



ELSEVIER

Contents lists available at SciVerse ScienceDirect

## Journal of Phonetics

journal homepage: [www.elsevier.com/locate/phonetics](http://www.elsevier.com/locate/phonetics)

# Native-language phonetic and phonological influences on perception of American English approximants by Danish and German listeners

Ocke-Schwen Bohn<sup>a,\*</sup>, Catherine T. Best<sup>b,c,1</sup>

<sup>a</sup> English Department and Center on Autobiographical Memory Research, Department of Psychology, Aarhus University, Aarhus, Denmark

<sup>b</sup> MARCS Auditory Laboratories, University of Western Sydney, Penrith South, Australia

<sup>c</sup> Haskins Laboratories, New Haven Connecticut, USA

## ARTICLE INFO

## Article history:

Received 25 August 2010

Received in revised form

8 August 2011

Accepted 9 August 2011

Available online 23 September 2011

## ABSTRACT

Perception of non-native consonant contrasts may be influenced by phonetic, as well as phonological, properties of the listener's native language. The impact of both factors on perception of American English /r l w j/ was investigated with native speakers of Danish and German, which have /r l j/ but lack /w/, thus employing /r/-/l/ but lacking /w/-/j/ and /w/-/r/ as phonological contrasts. However, while the three languages realize /j/ identically, Danish/German "light" alveolar [l] differs modestly from English "dark" [ɫ] (velarized), Danish pharyngeal and labiodental approximant realizations of /r, v/ are more similar to English /r, w/ than are German uvular and labiodental fricative realizations, and Danish is richer in approximants than English or German. Phonetic similarities perceptually outweighed phonological correspondences: Danish listeners' performance on /w/-/r/ and /r/-/l/ approached that of English speakers, and discrimination of /w/-/j/ was remarkably higher than English speakers', all largely irrespective of spoken English experience. German listeners' identification of all contrasts was highly categorical, but discrimination was poorer than English and Danish listeners' for /w/-/r/ and /r/-/l/ and fell in between those two groups for /w/-/j/. Thus, cross-language phonetic relationships among corresponding (or neighboring) phonemes strongly influence perception. Together with systemic consideration of English, Danish, and German vowel and approximant subsystems, our results indicate that non-native speech perception is affected not only by the phonological contrastiveness and phonetic realizations of the target phonemes in the listeners' language, but also by broader systemic factors such as phonological subclasses.

© 2011 Elsevier Ltd. All rights reserved.

## 1. Introduction

### 1.1. Background

It is by now well-established that adult listeners' perception of non-native consonant contrasts is systematically constrained by their native language experience. But which properties of the listener's language shape perception of unfamiliar speech contrasts? Is the source of perceptual constraints to be found in the abstract linguistic distinctions that comprise the infrastructure of the native phonological system, or is it traceable instead to the familiarity of the surface articulatory-phonetic patterning of native speech? Alternatively, if experience with *both* the abstract and the physical aspects of native speech combine in shaping the listener's response to non-native elements, how might these two types of information converge on perception? These are the

primary questions we addressed in novel comparisons of the perception of minimal pair distinctions among English approximants by native Danish and native German listeners. We begin by summarizing findings on the effects from both levels of native speech on the perception of these same contrasts by listeners of other non-native languages. From that vantage point we go on to consider several open issues that motivated our choice of Danish and German as the listener languages for the present investigation. We then discuss a range of potential outcomes for these two new groups, by comparison to those previous listener groups, from the viewpoint of current theoretical models of nonnative speech perception, which provides the rationale for the present paper.

Among the most frequently cited examples of dramatic native-language (L1) effects on perception is the difficulty that native Japanese and Korean listeners have in categorizing and discriminating either natural or synthesized tokens of the American English (AmE) approximant contrast /r/-/l/ in initial position (e.g., Gillette, 1980; Ingram & Park, 1998; MacKain, Best, & Strange, 1981; Miyawaki et al., 1975; Sheldon & Strange, 1982; Takagi & Mann, 1995; Yamada & Tohkura, 1992). A common claim

\* Corresponding author.

E-mail addresses: [engosb@hum.au.dk](mailto:engosb@hum.au.dk) (O.-S. Bohn), [c.best@uws.edu.au](mailto:c.best@uws.edu.au) (C.T. Best).

