The Perceptual Assimilation of Danish Monophthongs and Diphthongs by Monolingual Australian English Speakers

Mona M. Faris[1,2], Catherine T. Best[1,3], and Michael D. Tyler[1,2]

[1]The MARCS Institute, [2]School of Social Sciences and Psychology, [3]School of Humanities and Communication Arts, University of Western Sydney, Sydney, Australia
m.faris@uws.edu.au; c.best@uws.edu.au; m.tyler@uws.edu.au

ABSTRACT

This study explores how experience with native language (L1) diphthongs influences the assimilation of non-native diphthongs. To obtain a comprehensive understanding of L1 attunement, native Australian English (AusE) speakers categorized and rated the Danish diphthongs, in addition to the monophthongs, in relation to their entire native vowel inventory. Short Danish vowels were assimilated to both lax and tense AusE vowels, and long Danish vowels were primarily categorized to tense AusE vowels. Only two of the Danish diphthongs were assimilated to an AusE diphthong, while the rest were either assimilated to an AusE tense vowel or were uncategorized. This suggests that the perceptual assimilation of non-native diphthongs is not based purely on sensitivity to vowel dynamics, but also on the perceived phonetic similarity between non-native diphthongs and L1 vowels. Implications for modelling cross-language speech perception are discussed.

Keywords: Cross-language speech perception, perceptual assimilation, vowels, diphthongs.

1. INTRODUCTION

Adults' perception of foreign consonants and vowels is influenced by their native language (L1) experience [2]. An adult's phonological system has often been likened to a 'sieve', which filters out phonetic information that is not phonologically meaningful in the L1 [12]. There have been no studies to date that have examined how experience with L1 diphthongs affects the perception of non-native diphthongs, so the primary focus of this study will be on the perceptual assimilation of non-native diphthongs by naïve listeners. Additionally, to gain a more complete understanding of perceptual attunement to the L1, the non-native listeners will also be asked to perceptually assimilate both diphthongs and monophthongs from a foreign language to their entire vowel inventory.

The Perceptual Assimilation Model (PAM; [2]) was developed to account for the effect of L1 attunement on speech perception. According to PAM, non-native phones may assimilate to an adult's phonological system in three different ways. A non-native phone may be *categorized* as an identical or deviant exemplar of an L1 category. If a non-native phone is perceived as speech but not as similar to any particular L1 category, it is considered *uncategorized*, and if it is not perceived as speech, it is deemed *non-assimilable*. PAM predicts discrimination accuracy of pairs of non-native phones based on these assimilation patterns. This study will only focus on the assimilation of individual non-native phones from an entire non-native vowel system in order to provide insight into how L1 attunement shapes the L1 phonological system.

The perception of non-native monophthongs has been shown to be influenced by L1 experience [13]. For instance, identifying and discriminating front rounded vowels is difficult for those who do not distinguish between front rounded versus unrounded vowels in their L1. In [11], American English (AE) listeners categorized and rated the North German vowels /yː, ʏ, øː/ and the Parisian French vowels /y, ø/ to L1 monophthongs. These vowels were primarily categorized to the AE back rounded vowels, despite being more acoustically similar to the front unrounded vowels. In [8], monolingual Spanish speakers categorized AE /i/ and /u/ to their /i/ and /u/, respectively, while the AE /ɛ, ɪ, æ/ were assimilated to Spanish /e/. Such studies demonstrate that listeners perceive non-native monophthongs in relation to those in their L1 inventory.

In contrast, little is known about the effect of L1 attunement on the perception of non-native diphthongs. Although diphthongs and monophthongs are both classified as vowels, they differ markedly from one another articulatorily and acoustically. While monophthongs are produced with the tongue and jaw in a relatively stable position throughout production, diphthongs are much more dynamic as they are produced with substantial movement of the tongue, jaw, or lips during articulation [6]. Diphthongs are often described as a combination of two monophthongs, but native diphthongs are nonetheless perceived as a single phone. It has been shown that listeners make fewer errors identifying diphthongs than monophthongs due to the rich dynamic information available in diphthongized

vowels [1]. Given such differences, the perception of non-native diphthongs is likely to differ from monophthongs.

