

# Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages

Catherine T. Best<sup>1,2</sup>, Pierre Halle<sup>3</sup>, Ocke-Schwen Bohn<sup>4</sup>, and Alice Faber<sup>2</sup>

<sup>1</sup> Wesleyan Univ. (USA), <sup>2</sup> Haskins Labs (USA), <sup>3</sup>LPE/CNRS-Paris5 (France), <sup>4</sup>Aarhus Univ. (Denmark)

E-mail: cbest@wesleyan.edu, halle@psycho.univ-paris5.fr, engosb@hum.au.dk, faber@haskins.yale.edu

## ABSTRACT

Several theoretical models predict that nonnative speech discrimination depends on phonetic fit as well as phonological correspondence to native phonemes. Languages differ both phonologically and phonetically, e.g., high vowels differ among English, French, Danish, Norwegian. All use /i u/, all except English use /y/, but only Norwegian has /ʉ/. These languages realize /i y u/ differently. American, Danish and French listeners categorized and discriminated Norwegian /i/-/y/, /y/-/u/, /y/-/ʉ/, /ʉ/-/u/. Danes assimilated the first three to native contrasts, /ʉ/-/u/ as an /y/ goodness difference, and discriminated all near ceiling. French listeners assimilated /y/-/u/, /y/-/ʉ/, /ʉ/-/u/ to native contrasts, Americans assimilated /y/-/u/ and /y/-/ʉ/ to native contrasts, /ʉ/-/u/ to allophones of native /u/. They discriminated those contrasts near ceiling. French listeners assimilated /i/-/y/ as an /i/ goodness difference, discriminating it worse than Danes but better than Americans, who assimilated it as equally-good /i/s. Results coincide with the languages' phonological and phonetic properties.

## 1. INTRODUCTION

Nonnative speech perception is influenced not only by the phonological system of the listener's native language (L1), but also by experience with the fine-grained phonetic details of L1 phonemes [1, 2]. Thus, data on perception of the same nonnative phonological contrasts by listeners of languages that differ systematically on both phonological and phonetic dimensions could provide insights about the relative roles of contrastive and non-contrastive phonetic properties in listeners' knowledge of the L1 sound system. Findings on this issue could inform phonological theory by illuminating the extent to which speech perception is constrained by the L1's contrastive phonology as defined by abstract phonetic features, versus guided by its concrete articulatory patterns [3]. Such findings could also provide an index of listener sensitivity to non-contrastive phonetic variation within native phonemic categories.

Two prior studies have systematically investigated the comparative effects of L1 phonological contrasts and of L1 phonetic realizations on perception of nonnative consonant contrasts. Both examined categorical perception of the American English approximant consonant contrasts /r l/, /w r/ and /w j/. One tested Japanese listeners [2], the other, French listeners [1]. Although Japanese lacks an /r l/ phonological contrast, it does have an /r/ and contrasts

/w/-/r/ and /w/-/j/. Notably, however, the Japanese /r/ and /w/ are realized differently (tapped [ɾ] and unrounded velar approximant [ʋ], respectively) than the corresponding AE phonemes (liquid [ɹ] and semivowel [w]). French, on the other hand, has all three phonological contrasts, but its /r/ is realized differently (uvular fricative [ʁ] or trill [ʀ]) than AE /r/. The French /l/ (non-velarized/light [l]) is also somewhat different from the AE velarized/dark [ɫ].

Japanese listeners who were relatively inexperienced with spoken English, not surprisingly, had notable difficulty categorizing and discriminating the /r/-/l/ continuum, but not /w r/ and /w j/ continua. However, consistent with the difference between English and Japanese /w/ and /r/, their category boundary for /w r/ differed significantly from that of Americans' [1]. On the other hand, even though French has all three phonological contrasts, the difference between the French and AE /r/s led French listeners to display marked difficulties with the /w r/ continuum [2]. Thus, in both studies, not only the native language's phonological structure, but also its non-contrastive phonetic details strongly affected perception of both nonnative contrasts and nonnative realizations of native phonemes.

Nonnative vowel contrasts may be particularly useful for extending the investigation of L1 phonological versus phonetic influences on perception [4]. Numerous vowel characteristics suit them especially well for examining this issue. Vowels are usually higher in intensity and longer in duration than consonants. They also involve different articulatory gestures (e.g., less vocal tract constriction, different tongue muscles) than consonants. Normally, vowels are voiced throughout whereas many consonants have some aperiodic noise, often involving a non-glottal source. Of special interest to the present issue, the phonological inventories of most languages include many fewer vowels than consonants, which could influence the nature of contrastivity between vowels. Perhaps relatedly, the articulatory and acoustic properties of a given vowel can vary greatly among languages and even among dialects. Interestingly, while isolated vowels may be less categorically perceived than consonants, identification of vowels in context depends more on their dynamic properties than on their quasi-steady-state nuclei [5], consistent with the Dynamic Specification theory of vowel perception [6]. Vowels vary among languages and dialects not only in their "target" (nucleus) formant values but also in their dynamic properties, e.g., diphthongization.

