental and tonal contrasts was relevant. Listeners were more sensitive to the former than the latter. In addition, there was a main effect of tone, with better discrimination of pairs with a tonal contrast than without, other things being equal. This shows that Portuguese native speakers are not “tone deaf”. Differences in reaction time between the same and different tone conditions further support this finding.

It was also found that native speakers of Portuguese have more difficulties in discriminating tone pairs with the vowel 'A' than with the vowel 'I'. This result was not expected. A possible explanation is that the production of the vowel 'I' with the various tones is more stable than the vowel 'A', in Mandarin.

Keywords: perception, tone, Mandarin, European Portuguese

Lista de Figuras

Figura 1. Descrição dos tons em Mandarim (retirado de Pelzl, 2019)	4
Figura 2. Efeito da ordem: a ordem em que o estímulo é apresentado é 1-2 em (a) e 2-1 em (b) (retirado de Karypidis, 2007).....	17
Figura 3. Interação entre tom e ordem.	34
Figura 4. Interação entre Segmento e Tom-I.	36
Figura 5. Interação entre Tom, Ordem e Segmento.	37
Figura 6. Interação entre Tom e Ordem.	40
Figura 7. Tempo de reação de Segmento Igual-Tom Igual, Segmento Diferente-Tom Igual, Segmento Igual-Tom diferente e Segmento Diferente-Tom Diferente.	44

Lista de Tabelas

Tabela 1. Contornos nucleares por tipo frásico descrito.....	25
Tabela 2. <i>Corpus</i>	28
Tabela 3. Combinações com ba1. “I” indica igual, “D” indica diferente.	28
Tabela 4. Média e desvio padrão das diferentes sequências tonais	32
Tabela 5. Tempo de reação para diferentes combinações de tons.	39
Tabela 6. Tempo de reação para Ordem A e Ordem B.....	39
Tabela 7. Tempo de reação para Segmento A e Segmento I.....	40
Tabela 8. Média e desvio padrão das diferentes sequências tonais: Segmento Igual/Diferente, Tom Igual/Diferente.....	42
Tabela 9. Tempo de Reação médio e desvio padrão de Segmento Igual-Tom Igual (Seg I-Tom I), Segmento Igual -Tom D (SegI-Tom D), Segmento Diferente-Tom Igual (Seg D-Tom I) e Segmento Diferente-Tom Diferente (Seg D-Tom D).	43

1. Introdução

As línguas tonais são línguas que se usam as configurações melódicas para veicular significados lexicais, ou seja, uma palavra pode ter diferentes significados dependendo do tom. Línguas como o Tailandês e o Mandarim são línguas tonais. As línguas entoacionais são línguas em que as configurações melódicas não estão associadas aos significados lexicais, mas ao nível frásico. O Inglês e a maioria das línguas europeias são línguas entoacionais.

O Mandarim é uma língua tonal. O Mandarim tem 4 tons: Tom 1 (alto), Tom 2 (ascendente), Tom 3 (descendente-ascendente) e Tom 4 (descendente). As palavras têm significados diferentes quando estão com tons diferentes (e.g *ma* significa ‘mãe’ com o tom 1 e ‘cavalo’ com o tom 3).

A percepção dos tons tem sido o alvo de muitos estudos (e.g Gandour, 1983; Lee & Wurm, 1996; Li, 2016, entre outros). Os falantes nativos de Mandarim usam o hemisfério esquerdo ao processar os tons do Mandarim, enquanto falantes de Inglês usam o hemisfério direito (Wang, Jongman, & Sereno, 2001).

Entre todas as combinações de tons, algumas combinações são consideradas de discriminação mais difícil que outras. Por exemplo, o par T2T4 é mais fácil de discriminar para os falantes nativos de Tailandês, uma língua tonal, do que para falantes australianos do Inglês (Tsukada, 2019). A par T2T3 é referido como um par mais difícil de discriminar para falantes nativos de línguas tonais (e.g Tailandeses) e não tonais (e.g Inglês, Francês, Coreano; So & Best, 2014; Tsukada & Han, 2019; Rungruang & Mu, 2017).

Relativamente aos factores que condicionam a percepção dos tons, vários estudos prévios apontam para a influência de L1 como um factor importante. Falantes de línguas tonais têm mais experiência com tons, logo é esperado que possam ter melhor desempenho que falantes de línguas não tonais (Schaefer & Darcy, 2014; Wayland & Guion, 2004; Li, 2016 entre outros). No entanto, as semelhanças fonéticas em termos de tons ou segmentos (vogais, consoantes) entre duas línguas tonais podem confundir em vez de ajudar na percepção da outra língua tonal (Tsukada et al., 2013; Hao, 2012; Levy, 2009, entre outros). Recentemente, Li, Shao, & Bao (2017) observaram que um mesmo tom pode ser alto e contínuo para os falantes nativos de Mandarim, mas é percebido como um tom descendente por falantes de Vietnamita.

Vários estudos confirmam que falantes de L1 diferentes recorrem a pistas diferentes na percepção dos tons. Por exemplo, falantes de línguas entoacionais, como Inglês e Francês, recorrem mais à altura de F0 que ao contorno de F0 na percepção (Gandour, 1983; Qin & Mok, 2011).

Para além da L1, outros factores como a ordem em que os tons são apresentados, a aprendizagem de L2, ou factores metodológicos podem influenciar a percepção (cf., para o efeito da ordem, Francis & Ciocca, 2003; para a aprendizagem de L2, Lee, Tao, & Bond, 2009; para outros fatores metodológicos, Hung & Chung, 2016).

No que diz respeito à interação entre o processamento segmental (vogal, consoante) e suprasegmental (tom), alguns estudos consideram que os falantes são mais sensíveis às diferenças segmentais que suprasegmentais (Sereno & Lee, 2015; Yip, Leung, & Chen, 1998), mas outros defendem que a interferência do tom faz a classificação dos segmentos ser mais lenta e vice-versa (Lee & Nusbaum, 1993). Hao (2018) observa que ter mais experiência em Mandarim faz os falantes do Inglês serem mais sensíveis ao tom que ao elemento segmental.

Resumidamente, os tons são percebidos da forma diferente entre falantes com experiências de L1 diferentes. Tons de fácil percepção para uns podem ser difíceis para outros. Vários estudos prévios sobre a percepção dos tons no Mandarim focaram-se em fatores, como a duração, contorno de F0 e direção de F0, que podem condicionar a discriminação dos tons (Li, Shao, & Bao, 2017; Massaro, Cohen, & Tseng, 1985; Fu & Zeng, 2000, entre outros). O objectivo geral do presente estudo é investigar a percepção dos tons do Mandarim por falantes nativos de Português Europeu, uma língua entoacional. O presente trabalho tem por base três objetivos centrais:

1. Estudar a percepção de tons por falantes de uma língua não tonal, em que diferenças melódicas assumem uma função não lexical, mas antes discursiva ou pragmática.
2. Determinar os fatores que condicionam a percepção dos tons, nomeadamente a natureza dos contrastes tonais (tipo de tom), a posição do tom (ordem) e a natureza do segmento em que o tom ocorre. Isto permitirá verificar quais são os tons mais facilmente distinguidos, e quais são os mais difíceis de distinguir, para falantes portugueses.
3. Explorar a interação entre variação segmental e variação tonal na percepção dos tons.

Considerando estudos anteriores sobre a percepção dos tons, interessa-nos identificar os pares de tons mais difíceis/fáceis de discriminar para falantes nativos do Português, perceber os factores que poderão condicionar a sua percepção dos tons do Mandarim e comparar a interação entre o processamento segmental e suprasegmental. Partimos das seguintes hipóteses:

- **Hipótese 1** - O par **T2-T3** é o par mais difícil para os falantes portugueses, dado que os dois tons possuem características semelhantes, quer ao nível do pitch inicial registado, quer ao nível da duração. (Gandour 1978: 43). O par T2T4 é o par mais fácil para os falantes portugueses, uma vez que os dois tons têm características fonéticas distintas (e.g, contorno e altura de F0) e os falantes não nativos de língua tonal são mais sensíveis aos tons quando estes assumem função pragmática na sua língua (Braun & Johnson, 2011). Como o Português é uma língua entoacional e uma subida melódica assinala uma pergunta e uma descida uma afirmação, espera-se melhor desempenho na percepção do T2 (tom ascendente) versus o T4 (tom descendente).
- **Hipótese 2** - A posição do tom pode influenciar a identificação dos tons, uma vez que os tons têm níveis de F0 diferentes e a ordem em que essa diferença de F0 é apresentada (ouvir a sílaba com F0 alto primeiro ou depois) afecta o desempenho dos falantes (por exemplo, em Francis & Ciocca (2003) é observado que falantes do Cantonês são mais sensíveis à diferença de F0 quando a primeira sílaba tem F0 mais alto que a segunda).
- **Hipótese 3** - Os ouvintes são mais sensíveis à variação em segmentos que em tons, ou seja, os pares que se distinguem apenas em segmentos têm um grau de semelhança mais baixo que os se distinguem em tons, devido ao papel distintivo dos segmentos na língua nativa e ao papel não distintivo dos tons.

O presente estudo organiza-se em 6 capítulos. No capítulo 2, apresenta-se o enquadramento teórico. Dado que o estudo envolve a percepção do Mandarim por falantes de uma língua entoacional, é apresentada uma breve descrição dos tons do Mandarim (secção 2.1) e da percepção dos tons por falantes nativos e não nativos (secção 2.2), bem como dos fatores que podem influenciar a percepção dos tons (secção 2.3). A percepção do tom 2 e tom 3 é tratada como um caso especial apresentado na secção 2.4. Na secção seguinte, apresenta-se uma descrição entoacional do português europeu, focando principalmente as frases declarativas e interrogativas. No capítulo 3,

é descrita a metodologia adotada no estudo. Os capítulos 4 e 5 são dedicados à descrição dos resultados. Uma discussão dos resultados obtidos conclui o estudo.

2. Enquadramento teórico

2.1. Os tons em Mandarim

Uma língua tonal é definida como “*a language in which pitch is used to contrast individual lexical items or words*”. (Gandour 1978: 41). As línguas em que o pitch é usado para marcar diferenças sintáticas ou semânticas em frases são categorizadas como línguas entoacionais.

O Mandarim é uma língua tonal. A língua tem 4 tons lexicais, que se diferenciam no contorno de F0. O primeiro tom é alto e contínuo, o segundo é considerado como um tom ascendente, o terceiro desce primeiro e sobe novamente, o quarto é descendente. A descrição mais usada é a de Chao (1968: 26), que descreveu os 4 tons com os valores de pitch numa escala de 1 a 5: T55 (tom1), T35(tom2), T214(tom3), T51(tom4).

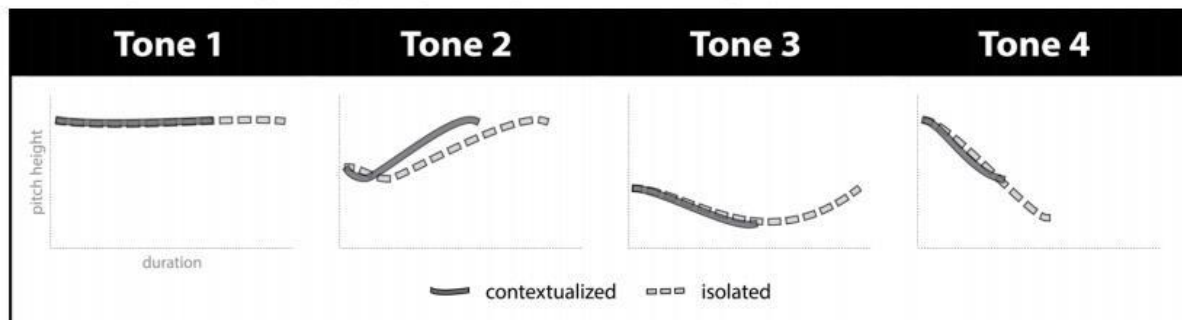


Figura 1. Descrição dos tons em Mandarim (retirado de Pelzl, 2019)

Entre os 4 tons, o tom 3 é o que tem mais variações contextuais. O T3 é [21] quando não está numa posição final e é [214] quando está em monossílabo (Duanmu 2007: 259; Chen 2000: 20-23) descreve que [214] passa a [35] quando está seguido por [214], processo designado por sândi de T3. Este processo é ilustrado nos exemplos 1-2, retirados de Chen (2000: 20-21).

1. xiao gou “*small dog, puppy*” (cachorro)

214. 214 (forma de base)

35. 214 (forma com efeito de sândi)

2. mai ma “*to buy a horse*” (comprar um cavalo)

214. 214 (forma de base)

35. 214 (forma com efeito de sândi)

Ainda segundo o autor, T35 muda para T55 quando está seguido por T55, T35 Chen (2000: 21).

3. tian wen tai “*observatory*” (observatório de astronomia)

55. 35. 35 (forma de base)

55. 55. 35 (forma com efeito de sândi)

4. ren minbi “*renminbi*” (moeda Chinesa)

35. 35. 51 (forma de base)

35. 55. 51 (forma com efeito de sândi)

No que diz respeito à duração dos tons, os tons 2 e 3 são os que apresentam uma duração maior comparativamente com o tom 4, que é o que apresenta uma duração menor (Jongman et al., 2006: 6). A duração poderá ser alterada num contexto diferente de produção. Yang et al. (2017) demonstraram que quando as sílabas são apresentadas isoladamente, o tom 3 tem a maior duração, enquanto o tom 4 tem a menor duração. O tom 1 e o tom 2 apresentam durações semelhantes. Os autores observaram que as durações dos tons se diferenciam em monossílabos isolados, discurso de leitura de texto e conversação. Na primeira condição, T3 tem uma duração mais longa, enquanto o T4 tem a duração mais curta. Em geral, os tons possuem uma duração maior quando são apresentados isoladamente, em contraste com o que acontece em conversação ou durante a leitura de um texto.

2.2. Percepção dos tons do Mandarim

A percepção dos tons em Mandarim tem sido uma área importante de investigação. Consideremos primeiro a percepção por falantes nativos.

Segundo Huang & Johnson (2010), T1-T2 e T2-T3 são percebidos como tons semelhantes e T1-T4 e T2-T4 como os menos semelhantes pelos falantes nativos do Inglês. Para estes autores, a discriminação dos tons nem sempre necessita de uma sílaba inteira. Em Lee, Tao, & Bond (2009), foram comparadas sílabas dotadas apenas de uma parte central, que resultaram da remoção dos primeiros seis e dos últimos oito períodos de pitch, sílabas silenciosas no centro que têm apenas os primeiros seis e os últimos oito períodos de pitch e sílabas dotadas apenas de parte inicial, que possuem só os primeiros seis períodos de pitch. Os falantes nativos de Mandarim são melhores a perceber sílabas que têm apenas a parte central, seguida pela parte central silenciosa, e são piores com sílabas que apenas possuem a parte inicial apresentada

Entre os quatro tons, a identificação do tom 2 depende mais da parte central da sílaba. O tom 1, por sua vez, depende mais da parte inicial e final da sílaba. Quando uma parte da sílaba é apresentada, os quatro tons não são tratados igualmente pelos falantes nativos de Mandarim. Os falantes nativos identificam melhor T2 do que os outros tons quando apenas a parte central é apresentada; identificam melhor T3 quando estão perante sílabas com parte central silenciosa e favorecem T4 como resposta se apenas a parte inicial for apresentada (Lee, Tao, & Bond, 2009).

Tem sido observado que os falantes nativos são melhores na discriminação dos tons da sua própria língua, mesmo perante itens não lexicais. Em Lee & Wurm (1996) os falantes nativos de Mandarim apresentaram menos erros na identificação de tons em comparação com os falantes de Cantonês e Inglês. Huang (2001) realizou uma pesquisa de identificação de tons em sequências dissílabas com falantes de Mandarim e falantes nativos de Inglês. Os nativos obtiveram melhores resultados. Os falantes de Inglês, por sua vez, encontraram dificuldades na identificação quando os tons tinham o *onset* e o *offset* de pitch com valores próximos. (e.g., tom 1 e tom 2), e quando os tons tinham contornos tonais semelhantes (e.g., tom 3 e tom 4).

