PREDICTING NONNATIVE CONSONANT IDENTIFICATION: THE CASE OF ENGLISH CONSONANTS AS PERCEIVED BY NATIVE SPEAKERS OF KALAALLISUT (WEST GREENLANDIC)

Ocke-Schwen Bohn[1] and Kulunnguaq Korneliussen[2]

[1]Aarhus University, Denmark; [2]National Advisory Center, Nuuk, Greenland
[1]ocke.bohn@cc.au.dk; [2]kkor@aqqut.gl

ABSTRACT

We report two experiments which examine the perception of English initial consonants by native speakers of Kalaallisut (West Greenlandic). English allows 23 consonants in initial position, with voicing contrasts for all obstruents, whereas Kalaallisut has 12 initial consonants with no voicing contrasts. We generated predictions for the identification of English consonants from a comparison of the consonant inventories (initial position) of the two languages, and from Experiment 1, in which 12 native speakers of Kalaallisut perceptually assimilated English consonants, presented in [Ca] syllables, to their native inventory. The predictions were tested in Experiment 2, in which the participants of Experiment 1 identified the same consonants using English labels. Neither the comparison of the consonant inventories nor the results of the perceptual assimilation experiment account satisfactorily for the patterns of (mis-) identification observed in Experiment 2. We conclude our presentation with a discussion of factors which contribute to the observed patterns.

Keywords: Cross-language consonant perception, West Greenlandic, perceptual assimilation.

1. INTRODUCTION

A fairly large number of studies have been inspired by Best's Perceptual Assimilation Model (PAM) [1] to predict the level of discrimination of nonnative contrasts from cross-language perceptual assimilation patterns (e.g., [5, 8]). However, even though PAM was originally developed to predict cross-language *discrimination* by naïve listeners, this model has very rarely been used and adapted to predict the *identification* of nonnative speech sounds, with [10] being a recent exception.

The present study applies the predictions for cross-language identification as generated in [10] (and inspired by PAM predictions for discrimination) to the identification of English initial consonants by native (L1) speakers of Kalaallisut (West Greenlandic). These languages differ in that English has a fairly large consonant inventory (23 in initial position) in which all obstruents differ in voicing, whereas Kalaallisut has a smaller initial consonant inventory (12) with no voicing contrasts.

Table 1: Kalaallisut initial consonants.

	Plosive	Nasal	Fricative	Approx	Lateral
Labial	p	m		ʋ	
Alveolar	t	n			l ~ ɬ
Alv.-pal			ɕ		
Palatal					
Velar	k		ɣ	j	
Uvular	q		ʁ		

Table 2: English initial consonants.

	Plosive	Affr	Nasal	Fricat	Appr	Lateral
Labial	pʰ p		m	f v	w	
Dental				θ ð		
Alveolar	tʰ t		n	s z	ɹ ~ ɻ	l ~ ɫ
Alv.-pal		tʃ dʒ		ʃ ʒ		
Palatal					j	
Velar	kʰ k					
Glottal				h		

Specifically, Kalaallisut plosives are short-lag unaspirated, whereas English has aspiration contrasts. Only English has affricates, and it has a rich fricative inventory. The inventory of initial nasals does not differ between the languages, and the approximant inventories differ only with respect to the phonetic realization of /r/, and in that English has the velarized bilabial [w], whereas Kalaallisut has a labiodental [ʋ].

This leads one to expect that L1 Kalaallisut speakers would encounter problems especially in the identification of English stops (with respect to the voicing contrasts) and fricatives and affricates with respect to both voicing and place of articulation. The perception of English sonorants, however, could be expected to be comparatively unproblematic because both languages share [m, n, j, l ~ ɬ], but differ with respect to the phonetic realization of /r/ and with respect to the place of articulation of the approximants [w] (English) and [ʋ] (Kalaallisut).

As discussed and demonstrated by, e, g, [4], comparisons of sound inventories are, at best, a first step in attempts to predict cross-language perception. In an attempt to come up with more specific and testable predictions regarding the identification of English consonants by L1 Kalaallisut listeners, we conducted Experiment 1 in which L1 Kalaallisut listeners perceptually assimilated English consonants to Kalaallisut categories. The results from Experiment 1 were then used to generate predictions, based on the general hypotheses presented in [10], for the identification of English consonants by L1 Kalaallisut listeners (Experiment 2).