The experiment reported here examined L1 phonological

and phonetic influences on perception of contrasts among Norwegian high vowels by listeners of American English, French, and Danish. These four languages have large but distinct vowel systems. Norwegian has four high vowels: front unrounded /i/, out-rounded /y/, in-rounded /ɥ/ and back rounded /u/. French and Danish share three of those vowels (/i y u/), phonologically speaking, whereas English shares only two of them (/i u/). The distribution of vowels differs among the listener languages, being weighted heavily toward the upper half of the vowel space in Danish, and more evenly distributed in French and English. However, English, unlike the other three languages, is lacking in front rounded vowels. We tested perception of four Norwegian vowel contrasts (/i-/y/, /y-/ɥ/, /y-/u/, /u-/ɥ/) by American, French and Danish listeners.

There are also notable differences in the phonetic realizations of the shared vowels across these four languages. Norwegian (and Danish) /y/ is more fronted and less lip-protruded than French /y/, which may be more similar in tongue and lip configuration to the Norwegian /ɥ/. English /u/ is more fronted than in the other three languages, especially the advanced [ɥ] allophone that occurs in coronal consonant contexts (/t d s z/). The vowels /i/ and /u/ are diphthongized in American English, whereas Danish and Norwegian vowels are monophthongs, as are most French vowels including /i u/. English vowels are also subject to phonotactic constraints not found in the other languages. Specifically, the English lax vowels (ɪ ɛ ʊ ʌ æ/ generally do not occur in open syllables, but all vowels can occur in open syllables in Danish and Norwegian, and nearly all in French. While this constraint is not directly involved in our stimuli, it could limit English listeners' assimilation possibilities for nonnative /y ɥ/. To maximize the impact of cross-language phonetic and phonotactic differences in perception of the Norwegian vowel contrasts, the stimulus vowels were recorded and presented in open syllables with a coronal consonant onset (/sV/).

Several models of nonnative speech perception suggest possible differences among the listener groups' categorization and discrimination of the Norwegian vowel contrasts. The Speech Learning Model (SLM) focuses on variations in the ease with which different native (L1) phonemes can be acquired by second language (L2) learners [7]. SLM, posits that the relative difficulty of acquiring an L2 phoneme depends on its degree of similarity to the closest native one. An L2 phoneme may be either identical or similar to a native one, or it may be new, i.e., dissimilar from any L1 phoneme. Forming an L2 category should be easy for new phonemes but difficult for similar ones, and unnecessary for identical ones. Based on these principles, we could infer that SLM would expect Norwegian /i u/ to be identical or similar to /i u/ in each of the listener languages, and therefore perceived the same way by the three listener groups. The same prediction would hold for perception of Norwegian /y/ by French and Danish listeners, but /y/ should serve as a new phoneme for English listeners. Because it would be new for English listeners, and similar to native /y/ for Danish and French listeners, Norwegian /i-/y/ and /y-/u/ should be discriminated by all

three groups, though probably less well by naïve English speakers because they are not experienced L2 learners of Norwegian /y/. Conversely, /ɥ/ should be similar to French and Danish /y/, thus Norwegian /y-/ɥ/ should be difficult for those groups to discriminate. For English listeners, however, the story for /ɥ/ is less clear. It may be perceived as a new phoneme, but is more likely to be similar to the fronted English /u/, especially in the coronal consonant context used in this study (/sV/). If it is indeed heard as similar to English /u/, then Americans should easily discriminate /y-/ɥ/ but should have difficulty with /u-/ɥ/. SLM also assumes that native phonotactic constraints have an impact on perception of nonnative phonemes, and has provided evidence favoring such influences [8]. Thus, English listeners should be less likely to hear the nonnative /y/ as any English lax vowel because of the open syllable context of the present stimuli. This should increase the likelihood that /y/ will be heard as a new vowel and easily discriminated from the other three vowels (this would hold as well for /ɥ/, if it is heard as new rather than as being similar to English /u/).