Estudos sobre qual será a pista mais proeminente na percepção de tons mostraram que a inclinação do pitch (pitch slope) e o contorno de pitch (direção, altura) são pistas importantes na discriminação dos tons (e.g., Li, Shao, & Bao, 2017; Massaro, Cohen, & Tseng, 1985). Por outro lado, vários outros estudos focaram-se nas diferentes combinações de pistas quando o F0 não está

disponível. Liu & Samuel (2004) referem que a duração não exerce um efeito importante caso os itens tenham sido produzidos naturalmente. Todavia, assume um papel importante para os falantes nativos quando estes percebem os itens em Mandarim sussurrado, ou seja, quando o F0 não está a ser apresentado. Observa-se que uma curta duração é associada a T1 e T4, e uma longa duração é associada a T2 e T3. Os itens com T3 foram identificados como sendo T2 e os itens com T1 foram identificados como sendo T4, em Mandarim sussurrado. Fu & Zeng (2000) estudaram combinações diferentes de amplitude tonal (mudança entre 2 Hz e 50 Hz na amplitude tonal), duração e periodicidade (mudança entre 50 Hz e 500 Hz na amplitude geral). Os participantes tiveram pior performance quando apenas se apresentou a duração como pista. Adicionar a duração como uma pista não melhora a percepção se a amplitude ou periodicidade já estiverem disponíveis. A amplitude tonal, por sua vez, contribui mais na identificação de T3 e T4, enquanto que a periodicidade contribui significativamente para a discriminação dos 4 tons.

Quanto à percepção dos tons para os falantes não nativos, vários estudos demonstram que alguns tons são considerados mais difíceis que outros. Segundo Pelzl (2019: 51-78), o T2 é considerado como um tom difícil de ser identificado, nomeadamente quando é seguido pelo T3. Os falantes que se focam no *onset* do pitch, podem apresentar uma maior dificuldade na identificação do T1-T4, mas nem tanto na identificação do T1-T2. Segundo o autor, as identificações dos tons são influenciadas por 4 fatores: experiência linguística, experiência musical, aptidão na percepção do pitch e a L1 os falantes. Embora tenham mais experiência com os tons, os falantes de línguas tonais nem sempre têm vantagens sobre os falantes de línguas não tonais). Entre os falantes de línguas não tonais, o papel desempenhado pelo pitch na sua língua materna influencia a percepção dos tons lexicais. Falantes que tiveram a experiência de aprender uma segunda língua, mesmo que não seja uma língua tonal, poderão ter vantagens quando aprendem tons de uma nova língua. Por outro lado, os falantes com aptidão para identificar o tom mais alto conseguem aprender mais rapidamente e ter melhores resultados na aquisição dos tons que os outros.

É observado nos vários estudos que as dificuldades na percepção dos tons variam entre os falantes não nativos. Os falantes nativos do Inglês, Cantonês e Japonês tiveram dificuldades em identificar os pares T1-T4, T2-T3 e T1-T2, que partilham características fonéticas semelhantes, tais como o contorno de F0 ou a altura de pitch. Por outro lado, os tons que têm características fonéticas distintas são mais fáceis de perceber para os falantes de outra língua. Os falantes do Inglês, por exemplo,

confundiram o T1 com o T4 e o T4 com o T2 com mais frequência que os falantes de Japonês e os falantes de Cantonês. Mas tiveram melhor desempenho relativamente aos pares T1-T3, T2-T4 e T3-T4, os quais possuem características fonéticas distintas (e.g. os contornos de F0 e a duração entre T1 e T3 são diferentes; So & Best, 2010). Falantes de L1 diferentes são sensíveis a pistas diferentes na percepção.

Li, Shao, & Bao (2017) confirmaram que os falantes nativos de Mandarim e o do Cantonês são sensíveis à direção de F0 na distinção dos 4 tons. Segundo os autores, o *onset* de F0 é utilizado como outra pista importante para distinguir o T4, mas não é muito relevante na identificação de T1: os estímulos com a diferença de F0 menor que 12 HZ são principalmente identificados como T1, independentemente do *onset* de F0. Para os falantes de línguas tonais, ter mais experiência com tons nem é sempre uma vantagem. A queda do tom nem sempre está associada à identificação do tom descendente para os falantes nativos do Tailandês. Os falantes tailandeses identificaram o pitch mais alto, acima de 106 HZ, com uma queda menor que 24 HZ como T1 e consideraram apenas estímulos com *onset* mais alto que 112 HZ e uma queda de 30HZ como um tom descendente. Os Vietnamitas, por sua vez, são muito influenciados pelo *onset* de F0 e a diferença de F0 na identificação dos tons e mostram preferência pelo T4 comparativamente com os outros tons, uma vez que há mais tons descendentes em Vietnamita (Li, Shao, & Bao, 2017). Embora nem sempre tenham vantagens, os falantes de línguas tonais poderão beneficiar de ter mais experiência com tons, o que se traduz num melhor desempenho ao nível da identificação dos tons lexicais: os Cantonenses, por exemplo, têm melhores desempenhos na identificação dos tons lexicais em comparação com os falantes do Inglês (Lee, Vakoch, & Wurm, 1996: 533)

2.3. Factores que influenciam a percepção dos tons

Nesta secção focamo-nos em quatro fatores que influenciam a percepção dos tons, para além das características fonéticas já mencionadas acima. Primeiro, apresenta-se a influência da experiência de L1 (língua tonal ou língua entoacional). Em segundo lugar, apresenta-se a discussão sobre o processamento segmental e suprasegmental. Segue-se em terceiro lugar a apresentação do efeito da assimilação dos tons na percepção. Em quarto lugar, revisitamos as pistas fonéticas a que os falantes recorrem na percepção, em função da natureza tonal ou não tonal de L1. Finalmente, em

quinto lugar, mostra-se como a tarefa de percepção e a ordem de apresentação dos tons poderá influenciar os resultados.

2.3.1. A experiência de L1

A experiência de L1 é considerada como um dos fatores importantes ao nível da percepção. Vários estudos mostram que os falantes nativos são melhores na discriminação da sua língua materna, mesmo com palavras não lexicais, e os falantes de língua tonal são melhores que os falantes nativos de língua entoacional na discriminação dos tons (Lee, Vakoch, & Wurm, 1996).

Cao & Wang (2011: 404-407) notam que os falantes percebem os tons descendentes de formas diferentes: os falantes do Tailandês e os falantes de línguas entoacionais (falantes da França, Alemanha, Rússia e Coreia) percebem o “*low-falling*” tom como T3. Já os falantes de outras línguas entoacionais são mais influenciados, pela ordem dos tons do que os falantes do Tailandês.

O Tailandês é uma língua tonal. O Tailandês tem 5 tons: tom médio (*mid tone*), baixo (*low tone*), alto (*high tone*), descendente (*falling tone*) e ascendente (*rising tone*). Abramson (1978) classificou os tons em duas categorias: tom estático (*static tones*, como tom médio, tom baixo e tom alto) e tom dinâmico (*dynamic tones*, como tom descendente e ascendente). Para além de serem ambas línguas tonais, o Tailandês partilha mais características com o Mandarim: o T4 em Tailandês é semelhante ao T2 em Mandarim, sendo que ambos são tons ascendentes. T1, T3 e T5 em tailandês encontram também semelhanças fonéticas, tais como o valor de pitch e o contorno, com o T1, T4, T3 em Mandarim. Para ambas as línguas, os falantes têm dificuldades em discriminar tons quando estes partilham um contorno fonético semelhante. Os falantes nativos de Mandarim têm mais dificuldades em distinguir o T2 e T3, em comparação com os outros tons. Os falantes do Tailandês, por sua vez, confundem o tom médio e o tom baixo. O estudo feito por Li (2016) revela que os falantes nativos do Tailandês têm melhores desempenhos na discriminação dos 4 tons em Mandarim em comparação com os falantes nativos do Inglês, nomeadamente na identificação do T1 e T4, mas não com T2 e T3.

É reportado que os falantes de Mandarim apresentam melhores resultados e um tempo de reação menor que os falantes nativos de Japonês, Inglês e Coreano, línguas não tonais, na discriminação dos tons do Tailandês (Schaefer & Darcy, 2014). Wayland & Guion (2004) mostraram que os

falantes nativos de Mandarim tiveram melhor desempenho que os falantes de Inglês na discriminação dos tons do Tailandês antes e após treino auditivo. Os falantes nativos do Mandarim são melhores especialmente na identificação do tom baixo e médio. O estudo concluiu que a capacidade de seguir a direção e movimento de F0 poderá ser aplicada pelos falantes chineses na tarefa de discriminação dos tons do Tailandês. Do mesmo modo, os falantes de Tailandês são melhores que os falantes indonésios (falantes de língua não tonal) na discriminação dos tons em Mandarim (Chow, Liu, & Ning, 2018).

No entanto, alguns estudos defendem que ter uma língua tonal como L1 nem é sempre uma vantagem na identificação dos tons. Resultados reportados pelo Tsukada et al. (2013), por exemplo, revelam que os falantes nativos do Tailandês não têm melhor performance que os falantes do Japonês na discriminação dos tons em Mandarim. Segundo Wang (2013: 157), *“tone language speakers' familiarity with lexical tones in general may not always give them advantages over the non-tone language speakers in contrasting L2 tones because tone language speakers may also attach different weight to F0 contour or height in L2 tone perception based on their L1 tone experience”*. Na tarefa de discriminação dos tons em Mandarim, falantes de Japonês e de Inglês tiveram melhor desempenho que falantes de Hmong, uma língua tonal. Os falantes de Hmong tiveram dificuldades em combinar experiências de L1 com os tons em Mandarim. Igualmente, Tsukada (2019: 329-346) reportou a ausência de diferenças de desempenho significativas entre australianos, tailandeses e vietnamitas. Se ter uma língua tonal como L1 é uma vantagem na discriminação da outra língua tonal, seria de esperar que os falantes tailandeses e vietnamitas tivessem melhor performance. No entanto, os resultados mostram que o grupo de ingleses não se diferencia dos grupos de tailandeses e vietnamitas no que concerne à discriminação dos tons, com exceção da discriminação do T2-T4, em que os australianos tiveram pior performance que os tailandeses. O autor considerou que *“Tonal language speakers may be sensitive to and, at the same time, tolerant of pitch variations in their L1. This is useful for processing native tones for general communicative purposes, but unless the native and non-native tonal systems are identical or closely matched, it may pose a challenge to adapt to the unknown tonal system”*.

Outros estudos observaram que os falantes nativos do Mandarim tiveram pior desempenho relativamente ao processamento do pitch de itens não lexicais (*nonspeech*) (e.g., Bent, Bradlow, & Wright, 2006). Também na aquisição do Mandarim como segunda língua, os falantes da língua

tonal nem sempre têm melhores desempenhos que falantes da língua não tonal. Hao (2012), por exemplo, mostra que os falantes do Cantonês e Inglês não se diferenciam significativamente na identificação dos tons.

2.3.2. Processamento do tom e do segmento

A interação entre o processamento segmental (consonante, vogal) e suprasegmental (tom) têm sido alvo de vários estudos.

Sereno & Lee (2015) verificaram que os ouvintes são mais sensíveis à diferença segmental que suprasegmental na decisão lexical. Os autores pediram a falantes nativos de Mandarim para ouvir as sílabas lexicais e não lexicais em quatro condições - T (igual em tons), UR (diferente em tom e segmento), ST (igual em tom e segmento) e S (igual em segmento) numa tarefa de decisão lexical. Observaram que os falantes nativos cometeram significativamente mais erros quando foram apresentados os itens que se diferenciam em tom, em relação às outras combinações. Os participantes são mais rápidos na condição ST que UR. Existem também diferenças significativas em tempo de reação entre S que UR, T e UR. Em geral o tempo de reação é: $ST < S < UR < T$. Quando o *offset* de F0 da primeira sílaba se diferencia do *onset* de F0 da sílaba seguinte, o tempo da reação é mais curto: no par T3 (*offset* baixo) T1(*onset* alto) e no T4(*offset* baixo)T1(*onset* alto), mas mais lento no par T1(*onset* alto)T2 (*offset* alto). Segundo os autores, os participantes demoram mais tempo para responder quando existem diferenças apenas em tom, mas não segmento.

Gottfried & Suiter (1997) apresentaram as sílabas /di/, /da/, /duo/ e /du/ em Mandarim a falantes nativos do Inglês que têm algum conhecimento de Mandarim. Os itens foram apresentados em quatro condições: sílabas inteiras (*intact syllables*), sílaba com apenas a parte central (*center-only*), sílaba sem parte central (*silent-center*) e sílaba com apenas a parte inicial. Os resultados mostram que os falantes não nativos de Mandarim têm mais confusões em tons que em vogais em todas as quatro condições. A confusão é mais comum entre o Tom 2 e 3 quando as sílabas estão completas.

Um estudo feito por Cutler & Chen (1997) com o Cantonês revela também que a diferença em vogal e tom é mais provável de ser ignorada na tarefa de decisão lexical. Notaram que a probabilidade de cometer erros é mais alta quando o *onset* de F0 é semelhante entre duas sílabas apresentadas. Tanto os falantes nativos do Cantonês, como os falantes de Holandês que não têm

nenhum conhecimento de Mandarim, cometem mais erros e ao mesmo tempo precisam de mais tempo para responder quando as sílabas apenas se diferenciam em tom.

Yip, Leung, & Chen (1998) sugerem que embora os falantes nativos do Cantonês sejam atentos à informação tonal e segmental, são menos eficientes em utilizar a informação do tom que utilizar a informação segmental.

Por outro lado, ter ou não contexto pode influenciar o processamento do tom e segmento pelos falantes nativos de Mandarim. Ye & Connine (1999) observaram que para as sílabas individuais, a diferença em vogal é mais rapidamente detetada que a diferença em tom. No entanto, o contrário foi reportado quando os itens a ser processados estão num contexto idiomático.

A presença de [i], [u], [y], consoante fricativa e oclusiva afeta a percepção de T2 e T3. É reportado que entre os falantes nativos do Mandarim, o T2 é preferível ao T3 quando as consoantes são fricativas [x] em vez de africadas [ts]. As sílabas construídas pelas consoantes oclusivas [t^huan] assim como sílabas com consoantes africadas [tʂ^han] são mais provavelmente percebidas como T3 que T2, em comparação com consoantes fricativas [xuan]. Do mesmo modo, o T2 é a escolha preferível quando a sílaba é oclusiva não seguida de [u], [t^han], enquanto com fricativa [xan] é mais provável a percepção como T3. As consoantes aspiradas baixam o *onset* de F0, o que poderá resultar na preferência pelo T3 em relação a T2. Ao contrário, as sílabas construídas com consoantes fricativas aumentam o *onset* de F0 o que faz com que as escolhas com T2 sejam mais prováveis (Cao et al., 2016). O efeito da consoante aspirada no *onset* de F0 também foi confirmado em Xu & Xu (2003): as consoantes oclusivas estão associadas aos valores de F0 mais baixos e as não aspiradas aos valores mais altos. A presença de consoantes aspiradas tem um maior impacto no tom ascendente que no tom descendente.

Há estudos que consideram que a interação entre vogal e tom é “assimétrica”, ou seja, o segmento tem maior efeito que o tom na percepção. É observado que os falantes nativos do Mandarim e do Inglês cometeram mais erros e tiveram um tempo de reação mais longo na classificação tonal do que na classificação segmental (Repp & Lin, 1990). Quando lhes é pedido para prestar atenção a uma das dimensões (segmento, ou tom) e ignorar a interferência da outra, a variação segmental interfere mais na classificação dos tons que o contrário. Entre consoante, vogal e tom, a vogal exerce a maior interferência (Tong, Francis, & Gandour, 2008).

Por outro lado, há estudos que defendem uma interação “simétrica”: a interferência do tom faz a classificação dos segmentos ser mais demorada, e vice-versa. Lee & Nusbaum (1993), por exemplo, mostram que os falantes do Inglês são influenciados pela variação suprasegmental quando querem atender aos segmentos. Igualmente, Lin & Francis (2014) verificam que os falantes nativos do Mandarim são sensíveis às diferenças segmentais e tonais. Mas os falantes nativos do Inglês não. A sensibilidade para os contrastes tonais pode ser melhorada com a aprendizagem da língua. Hao (2018) observa que os falantes nativos do Inglês que têm conhecimento de Mandarim são mais sensíveis aos contrastes tonais. Em contrapartida, os falantes nativos do Inglês que não possuem nenhum conhecimento de Mandarim são mais sensíveis aos contrastes entre vogais que os contrastes tonais.

