

PERCEIVING THROUGH THE LENS OF NATIVE PHONETICS: ITALIAN AND DANISH LISTENERS' PERCEPTION OF ENGLISH CONSONANT CONTRASTS

Ocke-Schwen Bohn^a, Catherine T. Best^b, Cinzia Avesani^c & Mario Vayra^d

^aEnglish Department and ConAmore, Department of Psychology, Aarhus University, Denmark;

^bMARCS Auditory Laboratories, University of Western Sydney, Australia; Haskins Laboratories, USA;

^cInst. Cognitive Sciences and Technologies - CNR, Padua, Italy;

^dDepartment of Linguistics, University of Bologna, Italy

engosb@hum.au.dk; c.best@uws.edu.au; cinzia.avesani@pd.istc.cnr.it; mario.vayra@unibo.it

ABSTRACT

This study examined how the phonetic details of the phonological systems for two different native-language listener groups interact with their perception of the phonetic details of three non-native consonant contrasts. Current theoretical models of cross-language and second language speech perception are evaluated by relating native Italian and native Danish listeners' perceptual assimilation of audio tokens of English /b v w θ¹ to how well the two groups discriminate the corresponding English contrasts /b/-/v/, /w/-/v/, and /θ/-/v/. Results indicate some support for the models, but also performance differences between the groups that are unexpected by any existing models. Implications for existing hypotheses about non-native speech perception are discussed.

Keywords: cross-language, nonnative consonants, perceptual assimilation, native phonetics

1. INTRODUCTION

Adult listeners are well known to have difficulty in perceiving non-native speech contrasts, but the degree of perceptual difficulty varies considerably across contrasts. Several models of cross-language and second language speech perception attempt to account for these variations across non-native contrasts; the models differ primarily with respect to the level of analysis (phonetic, phonological), degree of phonetic detail, and type of phonetic information (acoustic, auditory, articulatory) they postulate to be involved in cross-language speech perception.

Best's Perceptual Assimilation Model (PAM) [1] posits that listeners strongly tend to assimilate a non-native phone to the closest native phonological category on the basis of phonetic (articulatory) similarity, such that it will be categorized as a token of that native phoneme. However, if a non-native

phone is instead moderately similar to more than one native phoneme, it will be heard as an uncategorized speech sound; in rare cases the phone's articulatory properties may be so deviant from any native phonemes that it will not be assimilated to the native phonological system at all (non-assimilable). Discrimination levels for non-native contrasts, accordingly, are predicted to depend on how the contrasting phones are assimilated. When each is assimilated to a different native category (Two Category: TC type) or one is Uncategorized and the other Categorized (UC type), discrimination will be excellent, on the reasoning that in this case the native system of *phonological contrasts* supports the distinction. But if the contrasting phones are both assimilated to the same native category, discriminability will depend on each non-native phone's similarity to the *phonetic* properties of the native phoneme. Discrimination will be poor if the two non-native phones are equivalent in degree of phonetic fit to the native category (Single Category: SC type), but it will be fairly good if one is a phonetically better example of the native category than the other (Category Goodness difference: CG type). Extension of PAM to second language (L2) learners [2] adds the predictions that L2 learners will have difficulties learning new L2 categories both for SC contrasts, where neither L1 phonology *nor* phonetics will assist them, and for TC contrasts, which engender no pressure to establish a new contrast. However, L2 learners should easily establish a new L2 category for the phonetically less-good member of a CG assimilation.

Flege's Speech Learning Model (SLM) [3] focuses instead on L2 speech production, positing it to be constrained by L1 phonetic influences on the perception of L2 phones. SLM views the relationship of L2 phones to L1 phonetic categories as falling along a dimension that runs from "identical" through "similar" to "new." Similar

phones of the L2 are perceived to be more or less deviant exemplars of L1 categories, and hence are subject to equivalence classification, which blocks L2 category formation. "New" phones, however, are phonetically different enough from any existing L1 category that L2 learners will eventually establish new categories for them.