3. EXPERIMENT 1

Experiment 1 examined the perceptual assimilation of English initial consonants to Kalaallisut. We used the results from Experiment 1 to generate predictions for L1 Kalaallisut listeners' identification accuracy of English initial consonants. Our specific predictions are derived from the general predictions for the identification of nonnative speech sounds from assimilation patterns as developed in [10] and presented below. These are related, respectively, to PAM's assimilation types a) Two-Category, b) Single Category, c) Category Goodness, and d) Uncategorized:

a) L2 phones which have a unique match in an L1 category irrespective of the goodness of fit will be accurately identified by all L2 learners, irrespective of L2 experience.
b) Assimilation of two or more L2 phones to one L1 category with comparable goodness ratings will result in inaccurate identification of these phones, irrespective of L2 experience.
c) Assimilation of two or more L2 phones to one L1 category with different goodness ratings will result in accurate identification of the better match(es) irrespective of L2 experience, and in identification accuracy depending on L2 experience for the poorer match(es).
d) The identification accuracy of L2 phones which are not consistently assimilated to any L1 category (assimilation frequency of < 50%, see [6]) will depend on L2 experience.

The predictions a), b), and the first part of c) can be tested with any nonnative listeners. Testing the second part of prediction c) and prediction d) requires listener groups differing in L2 experience. Participant availability for the present study did not allow for testing the effect of English language experience.

3.2. Methods

Twelve L1 speakers of Kalaallisut (6 f, 6 m; mean age: 22.7 years) participated. All participants were fluent speakers of their second language, Danish, and none had spent any extended period of time in an English-speaking country.

The stimuli were drawn from the Shannon et al. (1999) corpus [11] as used by the Alvin3 software [9]. Twenty-one tokens each of [Ca] syllables with [C] = [p^h, p, w, f, v, θ, ð, t^h, t, s, z, ʃ, ʒ, k^h, k, ɹ, ɫ, tʃ, dʒ, h, j, h], produced by two L1 English speakers (1f, 1m) were presented in two randomizations for a total of 84 trials with praat's *ExperimentMFC* [3].

Participants were presented with each stimulus over high-quality headphones and they selected, via mouse click, one of 11 orthographically presented response alternatives corresponding to graphemes used for Kalaallisut, i.e., *pa, ta, ka, qa, fa, sa, ga, ra, va, la, ja*. (Selection of these alternatives was based on piloting in which participants used *ka* and *ga*, and *fa* and *va*, even though the graphemes in each pair do not correspond to different Kalaallisut categories.) Immediately after selecting the Kalaallisut category corresponding to the English consonant, participants indicated the goodness of their match on a 5-point Likert scale, whose endpoints were labeled (in Kalaallisut) as *Ajorpoq* ("bad") and *Ajunngilaq* ("good"). Instructions were provided in Kalaallisut by the second author, a native Kalaallisut speaker.

3.3. Results

Tables 3a-e present the results of the perceptual assimilation (PA) task in terms of fit indices, which are derived by multiplying the proportions of identifications with the goodness rating for this match (see [8]). For example, the proportion of English [p] matched with Kalaallisut <p> was 0.583, with a mean goodness rating of 4.8, resulting in a fit index of (0.583 x 4.8 =) 2.8. In the tables below, cells of interest are highlighted with bold frames. Cells with significantly different fit indices are separated by a bold line, and cells with Single Category assimilations are not separated by a line.

Table 3a: Fit indices for English initial plosives to Kalaallisut categories.

EN Stm	Kalaallisut response				
	p	v	t	K	g
p^h	2.7				
p	2.8	1.5			
t^h			2.8		
t			4.6		
k^h				3.1	0.3)
k				3.0	0.5

Table 3a suggests problems for the identification of English [p^h, p] and [k^h, k], which are each assimilated to just one Kalaallisut category with comparable fit indices, and it predicts that [t] will be identified more correctly than [t^h] because of a significant difference in fit indices for [t] (high) vs [t^h] (low).

Table 3b: Fit indices for English initial approximants to Kalaallisut categories.

EN Stm	Kalaallisut response			
	v	l	j	r
v	4.1			
w	3.3			
ɫ	0.4	3.8		
ɻ				1.7
j			4.4	

Table 3b suggests identification problems for [v, w], which are assimilated to Kalaallisut /v/ without a significant difference in goodness of fit. [j, ɻ, ɫ] are predicted to be identified accurately because they are each uniquely assimilated to Kalaallisut categories.