Embora os tons não tenham significados lexicais para os falantes da língua entoacional, os tons podem assumir função pragmática: a subida do tom no fim significa pergunta e a descida implica afirmação. Uma questão interessante é se o pitch é atendido de forma diferente quando assume ou não função linguística. Braun & Johnson (2011) comparam a performance de falantes holandeses e chineses na percepção de itens não lexicais na tarefa ABX. Concluíram que os falantes nativos do Holandês são mais atentos aos tons quando estes assumem função pragmática. Interessantemente, mesmo que os tons não tenham significado lexical, os holandeses demoram mais tempo com os pares incongruentes, ou seja, quando os estímulos X se diferenciam de A em segmento e B em tom, ou vice-versa. (e.g. rising mova – falling noba – falling mova). Em comparação com os falantes do Holandês, os falantes Chineses classificam mais facilmente os itens de acordo com os aspetos suprasegmentais e demoram ainda mais tempo que os holandeses perante pares incongruentes, sendo que tanto pitch como segmento são critérios importantes. A análise do tempo de reação mostra que os sujeitos estão mais atentos ao pitch quanto mais informações linguísticas o pitch assume (informação lexical > informação pragmática > não assume nenhuma informação linguística).

2.3.3. Assimilação dos tons

A assimilação dos tons é usada para explicar os resultados ao nível da percepção dos tons por falantes não nativos.

Segundo o modelo da assimilação perceptiva (*Perceptual Assimilation Model*), os segmentos não nativos são percebidos de acordo com as suas semelhanças e diferenças em relação a segmentos fonéticos nativos mais próximos (Best 1995: 193).

Segundo So & Best (2010), a assimilação de T1 (alto e contínuo) e T4 (tom descendente) de Mandarim ao T1 de Cantonês (alto e contínuo), assim como o T2 (tom ascendente) e T3 (sobe primeiro e desce) de Mandarim ao T2 de Cantonês (alto e ascendente), leva à confusão entre T1-T4 e T2-T3. Os falantes nativos de Japonês, por outro lado, maioritariamente optaram por T2 na tarefa de discriminação dos 4 tons, devido à assimilação do T2 de Mandarim ao acento tonal LH, encontrado nas palavras de 2 moras no Japonês (e.g., *ame*, essas palavras podem ser pronunciadas com LH ou HL)

Relativamente às línguas entoacionais, a assimilação dos tons é mais complicada. Hallé, Chang, & Best (2004) defendem que, se os tons percebidos fazem parte do sistema fonológico dos falantes, então serão percebidos como categorias fonémicas (*phonemic categories*). Pelo contrário, se não, os tons são percebidos como uma variação melódica não linguística (*nonlinguistic melodic variations*). A entoação, para os falantes do Inglês, assume funções pragmáticas. Por exemplo, o pitch descendente implica declarativas neutras e o pitch ascendente é associado às interrogativas. Quando percebem os tons de Mandarim, os ingleses categorizaram o T1, T3 e T4 como declarativa neutra, e o T2 como interrogativa. Para os franceses, o T3 foi considerado como declarativa neutra e o T4 como exclamação (So & Best, 2014). A associação dos tons lexicais à entoação frásica possivelmente leva à confusão tonal na discriminação dos tons de Mandarim. Por exemplo, os falantes coreanos, que associam o T4 à entoação das frases afirmativas, têm dificuldades em distinguir o T4 e o T1 (Tsukada & Han, 2019)

2.3.4. Pistas fonéticas: falantes de línguas tonais e falantes de línguas não tonais

Devido a experiências de L1 diferentes, falantes de línguas tonais e não tonais focalizam a sua atenção em pistas fonéticas diferentes na percepção dos tons.

Em relação aos falantes de línguas entoacionais, a altura de pitch e a média de F0 são referidas como pistas importantes na identificação dos tons (e.g., Guion & Pederson, 2007). Em Gandour (1983), os falantes de Inglês dependem mais da altura de F0 e menos da direção de F0 na

discriminação dos tons, em comparação com os falantes nativos do Mandarim, Taiwanês e Tailandês, as quais são consideradas línguas tonais. Resultados semelhantes foram observados em Qin & Mok (2011). Os autores observaram que na identificação do Cantonês, os falantes de Mandarim tiveram menos erros quando os pares apresentavam contornos de F0 distintos. Em contrapartida, os falantes nativos do francês tiveram melhor desempenho na discriminação dos pares que se diferenciam ao nível da altura do pitch, mas não no contorno. Isto sugere que os falantes nativos do Francês recorrem mais à altura, não recorrendo tanto ao contorno de F0, aquando da tarefa de discriminar tons.

Os falantes de línguas tonais recorrem ao F0 médio, à direção do contorno e à inclinação de F0 na identificação dos tons. Na identificação dos tons do Tailandês, Abramson (1978) considera que o F0 médio é a pista que leva mais informações na identificação do tom estático (tom médio, tom baixo e tom alto). Segundo Gandour (1983), os falantes nativos do tailandês dão mais importância à direção que os falantes nativos do Mandarim.

2.3.5. Influência da tarefa e outras condições metodológicas

Hung & Chung (2016) exploram um outro aspeto: a influência da tarefa e outros aspetos metodológicos na eficiência da percepção. O estudo pretende comparar várias condições, tais como tarefa (discriminação categórica, discriminação auditiva), natureza dos estímulos (itens lexicais, itens gerados pela onda sinusoidal) e a novidade (itens com contorno inventado, ou com contorno conhecido), que afetam a eficiência da percepção e deste modo encontrar a melhor combinação possível que resulte numa eficiência máxima. Os autores pediram a falantes nativos do Mandarim para discriminar os itens sob as seguintes condições: itens da forma de onda sinusoidal (sine-wave tone) (SW), itens com tom lexical (LW); itens com dois contornos tonais conhecidos (FF), itens com um contorno conhecido e o outro inventado (NF). Os sujeitos participaram em duas tarefas: a tarefa de discriminação auditiva em que os participantes responderam se os dois estímulos são idênticos (AUD) e a tarefa de discriminação categórica em que os participantes indicaram se os dois estímulos pertencem à mesma categoria (CAT). Concluíram que de entre as três condições, a natureza da tarefa (envolver ou não conhecimento categórico) é a condição que mais capacidade tem de maximizar a eficiência do processo. Segundo os autores, a novidade, por sua vez, também é um fator positivo sendo que ter um estímulo

inventado na tarefa de discriminação, aumenta a sensibilidade auditiva dos participantes. Por fim, a condição de ter ou não ter itens de fala não se revela um fator significativo aquando da percepção. Portanto, a combinação NF x SW x AUD é considerada como a melhor combinação possível, à qual se pode recorrer quando se pretende maximizar a eficiência.

Chen et al.(2019) observam que um intervalo de resposta longo na tarefa de categorização faz os ouvintes dependerem mais da semelhança fonológica dos tons não nativos. Mas a tarefa de categorização não é influenciada por variabilidade de falantes e de vogais. Ao contrário, a tarefa de discriminação é menos influenciada pelo intervalo de resposta longo ou curto, mas os resultados são afetados pela variabilidade de falantes e vogais.

2.3.6. Efeito da Ordem

O efeito da ordem, ou “*vowel perception asymmetry*” foi definido em Polka & Bohn (2003) como “*Asymmetries in vowel perception occur such that discrimination of a vowel change presented in one direction is easier compared to the same change presented in the reverse direction.*” Os autores observaram que os bebés alemães e ingleses entre 6-8 e 10-12 meses são melhores em discriminar vogais alemães /y/-/u/ quando ouvem a vogal /y/ primeiro. Do mesmo modo, entre Inglês /æ/e /ε/, a discriminação foi mais fácil se /ε/ vem primeiro. Repp, Healey, & Crowder (1979) pediram aos participantes para ouvir /i/, /I/ e /ε/ e indicar se os estímulos são iguais ou diferentes. Os resultados demonstram que os estímulos são considerados mais “diferentes” quando o primeiro estímulo é mais parecido com /i/.

Cowan & Morse (1986) sugeriram que ao ouvir duas vogais seguidas, a primeira vai para a posição neutra /ə/ quando é guardada na memória. Esta hipótese foi lustrada em Karypidis (2007), como representado na Figura 2:

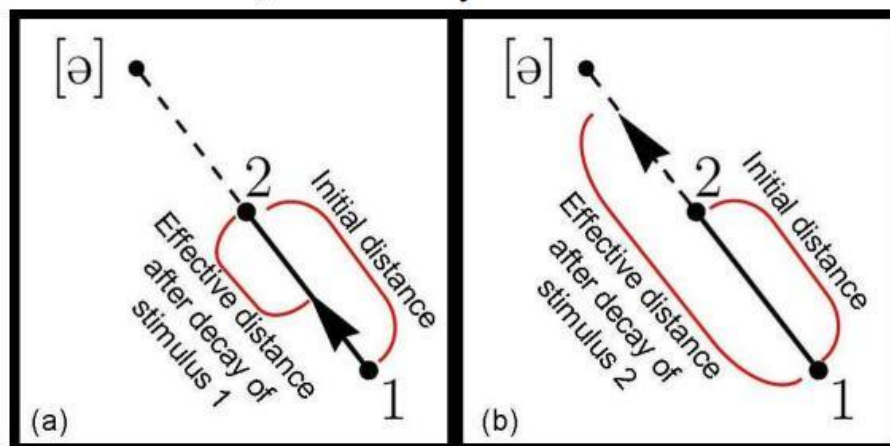


Figura 2. Efeito da ordem: a ordem em que o estímulo é apresentado é 1-2 em (a) e 2-1 em (b) (retirado de Karypidis, 2007)

Em (a), quando a ordem é vogal 1- vogal 2, a primeira vogal vai para a posição neutra /ə/ e ao mesmo tempo, torna-se mais semelhante à segunda vogal. Em (b), quando a segunda vogal é apresentada primeiro, também vai em direção a /ə/, daí resultando uma distância inicial mais afastada de 1. Por isso, segundo esta hipótese, a discriminação entre 2-1 vai ser mais fácil que 1-2, uma vez que a distância final é mais longa. Resultados obtidos em Macmillan, Braid, & Goldberg (1987) ou Polka & Werker (1994) estão alinhados com esta hipótese, mas não se encontra o mesmo em Karypidis (2007).

Para além das vogais, o efeito da ordem também foi reportado para a discriminação dos tons. Francis & Ciocca (2003) observam que os falantes nativos do Cantonês são mais sensíveis à diferença de F0 se ouvirem a sílaba em Cantonês com F0 mais baixo primeiro do que na ordem inversa. No entanto, não se observou o mesmo efeito entre falantes nativos do Inglês que não têm nenhum conhecimento de Cantonês. Os autores concluíram que o efeito de ordem é “*language specific*” (p.1615).

2.4. Identificação de Tom2 e Tom3: um caso especial

A confusão existente ao nível do Tom 2 e Tom 3 para os falantes não nativos e nativos é reportada em vários estudos. O par T2-T3 é considerado o mais semelhante em comparação com as outras

combinações e é o par que os falantes nativos demoram mais tempo para discriminar (Huang & Johnson, 2010; Tsukada & Kondo, 2019)

Entre os falantes de língua não tonal (e.g., falantes franceses, ingleses), observa-se a mesma dificuldade (e.g., So & Best, 2014). Também falantes com algum conhecimento de Mandarim demonstram dificuldades ao nível da discriminação de T2-T3. Por exemplo, no estudo de Tsukada & Han (2019) registou-se a pior performance na discriminação de T2-T3, tanto para os falantes nativos do coreano como para os falantes coreanos que já tinham conhecimento de Mandarim. A confusão entre T2 e T3 também foi reportada entre falantes nativos de língua tonal (e.g. Wang, 2013; Chow, Liu, & Ning, 2018). So & Best (2010) observam que falantes cantoneses confundem T4 com T1 e T2 com T3, mas tiveram melhor performance na discriminação de T1-T2 que os falantes nativos do Inglês e do Japonês. Para os falantes L2 do Mandarim, a distinção de T2-T3 também é considerada uma das dificuldades durante a aquisição da língua. Em Rungruang & Mu (2017), os autores fizeram uma entrevista com os aprendentes tailandeses do Mandarim, verificando que a confusão entre T2 e T3 é uma das dificuldades mais comuns. Em contrapartida, o tom 4 e o tom 1 foram considerados como os mais fáceis.

Em relação às pistas para a desambiguação entre T2-T3, a variação de F0, a duração, a altura do pitch e o ponto de alteração da curva de pitch são consideradas como pistas robustas. Segundo Moore & Jongman (1997), uma das características do tom 3 é ter um ponto de viragem a partir da descida anterior para a subida seguinte. Segundo os autores, um item com ponto de viragem por volta de 140ms e uma variação de F0 de 20 HZ foi identificado como tom 2, enquanto que essa identificação desce para 50% com uma maior variação de F0. Para Shen, Deutsch & Rayner (2013) um ponto de viragem baixo é uma pista robusta na identificação do tom 3. Sarmah & Cao (2007) fizeram uma análise mais detalhada: os autores manipularam a duração do ponto de viragem no tom 3 (0-390ms) e o tempo onde essa viragem acontece. Quando o ponto de viragem ocorre mais tarde, ou seja, entre 170ms e 290ms, os itens foram identificados como tom 3. Se o mesmo ocorre antes de 170ms, são identificados como tom 2. Os itens são identificados como tom 4, caso esse ponto ocorra mais tarde que 290ms. Em termos de percentagem, um item é classificado como tom 3, caso o ponto de viragem ocorra entre 42.5% e 72.5% da sua duração total. Se ocorrer antes de 42.5%, tende a ser considerado como tom 2 e se o mesmo aparecer após 72.5% da duração total, o tom 4 será a resposta.

Surpreendentemente, os falantes nativos nem precisam de ter uma sílaba inteira para distinguir T2-T3. Resultados obtidos por Liu & Samuel (2004) mostram que quando a subida de T3 é substituída por ruído branco, os falantes nativos ainda conseguem ter uma taxa de correção de 91%. Mesmo quando a parte descendente é substituída, regista-se uma taxa de correção de 80%. Para o T2, quando a parte ascendente é tirada, ainda 74% dos itens são identificados corretamente. Verificou-se ainda que a parte descendente é a mais importante para identificação de T3, enquanto a subida é mais importante para a identificação do Tom 2.

A altura do pitch na parte final e inicial também influencia a percepção de T2-T3. Os itens são considerados como tom 2 caso tenham pitch alto no fim. Em contrapartida, são identificados como tom 3 caso apresentem um pitch baixo no fim. A altura do pitch também é uma das pistas importantes para distinguir T3-T1, mas menos robusta para desambiguação de T4-T2 (Shen, Deutsch, & Rayner, 2013; Shen, Deutsch, & Le, 2011). Por outro lado, os falantes nativos de Mandarim associam um F0 baixo na parte inicial com T3 e o contrário com T2.

Blicher, Diehl, & Cohen (1988) defendem que a duração total é uma pista importante na desambiguação entre tom 2 e tom 3. Segundo os autores, as sílabas com uma duração mais longa são mais provavelmente identificadas como tom 3. Entre os falantes L1 e os falantes L2 de Mandarim, a duração exerce uma maior influência para os falantes L2 que para os L1. Quando os itens têm uma duração normalizada, tanto os nativos como os não nativos necessitam de muito mais tempo para responder (Chang, 2011). A duração é uma pista importante na identificação do T3, especialmente quando o F0 não está disponível. A duração não melhora os desempenhos se a amplitude ou F0 forem disponibilizadas como pistas. No entanto, na discriminação de T3, os falantes quase identificaram 50% dos itens com sucesso quando apenas a duração está disponível (Fu & Zeng, 2000). Por outro lado, Whalen & Xu (1992) consideram que o contorno de F0 é uma pista sólida na identificação do tom 2 e 3.

Para além das pistas de desambiguação, o efeito de sândi influencia o desempenho dos falantes nativos na identificação dos tons. Os falantes nativos tendem para a semelhança entre T2-T3 e demoram mais tempo na identificação de T3-T2 que os falantes ingleses (Huang, 2001). Do mesmo modo, em Huang & Johnson (2010) os tons relacionados entre si por sândi (T35-T214, T55-T35) apresentam tempos de reação de discriminação mais longos do que os outros pares. No

entanto, para os falantes nativos do Inglês, não se observa essa diferença significativa ao nível do tempo de reação.

Em Chen, Liu, & Kager (2015), observou-se que a identificação do par T2-T3 é influenciada pela estrutura dos pares a serem discriminados: os holandeses e os Chineses são melhores a discriminar T2-T3 de T3-T3 do que a discriminar T3-T2 de T3-T3. A razão, segundo o estudo, reside em *memory load*: quando se ouve as sílabas iguais primeiro, há um *memory load* mais baixo, que resulta em melhores desempenhos. Segundo os autores, em comparação com os holandeses, os Chineses tiveram uma maior variação na percentagem de correção entre os pares T2T3-T3T3 e T3T2-T3T3. Isso significa que o conhecimento fonológico de neutralização entre T2-T3 e T3-T3 aumenta a preferência por esse par. Para além disso, a ordem em que as sílabas de referência (T3-T3) ocorre também influencia os resultados. Observaram que os participantes tiveram melhores desempenhos quando ouviram primeiro T3-T3 e piores desempenhos quando ouviram o mesmo par de tons no fim na tarefa de identificação.