Our study tests and extends the predictions of these models by examining how native speakers of Italian (IT) and of Danish (DK) perceive three English consonant contrasts: /b/-/v/, /w/-/v/ and /ð/-/v/. Both English and IT employ a three-way constriction parameter contrast among labial consonants with /b/, /v/, /w/ as phonologically contrastive segments, the only difference being that IT /b/ is always pre-voiced as [b], whereas English /b/ in initial position is short-lag voiceless [p]. Danish does not have /w/, but has /b/, realized as [p], and /v/, realized as a labiodental approximant [v], which differs in constriction degree from IT and English fricative [v]. DK and IT both lack /ð/. The IT /d/, though, is voiced and dental, making it articulatorily more similar to English /ð/ than is DK /d/, which is short-lag voiceless and alveolar *except* when it occurs postvocally as a dental approximant.

PAM/PAM-L2 predicts IT listeners will show TC assimilations to native contrasts for all three English contrasts, as /p/, /v/, /w/ are native phonemes and [ð] should be assimilated to IT /d/. Thus Italians' discrimination should be excellent for all three contrasts. For DK listeners, however, only /b/ and its phonetic realization as [p] are native. The English fricative realization of /v/ differs from DK approximant [v], and neither /w/ nor /ð/ are found in DK. Nor does phonetic [w] occur, though a dental approximant occurs as a postvocalic allophone of /d/. DK listeners should thus assimilate English /ð/ to DK /d/ and categorically distinguish it from [v] as a TC contrast. They should also show TC assimilation for English [b]-[v]. However, *phonetic* differences between English [v] and DK [v], and between English [ð] and the allophones of DK /d/ ([d] and the dental approximant) may be expected to reduce their discrimination somewhat on /ð/-v/ [4]. They should assimilate [w]-[v] to Danish [v] as either a CG contrast, with fair to good discrimination, or as an SC contrast and thus discriminate it poorly.

SLM predicts that IT and DK listeners will discriminate English [p]-[v] very well because IT listeners will treat English [p] as equivalent to IT [p] and because IT and English [v] are identical. DK and English [p] are identical, and DK listeners will treat English [v] as phonetically similar to DK [v].

Italian listeners will also discriminate [w]-[v] very well because IT has these phonetic categories. Danish listeners, however, should discriminate [w]-[v] less well because they are expected to equivalence-classify both labials to DK [v]. The SLM prediction for [ð]-[v] is that IT listeners should discriminate this contrast well because IT has [v], and English [ð] is treated as equivalent to Italian [d]. Danes should also discriminate [ð]-[v] well, if [ð] is treated as equivalent to L1 [d] and [v] to L1 [v]. If one or both [ð], [v] are treated as *new* L2 categories, they should still discriminate well, but perhaps slightly less well than the Italians.

2. METHODS

Native Italian (IT) (Tuscany, $n = 41$) and native Danish (DK) speakers (East Jutland, $n = 30$) participated (age = 16-25 yrs). English experience of members in each group was comparable and typical for young IT and DK adults.

A male speaker of American English produced 20 tokens each of [ba, wa, ða, va], from which 6 tokens of each category were selected; acoustic analysis ensured that they matched on all acoustic dimensions except those critical to the contrasts.

Participants first took an AXB discrimination test, in each trial of which A was either a token of [ba], [wa], or [ða], and B was a token of [va], or A was [va] and B was [ba], [wa], or [ða] (216 randomized trials in 72-trial blocks; ISI = 1.0 s, ITI = 3.5 s). They then completed an open-response categorization test in which each stimulus was randomly presented eighteen times. DK listeners labeled the stimuli twice, first using only DK orthographic labels, after which they were told they could use any orthographic labels they wished (including English). The IT listeners labeled the stimuli once, using IT orthographic labels.

3. METHODS

3.1. Labeling

Table 1 shows IT and DK assimilations of English [b, v, w, ð] to L1 consonants. A target item was considered assimilated to a native category if that category was used in $\geq 70\%$ of opportunities.

As expected, the labeling of English [b] and [v] by both groups was straightforward: IT listeners always labeled English [p] (/b/) as P, and DK listeners always labeled it B. Similarly, all DK and all but one IT listeners labeled English [v] as V. Thus, both groups show TC assimilation of [b] and [v], predicting excellent discrimination of [b]-[v].