The Native Language Magnet model (NLM) [9] posits that exposure to a native phoneme leads to the formation of a category prototype, which acts like a perceptual magnet, making discrimination more difficult within the vicinity of the prototype than of a non-prototype. Nonnative categories lack this perceptual prototype structure for naïve listeners. Thus, NLM should posit that if a nonnative vowel is acoustically similar to a native vowel, it will show a prototype, or perceptual magnet, effect like the native vowel. If it is not acoustically similar to any native vowel, it should behave like a non-prototype and be more easily discriminable from surrounding vowels. Because NLM's central hypothesis is that there is a single "ideal" or prototype of a native phoneme, it doesn't consider context effects on vowels or phonotactic constraints as potential influences on nonnative speech perception. Thus, we need not consider any such influences in generating NLM predictions. The implications of NLM for our investigation would seem to be that Norwegian /i u/ should behave like native prototypes for all three listener groups; /y/ should likewise behave as a native prototype for Danish listeners, but as a non-prototype of /u/ for English listeners. Norwegian /ɥ/ should be another non-prototype of /u/ for English listeners and a non-prototype of /y/ for Danish listeners. For French listeners, either Norwegian /y/ or /ɥ/ should fit their native /y/ prototype, the other being a non-prototype of the same vowel. But what do these characterizations predict about discrimination of the Norwegian vowel contrasts by the three listener groups? We can infer that prototypes of two different native vowels will be discriminated better than two non-prototypes of a single native vowel, which should in turn be discriminated better on average than a prototype versus a non-prototype of the same native vowel. Therefore, discrimination should be highest for Danish and French listeners on /i-/y/ and /y-/u/ (two prototypes), good but lower for English listeners on /y-/ɥ/ (two non-prototypes), lower still for Danish and French listeners on /y-/ɥ/ and for all three groups on /u-/

/ʉ/ (prototype vs. non-prototype).

The Perceptual Assimilation Model (PAM) [10] hypothesizes that listeners assimilate nonnative phones to the native phonemes that are perceived to be the most similar articulatorily. Discrimination of a nonnative contrast depends on whether its members are assimilated to the same or different L1 phonemes, and on their phonetic goodness of fit to the native categories. Discrimination is posited to be excellent for nonnative contrasts assimilated to two native phonemes, good but lower for contrasts assimilated as showing a category goodness difference in fit to the same native phoneme, and poor for contrasts assimilated as equivalent in fit to a single native phoneme. PAM hasn't overtly addressed allophonic and phonotactic effects, but is implicitly open to them because of its assumptions about articulatory gesture relationships, which are highly influenced by phonetic context. (The notion that assimilation and discrimination are influenced by the gestural dynamics of native and nonnative vowels is also consistent with the Dynamic Specification theory of vowel perception [6].) Based on these considerations, PAM predicts that English listeners will assimilate Norwegian /i u/ to native /i u/ with less than perfect fit because of the difference in diphthongization and fronting/backing of these vowels in the two languages. They should assimilate /ʉ/ in the target /sV/ syllables to the advanced allophone of English /u/ ([ʊ]) that occurs in coronal contexts (e.g., DUDE) rather than the plain allophone of /u/ ([u]) in non-coronal contexts (e.g., COOP). Norwegian /y/ may be heard as imperfect English /i/ rather than /u/ because of the Norwegian /i/'s strongly fronted tongue position and relatively non-protruded lip-rounding. Thus, they may discriminate /i/-/y/ rather poorly (imperfect fits to /i/), /u/-/ʉ/ substantially better (allophonic difference within /u/), and /y/-/u/ and /y/-/ʉ/ quite well (assimilated to native /i/-/u/). Danish and French listeners should assimilate Norwegian /i u/ as fairly good exemplars of native /i u/ (all monophthongs, highly fronted/backed). Danish listeners should assimilate Norwegian /y/ and/or /ʉ/ as a good Danish /y/, and the other as less-good /y/. French listeners should assimilate /ʉ/ to French /y/, which is more lip-protruded and less fronted than Norwegian /y/. They may assimilate Norwegian /y/ as either a poor French /y/ (insufficient lip-protrusion) or, perhaps more likely, as an imperfect /i/. Therefore, Danish listeners should discriminate /i/-/y/, /y/-/u/ and /u/-/ʉ/ excellently (as native contrasts /i/-/y/, /y/-/u/ and /u/-/y/, respectively). They should discriminate /y/-/ʉ/ well but less well (goodness difference within native /y/). French listeners should discriminate /y/-/u/, /y/-/ʉ/ and /u/-/ʉ/ excellently (as native /i/-/u/, /i/-/y/ and /u/-/y/), but /i/-/y/ somewhat less well (goodness difference in native /i/).