A laringalização de T3 é referida num conjunto de estudos. Em Cao, Wayland, & Kaan (2012), falantes L1 de Mandarim e falantes L2 ouviram os itens de T3 com presença de laringalização e sem laringalização. Verificaram que a presença de laringalização em T3 diminuiu a duração do ponto de viragem, resultando na redução da resposta T3, independentemente da proficiência na língua. Outros autores concluíram que a laringalização não é uma pista robusta, sendo que a alteração dos valores de F0, pelo contrário, é uma pista “necessária” na percepção do T3 (Garding et al., 1986).

Nos estudos descritos acima, é mostrado que a confusão entre T2 e T3 é observada tanto em falantes de línguas tonais, como em falantes de línguas entoacionais. Entre várias pistas de desambiguação, a variação de F0, duração, altura do pitch, o ponto de viragem da curva de pitch e a posição dos contrastes tonais são apontados como pistas robustas.

Embora a percepção dos tons do Mandarim já tenha sido amplamente estudada, pelo que sei, não há nenhum estudo feito sobre a percepção dos tons de Mandarim por falantes nativos do Português, uma língua entoacional. O estudo atual focaliza-se na percepção dos tons do Mandarim por falantes nativos do Português, analisa as suas dificuldades na identificação dos pares e os efeitos da posição dos tons na sua percepção.

2.5. Características Entoacionais do Português Europeu

Em Nespor & Vogel (2007: 11), entre outros autores, propõe-se que a estrutura prosódica é constituída por vários domínios. Por exemplo, as sílabas formam pés, os pés agrupam-se para formar palavras prosódicas (PW), que, por sua vez, formam o Sintagma Fonológico (PhP), que vem a formar o Sintagma Entoacional (IP). No Português Europeu (PE), um sintagma entoacional pode ser formado por um ou mais que um sintagma fonológico e os sintagmas fonológicos no domínio da mesma frase-raiz agrupam-se num único sintagma entoacional (Frota, 2000)

A realização da fricativa é uma das evidências do Sintagma Entoacional. Vejam-se os exemplos abaixo, retirados de Frota (2000: 14):

5. [a[z] aluna[z] obtiveram boa[z] avaliaçãoe[ʃ]I

6. [a [z] aluna[ʃ]I [até onde pensamos[ʃ] e sabemos[ʃ]]I [obtiveram boa[z] avaliaçãoe[ʃ]]I

Esta divisão é evidenciada pela realização [z] no final da palavra quando se segue uma vogal. (Mateus et al., 2003: 1070-1071). Em PE, os IPs são condicionados pelos seus tamanhos: os mais longos tendem a ser divididos, os constituintes balanceados ou o constituinte mais longo ocorre à direita na sequência. A proeminência de cada IP é à direita, ou seja, a cabeça prosódica situa-se mais à direita (Frota, 2000)

Para além disso, a construção do sintagma entoacional também é dependente de outros fatores como a velocidade da fala.

O principal elemento caracterizador do sintagma entoacional é a entoação. Nas línguas entoacionais, os eventos tonais constituem o contorno entoacional. A melodia dessa sequência tonal é constituída por dois grandes tipos de eventos tonais: os acentos tonais, que se associam ao elemento acentuado, e os tons de fronteira, que marcam as fronteiras dos constituintes prosódicos. Os eventos tonais podem ser constituídos pelo tom alto (H), ou tom baixo (L). Os acentos tonais são marcados por asterisco (*), enquanto os tons de fronteira são marcados por % (ou o sinal que indica o constituinte prosódico cujo limite é assinalado). Cada sintagma entoacional está associado

Por outro lado, Frota (2014) apresenta uma descrição entoacional mais detalhada de pedidos e ordens. Fonologicamente, os pedidos apresentam o contorno $H^* L^* L\%$. A ordem com o acento nuclear no verbo caracteriza-se por um contorno $H^* +L (L^*) L\%$ ou $L^* +H^*(L^*) L\%$; já se o acento nuclear for no objeto, temos o contorno $H^* +L L\%$.

Segundo Frota (2014), o que diferencia o contorno de pedido dos outros contornos é o L^* no acento nuclear e o alongamento da vogal.

Resumidamente, a Tabela 1 mostra a configuração dos contornos entoacionais nucleares no Português Europeu Standard. Importa salientar que os contornos nem sempre são os mesmos entre as variedades do Português Europeu (ver, por exemplo, Vigário & Frota, 2003; Cruz & Frota, 2011; Frota et al., 2015, para as outras variedades).

	Declarativa neutra	Interrogativa parcial	Interrogativa total	Pedido(multi-palavras)	Pedido(uma palavra)	Ordem(acento nuclear final)	Ordem(early focus)
Contorno entoacional	$H+L^*Li$	$H+L^* Li$	$H+L^*LHi$	$L^*L\%$	$\%H L^* L\%$	$H^* +L L\%$	$H^*+L(L^*) L\%$ $L^*+H(L^*)L\%$

Tabela 1. Contornos nucleares por tipo frásico descrito

2.5.4. Percepção de entoação dos tipos frásicos

No caso da identificação de frases interrogativas e declarativas em Português Europeu (PE), Falé & Faria (2007) afirmam que a subida de mais que 2 semitons no final da frase é percebida como uma frase interrogativa. Os dados mostram também que os estímulos com queda de F_0 no final da frase são identificados como declarativas, mas a subida de 1-2 semitons já faz com que a escolha entre a categoria de “interrogativa” e “declarativa” seja mais ambígua. Frota, Butler, & Vigário (2014) pediram aos falantes adultos de PE para avaliar os contornos de frases interrogativas ($H+L^*LH$) e frases declarativas ($H+L^*L\%$) numa escala de 1 (tem a certeza que a frase é declarativa) a 5 (tem a certeza que a frase é interrogativa). Os resultados mostram que a diferença entre $LH\%$ e $L\%$ no final é suficiente para a identificação de interrogativas e declarativas. Observaram também que as crianças de 5 meses já possuem a capacidade de distinguir entre contornos interrogativos ($H+L^*LH$) e declarativos ($H+L^*L\%$).

Quanto à distinção entre ordens e pedidos, Frota (2014) confirma que os falantes nativos distinguem os dois tipos de frases. Foi notado que os pedidos têm mais variações que as ordens, (e.g pedido gentil, pedido insistente e pedido normal). Um contorno com pico alto seguido por vogal prolongada pode ser um pedido insistente para alguns, mas pedido gentil para outros.

Cruz & Frota (2011) visa fornecer uma descrição da entoação dos tipos frásicos da região Alentejo (ALE) e Algarve (ALG). as autoras pediram aos ouvintes da variedade SEP para ouvir estímulos produzidos pelos falantes ALG e ALE. Os participantes podiam escolher entre três hipóteses de resposta: declarativa (A), interrogativa (B) e Não sei (C). Os resultados obtidos demonstram que as declarativas são mais facilmente identificadas do que as interrogativas. Embora um falante do estudo produza as interrogativas com o contorno declarativo (L*Li), mais de metade (66%) dos falantes do SEP conseguiram identificar esse tipo de frase com sucesso, indicando que a fronteira alta poderá não ser a pista única para os ouvintes do SEP identificarem o tipo frásico.

3. Metodologia

Neste capítulo será descrita a metodologia seguida e aplicada no presente estudo no sentido de testar as três hipóteses já previamente enunciadas, aqui repetidas para conveniência do leitor:

- **Hipótese 1** - O par T2-T3 é o par mais difícil para os falantes portugueses, dado que os dois tons possuem características semelhantes, quer ao nível do pitch inicial registado, quer ao nível da duração (Gandour 1978: 43). O par T2T4 é o par mais fácil para os falantes portugueses, uma vez que os dois tons têm características fonéticas distintas (e.g, contorno e altura de F0) e os falantes não nativos de língua tonal são mais sensíveis aos tons quando estes assumem função pragmática na sua língua (Braun & Johnson, 2011). Como o Português é uma língua entoacional e uma subida melódica assinala uma pergunta e uma descida uma afirmação, espera-se melhor desempenho na percepção do T2 (tom ascendente) versus o T4 (tom descendente).
- **Hipótese 2** - A posição do tom pode influenciar a identificação dos tons, uma vez que os tons têm níveis de F0 diferentes e a ordem em que essa diferença de F0 é apresentada (ouvir sílaba com F0 alto primeiro ou depois) influencia o desempenho dos falantes. No estudo feito pelo Francis & Ciocca (2003), é observado que os falantes do Cantonês, são mais

sensíveis à diferença de F0 quando a primeira sílaba tem F0 mais alto que a segunda. É esperado que a ordem em que os tons são apresentados influenciam o desempenho dos enunciadores.

- **Hipótese 3** - Os ouvintes são mais sensíveis à variação em segmentos que em tons, ou seja, os pares que se distinguem apenas em segmentos têm um grau de semelhança mais baixo que os se distinguem em tons, devido ao papel distintivo dos segmentos na língua nativa e ao papel não distintivo dos tons.

A secção 3.1 apresenta os participantes no estudo. Os materiais utilizados são descritos na secção 3.2. Na secção 3.3 apresentam-se os procedimentos experimentais e na secção 3.4 descreve-se o tratamento de dados e o modelo de análise estatística aplicado.

3.1. Participantes

Participaram no estudo 18 falantes nativos do português com idades compreendidas entre os 18 e 66 anos, 10 dos quais do sexo feminino. Cinco dos participantes têm conhecimentos linguísticos (são estudantes na Faculdade de Letras da Universidade de Lisboa). Nenhum dos participantes possui conhecimentos de Mandarim.

O consentimento informado foi obtido de todos os participantes antes da realização do estudo.

3.2. Materiais

Os materiais construídos para a tarefa de percepção têm as seguintes características.

Foram escolhidas sílabas que possuem a estrutura CV, em que C representa consoante e V vogal. Utilizaram-se as vogais ‘a’ e ‘i’ e as consoantes ‘b’, ‘p’, ‘d’, ‘y’, ‘sh’ e ‘m’. Estas consoantes e vogais foram escolhidas porque os sons a que estão associadas em Mandarim têm sons equivalentes em português. As sílabas têm 4 tons possíveis: T1, T2, T3, T4 (ver Tabela 2). A maioria dos estímulos constituem palavras no Mandarim.

Tom 1			Tom 2		
	ba1	bi1		ba2	bi2

	pa1	pi1		pa2	pi2
	da 1	di1		da2	di2
	ya1	yi 1		ya2	yi2
	sha1	shi1		sha2	shi2
	ma1	mi1		ma2	mi2
Tom 3			Tom 4		
	ba3	bi3		ba4	bi4
	pa3	pi3		pa4	p4,
	da3	di3		da4	di4
	ya3	yi3		ya4	yi4
	sha3	shi3		sha4	shi4
	ma3	mi3		ma4	mi4

Tabela 2. *Corpus*

As sílabas foram combinadas em pares de forma a obter-se as seguintes combinações: 1) tom diferente, segmento igual; 2) tom diferente, segmento diferente; 3) tom igual, segmento diferente; 4) tom igual e segmento igual. Apresenta-se um exemplo das combinações possíveis na Tabela 3.

ba1	
Tom D(2) Segmento I	ba2
Tom D(3) Segmento I	ba3
Tom D(4) Segmento I	ba4
Tom D(2) Segmento D	bi2
Tom D(3) Segmento D	bi3
Tom D(4) Segmento D	bi4
Tom I Segmento D	bi1
Tom I Segmento I	ba1

Tabela 3. Combinações com ba1. “I” indica igual, “D” indica diferente.

Os itens foram gravados pela própria investigadora, que é uma falante nativa de Mandarim e não possui conhecimento de nenhum dialeto do Chinês. As gravações foram feitas no Laboratório de Fonética e Fonologia da Faculdade de Letras da Universidade de Lisboa. Utilizou-se o microfone

LAPELA Earset micro DPA d:fine e o gravador MARANTZ PMD660. Cada item foi produzido isoladamente 2 vezes. As gravações foram registradas em ficheiros de áudio em formato .wav, com a frequência de amostragem de 44100z. A melhor produção em termos de naturalidade foi escolhida pela autora para este trabalho. No total, foram considerados 48 estímulos (12 sílabas x 4 tons) e construídos 768 pares de estímulos (48 x 8 x2 repetições).

3.3. Tarefa de percepção

Foi construída uma tarefa de percepção para avaliar o grau de semelhança/diferença percetiva dos contrastes tonais e segmentais, através da aplicação de uma escala de Likert com 5 valores (Bertram, 2007). Os participantes foram informados de que iriam ouvir duas sílabas seguidas em Mandarim, devendo indicar o nível de semelhança percepcionado aquando da audição das mesmas, recorrendo para tal a uma escala de 1 a 5 em que 1 significa que os dois itens não são semelhantes, 2 indica os itens têm uma semelhança baixa, 3 indica uma semelhança média, 4 indica uma semelhança alta, e 5 indica que os dois itens são idênticos.

Após a audição do par de sílabas, os participantes escolhem as teclas 1 a 5 no computador para responder. A seguir à resposta, passe-se automaticamente para o próximo item. Os itens são apresentados aleatoriamente aos participantes. Estes foram instruídos para responder da forma mais rápida e intuitiva possível.

Foram construídos 384 pares de sílabas: 144 para tom diferente, segmento igual; 144 para tom diferente, segmento diferente; 48 para tom igual, segmento diferente; 48 para tom igual e segmento igual. Todos os pares de itens foram apresentados duas vezes. Para este fim, a tarefa de percepção foi dividida em duas partes. A primeira parte corresponde à primeira apresentação dos itens, enquanto que a segunda parte envolve a segunda apresentação. As ordens de apresentação são aleatórias em cada parte. Existe um intervalo de 5 minutos entre as duas partes. O intervalo tem como função permitir aos participantes descansar. As duas repetições possibilitam avaliar a consistência das respostas dos participantes. Se as duas respostas para o mesmo item apresentarem uma diferença maior do que 2 valores na escala, essas respostas serão então excluídas da análise (por exemplo, se um item é avaliado como 1 da primeira vez e é avaliado como 5 da segunda vez, esse item é excluído da análise). Se as respostas apresentarem uma diferença inferior ou igual a 2,

serão incluídas na análise e é considerado o valor médio das duas respostas. No total, cada participante ouviu 768 pares.

A tarefa decorreu num local silencioso na biblioteca da Faculdade de Letras, tendo os participantes ouvido os estímulos por intermédio de auscultadores iphone. Utilizou-se o programa *Super Lab* (versão 5), para a montagem da experiência, bem como para o registo das respostas e do tempo de reacção dos participantes.

O estudo teve a aprovação da Comissão de Ética para a Investigação da Faculdade de Letras da Universidade de Lisboa (9_CEI2019).

3.4. Tratamento dos dados

As respostas e o tempo de reacção dos participantes foram automaticamente registados no *Super Lab* e exportados em formato .txt. Depois, os dados foram organizados em Excel. Obteve-se, para cada participante, a média das respostas para cada item e o respetivo tempo de reacção.

Foi utilizado o programa SPSS IBM para fazer a análise estatística. Na análise inicial para determinar a capacidade de discriminação dos tons, utilizou-se a ANOVA *repeated measures*, com os fatores intrasujeito Tom, Ordem e Segmento. O fator Tom tem os níveis T1-T2, T1-T3, T1-T4, T2-T3, T2-T4 e T3-T4 (6 níveis no total). O fator Ordem tem 2 níveis: ordem A e ordem B (por exemplo, para T1-T2, temos T1T2 como ordem A e T2T1 como ordem B, e assim sucessivamente para todos os pares de tons). O fator Segmento tem 2 níveis: segmento A (vogal ‘A’) e segmento I (vogal ‘I’). Uma segunda ANOVA *Repeated Measures* foi realizada com o objectivo de comparar os efeitos do tom e segmento. Para tal, foram utilizados os fatores intra-sujeito Tom (2 níveis, igual ou diferente) e Segmento (2 níveis, igual ou diferente).

4. Resultados

No presente capítulo, apresenta-se a análise das respostas e do tempo de reacção obtidos na tarefa de percepção.