<sup>1</sup> Both authors contributed equally.

has been that this difficulty results from the fact that these languages do not assign a contrastive function to /r/ versus /l/ as distinctive phonological elements. The problem for listeners of these L1s does indeed appear to be specific to /r/-/l/, by comparison to other English approximant contrasts: Japanese listeners show categorical perception of English /w/-/r/ and /w/-/j/, both of which do occur as phonological contrasts in their language (Best & Strange, 1992).

However, even when two languages both have the same functional *phonological* contrast, the physical *phonetic* properties of the contrasting phonemes may differ systematically between the languages, and this also affects perception. Indeed, differences between English and Japanese phonetic realizations of the “shared” phonological elements /r/ and /w/ also have significant impact on Japanese listeners’ categorical perception of synthetic continua for the related English contrasts. There are marked phonetic differences between AmE /r/, which is typically realized as a “bunched” central dorsal approximant [ɹ] (or less often, retroflex [ɻ]) with additional labial and pharyngeal constrictions (Boyce & Espy-Wilson, 1997; Delattre & Freeman, 1968; Westbury, Hashi, & Lindstrom, 1998; Zawadzki & Kuehn, 1980) versus Japanese /r/, realized as an alveolar flap [ɾ] (Bloch, 1950; Vance, 1987). There are also phonetic differences between AmE /w/, a rounded labio-velar approximant [w], and Japanese /w/, an unrounded velar approximant [ɰ] (Bloch, 1950; Vance, 1987). These phonetic differences for consonants that are generally thought to show cross-language phonological correspondence have systematic and reliable effects on Japanese listeners. Their identification of an AmE /w/-/j/ continuum, though highly categorical, includes significantly more /w/ responses than do those of AmE listeners. Those phonetic differences affect Japanese listeners’ perception of /w/-/r/ as well, for which the steepness of their categorization boundary and overall discrimination level are significantly lower than native English listeners’ (Best & Strange, 1992).

More recently, purely phonetic-level effects from the L1 have also been found in native Parisian French listeners’ categorical perception of these same contrasts (Hallé, Best, & Levitt, 1999). French differs from Japanese phonologically, in that it has /r l w j/ and employs all three of the previously examined AmE phonological contrasts. Yet as those authors describe, there are systematic phonetic differences in French versus English /r/ and /l/, whereas their realizations of /w/ and /j/ are essentially identical. These phonetic differences affected French listeners’ performance on the AmE /w/-/r/ continuum, where they categorized /r/ less consistently and showed poorer discrimination than either native AmE or Japanese speakers. Their post-test impressionistic descriptions of the stimuli indicated that they perceived AmE /r/ to be /w/-like yet not an ideal French /w/, consistent with their experimental results. This native articulatory-phonetic effect on the perception of /r/, and a similar effect for AmE /l/ (velarized/pharyngealized [ɹ̥]), which some French listeners perceived to be /w/-like, also influenced their performance on AmE /r/-/l/. Here, French listeners showed classic categorical perception, yet still their category boundary was significantly less steep, and discrimination performance lower and less “peaked,” relative to AmE listeners.

Conversely, however, and more surprisingly, French listeners’ discrimination of /w/-/j/ was substantially *better* than AmE listeners’. Because /w/ and /j/ contrast in French, and have

articulatory-phonetic realizations essentially identical to AmE /w/-/j/, Hallé et al. (1999) speculated that this finding might be driven by French experience with a third, non-English onset glide, the rounded labio-palatal semivowel approximant /ɥ/ (e.g., <huile> [ɥil], ‘oil’). Specifically, they postulated a broader systemic factor, i.e., a richer approximant system in French ([w, ɥ, j]) than in English ([w, j]).

However, this posited factor was not directly evaluated in their study. There is, moreover, another potential systemic factor not considered by Hallé and colleagues: a difference in vowel systems. French has a series of front-rounded vowels, whereas English (and Japanese) lacks front-rounded vowels. Although the L1 vowel system may seem of low relevance to perception of syllable-initial consonants, approximants and especially semivowels (nonsyllabic vowels) do have a number of vowel-like features, at both phonetic and phonological levels. Indeed, French listeners’ discrimination of the initial contrast /w/-/j/ is reminiscent of findings in classic categorical perception tests with vowel continua, as both show high discrimination between and within categories, i.e., flatter functions than for stop consonants, with only a small disadvantage within categories (e.g., Fry, Abramson, Eimas, & Liberman, 1962; Fujisaki & Kawashima, 1969; Pisoni, 1973; Repp, Healy, & Crowder, 1979; Stevens, Liberman, Studdert-Kennedy, & Öhman, 1969).