## 2. METHOD

A native Norwegian speaking male was recorded producing multiple tokens of the vowel contrasts /i/-/y/, /y/-/ʉ/, /ʉ/-/u/, and /y/-/u/ in each of four syllable contexts. We report here only on the perceptual findings for the /sV/ stimuli. Four tokens each of /si/, /sy/, /su/ and /sʉ/ were

selected for perceptual tests, matched as closely as possible on non-critical acoustic properties (fricative duration, intensity and spectrum, vowel intensity). The tokens were waveform-edited to equalize syllable intensities and to adjust the range of syllable durations ( $M_{\text{modified}} = 784$  ms, range = 676-899 ms). Our speaker verified the identity of the final tokens. (See Table 1 for formant values of stimuli.)

	/si/	/sy/	/sʉ/	/su/
<b>F1</b>	269	269	296	362
<b>F2</b>	2214	2267	1658	951
<b>F3</b>	3008	3376	3156	3444

Table 1. Mean formant values at mid-vowel.

The participants were 16 native speakers each of northeast American English and of western Danish, and 24 native speakers of Parisian French. All were college students tested in their native language and country; all had normal hearing and speech/language/reading abilities. None had experience with Norwegian. The Americans had no experience with any languages that have front-rounded vowels.

Listeners completed categorial AXB discrimination tests involving the multiple tokens of each syllable, for each stimulus contrast [for procedural details: 11]. They then completed a categorization task on the vowels of all stimulus tokens, judged with respect to native vowels presented in a list of native keywords [4]. Following categorization, they rated the token's goodness of fit to the native vowel they chose (1=poor match, 5=excellent match). All tokens were presented multiple times in random order.

## 3. RESULTS

The discrimination data were analyzed by a language (3) x contrast (4) ANOVA. The significant language effect,  $F(2, 53) = 7.421$ ,  $p < .002$ , indicated that Danish listeners performed best overall (98% correct), followed by French (96%), and then American listeners (92%). The contrast effect,  $F(3, 159) = 52.64$ ,  $p < .0001$ , showed that discrimination of /i/-/y/ was worse overall (88.5%) than for the other three contrasts (97-99% correct). The most informative finding was a significant language x contrast interaction,  $F(6, 159) = 11.85$ ,  $p < .0001$ . Danish listeners discriminated all four contrasts near ceiling (97-98% correct). French listeners discriminated /y/-/u/, /y/-/ʉ/ and /ʉ/-/u/ near ceiling (98-99% each), as did Americans (96-98%). The French listeners discriminated /i/-/y/ significantly less well (89%), and Americans showed substantially lower discrimination than the French (79%).

The categorization and rating data (Table 2) show that all listener groups assimilated Norwegian /y/-/u/ to a two-category native contrast (possibly, for some listeners, to a native versus an uncategorized vowel). Danish and French listeners also assimilated /u/-/ʉ/ as a two-category or categorized-uncategorized contrast. The Danish also assimilated /i/-/y/ to two categories, but American and French listeners assimilated it to the single category /i/. However, French listeners reported a sizeable difference in goodness of fit (i.e., category goodness difference assimilation),

whereas Americans gave a much smaller rating difference (single-category assimilation). Norwegian /y/-/ʏ/ was assimilated as a two-category native contrast by both the French and American English listeners. Danish listeners assimilated it primarily to a native long /y/ (also, less often and less well, to its short cognate /ʏ/), though there is a suggestion of a category goodness difference in that /y/ appeared to sound somewhat less rounded to them than /ʏ/ (inferred from /i ɪ/ assimilations and lower /y ʏ/ ratings).