Na secção 4.1, apresentam-se os resultados da análise das respostas segundo a primeira ANOVA *repeated measures*, incluindo a interação entre Tom e Ordem (secção 4.1.1), interação entre

Segmento e Tom (4.1.2) e Interação entre Tom, Ordem e Segmento (secção 4.1.3). Na secção 4.2, apresentam-se os resultados do tempo de reação. Por último, na secção 4.3, apresenta-se a análise dos efeitos de Segmento e Tom, de acordo com a segunda ANOVA *repeated measures*.

4.1. Análise das respostas: discriminação dos pares de tons diferentes

A ANOVA *repeated measures* mostra um efeito significativo do par de tons ($F(1.794, 30.499)=5.203, p=0.014, \eta^2=0.234$) e do segmento ($F(1, 17)=99.726, p=0.000, \eta^2=0.854$). Porém, o efeito da Ordem não é significativo ($F(1, 17)=0.322, p=0.578, \eta^2=0.019$).

Entre todas as combinações de itens diferentes, o par T3-T1 e o par T3-T4 com segmento I são os pares mais fáceis de discriminar para os participantes (Média=2.763, Média=2.893, respetivamente). Em contrapartida, o par T1-T4 e o par T4-T2 com segmento A são os pares mais difíceis de discriminar (com médias de resposta acima de 4.4). Os outros pares encontram-se dentro do intervalo de 3-4.3. Interessantemente, a maior parte dos pares que têm segmento I, independentemente da sua ordem e tom, têm médias mais baixas, logo são mais fáceis de discriminar, do que os pares que possuem o segmento A. (ver Tabela 4).

A análise de *pairwise comparison* dos tons mostra que quando se compara T1T2 com T1T3, o T1T3 é mais fácil para os participantes ($p=0.019$); entre T1T3 e T1T4, o T1T3 é mais fácil ($p=0.047$); do mesmo modo, o par T3T4 é menos parecido do que T2T4 ($p=0.026$).

	Média	Desvio Padrão
T1/T3_OrdemB_Segmento_I	2.763	0.904
T3/T4_OrdemA_Segmento_I	2.893	1.112
T3/T4_OrdemB_Segmento_I	3.041	1.029
T2/T3_OrdemB_Segmento_I	3.071	0.667
T1/T3_OrdemA_Segmento_I	3.189	0.755
T2/T4_OrdemA_Segmento_I	3.219	1.090
T2/T3_OrdemA_Segmento_I	3.456	0.539

T1/T4_OrdemB_Segmento_I	3.549	0.573
T1/T2_OrdemB_Segmento_I	3.607	0.745
T1/T4_OrdemA_Segmento_I	3.632	0.833
T3/T4_OrdemA_Segmento_A	3.652	1.322
T3/T4_OrdemB_Segmento_A	3.706	1.180
T2/T4_OrdemB_Segmento_I	3.844	0.559
T1/T2_OrdemA_Segmento_I	3.950	0.470
T1/T3_OrdemA_Segmento_A	4.005	1.015
T1/T3_OrdemB_Segmento_A	4.104	0.567
T1/T4_OrdemB_Segmento_A	4.181	0.742
T1/T2_OrdemB_Segmento_A	4.199	0.703
T2/T4_OrdemA_Segmento_A	4.224	0.935
T2/T3_OrdemA_Segmento_A	4.241	0.649
T2/T3_OrdemB_Segmento_A	4.280	0.587
T1/T2_OrdemA_Segmento_A	4.289	0.609
T1/T4_OrdemA_Segmento_A	4.413	0.538
T2/T4_OrdemB_Segmento_A	4.486	0.473

Tabela 4. Média e desvio padrão das diferentes sequências tonais

No que diz respeito às interações, há um efeito significativo da interação entre Tom e Ordem ($F(2.853, 48.509)=3.106, p=0.037, \eta^2=0.154$), entre Tom e Segmento ($F(5, 85)=4.547, p=0.001, \eta^2=0.211$), bem como entre Tom, Ordem e Segmento ($F(2.967, 50.447)=3.277, p=0.029, \eta^2=0.162$) No entanto, não se observa efeito significativo da interação entre Ordem e Segmento ($F(1, 17)=1.127, p=0.303, \eta^2=0.062$).

Perante os resultados obtidos e retomando as hipóteses 1 e 2 colocadas no início deste trabalho, considera-se que os resultados gerais não comprovam a hipótese 2 uma vez que não se observou um efeito significativo da Ordem (mas veja-se a secção 4.1.1). Em relação à hipótese 1, o par T2T3/T3T2 (independentemente do Segmento A/Segmento I) apresenta respostas entre 3.071 e 4.28 (Média=3.762), com valores mais baixos do que, T1T2/T2T1 (Média=4.01125), T1T4/T4T1

(Média=3.94375), T2T4/T4T2 (Média=3.94325). Assim, conclui-se que o par T2T3 não é o par mais difícil de perceber para os falantes nativos de PE, ao contrário da hipótese 1. Por outro lado, o par T2T4/T4T2 também não é o mais fácil de perceber. Vemos que T2T4/T4T2 (Média=3.94325) não é considerado mais fácil do que o T3T4/T4T3 (Média=3.323), T2T3/T3T2 (Média=3.762) ou T1T3/T3T1 (Média=3.515). Portanto, os resultados não confirmam a hipótese 1.

Verificou-se um efeito geral do segmento, pois os pares de tons com a vogal I são mais fáceis de discriminar do que os pares com a vogal A. Este efeito, não previsto inicialmente, será discutido no capítulo 5.

Explora-se, em seguida, as interações significativas encontradas, nomeadamente entre Tom e ordem (secção 4.1.1), entre Tom e Segmento (secção 4.1.2) e entre Tom, Ordem e Segmento (secção 4.1.3).

4.1.1. Interação entre Tom e ordem

A Figura 3 mostra a interação entre o tom e a ordem. De um modo geral, é mais fácil discriminar os pares com T1 se este ocorrer na segunda posição. O par T2T4 é o par que se diferencia mais entre duas ordens (a diferença atinge 0.443). Nota-se que o par T3T4 tem as médias mais baixas, independentemente da ordem, ou seja, T3T4 e T4T3 são considerados mais fáceis para os participantes independentemente da sua posição.

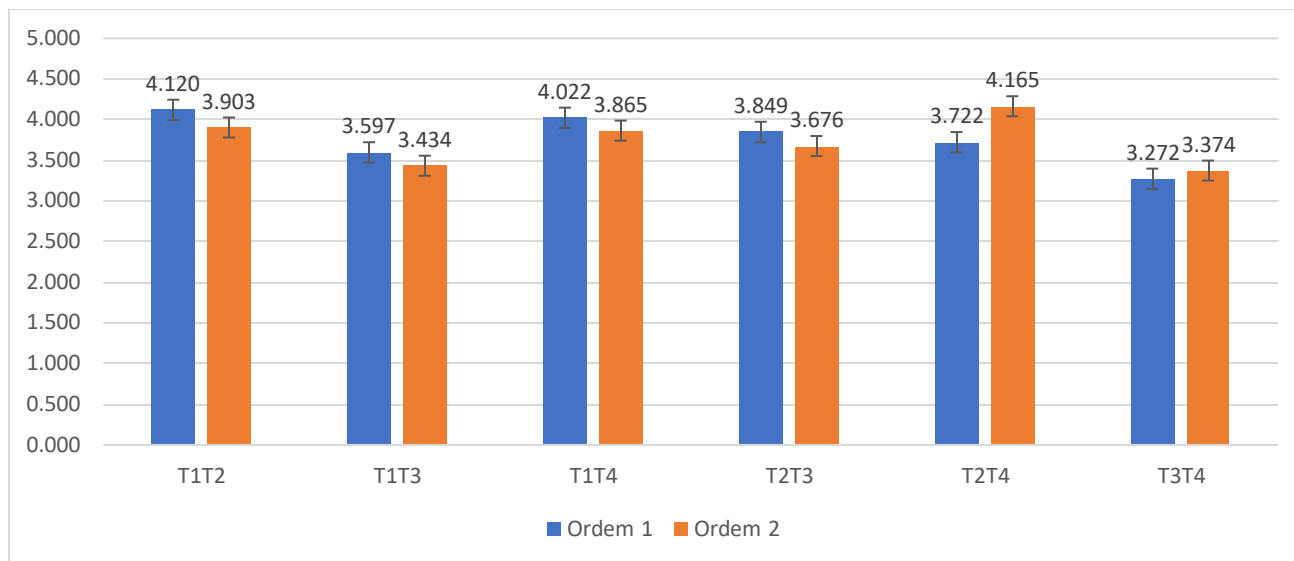


Figura 3. Interação entre tom e ordem.

Em relação aos pares que envolvem T3, constata-se que os itens têm respostas mais baixas quando o T3 surge primeiro. Isto sugere que o T3 possui características fonéticas diferentes em comparação com os outros tons, e por isso é possível que o processo de distinção seja facilitado ao ouvir primeiro um tom com características muito diferentes. Mais uma vez, observa-se que o par T2T3 não é considerado o mais difícil em comparação com os outros pares. No entanto, o T2T3 é o par mais confuso em comparação com os outros pares que envolvem T3 (T1T3 e T3T4). Esse resultado é esperado, uma vez que a confusão entre o tom 2 e o tom 3 tem sido reportada entre falantes de línguas tonais e não tonais (vê secção 2.4).

Os participantes consideram os pares mais difíceis de discriminar quando o valor do pitch do início da primeira sílaba coincide com o final da segunda sílaba: os pares T1T2, T2T1 e T4T2 são mais difíceis que T3T4, T4T3, T3T1 ou T1T3. O par T1T4, que tem o valor do *onset* de pitch próximo nas duas sílabas é também difícil para os participantes. Um resultado semelhante foi encontrado em So & Best (2010): os falantes nativos do Inglês confundem T1 com T2, T1 com T4 e T4 com T2 mais frequentemente do que os falantes do Japonês.

Estes resultados sugerem que quando os tons partilham características semelhantes, como é o caso da altura do pitch (*onset* e *offset*), estas semelhanças podem dificultar a sua identificação. Pelo

contrário, os pares de tons com características distintas, como é o caso de T3T4, ou T1T3, os quais se diferenciam em contorno e altura do pitch, são mais fáceis de identificar. Especificamente, o contorno de T3 (descendente e depois ascendente) é muito diferente em comparação com T4 (tom descendente) ou T1 (alto e contínuo). Acresce que T3 é o tom mais longo enquanto que T4 é o mais curto. (e.g Jongman et al., 2006).

Assim, os resultados do presente estudo confirmam que as características fonéticas podem influenciar a identificação dos tons. Os pares com características semelhantes entre si são mais difíceis de distinguir, pelo contrário os que apresentam mais características diferenciadoras entre si, são mais fáceis de distinguir. Por outro lado, a posição inicial de T3, pelas características diferenciadoras deste tom, facilita a identificação; no caso de T1, é a posição final que facilita a identificação. Isto pode dever-se à percepção da declarativa em Português. A descrição entoacional da declarativa é H+L*Li. A distinção prosódica entre a frase declarativa e, por exemplo, a frase interrogativa ocorre na fronteira. Portanto, é possível que os Portugueses sejam mais atentos ao tom alto quando ocorrem no final do que no início. Todavia, note-se que o intervalo de respostas se situa sempre acima de 3, mostrando que, na generalidade, a discriminação dos tons do Mandarim é difícil para os falantes de PE.

4.1.2. Interação entre Segmento e Tom

A Figura 4 mostra a interação entre o segmento e as diferentes combinações dos tons. Em todos os pares, as médias de resposta são mais altas quando o segmento é A. Regista-se a maior diferença entre segmentos no par T1T3 e a menor no par T1T2, ou seja, o segmento tem maior efeito no par T1T3 e menor no T1T2.

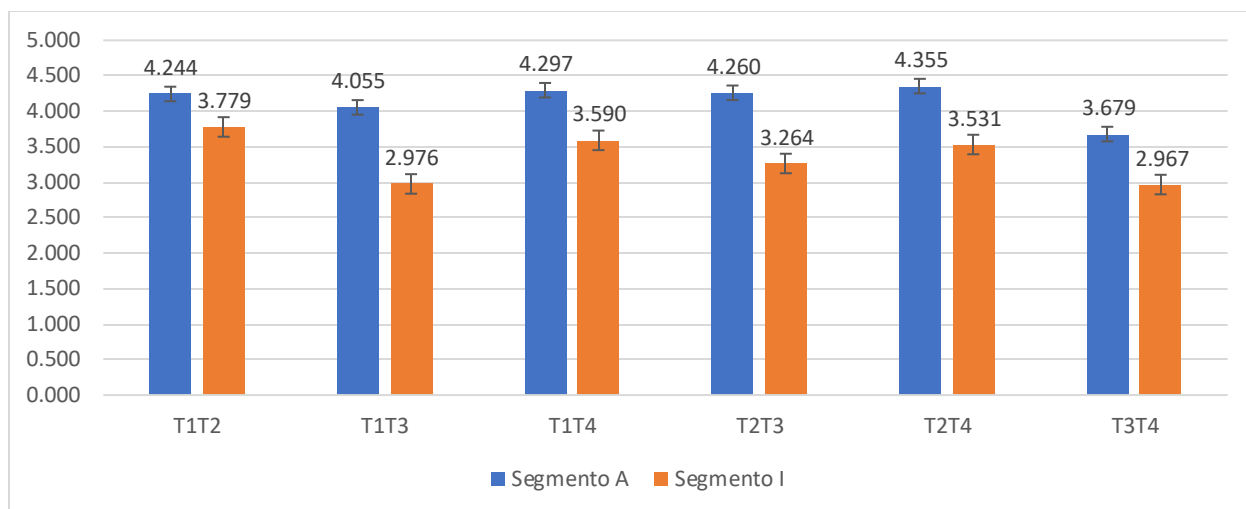


Figura 4. Interação entre Segmento e Tom-I.

Para os pares com segmento A, o par T2T4 é o par mais difícil de discriminar (Média=4.355), enquanto que o par T3T4 é o par mais fácil (Média=3.679). Por outro lado, para o segmento I o par T1T2 é o mais difícil para os participantes (Média=3.779) e o T3T4 o mais fácil (Média=2.967).

Em comparação com os outros pares, o par T3T4 apresenta valores mais baixos, tanto com o segmento A, como o segmento I. Interessantemente, na secção anterior vimos que o par T3T4 também apresenta valores mais baixos independentemente da ordem dos tons, em comparação com os outros pares. Portanto, o T3T4 é considerado como o par mais fácil para os participantes independentemente da ordem dos tons e do segmento. De destacar que os dois tons possuem características distintas, como o contorno fonético e duração, o que pode facilitar a sua identificação.

O par T1T3, por outro lado, apresenta valores bastante diferentes entre dois segmentos. Apesar de os tons apresentarem características fonéticas distintas (contorno de F0) a discriminação é dificultada quando o segmento é A.

4.1.3. Interação entre Tom, Ordem e Segmento

Por fim, apresenta-se os resultados relativos à interação entre tom, ordem e segmento. A Figura 5 mostra que, de entre todas as combinações, o par T4T2 com segmento A (Média=4.486) é o par mais difícil para os falantes nativos do Português, seguido pelo par T1T4, com segmento A

(Média=4.413). Em contrapartida, o par T3T1 com segmento I (Média=2.763) e o par T3T4 com segmento I (2.893) são os mais fáceis de discriminar. Assim, os pares de tons são diferentemente afetados pela ordem/segmento.

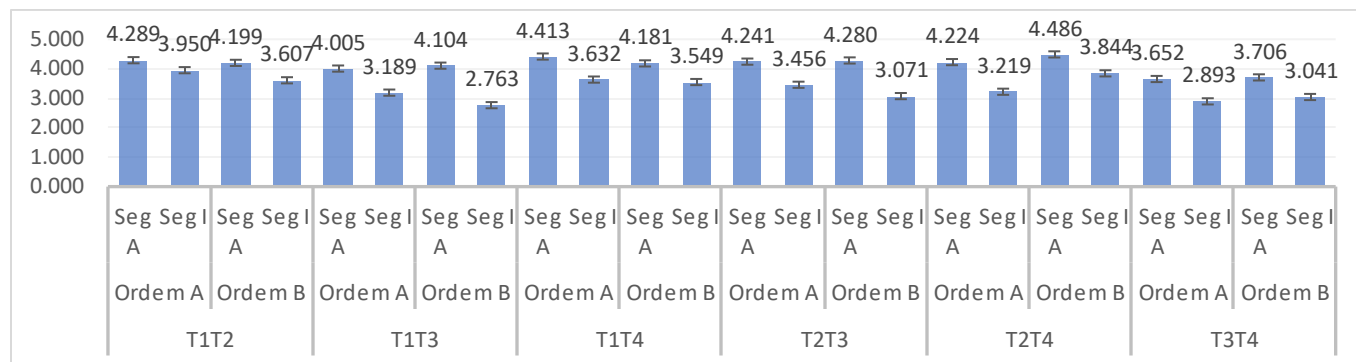


Figura 5. Interação entre Tom, Ordem e Segmento.