Table 3c: Fit indices for English initial (labio-) dental fricatives to Kalaallisut categories.

EN Stm	Kalaallisut response		
	f	v	t
f		4.3	
v		4.1	
θ	3.1		0.3
ð		0.8	1.5
t^h			2.8
t			4.5
ʧ			1.3
ʤ			1.1

Table 3c suggests low identification accuracy for English [f, v], and higher identification accuracy for [θ] than [ð].

Table 3d: Fit indices for English initial sibilants and affricates to Kalaallisut categories.

EN Stm	Kalaallisut response			
	t	s	k	g
s		4.3		
z		2.6		
ʃ		2.4		
ʒ		1.2	0.3	0.4
ʧ	1.3	0.5	0.3	0.4
ʤ	1.1	1.2		0.3

Table 3d suggests higher identification accuracy for [s, ʃ] than their voiced counterparts [z, ʒ], respectively. The affricates [tʃ, dʒ] will not be identified correctly.

Table 3e: Fit indices for English [h] to Kalaallisut categories.

EN Stm	Kalaallisut response			
	p	v	j	g
h	0.2	0.1	0.4	0.6

Table 3e suggest that [h] is uncategorized for Kalaallisut listeners. Given their extensive experience with L2 Danish, which has a /h/ category, we predict that [h] will be identified well.

4. EXPERIMENT 2

Experiment 2 examined the identification of English initial consonants by the same L1 Kalaallisut listeners as in Experiment 1, using the same stimuli as in Experiment 1. Each stimulus was presented, in random order, three times each, resulting in (2 talkers x 23 consonants x 3 randomizations =) 138 trials. Stimuli were presented via high-quality headphones using the Alvin3 software [9]

4.1. Results and discussion

The results of Experiment 2 and their relation to the predictions are presented separately for each class of consonants in Tables 4 a-e. The tables list the stimuli on the vertical axis and the responses on the horizontal axis. Only responses of > 10% are listed.

Table 4a: Identification of English stop consonants by L1 Kalaallisut listeners (rounded % correct).

	p^h	p	t^h	t	k^h	k	v
p^h	94	6					
p	14	50					26
t^h			67				
t				81			
k^h					90	6	
k					39	61	

The prediction from the comparison of the inventories was that identification accuracy would be compromised because Kalaallisut obstruents do not differ in voicing. The PA experiment suggested that [p^h, p] and [k^h, k] would be less accurately identified than [t]. Contrary to predictions, the aspirated [p^h, k^h] are identified highly and more accurately than [t].

Table 4b: Identification of English approximants by L1 Kalaallisut listeners.

	w	v	ɫ	ɻ	j
w	71	29			
ɫ			91		
ɻ				97	
j					97

As predicted from the PA experiment and, in part, from the comparison of the inventories, the identification accuracy for [ɫ, ɹ, j] was high. These approximants were each assimilated uniquely to an L1 category. However, [w] was identified less accurately, which was predicted because [w] and [v] were both assimilated with comparable goodness ratings to Kalaallisut <v>.

Table 4c: Identification of English (labio-) dental fricatives by L1 Kalaallisut listeners.

	f	v	θ	ð	t	w
f	96					
v		70				30
θ	39		36	18		
ð			33	61		

The comparison of the inventories resulted only in the general prediction that identification of English fricatives and approximants would be difficult because of the voicing contrasts and comparably large number of places of articulation. - Contrary to predictions from the Experiment 1, [f] was identified highly, and [v] somewhat accurately even though both were assimilated with comparable goodness ratings to L1 <v>. Identification rates for both [θ] and [ð] were low. This was expected for [ð] but not for [θ], which was assimilated uniquely to <f>.

Table 4d: Identification of English sibilants and affricates by L1 Kalaallisut listeners.

	s	z	ʃ	ʒ	ʧ	ʤ
s	73		24			
z	38	33	29			
ʃ	12		79	6		
ʒ			62	10	5	
ʧ			20		39	15
ʤ					35	24

As expected, identification rates for the voiceless sibilants [s, ʃ] were higher than for their voiced counterparts [z, ʒ] because [s, ʃ] were assimilated with significantly higher fit indices to one native category than their voiced counterparts. As expected, the identification rates for the affricates were low because both [ʧ] and [ʤ] were assimilated to one native category with comparable fit indices. – Finally, [h] was identified highly accurately at 98%, consistent with its uncategorized assimilation pattern.