In the only study to date to investigate the perception of non-native/second-language (L2) diphthongs under a PAM framework, native Japanese speakers' perceptual assimilation of all the Australian English (AusE) monophthongs and diphthongs was examined [5]. Since diphthongs are not employed in Japanese, listeners were asked to identify the AusE vowels in terms of all the permissible combinations of the monomoraic vowels (/ie, ia, iu, io, ei, ea, eu, eo, ai, ae, au, ao, ui, ue, ua, uo, oi, oe, oa, ou/), in addition to their monomoraic (/a, e, i, o, u/) and bimoraic vowels (/aː, eː, iː, oː, uː/). As predicted, the majority of the AusE diphthongs were categorized to the Japanese bimoraic vowel combinations. The diphthongs /ɑe, æɪ, oɪ/ were categorized as /ai, ei, /oi/, respectively, while /æɔ, əʉ/ were uncategorized. However, the listeners identified the AusE diphthongs in terms of their monomoraic vowel combinations, which are not classified as diphthongs [10]. It is yet to be determined how experience with vowel dynamics shapes the perception of non-native diphthongs for speakers of languages with diphthongs in their vowel inventories.

For PAM to be a comprehensive model of cross-language speech perception, it must be able to account for the effect of L1 attunement on the perception of all types of speech sounds. Therefore, the current study assessed the perceptual assimilation of non-native monophthongs and diphthongs in listeners who have L1 experience with both types of vowels – Australian English (AusE) speakers. Danish is diphthong-rich, so it served as the stimulus language. The AusE speakers perceptually categorized and rated the goodness-of-fit of all the Danish vowels in relation to their full AusE vowel inventory. It was predicted that short and long Danish vowels would be assimilated to AusE lax and tense monophthongs, respectively. There are two possible scenarios for the assimilation of the Danish diphthongs. It may be that due to the AusE speakers' experience with vowel dynamics, Danish diphthongs will be assimilated only to L1 diphthongs. Alternatively, sensitivity to vowel dynamics may not necessarily result in the assimilation of Danish diphthongs to L1 diphthongs. That is, listeners may only assimilate Danish diphthongs to L1 diphthongs if they are perceived as being phonetically similar, as opposed to being based on vowel dynamics *per se*. The latter possibility is more consistent with PAM, which states that the perceptual assimilation of non-native phones depends on their perceived similarities and differences to L1 categories. Danish diphthongs that are not assimilated to L1 diphthongs may instead be categorized to L1 tense vowels or may be uncategorized.

2. METHOD

2.1. Participants

Forty monolingual Australian English speakers (31 females, M_{age} = 25 yrs, age range: 18-55) were recruited from the student pool at the University of Western Sydney and from the Greater Western Sydney community. Participants who were enrolled in Introductory Psychology courses received course credit for their participation and those recruited from the community received monetary reimbursement. All participants reported normal or corrected-to-normal vision and no hearing or language impairments.

2.2. Stimuli and Apparatus

2.2.1. Stimulus recording

Three native Danish female speakers were recorded in a sound-attenuated booth at The MARCS Institute. All three speakers were 24 years of age and were born and raised in Copenhagen, Denmark. The speakers produced the Danish short vowels /i, e, ɛ, ɑ, u, o, ɔ, y, ø, œ/, long vowels /iː, eː, ɛː, aː, uː, oː, ɔː, yː, øː, œː/, diphthongs /uj, iw, ew, ɛw, ɒw, yw, øw, œw, œw, i͡ɐ, e͡ɐ, æ͡ɐ, y͡ɐ, ø͡ɐ, œ͡ɐ, u͡ɐ, o͡ɐ/, and /ə/ [9] in /ˈhVbə/ nonsense syllables.[1] They were instructed to speak in a natural, conversational manner, and to produce the tokens with a falling intonation. A female AusE speaker also produced all AusE vowels in /ˈhVbə/ context for a familiarization task. One auditory token per AusE vowel (18 in total) was selected.

Nonsense words were presented one at a time on a computer monitor situated in front of the speaker. The informants were instructed to speak in a normal, conversational manner, and to produce the nonsense words with a falling intonation. Approximately 8 randomized blocks of the 38 nonsense words were presented and speech production was recorded using a Shure SM10A headset microphone connected to a computer via an Edirol UA-25EX external USB sound card, with a sampling rate of 44.1 kHz.

Stimulus presentation and response collection was controlled using PsyScope X B57 on a MacBook laptop, Sennheiser HD 650 headphones, and an Edirol UA-25EX external USB sound card.

2.2.2. Stimulus preparation

Auditory recordings were high-pass filtered in Praat [3] at 70 Hz to attenuate any unwanted low-frequency rumble and to correct for the DC

component. Each token was ramped so that there was a 10 ms fade-in for the token onset and a 20 ms fade-out for the token offset. Audible clicks were excised from the tokens using Praat.