Observa-se que o par T3T4 tem médias quase sempre mais baixas, indiferentemente da ordem e do segmento. O par T1T2, por outro lado, tem médias relativamente altas, sendo pouco afetado pelos factores ordem e segmento.

A dificuldade com T1 é esperada, uma vez que foi verificada em falantes nativos de língua tonal e não tonal: por exemplo, a dificuldade em distinguir T1T2, T1T4, T2T4 foi verificada em Huang (2001) e So & Best (2010) entre os falantes nativos do Inglês; o estudo feito por Tsukada et al. (2013) mostra também que em geral os falantes do Tailandês e do Japonês têm dificuldade em identificar os tons quando o T1 está envolvido.

Para o par T1T3, a diferença em comparação com T3T1 atinge 1,34 quando o segmento muda de A para I, mas quando se compara T1T3 e T3T1 com segmento A, a diferença é muito pequena.

A diferença entre T1T4 (média=3.943, independentemente do segmento e ordem) e T1T3 (média=3.515, independentemente do segmento e ordem) é significativa ($p=0.047$). Esse resultado também é esperado. O T1T4 partilha mais características fonéticas semelhantes do que o T1T3. O *onset* de T1 e o *offset* de T4 é semelhante, o que pode causar confusão tornando os tons mais parecidos. Por outro lado, o T3 tem duração mais longa sendo que os dois tons apresentam contorno fonético diferente.

O T2T3, por sua vez, tem valores altos quando o segmento é A, mas menores que 3,5 quando o segmento é I. Isto não é esperado, uma vez que T2T3 tem sido identificado como o par mais difícil para falantes de línguas tonais e não tonais (ver a secção 2.4). Vemos que quando o segmento é I, o par T2T3 é considerado mais difícil em comparação com o T1T3, T3T1, T2T4, T3T4, T4T3, mas é mais fácil quando comparado com o T1T2, T2T1, T1T4, T4T1 e T4T2.

Para o par T2T4, a ordem T4T2 é mais difícil do que T2T4 para ambos os segmentos. O efeito de ordem também pode ser observado em T3T4, em que o T4T3 é considerado mais difícil quando comparado com T3T4.

Os resultados sugerem que os participantes têm melhor performance quando a altura do pitch do *offset* da primeira sílaba não coincide com o *onset* da segunda: *offset* do T4 é mais próximo do *onset* de T3 do que o contrário. Isso também pode ser verificado no caso de T1T4 (Média=4.0225, independentemente do segmento) e T4T1 (Média=3.865, independentemente do segmento), pois o *offset* de T1 é mais próximo do *onset* de T4, que dificulta a sua identificação. Igualmente, em comparação com o T2T4, o T4T2 é mais desafiante para os enunciadores.

4.2. Análise do tempo de reação: discriminação de pares de tons diferentes

Nesta secção apresentam-se os resultados do tempo de reação. A ANOVA *repeated measures* mostra que não existem diferenças significativas quanto aos fatores Tom ($F(3, 132, 53.246)=1.530$; $p=0.216$, $\eta^2=0.083$), Ordem, ($F(1, 17)=0.025$; $p=0.877$, $\eta^2=0.001$); Segmento ($F(1, 17)=1.9$; $p=0.186$; $\eta^2=0.101$). As interações Tom*Segmento ($F(5, 85)=1.827$; $p=0.116$, $\eta^2=0.097$), Ordem*Segmento ($F(1, 17)=1.662$; $p=0.215$; $\eta^2=0.089$), ou Tom*Ordem*Segmento ($F(2.850, 48.449)=1.378$; $p=0.261$; $\eta^2=0.075$) também não são significativas. O único resultado significativo é a interação Tom*Ordem ($F(5, 85)=4.332$, $p=0.001$, $\eta^2=0.203$).

A Tabela 5 apresenta os valores do tempo de reação para os diferentes pares de tons.

Estimates				
Measure:	MEASURE_1			
Tom	Média	Erro Padrão	95% Confidence Interval	
			Lower Bound	Upper Bound
T1-T2	701,626	49,044	598,153	805,099
T1-T3	710,275	49,331	606,196	814,354
T1-T4	760,464	46,386	662,598	858,329
T2-T3	707,570	49,929	602,229	812,911
T2-T4	728,735	42,030	640,059	817,411
T3-T4	744,286	50,246	638,278	850,295

Tabela 5. Tempo de reação para diferentes combinações de tons.

Vemos que todos os pares têm tempo de reação maior do que 700 ms. A diferença entre o par com tempo de reação mais longo (T1T4) e o com tempo de reação menor (T1T2) é de 59 ms.

Os valores para os pares com a ordem A e a ordem B são muito semelhantes (Tabela 6).

Estimates				
Measure:	MEASURE_1			
Ordem	Média	Erro Padrão	95% Confidence Interval	
			Lower Bound	Upper Bound
A	724,213	47,579	623,830	824,597
B	726,772	43,237	635,551	817,994

Tabela 6. Tempo de reação para Ordem A e Ordem B.

Os tempos de reação para os pares com segmentos diferentes estão também muito próximos (Tabela 7).

Estimates

Measure:	MEASURE_1			
Segmento	Média	Erro Padrão	95% Confidence Interval	
			Lower Bound	Upper Bound
A	712,189	47,964	610,993	813,384
I	738,796	43,440	647,146	830,447

Tabela 7. Tempo de reação para Segmento A e Segmento I.

Conclui-se assim que o tempo de reação não revela diferenças de processamento entre os diferentes pares de tons, ou em função da ordem, ou da natureza da vogal ‘I’ ou ‘A’. No entanto, o resultado significativo da interação Tom*Ordem aponta para comportamentos diversos entre os tons em função da ordem.

4.2.1. Interação entre Tom e Ordem

A Figura 6 mostra que os pares T1T2, T1T3 e T2T3 apresentam maior tempo de reação na Ordem 2, enquanto que os pares T2T4 e T3T4 apresentam maior tempo de reação na Ordem 1. Nas secções anteriores vimos que T1T2, T1T4, T4T2 são pares mais difíceis de discriminar para os participantes (veja-se a secção 4.1.1). Todavia, os pares T1T2 e T4T2 têm tempo de reação relativamente baixo em comparação com os outros pares, mas o T1T4 tem tempo de reação mais longo.

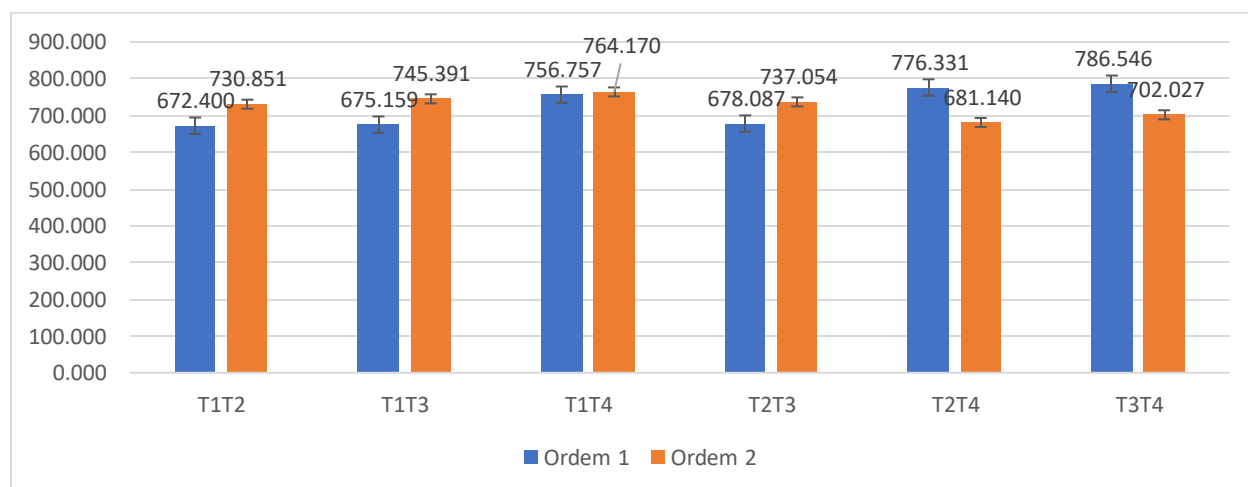


Figura 6. Interação entre Tom e Ordem.

Os pares T2T4 e T4T2 apresentam uma maior diferença relativamente à resposta de discriminação e ao tempo de reação. O par T2T4 tem uma média de resposta mais baixa (é mais fácil distinguir entre os tons), mas tempo de reação mais longo, ao contrário do que se verifica no par T4T2.

Conclui-se assim que não parece existir uma relação consistente entre o sentido das respostas e o tempo de reação.

4.3. Efeitos de Segmento e Tom

4.3.1. Segmento e Tom: análise das respostas

Na presente secção testamos a hipótese 3, procurando determinar os efeitos de Tom e de Segmento:

Hipótese 3 - Os ouvintes são mais sensíveis à variação em segmentos que em tons, ou seja, os pares que se distinguem apenas em segmentos têm um grau de semelhança mais baixo que os se distinguem em tons, devido ao papel distintivo dos segmentos na língua e ao papel não distintivo dos tons.

A segunda ANOVA *repeated measures* mostra um efeito significativo de Segmento ($F(1, 17)=109.072$, $p=0.000$, $\eta^2=0.865$) e de Tom ($F(1, 17)= 30.884$, $p=0.000$, $\eta^2=0.645$), sem interação entre os dois fatores ($F(1,17)=0.133$, $p=0.125$, $\eta^2=0.133$). Estes resultados indicam que os participantes são sensíveis tanto à diferença de segmento, como à diferença de tom. A média de resposta é mais alta quando o segmento é igual (4.242) do que quando o segmento é diferente (1.985). Da mesma forma, os pares com tons diferentes têm média de resposta mais baixa (Média=2.741) do que os pares com tons iguais (Média=3.485). Todavia, a diferença das respostas entre tom igual e tom diferente é de 0.744, enquanto a diferença para segmento igual e segmento diferente é de 2.257.

Pairwise comparisons para os SegI-TomI, SegI-TomD, SegD-TomI e SegD-TomD mostram que existem diferenças significativas entre todas as combinações, com exceção de SegD-TomI e SegD-TomD ($p=0.22$). Assim, quando o segmento é diferente, a diferença entre tons já não faz com que os pares sejam percebidos como mais “diferentes”. Quando o segmento é igual, existe diferença significativa entre SegI-TomI e SegI-TomD ($p=0.002$). Em outras palavras, os falantes nativos do Português são sensíveis às diferenças de tons e não são “*tone deaf*” em linha com outros

estudos feitos para falantes de línguas entoacionais (e.g., Hao, 2018; Hu, Chen, & Kager, 2018; Hallé, Chang, & Best, 2004).

A Tabela 8 apresenta as médias de resposta para de Segmento Igual-Tom Igual (SegI-TomI), Segmento Igual-Tom Diferente (SegI-TomD), Segmento Diferente-Tom Igual (Seg D-TomI) e Segmento diferente-Tom diferente (Seg D-TomD):

	Média	Desvio Padrão
SegI-TomI	4.71997	0.341970
SegI-TomD	3.76442	0.475717
SegD-TomI	2.25077	1.178526
SegD-TomD	1.71838	0.715793

Tabela 8. Média e desvio padrão das diferentes sequências tonais: Segmento Igual/Diferente, Tom Igual/Diferente.

A Tabela 8 mostra que os pares que têm segmento e tom iguais são considerados idênticos, seguidos pela combinação Segmento Igual-Tom Diferente, Segmento Diferente-Tom Igual e Segmento Diferente-Tom Diferente. De notar que os pares têm médias de resposta superiores quando o tom é diferente e o segmento igual, do que quando o segmento é diferente e o tom igual.

Em suma, para os falantes nativos do Português, é mais fácil distinguir entre segmentos do que entre tons. Este resultado alinha-se com a hipótese 3, a qual assume que os pares que se distinguem em segmentos têm um grau de semelhança mais baixo do que os que se distinguem em tons.

Face à análise feita, pode concluir-se que os participantes são melhores em discriminar segmentos que tons e são mais sensíveis às diferenças entre tons quando os segmentos são iguais do que diferentes.

4.3.2. Segmento e Tom: análise do Tempo de Reação

No que diz respeito ao tempo de reação, os resultados da ANOVA mostram que não há um efeito do Segmento $F(1,17)=2.959$, $p=0.104$, $\eta^2=0.148$, mas há um efeito do Tom $F(1, 17)=10.690$ $p=0.005$, $\eta^2=0.386$, bem como uma interação entre Tom e Segmento ($F(1, 17)=30.670$, $p=0.000036$, $\eta^2=0.643$). Estes resultados sugerem que o processamento das condições de

segmento igual (média=671.67) e segmento diferente (média=741.61) é equivalente para os participantes, o mesmo não acontecendo com as condições tom igual (média=682.537) e tom diferente(média=730.742). Os participantes são mais lentos quando tom é diferente, mostrando uma dificuldade em processar a diferença entre os tons.

A Tabela 9 apresenta as médias de tempo de reação para Segmento Igual-Tom Igual (SegI-TomI), Segmento Igual-Tom Diferente (SegI-TomD), Segmento Diferente-Tom Igual (Seg D-TomI) e Segmento diferente-Tom diferente (Seg D-TomD):

	Média	Desvio Padrão
SegI-TomI	617.00222	169.603759
SegI-TomD	726.32967	189.110595
SegD-TomI	748.07193	241.072389
SegD-TomD	735.15424	238.314867

Tabela 9. Tempo de Reação médio e desvio padrão de Segmento Igual-Tom Igual (Seg I-Tom I), Segmento Igual -Tom D (SegI-Tom D), Segmento Diferente-Tom Igual (Seg D-Tom I) e Segmento Diferente-Tom Diferente (Seg D-Tom D).

Vê-se que o tempo de reação é mais baixo quando os pares têm segmento igual e o tom igual. O tempo de reação não varia muito entre as outras três condições. Para além disso, os participantes responderam mais rapidamente quando apresentado o Segmento I-Tom I em comparação com o tempo de resposta ao ser apresentado o Segmento D-Tom D. Todavia, as *pairwise comparisons* não revelaram diferenças significativas entre as condições.

A interação entre Tom e Segmento deve-se a comportamentos diferentes entre as condições com segmento diferente e segmento igual em relação ao tom. A Figura 7 mostra o tempo de reação médio das 4 condições: Tom Igual-Segmento Igual, Tom Diferente-Segmento Diferente, Tom Igual-Segmento Diferente e Tom diferente-Segmento Igual. Quando o segmento é diferente, não há diferenças entre tom igual ou tom diferente; quando o segmento é igual, os participantes são mais rápidos quando o tom é igual do que quando o tom é diferente.

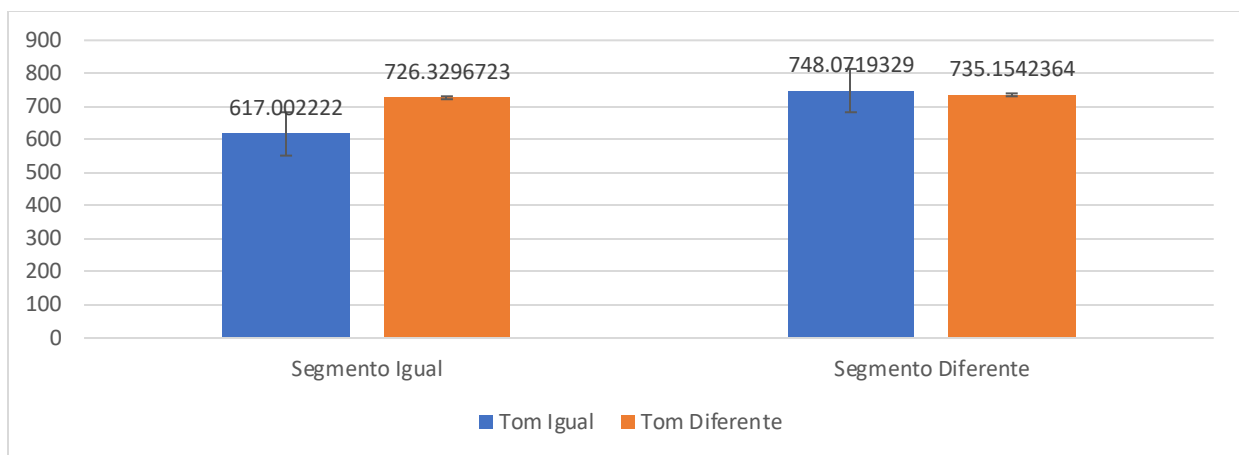


Figura 7. Tempo de reação de Segmento Igual-Tom Igual, Segmento Diferente-Tom Igual, Segmento Igual-Tom diferente e Segmento Diferente-Tom Diferente.