## 1.2. The present study

Our goals in the present study were to further elucidate the perceptual relationship between native phonological constraints and native phonetic influences, and to gain further insight on the unexpected cross-language differences in /w/-/j/ discrimination. To achieve them, we needed listener languages that deviate from English with respect to both phonological and phonetic properties of approximant contrasts in ways that differ from Japanese and French. A phonological gap for /w/ would address the two contrasts on which French (FR) and Japanese (JA) listeners’ performance differed from AmE listeners in *opposite* directions: /w/-/r/ (JA > FR) versus /w/-/j/ (FR ≫ JA and AmE). From a phonological perspective, listeners of languages that lack /w/ should instead do relatively poorly with both contrasts.

To probe these issues in more detail, we examined listeners of Danish (DK) and German (GE) using the same stimuli and procedures as in the previous studies with Japanese and American English listeners (Best & Strange, 1992) and French listeners (Hallé et al., 1999). We also compare the results from the DK and GE listeners to those of the two previous studies. The phonological systems of DK and GE include /r l j/ but lack /w/, and hence also lack the critical phonological contrasts /w/-/r/ and /w/-/j/. Thus, Danish and German are phonologically analogous to Japanese, in the sense that each lacks one of the AmE approximants (DK and GE: /w/; JA: /l/) and two associated English contrasts (DK and GE lack /w/-/r/, /w/-/j/; JA lacks /r/-/l/, /w/-/l/<sup>3</sup>).

There are three additionally relevant phonological properties of Danish and German. The first relates to Hallé et al. (1999) speculation that French listeners’ near-ceiling discrimination of /w/-/j/ might be related to the existence of a third approximant that does not occur in English, the high front-rounded semivowel /ɥ/ (as in <huard> [ɥaʁ]) that falls between /w/ and /j/. Unlike French, which is richer than English in initial semivowels, Danish and German are both *poorer* in word-initial semivowels. Each of these languages lack both /w/ and /ɥ/ word-initially; their only true semivowel onset is /j/. An alternative vowel-system possibility is raised, however, by the observation that like French, Danish

<sup>2</sup> We refer here to the *phonological* entity /r/, i.e., “rhotic” as defined by its similar phonological functions and/or phonological features across languages (e.g., Wiese, 2001). On the *phonetic* level, rhotics show a wide range of realizations across languages: No single acoustic or articulatory feature holds across the lot, though low F3 and posterior constriction are fairly common (e.g., Lindau, 1985).

<sup>3</sup> This contrast was not examined in the original studies, nor is it studied here.

and German both employ front-rounded vowels that English lacks, including the high front-rounded /y/. This allows us to probe whether native experience with these vowels may enhance sensitivity to the vowel-like properties of word-initial glides. The lower number of initial semivowels in Danish and German relative to both French and English, yet their similarity to French in having high front rounded vowels that are lacking in English, allowed us to test the semivowel hypothesis (Hallé et al., 1999) against our alternate front-rounded vowels hypothesis that the existence of front-rounded vowels in listeners' native languages leads to high, flat, vowel-like discrimination levels for /w/-/j/ that outpaces the more categorical performance of listeners of languages that lack front rounded vowels (English and Japanese). Of additional interest for probing the semivowel hypothesis of Hallé and colleagues is a difference between the Danish and German approximant subsystems. As described in Experiment 1, Danish is much richer in approximants than the other languages, especially German, which has only three approximants (/r l j/). Moreover, Danish and German also differ from French, English, and each other in several key *phonetic* and *phonotactic* respects that could provide a window to better understanding of non-native perception of the approximant subclass of consonants.

Table 1 summarizes the key phonological and phonetic characteristics of the five languages we have discussed. All five languages have /r/ and /j/ as phonological categories. English, French, Danish and German also have /l/, which Japanese lacks. English, Japanese and French have /w/, which Danish and German lack. The phoneme /v/ is included because, as we will address, it may serve as an L1 assimilation target for a nonnative approximant in some cases (DK, GE); /v/ is found in four of the languages, i.e., all except Japanese. Even for the phonemes shared across sets of languages, however, the most typical phonetic realizations differ, as shown by the narrow phonetic transcriptions in the individual language columns.