To select the most suitable tokens for the final set of stimuli, five native Danish speakers verified all potential tokens in an identification task with goodness ratings. The tokens for the vowels /eˌ̂, ɐw, øˌ̂, œw, øw, ə/ were inconsistently and/or incorrectly identified, and were therefore excluded from the present study. Three tokens per speaker for each vowel were selected, resulting in 288 tokens in total. Table 1 displays the acoustic measurements.

Table 1: Mean Danish target vowel duration (ms), and F1, F2, and F3 at the 25%, 50%, and 75% of the vowel duration.

Danish	Vowel	F1 (Hz)			F2 (Hz)			F3 (Hz)		
vowel	duration	25%	50%	75%	25%	50%	75%	25%	50%	75%
i	79	342	357	368	2384	2373	2310	3347	3277	2996
e	89	457	480	487	2341	2277	2143	3075	3039	2887
ɛ	87	512	537	540	2148	2074	1972	2930	2904	2760
a	107	915	910	854	1475	1452	1418	2754	2723	2700
y	93	362	377	382	2101	2079	1973	2609	2579	2422
ø	108	472	486	486	1852	1860	1784	2594	2568	2511
œ	113	508	529	523	1774	1754	1684	2605	2613	2566
u	96	418	425	424	923	899	839	2560	2555	2528
o	95	482	489	479	878	873	861	2681	2696	2690
ɔ	100	730	731	689	1208	1234	1229	2635	2629	2602
iː	169	330	325	338	2484	2498	2502	3536	3544	3473
eː	182	443	443	445	2465	2495	2505	3259	3308	3240
ɛː	176	476	482	503	2300	2289	2224	3039	3045	2950
aː	155	615	607	595	2214	2218	2171	2967	2975	2915
yː	171	348	339	345	2150	2150	2123	2534	2575	2548
øː	176	449	459	471	1890	1896	1866	2531	2537	2523
œː	178	484	490	498	1838	1840	1826	2607	2593	2591
uː	165	406	404	404	839	802	767	2490	2502	2511
oː	173	454	458	459	769	741	732	2740	2781	2802
ɔː	186	484	490	504	1088	1068	1048	2572	2603	2637
iw	157	377	427	428	2257	1751	1107	2825	2438	2508
uj	170	449	463	445	1036	1566	2040	2533	2489	2588
ɛw	166	511	545	514	2090	1632	1161	2759	2598	2597
yw	152	411	435	439	1919	1575	1112	2414	2370	2435
œw	162	535	534	500	1635	1352	1078	2496	2480	2539
ɒw	148	634	608	538	994	911	843	2692	2761	2822
iˌ̂	167	392	519	671	2345	2116	1723	3242	2965	2790
uˌ̂	176	466	530	691	770	1030	1221	2694	2653	2660
yˌ̂	176	390	502	615	2017	1797	1512	2541	2534	2611
oˌ̂	166	535	639	721	835	998	1151	2718	2738	2734
œˌ̂	164	689	751	773	1625	1505	1393	2602	2597	2606
æˌ̂	162	702	793	824	1985	1741	1543	2906	2765	2707

2.3. Procedure

Participants were tested in groups of up to three at the University of Western Sydney. They first completed a familiarization task to help make them aware of the vowel sounds in the English keywords. On a given trial, participants heard an AusE vowel in /ˈhVbə/ context. A 6 x 3 grid was displayed containing the AusE vowels /ɐ, e, ɪ, ɔ, ʊ, æ, ɐː, eː, iː, oː, ʉː, ɜː, ɑe, æɔ, æɪ, əʉ, ɪə, oɪ/ [7] presented in the English keywords *up, bet, hid, hot, hood, had, hard, hair, heed, hoard, boot, her, hide, how, bay, boat, here,* and *hoist,* respectively, with the target vowel sound highlighted in red. They were asked to select a keyword containing the same vowel that they had heard. If the incorrect label was selected, the correct response was given as feedback. The familiarization ended when they reached a score of 60 or they completed 100 trials. Each participant was presented with a different randomized order of the trials.

Participants then completed a category assimilation task with goodness rating using the Danish vowels. On a given trial, participants heard a /ˈhVbə/ nonsense word and were again asked to attend to the first vowel. The same grid from the familiarization task was presented and participants used a computer mouse to select an English keyword containing the closest sounding vowel to the one they heard. The same auditory token was presented a second time for participants to rate how well it matched the chosen vowel on a 7-point scale (1 = sounds very poor to 7 = sounds perfect). No feedback was provided. There were 288 trials (32 vowels x 3 speakers x 3 repetitions) in total, and each participant received a different randomized order of the trials. They were given 6 s to select a category label and 3.5 s to rate the vowel, and if no response was registered for either part of the trial, that same trial was randomly repeated later. The ITI was 480 ms. The session duration was 50 min.