5. Discussão

O presente trabalho investiga a percepção de tons do Chinês mandarim por falantes nativos do Português e pretende determinar se a natureza dos contrastes tonais (tipo de tom), a posição do tom (ordem) e a natureza do segmento (vogal) condicionam a percepção.

Partimos das seguintes hipóteses:

- **Hipótese 1** - O par T2-T3 é o par mais difícil para os falantes portugueses, dado que os dois tons possuem características semelhantes, quer ao nível do pitch inicial registado, quer ao nível da duração. (Gandour, 1978: 43) O par T2T4 é o par mais fácil para os falantes portugueses, uma vez que os dois tons têm características fonéticas distintas (e.g, contorno e altura de F0) e os falantes não nativos de língua tonal são mais sensíveis aos tons quando estes assumem função pragmática na sua língua (Braun & Johnson, 2011). Como o Português é uma língua entoacional e uma subida melódica assinala uma pergunta e uma descida uma afirmação, espera-se melhor desempenho na percepção do T2 (tom ascendente) versus o T4 (tom descendente).
- **Hipótese 2** - A posição do tom pode influenciar a identificação dos tons, uma vez que os tons têm níveis de F0 diferentes e a ordem em que essa diferença de F0 é apresentada (ouvir sílaba com F0 alto primeiro ou depois) influencia o desempenho dos falantes. No estudo feito pelo Francis & Ciocca (2003), é observado que os falantes do Cantonês, são mais

sensíveis à diferença de F0 quando a primeira sílaba tem F0 mais alto que a segunda. É esperado que a ordem em que os tons são apresentados influenciam o desempenho dos enunciadores.

- **Hipótese 3** - Os ouvintes são mais sensíveis à variação em segmentos que em tons, ou seja, os pares que se distinguem apenas em segmentos têm um grau de semelhança mais baixo que os se distinguem em tons, devido ao papel distintivo dos segmentos na língua nativa e ao papel não distintivo dos tons.

A hipótese 1 não foi confirmada. Os resultados revelam que os falantes nativos do Português consideram que os pares que partilham características fonéticas parecidas, como é o caso dos pares T1T2, T2T1 e T4T2, são mais difíceis de discriminar do que os restantes pares de tons, como T3T4, T4T3, T3T1 e T1T3. Os pares T1T2, T1T4, T4T2 com segmento A são os que apresentam médias de resposta mais altas, logo são percebidos como semelhantes. A dificuldade com os pares T1T4 e T1T2 também foi reportada em So & Best (2010) e Huang (2001) entre os falantes nativos de Inglês. No entanto, ao contrário de estudos anteriores (e.g Rungruang & Mu, 2017; Tsukada & Han, 2019; Wang, 2013; Chow, Liu, & Ning, 2018), o par T2T3 não é considerado como o par mais difícil entre os falantes de línguas tonais e não tonais. Mais na linha do presente estudo, em Hao (2018) também não se observa uma diferença significativa ao nível da taxa de casos corretos entre T1T4 e T2T3 entre os falantes nativos do Inglês. Já a dificuldade com T2T4 também foi reportada, por exemplo, em Tsukada (2019) entre falantes australianos de Inglês. O autor observa que os falantes australianos não se diferenciam dos falantes tailandeses e vietnamitas quanto à discriminação dos tons do Mandarim, com exceção de T2T4. No nosso estudo, a dificuldade com T2T4/T4T2 sugere que uma eventual associação a um contraste entoacional com função pragmática na língua (na linha do sugerido em Braun & Johnson, 2011), não constituiu um factor facilitador da percepção.

A dificuldade com T1T4 e T4T2 indica que os falantes nativos do Português provavelmente se focam no *onset* do pitch ao processar os contrastes tonais. Segundo Pelzl (2019: 51-78), T1T4 é mais difícil do que T1T2 se os participantes se focarem no início do pitch. A outra observação relevante é a de que quando o valor do pitch do início da primeira sílaba coincide com o final da segunda sílaba, como no caso de T1T2, T2T1, T4T2 e T1T4 esses pares revelam ser mais difíceis

de discriminar. Isso também foi observado entre os falantes nativos de Inglês. Por exemplo, Guion & Pederson (2007) concluíram que os falantes nativos de Inglês recorrem mais ao F0 médio e à altura de F0 na percepção dos tons, mas não ao contorno de F0.

A hipótese 2 também não foi confirmada. O fator Ordem não é significativo em nenhuma das análises realizadas. As interações verificadas entre ordem e tom explicam-se pelo efeito das características fonéticas dos tons na sua percepção, já referido acima: quando os tons partilham características semelhantes, como é o caso da altura do pitch (*onset* e *offset*), estas semelhanças podem dificultar a sua identificação.

A hipótese 3 foi confirmada. Os pares que se distinguem em segmentos foram mais facilmente discriminados do que os pares que se distinguem em tons. Este resultado esperado foi também verificado em Gottfried & Suiter (1997), que observam que os falantes de línguas não tonais têm mais dificuldades com diferenças tonais do que com diferenças entre vogais. Hao (2018) sugere que os falantes nativos do Inglês sem conhecimento de Mandarim têm melhor desempenho ao discriminar contrastes não tonais do que ao discriminar contrastes tonais e são menos sensíveis à variação de pitch do que a diferenças entre vogais. No entanto, isso não implica uma grande dificuldade na aquisição dos tons. Depois de ganhar mais experiência com a língua tonal, os falantes de línguas não tonais conseguem discriminar os tons mais rapidamente do que contrastes entre vogais. Por exemplo, foi observado que os falantes com conhecimento avançado de Mandarim são mais rápidos na identificação do contraste tonal T1T4 do que do contraste não tonal do Mandarim /lu-ly/ (Hao, 2018)

Note-se que o efeito do tom também é significativo, para além do segmento. Os pares com segmentos diferentes e tons diferentes, foram percebidos como sendo mais diferentes que os pares com segmentos diferentes e tons iguais. Do mesmo modo, os pares com segmentos iguais e tons diferentes foram percebidos como mais diferentes do que os pares com segmentos iguais e tons iguais. Assim, apesar de mais sensíveis aos segmentos, os falantes nativos de PE foram todavia também sensíveis ao tom.

No entanto, foi inesperado o facto de os pares com tons diferentes serem considerados mais fáceis de discriminar quando o segmento é I em vez de A. Estudos prévios revelam que algumas vogais podem ser mais difíceis para falantes L2. Polka (1995) verifica que os falantes nativos do Inglês conseguem discriminar as vogais alemãs /u/ vs /y/ muito bem, mas tal já não se verifica com /ʊ/ e

/ɲ/. Discriminar contrastes não tonais pode ser mais desafiante do que contrastes tonais. Burnham (2000) apresentou um estudo relativamente à percepção dos contrastes tonais e consoantes do Tailandês por falantes australianos que não possuem nenhum conhecimento de Tailandês. Os participantes tiveram piores desempenhos com contrastes não tonais [ba]-[pa] do que com contrastes tonais. Levy & Strange (2008) investigou a percepção das vogais do Francês por falantes do Inglês. Segundo os autores, o grupo que tem algum conhecimento do Francês, tem piores desempenhos em discriminar vogais: /i-y/ (16% erro), /u-œ/ (27% erro), e /y-œ/ (29% erro) em comparação com o outro grupo que consiste em falantes com muitos anos de experiência (3%, 3% e 8% erro, respetivamente). Porém, mesmo com muito mais experiência, a taxa de erro com o par /u-y/ continua a ser muito alta para ambos os grupos (30% vs. 24%).

Todavia, as vogais I e A existem no Português. Uma possível explicação para a maior dificuldade de discriminação dos tons com a vogal A é o facto de a produção de “I” ser mais estável com os vários tons, enquanto o “A” apresenta variação considerável na sua articulação. Erickson et al. (2004) demonstrou que a língua está mais recuada quando produzem “A” com Tom 3. Segundo Hoole & Hu (2004), para “A”, a posição da parte da frente da língua diferencia-se significativamente entre Tom 1, 3 e 4. Para “I”, encontra-se apenas uma diferença significativa em forma da língua entre Tom 3 e 1. Em acréscimo, foi reportado que o segmento pode influenciar a identificação dos tons. Shaw et al. (2013) mostram diferenças na percepção do Tom 2 e Tom 3 quando é produzido com /a/ em relação a outras vogais e na percepção do Tom 4 quando é produzido com /u/ em relação a outras vogais.

Em relação ao tempo de reação, não se registou efeito significativo do Segmento, mas sim do Tom. Ao contrário do que acontece com os segmentos, os participantes são mais rápidos quando o tom é igual e mais lentos quando o tom é diferente, mostrando uma dificuldade em processar a diferença entre os tons. Isso implica que os participantes não sejam “*tone deaf*” sendo sensíveis às variações tonais. Em Hao (2018), não se regista diferença significativa ao nível do tempo de reação entre falantes sem conhecimento de Mandarim quer na percepção de contrastes tonais quer na percepção de contrastes não tonais. Mas foi observado que os participantes com mais experiência em Mandarim têm tempo de reação mais curto em comparação com aqueles que não têm conhecimento de Mandarim ou têm pouco conhecimento. O efeito do tom no tempo de reação

vem confirmar que os falantes são sensíveis à variação do tom, mas têm dificuldades na identificação dos contrastes tonais.

6. Conclusão

O presente estudo focou-se na percepção dos tons em Mandarim por falantes nativos do Português Europeu. Tem como objetivo estudar a percepção de tons por falantes de uma língua não tonal, determinar os fatores que condicionam a percepção dos tons, nomeadamente o tipo de contraste tonal, a ordem de apresentação dos tons e o segmento em que o tom ocorre, bem como explorar a relação entre os contrastes tonais e os contrastes entre segmentos. Partimos de 3 hipóteses: i) O par T2-T3 é o par mais difícil para os falantes portugueses, dado que os dois tons possuem características semelhantes, quer ao nível do pitch inicial registado, quer ao nível da duração. (Gandour, 1978: 43) O par T2T4 é o par mais fácil para os falantes portugueses, uma vez que os dois tons têm características fonéticas distintas (e.g, contorno e altura de F0) e os falantes não nativos de língua tonal são mais sensíveis aos tons quando estes assumem função pragmática na sua língua (Braun & Johnson, 2011). Como o Português é uma língua entoacional e uma melódica assinala uma pergunta e uma descida uma afirmação, espera-se melhor desempenho na percepção do T2 (tom ascendente) versus o T4 (tom descendente). ii) A posição do tom pode influenciar a identificação dos tons, uma vez que os tons têm níveis de F0 diferentes e a ordem em que essa diferença de F0 é apresentada (ouvir sílaba com F0 alto primeiro ou depois) influencia o desempenho dos falantes. No estudo feito pelo Francis & Ciocca (2003), é observado que os falantes do Cantonês, são mais sensíveis à diferença de F0 quando a primeira sílaba tem F0 mais alto que a segunda. É esperado que a ordem em que os tons são apresentados influenciam o desempenho dos enunciadores; iii) Os ouvintes são mais sensíveis à variação em segmentos que em tons, ou seja, os pares que se distinguem apenas em segmentos têm um grau de semelhança mais baixo que os se distinguem em tons, devido ao papel distintivo dos segmentos na língua nativa e ao papel não distintivo dos tons.

A análise de dados mostra que há um efeito significativo, quando às respostas de discriminação, dos factores Tom e Segmento, da Interação entre Tom e Segmento e da interação entre Tom, Ordem e Segmento, mas não existe efeito significativo do factor de ordem (contra a hipótese 2).

Os resultados relativos ao par T2T3 foram surpreendentes, uma vez que T2T3/T3T2 não foram mais difíceis nem o par com tempo de reação mais longo. Os pares T2T4/T4T2 também não foram os pares mais fáceis para os falantes nativos do Português (contrariamente à hipótese 1)

Comparando segmento e tom, verifica-se que os participantes são mais sensíveis ao contraste segmental do que a contrastes tonais (em linha com a hipótese 3). Os pares que se diferenciam em segmentos foram mais facilmente discriminados do que os pares que se distinguem em tons (Do mais difícil ao mais fácil: Seg I- Tom I > Seg I-Tom D > Seg D-Tom I > Seg D-Tom D.) A análise de dados confirma que os falantes são sensíveis às diferenças entre tons quando o segmento é igual, mas não quando o segmento é diferente. Foi observado que os pares com tons e segmentos diferentes não são percebidos como mais diferentes que pares com segmentos diferentes e tons iguais.

Assim, apesar de a diferença entre tons ser mais difícil de discriminar, as falantes de Português não são “*tone-deaf*”. Os resultados relativamente ao tempo de reação vêm acentuar este resultado, uma vez que não existem diferenças significativas quanto ao fator Segmento, mas sim quanto ao Tom (ou seja, a condição tom igual e tom diferente).

Os resultados apresentados neste trabalho revelam que os falantes nativos de Português identificaram o T1T2 e T4T2 como os pares de tons mais difíceis de discriminar e T3T4 como o par mais fácil. Isso também foi reportado entre falantes nativos do Inglês (So & Best, 2010), mas o mesmo não se observa entre, por exemplo, falantes do Japonês, que têm pior desempenho na identificação do par T2T3 (Tsukada, Kondo, & Sunaoka, 2016). Assim, tal como foi descrito previamente (e.g, Schaefer & Darcy, 2014; Wang, 2013; Braun & Johnson, 2011 entre outros), conclui-se que a percepção dos tons varia entre falantes e a L1 influencia a percepção dos tons.

Os falantes de Português mostraram dificuldades em discriminar os tons quando estes partilham uma altura do F0 (*onset* e *offset*) semelhante, como no caso de T1T2, T2T1, T4T2 e T1T4. É possível que os falantes de PE tendam a usar a altura de F0 para distinguir os tons. Portanto, têm mais dificuldades quando as sílabas têm valores de F0 (*onset*, *offset*) mais próximos. Por outro lado, o T1, o tom alto e contínuo, é mais facilmente identificado quando ocorre na posição final. Relembremos que a distinção prosódica entre uma frase interrogativa (H+L*LHi) e uma frase declarativa (H+L*Li) ocorre na fronteira final (Frota, 2002a). Portanto, os falantes podem reconhecer melhor o tom alto quando este ocorre no fim. Além disso, os falantes de PE revelaram-

se mais atentos à altura de F0 que ao contorno de F0: o par T4 (tom descendente) T2 (tom ascendente), que tem contornos distintos, mas altura de F0 próxima (*onset* e *offset*), é o mais difícil para os falantes de PE.

Em acréscimo, os falantes nativos do Português têm mais dificuldades na discriminação quando o segmento é A do que quando é I. Este resultado não era esperado. Uma possível explicação é a produção de “I” é mais estável que “A”, como sugerido por Hoole & Hu (2004). Note-se que nesse trabalho foram selecionados duas vogais que também existem na língua portuguesa. De futuro será interessante verificar a influência de outras vogais e de consoantes na percepção dos tons em Mandarim.

No que respeita à aprendizagem do Mandarim, a identificação do tom tem sido desafiante para os falantes L2. O presente trabalho confirma que essa dificuldade existe também entre os falantes nativos do Português. Os resultados mostram que os falantes nativos do Português têm dificuldade em identificar os pares quando o *offset* da primeira sílaba coincide com o *onset* da segunda, e a natureza da vogal “A” ou “I” tem um impacto significativo na discriminação dos tons.

Vários estudos já confirmaram que usar gestos com as mãos, ver movimentos de lábios na produção do tom e escrever sílabas em Pinyin, um sistema de transcrição alfabética dos caracteres Chineses, podem melhorar significativamente a aprendizagem de tons (Eng et al., 2013; Hirata & Kelly, 2010; Liu et al., 2011). Seria também interessante utilizar algumas destas metodologias e, posteriormente, aplicar o estudo perceptivo a fim de verificar o impacto destas metodologias na percepção dos tons do Mandarim por falantes nativos de Português.