English [w] was assimilated, without exception, to W by Italians. The assimilation of English [w] by DK listeners suggests they are quite aware of the difference between DK [v] and the non-native [w]. When instructed to use only DK orthography, six (of 30) labeled [w] as V, 11 labeled it variably as B or V, and 13 responded either exclusively with W (a non-Danish label) or with W and V. Upon debriefing, the DK listeners justified their use of the foreign orthographic symbol W with English loanwords such as *weekend*. When given a second opportunity to label English [w], this time using any orthographic symbol they saw fit, 24 used W exclusively, while the remaining six additionally used B and V. This suggests that DK listeners are aware of constriction location differences between DK [v] and English [w], suggesting they should discriminate [w]-[v] very well, like IT listeners, who show TC-assimilation of this English contrast.

Table 1: Assimilation of English consonants to L1 categories, in percent choice per L1 orthographic item.

English Consonants	Native Categories							
	b	p						
[b]	IT	100						
	DA	100						
[v]		va	v					
	IT	97	3					
	DA	100						
[ð]		d	t	“th”	l	v	W	other
	IT	78		8			3	11
	DA	50	17		10	10		13
[w]		b	v	w				other
	IT			100				
	DA	4	30	36				30

Almost all IT listeners (34 of 41) assimilated English [ð] to native D. Of the remaining IT participants, three responded exclusively with the non-Italian TH, and four gave variable responses of D, TH or L. Together with their exceptionless assimilation of English [v] to L1 V, this suggests that IT listeners TC-assimilate English [ð]-[v], which should result in excellent discrimination.

The DK participants were more variable in their responses. When told to use DK orthographic labels, five (of 30) DK listeners assimilated English [ð] to V, 15 used D, six used T, and four used several DK categories. When permitted to use any label, 23 DK listeners labeled English [ð] with the non-Danish TH, and seven gave variable responses (mostly including V). This variable assimilation pattern including the use of V by DK listeners suggests that their discrimination of [ð]-[v] will not be as good as native IT listeners'. Several DK listeners showed assimilation patterns

for this contrast which suggest English [ð] and [v] are assimilated to the same DK category (SC or CG), which should compromise discriminability.

Table 2 summarizes the pairwise assimilation types for the English contrasts /b-/-v/, /w-/-v/, and /ð-/-v/, based on IT and DK listeners' labeling of each consonant. We distinguish UC assimilations in which the split of responses shows no overlap of the native categories used for each of the two non-native phones (UC-N), from those in which one or more native categories were used in response to *both* non-native phones (UC-O). Table 3 displays these assimilation types using a 3-way distinction among assimilations to separate L1 categories (which result in excellent discrimination), overlapping categories (reduced discriminability), and the same category(s) (further reduced).

Table 2: Assimilation types in percentages: TC = Two category; UC-N = Uncategorized-Categorized/non-overlapping labels; UC-O = Uncategorized-Categorized/overlapping labels; CG = Category Goodness; SC = Single Category.

	English contrast	TC	UC-N	UC-O	C G	S C
Italian with Italian orthography	b-v	100				
	ð-v	86	14			
	w-v	97	3			
Danish with Danish orthography	b-v	100				
	ð-v	70	3	10	10	6
	w-v	40	3	27	30	
Danish with English orthography	b-v	96	4			
	ð-v	74	7	14	2	3
	w-v	80	10	10		

Table 3: Assimilation types in percentages. Separate native categories for the contrasting non-native phones: TC + UC-N; Overlapping categories: UC-O; Same category: SC + CG.

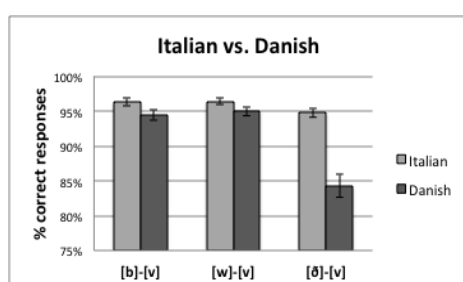
	English contrast	separate	overlap	same
Italian with Italian orthography	b-v	100	0	0
	ð-v	100	0	0
	w-v	100	0	0
Danish with Danish orthography	b-v	100	0	0
	ð-v	74	10	16
	w-v	44	26	30
Danish with English orthography	b-v	100	0	0
	ð-v	80	14	4
	w-v	90	10	0