3. RESULTS

Table 2 displays the mean percent categorization and goodness ratings of each Danish vowel to an AusE category label. Using a 50% assimilation criterion (following [5]), two of the Danish short vowels were uncategorized, while the remaining short vowels were categorized, with the responses split between the AusE lax and tense vowels. For example, the Danish /e/ and /i/ were both categorized to the AusE lax vowel /ɪ/, while /œ/ and /ø/ were both categorized to the AusE tense vowel /ɜː/.

All Danish long vowels were assimilated to tense AusE vowels except for three long vowels that were uncategorized. Similar to the assimilation pattern observed for some Danish short vowels, there were instances where two Danish tense vowels were categorized to a single AusE vowel. For instance, /eː/ and /iː/ were both assimilated to AusE /iː/, and /œː/ and /øː/ were both categorized to /ɜː/. In some instances, Danish vowels differing only in length were categorized to the same AusE vowel. For example, /y/ and /yː/ were both categorized to /ʉː/, and /œ/ and /œː/ were both categorized to AusE /ɜː/.

The only Danish diphthongs that were assimilated to AusE diphthongs were /uj/ and /iˌ̂/, which were categorized as /oɪ/ and /ɪə/, respectively. Five Danish

diphthongs were categorized to an AusE tense vowel, while the remainder of the diphthongs were uncategorized. Some of the Danish long vowels and diphthongs were categorized to the same L1 category. For instance, the Danish long vowels /œː/ and /øː/, and the diphthongs /œɐ̯/ and /œw/ were all categorized as AusE /ɜː/. Similarly, /ɛː/ and /æɐ̯/ were assimilated to AusE /eː/.

Table 2: Mean percent categorization (%) and goodness rating (GR), both averaged across participants, of Danish vowels by Australian English speakers. The top three responses (R) are presented for the uncategorized vowels.

Danish	1			2			3		
vowel	R	%	GR	R	%	GR	R	%	GR
ɔ	ɔ	57%	5.64						
e	ɪ	55%	5.40						
i	ɪ	57%	5.43						
u	ʊ	63%	5.15						
a	ɐː	53%	5.52						
œ	ɜː	89%	5.87						
ø	ɜː	60%	5.26						
y	ʉː	58%	5.01						
ɛ	e	40%	5.10	eː	23%	4.90	ɜː	12%	3.79
o	ɔ	33%	4.32	ʊ	22%	4.90	oː	21%	4.82
eː	iː	51%	5.30						
iː	iː	59%	5.58						
ɛː	eː	51%	5.28						
œː	ɜː	86%	5.83						
øː	ɜː	53%	5.07						
yː	ʉː	58%	5.35						
uː	ʊ	56%	5.04						
aː	eː	48%	5.42	æɪ	24%	5.46	e	14%	5.14
ɔː	oː	36%	4.61	əʉ	15%	4.45	ɔ	15%	3.73
oː	oː	44%	5.08	ʊ	21%	4.81	əʉ	9%	4.24
uj	oɪ	78%	4.63						
iɐ̯	ɪə	80%	6.14						
æɐ̯	eː	55%	5.36						
œɐ̯	ɜː	68%	5.37						
œw	ɜː	54%	4.19						
uɐ̯	oː	57%	4.77						
oɐ̯	oː	74%	5.24						
ɒw	ɔ	36%	4.59	oː	23%	3.93	əʉ	18%	4.66
ɛw	ɜː	28%	3.69	e	18%	4.22	eː	18%	3.78
iw	ɐ	19%	3.41	ɪ	16%	4.04	ʉː	16%	3.41
yɐ̯	ʉː	20%	3.44	ɜː	18%	3.26	ʊ	16%	3.72
yw	ʉː	37%	3.83	ɐ	16%	2.79	ʊ	13%	4.06

4. DISCUSSION

This study examined how AusE listeners perceptually assimilate the Danish monophthongs and diphthongs to their entire vowel inventory. It was predicted that short and long Danish vowels would be assimilated to AusE lax and tense vowels, respectively. This prediction was largely upheld for the assimilation of long Danish vowels, but it was not the case for the short Danish vowels as many of them were also assimilated to tense AusE vowels.