### 1.2.1. Theoretical considerations and predictions

From a purely phonological perspective (i.e., concepts such as an L1 “phonological filter,” Polivanov, 1931; Trubetzkoy, 1958/1969, or “phonological deafness,” Dupoux & Peperkamp, 2002), native listeners of Danish, German and French should categorize and discriminate the AmE /r/-/l/ contrast much like native listeners because they all have an /r/-/l/ contrast. Yet Danish and German listeners should have difficulty categorizing and discriminating both AmE /w/-/r/ and /w/-/j/ because their L1s lack these contrasts. Japanese listeners, conversely, should discriminate AmE /r/-/l/ much less accurately than the other listener groups because Japanese has no /r/-/l/ contrast. But because Japanese and French have both /w/-/r/ and /w/-/j/ contrasts, listeners of Japanese and French should not differ from native English listeners in categorizing and discriminating /w/-/r/ and

/w/-/j/, on which they should outperform Danish and German listeners.

However, somewhat different predictions arise if we take into consideration how the finer-grained *phonetic* realizations of AmE approximants compare to their counterparts (or closest neighbors) in the other languages. Indeed, as described earlier, some purely phonological predictions have already been refuted or qualified, while those motivated by language-specific phonetic realizations have gained support (e.g., Best & Strange, 1992; Hallé et al., 1999). Effects of language-specific phonetic similarities are central to two widely considered models of non-native speech perception, the Perceptual Assimilation Model (PAM: e.g., Best, 1995; Best & Tyler, 2007) and the Speech Learning Model (SLM: e.g., Flege, 1995, 2003; Flege, Schirru, & MacKay, 2003).

PAM, especially its recent extension to L2 learning (Best & Tyler, 2007), generates predictions based on both phonological contrasts and phonetic relationships between the target stimuli and the most similar items in each listener language. Given PAM's emphasis on assimilation of *contrasts*, then, Japanese listeners' difficulty with AmE /r/-/l/ is consistent with perceptual assimilation of both items to Japanese /r/, either as equally poor realizations (Single Category assimilation), or possibly as differing slightly in goodness of fit (weak Category Goodness assimilation) in that AmE /l/ may be phonetically more similar than AmE /r/ to Japanese /r/ (Bloch, 1950). Instead, Danish, German, and French listeners should all label and discriminate AmE /r/-/l/ categorically either as corresponding to their native /r/-/l/ contrasts (Two Category assimilation: TC), or as a distinction between an Uncategorized (i.e., assimilations split among /r/, /w/ and/or /l/) vs. Categorized (/l/) (UC contrast). Both TC and UC assimilation predict excellent, categorical labeling and discrimination of the continuum. Importantly, however, PAM also predicts that their categorization boundary locations and slopes, and within- and between-category discrimination levels, should reflect phonetic differences in goodness of fit to their native /r l/ realizations.

AmE /w/-/r/ should yield either TC assimilation to native /w/-/r/ for French and Japanese listeners, or possibly Uncategorized (for AmE /r/) vs. Categorized (AmE /w/) assimilation (UC). Differences between AmE and the listeners' native phonetic realizations, however, are again likely to shift labeling and discrimination according to PAM. French performance should differ from English listeners on the /r/ side of the /w/-/r/ continuum (Hallé et al., 1999). Conversely, the Japanese and English /w/ and /r/ both differ phonetically, so Japanese and AmE performance should differ at both ends of the continuum. As for Danish and German listeners, substitution of native /v/ for English /w/ in loanword phonology and in L1-accented production of English words suggests that they detect phonetic similarities between AmE /w/ and their L1 /v/s. However, the Danish and German realizations of both native /v/ and /r/ differ phonetically from AmE /w/ and /r/. Thus, each group is likely to perceptually assimilate AmE /w/-/r/ to native /v/-/r/ as a TC contrast, or they may possibly assimilate AmE /w/ as Uncategorized to any single native consonant. These considerations should yield finer-grained differences from English listeners, and between Danish and German listeners, in perception of one or both sides of the /w/-/r/ continuum. Danish /v/ (labio-dental approximant) is phonetically more similar than German /v/ (labio-dental fricative) is to AmE /w/ (labio-velar approximant), so German listeners may differ more from English listeners in boundary, slope, and/or other aspects of the /w/-/r/ identification functions than Danish listeners do, particularly on the /w/ side of the continuum.