<u>/si/ assimilation:</u>		/i/	/ɪ/	/y/	/ʏ/	[ɥ]	/u/	/ʊ/	/o/	/ɔ/
<b>English</b>	<i>M</i> %	<b>93</b>								
	<i>M</i> <sub>rating</sub>	3.4								
<b>French</b>	<i>M</i> %	<b>99</b>								
	<i>M</i> <sub>rating</sub>	4.0								
<b>Danish</b>	<i>M</i> %	<b>66</b>	<b>32</b>							
	<i>M</i> <sub>rating</sub>	3.4	3.3							
<u>/sy/ assimilation:</u>										
<b>English</b>	<i>M</i> %	<b>93</b>								
	<i>M</i> <sub>rating</sub>	3.1								
<b>French</b>	<i>M</i> %	<b>97</b>								
	<i>M</i> <sub>rating</sub>	2.9								
<b>Danish</b>	<i>M</i> %	<b>8</b>	<b>9</b>	<b>58</b>	<b>23</b>					
	<i>M</i> <sub>rating</sub>	2.9	3.2	3.2	2.7					
<u>/su/ assimilation:</u>										
<b>English</b>	<i>M</i> %					<b>62</b>	<b>29</b>			
	<i>M</i> <sub>rating</sub>					3.1	3.0			
<b>French</b>	<i>M</i> %			<b>96</b>						
	<i>M</i> <sub>rating</sub>			3.2						
<b>Danish</b>	<i>M</i> %			<b>66</b>	<b>29</b>					
	<i>M</i> <sub>rating</sub>			3.4	3.0					
<u>/su/ assimilation:</u>										
<b>English</b>	<i>M</i> %					<b>38</b>	<b>36</b>		<b>8</b>	<b>11</b>
	<i>M</i> <sub>rating</sub>					3.0	3.0		2.8	3.6
<b>French</b>	<i>M</i> %						<b>53</b>		<b>39</b>	<b>8</b>
	<i>M</i> <sub>rating</sub>						3.0		2.7	2.1
<b>Danish</b>	<i>M</i> %			<b>6</b>			<b>19</b>	<b>9</b>	<b>53</b>	<b>7</b>
	<i>M</i> <sub>rating</sub>			3.1			3.2	3.1	3.1	3.0

**Table 2:** Mean percent assimilations (> 5%) and ratings. Note: [ɥ] = advanced English /u/ in coronal contexts (Americans could choose DUDE = [ɥ] or COOP = [u]).

#### 4. CONCLUSIONS

The results are quite compatible with the phonologically contrastive and non-contrastive phonetic-articulatory properties of the listener L1s. The pattern of findings within and across groups is most consistent with PAM predictions, though less than perfectly so. One unresolved puzzle is why the Danish listeners discriminated the /y/-/ʏ/ contrast so well, given that they detected only a modest goodness difference, at best, in their assimilations to native /y ʏ/. However, this finding and the results as a whole are substantially less consistent with SLM and NLM predic-

tions, as best as we could infer them. In conclusion, both phonological and phonetic properties of the native language effect strong, systematic differences in nonnative vowel perception by listeners of varying L1s.

**Acknowledgment.** Supported by NIH (U.S.A.) grant DC00403 (C. Best).

#### REFERENCES

- [1] P.A. Hallé, C.T. Best and A. Levitt, "Phonetic versus phonological influences on French listeners' perception of American English approximants," *Journal of Phonetics*, vol. 27, p. 281-306, 1999.
- [2] C.T. Best and W. Strange, "Effects of phonological and phonetic factors on cross-language perception on approximants," *Journal of Phonetics*, vol. 20, 1992.
- [3] C.P. Browman and L. Goldstein, "Articulatory gestures as phonological units," *Phonology*, vol. 6, p. 201-251, 1989.
- [4] W. Strange, R. Akahane-Yamada, R. Kubo, S.A. Trent and K. Nishi, "Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners," *Journal of the Acoustical Society of America*, vol. 109, p. 1691-1704, 2001.
- [5] W. Strange, J.J. Jenkins and T. Johnson, "Dynamic specification of coarticulated vowels," *Journal of the Acoustical Society of America*, vol. 74, p. 695-704, 1983.
- [6] W. Strange, "Evolving theories of vowel perception," *Journal of the Acoustical Society of America*, vol. 85, p. 2081-2087, 1989.
- [7] J.E. Flege, "Second-language speech learning: Theory, findings, and problems", in *Speech perception and linguistic experience*, W. Strange, Ed. Timonium MD: York Press, 1995.
- [8] J.E. Flege, "Chinese subjects' perception of the word-final English /t/-/d/ contrast: Before and after training," *Journal of the Acoustical Society of America*, vol. 86, p. 1684-1697, 1989.
- [9] P. Kuhl and P. Iverson, "Linguistic experience and the "perceptual magnet effect"", in *Speech perception and linguistic experience: Issues in cross-linguistic research*, W. Strange, Ed., p. 121-154. Baltimore MD: York Press, 1995.
- [10] C.T. Best, "A direct realist perspective on cross-language speech perception", in *Cross-language speech perception*, W. Strange and J.J. Jenkins, Eds., p. 171-204. Timonium, MD: York Press, 1995.
- [11] C.T. Best, G.W. McRoberts and E. Goodell, "American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology," *Journal of the Acoustical Society of America*, vol. 109, p. 775-794, 2001.