If listeners assimilated Danish diphthongs to AusE diphthongs purely based on dynamic vowel quality, the majority of the Danish diphthongs would be expected to have been categorized to AusE diphthongs. However, this prediction was not supported as only two of the Danish diphthongs were assimilated to an AusE diphthong. This finding is more consistent with our alternate hypothesis that despite experience with L1 vowel dynamics, non-native listeners do not assimilate Danish diphthongs to L1 diphthongs purely because of their sensitivity to vowel dynamics. According to PAM, this finding may be explained in terms of the perceived phonetic similarity between the native and non-native phones. For Danish diphthongs assimilated to AusE tense vowels, non-native listeners appear to have detected one of the targets in the Danish diphthongs and assimilated it based on the most similar AusE tense vowel. For instance, the Danish monophthong /œ/ and the diphthongs /œw/ and /œɐ̯/ were assimilated to AusE /ɜː/. It seems that listeners were sensitive to the rounded /œ/ element of both diphthongs. Similarly, the Danish vowels /oɐ̯/ and /uɐ̯/, both of which involve lip rounding during the first vowel target, were categorized to the rounded AusE /oː/. Additional studies within a PAM framework are required to further explore the factors determining assimilation of non-native diphthongs to the L1 phonological system.

While the current study has demonstrated the way in which non-native diphthongs assimilate to listeners' L1 phonological system, little is known about the extent to which PAM's discrimination predictions are applicable to diphthongs, particularly since studies testing PAM's predictions have been conducted using non-native consonants and monophthongs. As PAM was developed to account for all types of speech segments, it is necessary to examine PAM discrimination predictions for diphthongs. The only study to assess PAM's discrimination predictions using diphthongs is that by [4]. Native Japanese speakers were assessed on their perception of the AusE vowels. In a discrimination task, the two contrasts tested involving diphthongs were /iː/-/ɪə/ which were assimilated as single category, and /əʉ/-/oː/ which were assimilated as uncategorized-categorized. Consistent with PAM predications, the single-category contrast was discriminated poorly and discrimination for the uncategorized-categorized contrast was fair. These contrasts were composed of a tense vowel and a diphthong. But, it is still not known how well listeners are able to discriminate *between* non-native diphthongs. On the basis of these results, future studies will be able to test PAM discrimination predictions for contrasts where both phones are diphthongs. Also, given that the current study has established the assimilation of the Danish vowels by AusE speakers, there is also the opportunity for testing PAM discrimination predictions for various assimilation types within the same stimulus language.

5. REFERENCES

[1] Assmann, P. F., Nearey, T. M., Hogan, J. T. 1982. Vowel identification: Orthographic, perceptual, and acoustic aspects. *J. Acoust. Soc. Am.* 71, 975-989.

[2] 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*. Baltimore: York Press, 171-204.

[3] Boersma, P., Weenink, D. 2005. Praat: Doing phonetics by computer (version 4.3. 14). [Computer program], retrieved May 26, 2005, http://www.praat.org/.

[4] Bundgaard-Nielsen, R. L., Best, C. T., Tyler, M. D. 2011. Vocabulary size is associated with second-language vowel perception performance in adult learners. *SSLA* 33, 433-461.

[5] Bundgaard-Nielsen, R. L., Best, C. T., Tyler, M. D. 2011. Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *J. App. Psycholing.* 32, 51-67.

[6] Catford, J. C. 2001. *A Practical Introduction to Phonetics*. New York: Oxford University Press.

[7] Cox, F. 2006. The acoustic characteristics of /hvd/ vowels in the speech of some australian teenagers. *Aust. J. Ling.* 26, 147-179.

[8] Escudero, P., Chládková, K. 2010. Spanish listeners' perception of American and Southern British English vowels. *J. Acoust. Soc. Am.* 128, EL254-EL260.

[9] Grønnum, N. 1998. Danish. *JIPA* 28, 99-105.

[10] Labrune, L. 2012. *The Phonology of Japanese*. New York: Oxford University Press.

[11] Strange, W., Levy, E. S., Law II, F. F. 2009. Cross-language categorization of French and German vowels by naïve American listeners. *J. Acoust. Soc. Am.* 126, 1461-1476.

[12] Trubetzkoy, N. S. 1969. *Principles of Phonology*. Berkeley, California: University of California Press.

[13] Tyler, M. D., Best, C. T., Faber, A., Levitt, A. G. 2014. Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71, 4-21.

---

[1] We acknowledge that the Danish diphthongs could be transcribed as either vowel-vowel or vowel-glide. We employ the transcription approach used in [9].