As for AmE /w/-/j/, PAM predictions are that French listeners should perceive AmE /w/-/j/ identically to English listeners because the phonetic realizations for both languages are [w]-[j];

**Table 1**

Phonological and phonetic comparisons of English approximants with their counterparts in Japanese, French, Danish, and German. Phonemes are shown in phonological transcription at left, including /v/ as the closest phonological neighbor to /w/ for German and Danish. Blank cells indicate a gap in the language's phonological inventory.

Phoneme	Language				
	English	Japanese	French	Danish	German
/r/	[ɹ/ɻ]	[ɾ]	[ʁ/ʀ]	[ʁ]	[ʁ/ʀ]
/l/	[ɫ]		[l]	[l]	[l]
/j/	[j]	[j]	[j]	[j]	[j]
/w/	[w]	[w]	[w]		
(/v/)	[v]		[v]	[v]	[v]

however, we already know that they unexpectedly outperformed AmE listeners in discrimination of this contrast (Hallé et al., 1999). Japanese listeners should categorize and discriminate AmE /w/-/j/ quite well as a TC assimilation, but should differ from English (and French) listeners on the /w/ side of the boundary, as has indeed been found (Best & Strange, 1992). In the present study, Danish and German listeners should also perform well on /w/-/j/ as either a TC or a UC assimilation type, with /j/ being identical to AmE /j/, and AmE /w/ being assimilated either to their native /v/s (TC contrast) or heard as Uncategorized (UC contrast). Again, however, performance on the /w/ side of /w/-/j/ should be lower for German than for Danish listeners given the phonetic differences of their /v/s re: AmE /w/.

On /w/-/r/, however, both groups are expected to perform equivalently to AmE listeners, as they are expected to assimilate them as a TC /v/-/r/ native contrast. Again, German listeners' performance on the /w/ side of the continuum should be lower than Danish (DK) listeners' because of the closer similarity of AmE [w] to Danish [v] than German [v].

SLM differs from PAM in two substantive ways that are relevant to the present study: (1) SLM focuses on individual phonetic categories while PAM focuses specifically on pairwise contrasts; (2) SLM does not explicitly address how language-specific phonetic details relate to abstract phonological structure. These characteristics make SLM predictions about discrimination of non-native contrasts somewhat more difficult to derive. However, we can extrapolate some basic predictions from core SLM principles: If the phonetic realizations of an L2 phoneme are "similar" to those of a given native phoneme, they will be difficult to discriminate from the native phoneme or from another L2 phoneme whose realizations are "identical" or "similar" to that same native phoneme. Conversely, "new" phonemes should be easier to discriminate from all native phonemes, and hence from any L2 phonemes that are "identical" or "similar" to any native phonemes, as well as from other "new" phonemes.<sup>4</sup>

Based on the phonetic descriptions in Table 1, SLM predicts that for AmE /r/-/l/ the German, Danish and French listeners will equivalence-classify AmE /l/ to their native /l/s because the phonetic distances are modest. Danish listeners should also equivalence-classify AmE /r/ as "similar" to Danish /r/, whereas German and French listeners should instead perceive it as fairly different from their native /r/s and thus hear it as a "new" phone. Extrapolating, then, listeners of these three groups should discriminate AmE /r/-/l/ as a distinction between a "similar" phone for /l/ versus either a "similar" phone of a different native category (Danish /r/), or versus a "new" phonetic category (German/French /r/). Whether AmE /r/ is "similar" to native /r/ or is perceived as a "new" phonetic category, English-inexperienced listeners of all three languages should perform somewhat less consistently than AmE listeners on identification and discrimination near the endpoints of /r/-/l/, which are not as phonetically well-defined for them as they are for native English listeners. By SLM reasoning, Japanese listeners should perform more poorly on both tasks than the other groups because they equivalence-classify both AmE /r/ and /l/ as phonetically "similar" to Japanese /r/, which would block accurate perception of AmE /r/-/l/ as a phonetic distinction.