5. CONCLUSION

The present study examined the perception of the relatively large initial consonant inventory of English by listeners with an L1 (Kalaallisut) which has a comparably small consonant inventory. Predictions for the accuracy of English consonant identification by L1 Kalaallisut listeners were derived from a comparison of the inventories of the two languages and from a cross-language mapping experiment, in which L1 Kalaallisut listeners perceptually assimilated English initial consonants to native categories.

The comparison of the inventories was a first step in an attempt to predict identification accuracy by the nonnative listeners. The predictions were, in part, necessarily vague (for fricatives) and mostly unsuccessful in identifying identification problems (except for some of the approximants).

We had expected that perceptual assimilation patterns would provide valid predictions for cross-language identification accuracy because current speech learning models [2, 7] assume that nonnative speech perception is largely guided by how the sounds of the L2 are perceptually mapped on to the L1.

This expectation was only met, to some extent, for the identification of English approximants, sibilants, and the glottal fricative by L1 Kalaallisut listeners. The predictions for the identification of plosives and (labio-)dental fricatives were not borne out. We are currently unable to account for these unexpected results. The most puzzling result was that the aspirated [p^h, k^h], but not [t^h], were more accurately identified than their unaspirated counterparts, even though Kalaallisut has only unaspirated stops. Familiarity with the L2 Danish of our participants cannot help to explain this result, but it may be the reason for the surprisingly accurate identification of [f, v] since Danish has the very similar [f - ʋ] contrast. Several inaccurate predictions were derived from perceptual assimilations which differed in category goodness, and we suggest that the ability of nonnatives to detect a difference in goodness of fit to an L1 category is a necessary, but not sufficient prerequisite for accurate identification. Finally, the low identification accuracy for the dental fricatives could be due to confusable response labels for [θ, ð], i.e., *thigh* and *thy*.

We hope that future studies will be able to further explore the relation between the perceptual assimilation of nonnative sounds to native categories on the one hand and identification accuracy of these nonnative sounds on the other.

6. REFERENCES

[1] Best, C. T. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed.) *Speech Perception and Linguistic Experience*. York Press, 171-206.

[2] Best, C. T., Tyler, M. D. 2007. Nonnative and second-language speech. In Bohn, O.-S., Munro, M. J. (eds.) *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, 17, 13.

[3] Boersma, P., Weenink, D. 2016. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.14. Online: http://www. praat. org/retrieved.

[4] Bohn, O.-S. 2002. On phonetic similarity. In Burmeister, P. Piske, T., Rohde, A. (eds.), *An Integrated View of Language Development: Papers in Honor of Henning Wode.* Wissenschaftlicher Verlag Trier, 191-216

[5] Bohn, O.-S, Best, C. T., Avesani, C., Vayra, M. 2011. Perceiving through the lens of native phonetics: Italian and Danish listener's perception of English consonant contrasts. *Proc. 17th ICPhS* Hong Kong, 336-339.

[6] 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. *Appl. Psycholing.* 32, 51-67.

[7] Flege, J. E., Bohn, O.-S. 2021. The revised Speech Learning Model (SLM-r). In: Wayland, R. (ed.) *Second Language Speech Learning: Theoretical and Empirical Progress*. Cambridge, UK: Cambridge University Press, 3-83.

[8] Guion, S. G., Flege, J. E., Akahane-Yamada, R., Pruitt, J. C. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *J. Acoust. Soc. Am*. 107, 2711-2724.

[9] Hillenbrand, J., Gayvert, R. T. 2005. Open source software for experiment design and control. *J. Speech, Language, and Hearing Res.* 48, 45-60.

[10] Horslund, C. S., Bohn, O. S. (2022). Assimilation patterns predict L2 identification accuracy of English initial consonants. *J. Second Language Pronunciation*, 8, 218-247.

[11] Shannon, Robert V., Jensvold, Angela, Padilla, Monica, Robert, Mark E., Wang, Xiaosong. 1999. Consonant recodings for speech testing. *J. Acoust. Soc. Am*. 106, L71-L74.