3.2. Discrimination

Figure 1 shows percent correct discrimination by IT and DK listeners for English for /b-/-v/, /w-/-v/, and /ð-/-v/. As predicted by both PAM and SLM, the IT and DK listeners' discrimination of English [b]-[v] was excellent (IT: 96%, DK: 94 %) and did not differ between IT and DK, *n.s.* The groups were also equally good at discriminating /w-/-v/

(IT: 96%, DK: 95%; no difference, *n.s.*). The DK listeners' very high discrimination level here is somewhat surprising given that DK, unlike English and IT, lacks this continuant labial consonant contrast. However, the labeling responses of the DK listeners clearly indicated that they could hear a difference between [w] and [v]. The only contrast for which IT and DK listeners differed was [ð]-[v]: Again, the IT listeners' discrimination was excellent (95% correct), but the DK listeners' discrimination, although fairly good, was significantly worse (84% correct) than the IT group, $F(1,149) = 27.01, p < 0.0001$.

Figure 1: Percent correct discrimination of the three English contrasts by native speakers of Italian and Danish.



4. DISCUSSION

This study examined the predictions of two models of non-native speech perception, PAM and SLM, regarding assimilation patterns for L1 Italian and L1 Danish perception of English /b/, /v/, /w/, /ð/, and how well these assimilation patterns predicted discrimination by the non-native listeners of the English contrasts /b/-/v/, /w/-/v/, and /ð/-/v/.

Both PAM and SLM predicted correctly that IT listeners would assimilate (PAM), or treat as equivalent to native categories (SLM), all four English target consonants. This resulted in the IT listeners' expected high discrimination scores for /b/-/v/, /w/-/v/, and /ð/-/v/. The phonetic differences between English and IT /b/ (absence vs presence of glottal pulsing during labial closure) and between English /ð/ and IT /d/ (constriction degree) did not compromise IT listeners' discrimination of the /b/-/v/ and /ð/-/v/ contrasts.

The models' predictions for the DK listeners were correct only for the /b/-/v/ contrast, though. This contrast was assimilated or treated as equivalent to two non-overlapping DK categories and accordingly discriminated very well. However, both PAM and SLM predicted that DK listeners would not discriminate /w/-/v/ well because of SC assimilation (PAM) or equivalence classification with DK [v]. When restricted to their native orthography, the DK listeners' responses did indeed suggest that /w/-/v/

would be assimilated in part to overlapping or the same L1 category, but this masked their ability to attend to the phonetic *differences* between [v] and [w], which they demonstrated when given the opportunity to use the foreign grapheme W. Future research will be needed to determine whether TC-assimilation of [w]-[v] by DK listeners is due to fairly high English proficiency in Denmark, or whether common native strings with syllable-initial /uV/ aid DK listeners in discerning [v] and [w].

The only discrimination difference between the two listener groups was seen for /ð/-/v/, which DK listeners discriminated less well than IT listeners. PAM predicted TC assimilation for DK listeners, but suggested, along with SLM, possibly somewhat reduced discriminability due to the phonetic differences between /v/ and /ð/ and any DK counterparts. The DK assimilation pattern for /v/ and /ð/ showed some overlap, but it was less than (with DK orthography) or equal to (with English orthography) that for /w/-/v/, which DK listeners discriminated better than /ð/-/v/. Thus, the assimilation pattern for /ð/-/v/ fails to clearly predict reduced discriminability relative to the other contrasts. One possible reason for this surprising finding could be the auditory similarity of [ð] and [v]. Another could be the relatively poor phonetic fit of the non-native phones to each native category [4]. Either or both of these factors could perceptually hamper discrimination of this apparently TC-assimilated contrast by Danes.

5. REFERENCES

- [1] Best, C.T. 1995. A direct realist perspective on cross-language speech perception. In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium MD: York Press, 167-200.
- [2] Best, C.T., Tyler, M.D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, O.-S., Munro, M.J. (eds.), *Language Experience in Second language Speech Learning. In Honor of James Emil Flege*. Amsterdam: John Benjamins, 13-34.
- [3] Flege, J.E. 1995. Second language speech learning: Theory, findings, and problems. In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium MD: York Press, 233-277.
- [4] Hallé P.A., Best, C.T., Levitt, A. 1999. Phonetic versus phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics* 27, 281-306.

¹ Because we need to distinguish between abstract *phonological* units and contrasts versus *phonetic* details of realization, we heuristically use / / to represent the former, and [] to represent the latter.