7. Referências

- Abramson, A. S. (1978). Static and Dynamic Acoustic Cues in Distinctive Tones. *Language and Speech*, 21(4), 319–325. <https://doi.org/10.1177/002383097802100406>
- B. Moore, Corinne & Jongman, Allard. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America*. 102. 1864-77. 10.1121/1.420092. <https://doi.org/10.1121/1.420092>
- Burnham, D. K. (2000). Excavations in language development: Cross-linguistic studies of consonant and tone perception. In D. Burnham, S. Luksaneeyanawin, C. Davis, & M. Lafourcade (Eds.), *Interdisciplinary approaches to language processing: The International Conference on Human and Machine Processing of Language and Speech* (pp. 44–69). Bangkok, Thailand: NECTEC.
- Best, C. T. (1995). A Direct Realist View of Cross-Language Speech Perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.
- Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1988). Effects of syllable duration on the perception of Mandarin tones: A cross-language study. *The Journal of the Acoustical Society of America*, 84, S157. <https://doi.org/10.1121/1.2025902>
- Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, 39(4), 585-594. <https://doi.org/10.1016/j.wocn.2011.06.002>
- Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 32(1), 97–103. <https://doi.org/10.1037/0096-1523.32.1.97>
- Bertram, D. (2007). Likert scales. <http://poincare.matf.bg.ac.rs/~kristina/topic-dane-likert.pdf>
- Cowan, N., & Morse, P. A. (1986). The use of auditory and phonetic memory in vowel discrimination. *The Journal of the Acoustical Society of America*, 79(2), 500-507. <https://doi.org/10.1121/1.393537>

- Cao, W., & Wang, Q. (2011, August). Comparative Study on Chinese Tone Perception: A Report on Falling f0 Contours. *ICPhS XVII*, Hong Kong, pp. 404-407
- Cao, C., Xie, Y., Lin, J., Li, Q., & Zhang, J. (2016, October). The preliminary study of influence on tone perception from segments. In *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)* (pp. 1-5). IEEE.
- Cao, R., Wayland, R., Kaan, E. (2012). The role of creaky voice in Mandarin Tone 2 and Tone 3 perception. *13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012*. 1. 426-429.
- Chow, R., Liu, Y., & Ning, J. (2018). The Perception of Mandarin Tones by Thai and Indonesian Speakers. In *Proceedings of the 6th International Symposium on Tonal Aspects of Languages (TAL 2018)*.
- Chao, Y. R. (1968). *Grammar of spoken Chinese*. Berkeley, CA: University of California Press.
- Chen, J., Best, C. T., Antoniou, M., & Kasisopa, B. (2019). Cognitive factors in perception of Thai tones by naïve Mandarin listeners. *Proceedings of the 19th ICPHS*, Melbourne, 1684-1688.
- Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects (Vol. 92)*. Cambridge University Press.
- Chang, Y.-h. S. (2011). Distinction between Mandarin Tones 2 and 3 for L1 and L2 Listeners. In Z. Jing-Schmidt (Ed.), *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23)*. 1, pp. 84-96. Eugene: University of Oregon
- Chen, A., Liu, L., & Kager, R. (2015). Cross-linguistic perception of Mandarin tone sandhi. *Language Sciences*, 48, 62-69. <https://doi.org/10.1016/j.langsci.2014.12.002>
- Cruz-Ferreira, M. (1998). Intonation in European Portuguese. In D. Hirst & A. Di Cristo. (Eds.), *Intonation systems. A survey of twenty languages* (pp.167-178). Cambridge: Cambridge University Press.
- Cruz, M., & Frota, S. (2011). Prosódia dos tipos frásicos em variedades do português europeu: produção e percepção. In A. Costa, I. Falé & P. Barbosa (Eds.), *Textos seleccionados: XXVI*

Encontro nacional da Associação portuguesa de linguística. Porto: Associação Portuguesa de Linguística, 208–25.

Cruz, M., & Frota, S. (2012). Para a prosódia do foco em variedades do português europeu. In A. Costa, C. Flores e N. Alexandre (Eds.), *Textos Selecionados: XXVII Encontro nacional da Associação portuguesa de linguística*. Lisbon: Associação Portuguesa de Linguística, 196–216.

Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, 59(2), 165-179. <https://doi.org/10.3758/BF03211886>

Duanmu, San. (2007). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Erickson, D., Iwata, R., Endo, M., & Fujino, A. (2004). Effect of tone height on jaw and tongue articulation in Mandarin Chinese. In *International symposium on tonal aspects of languages: With emphasis on tone languages*, Beijing, China.

Eng, K., Hannah, B., Leung, L., & Wang, Y. (2013). Can co-speech hand gestures facilitate learning of non-native tones? In *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, p. 060225). ASA. <https://doi.org/10.1121/1.4799746>

Francis, A. L., & Ciocca, V. (2003). Stimulus Presentation Order and the Perception of Lexical Tones in Cantonese. *The Journal of the Acoustical Society of America*, 114(3), 1611-1621. <https://doi.org/10.1121/1.1603231>

Falé, I., & Faria, I. H. (2007). Imperatives, Orders and Requests in European Portuguese Intonation. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 1041-1044.

Frota, S. (2000). *Prosody and Focus in European Portuguese: Phonological Phrasing and Intonation*. New York: Garland.

Frota, S. (2002a). Nuclear falls and rises in European Portuguese: a phonological analysis of declarative and question intonation. *Probus*, 14(1), 113-146.

Frota, S. (2002b). Tonal association and target alignment in European Portuguese nuclear falls. In C. Gussenhoven and N. Warner (Eds.), *Laboratory Phonology 7*. Berlin: Mouton de Gruyter, 387–418

- Frota, S. (2014). The intonational phonology of European Portuguese. In S.-A. Jun (Ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 6–42.
- Frota, S., & M. Vigário (2000). Aspectos de prosódia comparada: ritmo e entoação no PE e no PB. In R.V. Castro & P. Barbosa (Eds.), *Actas do XV Encontro da Associação Portuguesa de Linguística*, Vol.1, 533-555. Coimbra: APL.
- Frota, S., Butler, J., & Vigário, M. (2014). Infants' perception of intonation: Is it a statement or a question? *Infancy*, 19(2), 194–213. <https://doi.org/10.1111/inf.12037>
- Frota, S., Cruz, M., Svartman, F., Collischonn, G., Fonseca, A., Serra, C., Oliveira, P., & Vigário, M. (2015). Intonational variation in Portuguese: European and Brazilian varieties. In Frota, S., & Prieto, P. (Eds.), *Intonation in Romance*. Oxford: Oxford University Press, 235–283
- Fu, Q. J., & Zeng, F. G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing*, 5(1), 45-57.
- Gårding, E., Kratochvil, P., Svantesson, J.-O., & Zhang, J. (1986). Tone 4 and Tone 3 Discrimination in Modern Standard Chinese. *Language and Speech*, 29(3), 281–293. <https://doi.org/10.1177/002383098602900307>
- Gottfried, T. L., Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, 25(2), 207–231. <https://doi.org/10.1006/jpho.1997.0042>
- Gandour, J. T. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey*, 41-76. New York: Academic Press. <https://doi.org/10.1016/B978-0-12-267350-4.50007-8>.
- Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11(2), 149-175.
- Guion, S. G., & Pederson, E. (2007). Investigating the role of attention. In phonetic learning O.-S. Bohn, M. Munro (Eds.), *Second-language speech learning: The role of language experience in speech perception and production: A festschrift in honor of James E. Flege*, Benjamins, Amsterdam, 57-77.

- Grønnum, N. & M. C. Viana (1999). Aspects of European Portuguese Intonation. ICPHS 99, vol.3, 1997-2000. San Francisco.
- Hu, S., Chen, A., & Kager, R. W. J. (2018). Influence of pitch dimensions on non-native tone perception by Dutch and Mandarin listeners. In Proceedings of TAL 2018, 6, 252-255.
- Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of phonetics*, 32(3), 395-421.
- Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 40(2), 269-279. <https://doi.org/10.1016/j.wocn.2011.11.001>
- Hao, Y. C. (2018). Second language perception of Mandarin vowels and tones. *Language and Speech*, 61(1), 135-152. <https://doi.org/10.1177/0023830917717759>
- Hirata, Y., Kelly, S. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, 53(2), 298–310. [https://doi.org/10.1044/1092-4388\(2009/08-0243\)](https://doi.org/10.1044/1092-4388(2009/08-0243))
- Hung, A. C. J., & Chung, R. F. (2016). Perception of Tone by Native Mandarin Chinese Listeners: Optimal Auditory Perception. In Proceedings of the Annual Meetings on Phonology (Vol. 2). <https://doi.org/10.3765/amp.v2i0.3756>
- Hoole, P., & Hu, F. (2004). Tone-vowel interaction in standard Chinese. In Proceedings of the International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages, Beijing, 89–92.
- Huang T. (2001). The interplay of perception and phonology in Tone 3 Sandhi in Chinese Putonghua. In E. Hume & K. Johnson, (Eds.), *Studies on the Interplay of Speech Perception and Phonology*, Ohio State University Working Papers in Linguistics 55, 23-42.
- Huang, T., & Johnson, K. (2010). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica*, 67(4), 243-267. <https://doi.org/10.1159/000327392>

- Jongman, A., Wang, Y., Moore, C., & Sereno, J. (2006). Perception and production of Mandarin tone. In P. Li, L. H. Tan, E. Bates & O.J.L. Tzeng (Eds.), *Handbook of East Asian psycholinguistics* (Vol. 1: Chinese) (pp. 209–217). Cambridge: Cambridge University Press.
- Karypidis, C. (2007, August). Order effects and vowel decay in short-term memory: the neutralization hypothesis. *In Proceedings of the 16th International Congress of Phonetic Sciences*, 657-660.
- Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics*, 53(2), 157-165.
- Lee, C. Y., Tao, L., & Bond, Z. S. (2009). Speaker variability and context in the identification of fragmented Mandarin tones by native and non-native listeners. *Journal of Phonetics*, 37(1), 1-15.
- Lee, Y. S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25(5), 527-542. <https://doi.org/10.1007/BF01758181>
- Levy, E. S. (2009). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America*, 125(2), 1138-1152.
- Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of phonetics*, 36(1), 141-157.
- Li, Y. (2016). English and Thai speakers' perception of Mandarin tones. *English Language Teaching*, 9(1), 122–132.
- Li, B., Shao, J., & Bao, M. (2017). Effects of phonetic similarity in the identification of Mandarin tones. *Journal of psycholinguistic research*, 46(1), 107-124.
- Lin, M., & Francis, A. L. (2014). Effects of language experience and expectations on attention to consonants and tones in English and Mandarin Chinese. *The Journal of the Acoustical Society of America*, 136(5), 2827-2838.

- Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and speech*, 47(2), 109-138.
- Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., & MacWhinney, B. (2011). Learning a tonal language by attending to the tone: An in vivo experiment. *Language Learning*, 61(4), 1119-1141.
- Massaro, D. W., Cohen, M. M., & Tseng, C. Y. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics*, 13(2), 267-289.
- Mateus, M. H. M., Brito, A., Duarte, I., Faria, I., Frota, S., Matos, G., Oliveira, F., Vigário, M. & Villalva, A. (2003), Gramática da língua portuguesa. Lisboa: Caminho, 449–60.
- Macmillan N.A., Braidia L.D., Goldberg R.F. (1987) Central and Peripheral Processes in the Perception of Speech and Nonspeech Sounds. In Schouten M.E.H. (Eds.), *The Psychophysics of Speech Perception*. NATO ASI Series (Series D: Behavioural and Social Sciences), Vol 39. Springer, Dordrecht. https://doi.org/10.1007/978-94-009-3629-4_2
- Nespor, M., & Vogel, I. (2007). *Prosodic phonology: with a new foreword* (Vol. 28). Walter de Gruyter.
- Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America*, 97, 1286–1296. <https://doi.org/10.1121/1.412170>
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human perception and performance*, 20(2), 421-435. <https://doi.org/10.1037/0096-1523.20.2.421>
- Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41(1), 221-231. [https://doi.org/10.1016/S0167-6393\(02\)00105-X](https://doi.org/10.1016/S0167-6393(02)00105-X)
- Pelzl, E. (2019). What makes second language perception of Mandarin tones hard?: A non-technical review of evidence from psycholinguistic research. *Chinese as a Second Language. The Journal of the Chinese Language Teachers Association, USA*, 54(1), 51-78.

- Qin, Z., & Mok, P. P. K. (2011). Perception of Cantonese Tones by Mandarin, English and French Speakers. In ICPhS, 1654-1657.
- Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: human perception and performance*, 5(1), 129.
- Repp, B. H., & Lin, H. B. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics*, 18(4), 481-495.
- Rungruang, Apichai & Mu, Yanhong. (2017). Mandarin Chinese Tonal Acquisition by Thai Speakers. *Asian Social Science*, 13(5), 107-115.
- Sarmah, P., & Cao, R. (2007). A perception study on the third tone in Mandarin Chinese. *UTA Working Papers in Linguistics*, 2, 51 - 66.
- Schaefer, V., & Darcy, I. (2014). Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. *Laboratory Phonology*, 5(4), 489-522.
- Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech*, 58(2), 131-151.
- Shaw, J. A., Tyler, D. A., Kasisopa, B., Ma, Y., Proctor, M. I., Han, C., Derrick, D., & Burnham, D. K. (2013). Vowel identity conditions the time course of tone recognition. *Interspeech*, 3142–3146.
- Shen, J., Deutsch, D., and Le, J. (2011). Overall pitch height as a cue for lexical tone perception, poster session presented at the 162nd Meeting of Acoustical Society of America, San Diego, CA.
- Shen, J., Deutsch, D., & Rayner, K. (2013). On-line perception of Mandarin Tones 2 and 3: Evidence from eye movements. *The Journal of the Acoustical Society of America*, 133(5), 3016-3029.
- So, C. K., & Best, C. T. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech*, 53(2), 273–293.

- So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, 36(2), 195–221.
- Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, 23(5), 689-708.
- Tsukada, K. (2019). Are Asian Language Speakers Similar or Different? The Perception of Mandarin Lexical Tones by Naïve Listeners from Tonal Language Backgrounds: A Preliminary Comparison of Thai and Vietnamese Listeners. *Australian Journal of Linguistics*, 39(3), 329-346. DOI: 10.1080/07268602.2019.1620681
- Tsukada, K., Roengpitya, R., Xu, H. L., & Xu, N. (2013). The perception of Mandarin lexical tones by native Japanese and Thai listeners. *The Journal of the Acoustical Society of America*, 134(5), 4245-4245.
- Tsukada, K., Kondo, M., & Sunaoka, K. (2016). The perception of Mandarin lexical tones by native Japanese adult listeners with and without Mandarin learning experience. *Journal of Second Language Pronunciation*, 2(2), 225-252.
- Tsukada, K., & Kondo, M. (2019). The Perception of Mandarin Lexical Tones by Native Speakers of Burmese. *Language and Speech*, 62(4), 625–640. <https://doi.org/10.1177/0023830918806550>
- Tsukada, K., & Han, J.-I. (2019). The perception of Mandarin lexical tones by native Korean speakers differing in their experience with Mandarin. *Second Language Research*, 35(3), 305–318. <https://doi.org/10.1177/0267658318775155>
- Vigário, M. & Frota, S. (2003). The intonation of standard and northern European Portuguese: a comparative intonational phonology approach. *Journal of Portuguese Linguistics* 2(2), 115–37.
- Wang, Xinchun. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal*, 97(1), 144-160. <https://doi.org/10.1111/j.1540-4781.2013.01386.x>

- Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and language*, 78(3), 332-348. <https://doi.org/10.1006/brln.2001.2474>
- Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, 54(4), 681-712. <https://doi.org/10.1111/j.1467-9922.2004.00283.x>
- Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49(1), 25-47. <https://doi.org/10.1159/000261901>
- Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association*, 33(2), 165-181. <https://doi.org/10.1121/1.4777032>
- Yang, J., Zhang, Y., Li, A., & Xu, L. (2017). On the Duration of Mandarin Tones. *Interspeech 2017, ISCA (2017)*, 1407-1411.
- Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and cognitive processes*, 14(5-6), 609-630. <https://doi.org/10.1080/016909699386202>.
- Yip, M. C., Leung, P. Y., & Chen, H. C. (1998). Phonological similarity effects in Cantonese spoken-word processing. In *Proceedings of ICSLP'98 (Vol. 5, pp. 2139–2142)*, Sydney, Australia.