SLM also differs from PAM in phonetic-level predictions for AmE /w/-/r/. SLM does not anticipate *partial* equivalence classification, as was seen in French listeners' judgments of AmE /r/ as only somewhat similar to French /w/ and thus inconsistently classified as /w/, a pattern that PAM does account for (Hallé et al.,

1999). In addition, SLM principles predict that Danish listeners should perceive /w/-/r/ more categorically than German listeners, and more like native English listeners at both endpoints because AmE /w/ and /r/ are phonetically more similar to both Danish /v/ and /r/, respectively, than they are to German /v/ and /r/. Hence /w/ and /r/ should be equivalence-classified as "similar" to contrasting Danish phonemes, whereas both should be classified as "new" by German listeners. By extension, SLM should predict that Japanese listeners' performance will be better at the /w/ end but worse at the /r/ end than is that of German listeners. Conversely, French listeners' performance at the /w/ end of the continuum should be better than Danish listeners and much better than German listeners, but at the /r/ end it should be equal to German and worse than Danish performance. That is, SLM should predict good discrimination of AmE /w/-/r/ as two categorically differing phonetic categories for all these listener groups, but their performance should differ around one or both endpoints in language-specific ways. SLM would not predict, as PAM does, that German and Danish listeners should perform *better* than French listeners on /w/-/r/.

SLM also predicts that AmE /w/ and /j/ will both be classified as "identical" to their French counterparts, so French listeners should perform essentially like AmE listeners across the continuum. However, AmE /w/ should be heard as "similar" to Danish [v] and Japanese [u] but "new" to German listeners, while AmE /j/ is "identical" to /j/ in these three languages. Therefore, by SLM reasoning Danish and Japanese listeners should categorize and discriminate AmE /w/-/j/ similarly, showing less perceptual consistency than native English listeners at the /w/ end. The German listeners should show lowest performance at the /w/ end, but none of these groups should differ from English listeners at the /j/ end.

Table 2 provides a schematic overview of the predictions generated by phonological viewpoints, by PAM, and by SLM, for perception of AmE approximant contrasts by Danish and German listeners (tested in Experiments 1–3), as compared to French, Japanese and native American English listeners (Experiment 3).

## 2. Experiment 1

Our first experiment examined perception of AmE approximant contrasts by native speakers of Danish, which has several interesting differences from AmE, as well as from French Japanese and our other new listener language, German (see Table 1; Experiment 3). Danish has a much more extensive set of approximant realizations than any of these other languages. In addition to /r l j/, Danish /v/ is realized as a labiodental approximant [ʋ] rather than a voiced fricative, and its intervocalic voiced stops undergo extreme lenition to bilabial, dental, and velar approximants [β, ɸ, ɣ] (Grønnum, 1998, 2003). Moreover, Danish /r/ is realized as an unrounded pharyngeal approximant [ʕ]<sup>5</sup> (e.g., Grønnum, 1998) rather than as a tap (JA [r]) or as a uvular fricative/approximant (FR, GE [ʁ/ʁ̥]). Thus, Danish speakers actively produce seven approximants ([β ʋ ɸ l j ɣ ʕ]). This is substantially more than the three (JA, GE) or four (AmE, FR) approximants found in the other languages under discussion.

With respect to the stimulus language, the phonetic characteristics of Danish /j/ are essentially identical to /j/ in English. Like Danish /r/, AmE /r/ ([ɹ]/[ɹ̥]) involves a pharyngeal constriction. However, AmE /r/ also generally has an alveolar/retroflex tongue

<sup>4</sup> Importantly, SLM regards perceived L1–L2 phonetic similarity as a continuum, not a tripartite identical-similar-new division.

<sup>5</sup> The IPA has no separate symbol for this; approximant is indicated by using a lowering diacritic below the homorganic voiced fricative symbol.



**Table 2**

Predictions for categorization and discrimination performance on the three AmE approximant contrasts by American-English (AmE), German (GE), Danish (DK), French (FR), and Japanese (JA) listeners, as generated by Phonological viewpoints, the Perceptual Assimilation Model (PAM), and the Speech Learning Model (SLM).

Contrasts	Models		
	Phonological	Perceptual assimilation	Speech learning
/r/-/l/	AmE=(DK,GE,FR) > JA	AmE > DK > (GE,FR) > JA	AmE > (DK,GE,FR) > JA
/w/-/r/	AmE=(FR,JA) > (DK,GE)	AmE > DK > GE > FR > JA	AmE > (DK,FR) > (GE,JA)
/w/-/j/	AmE=(FR,JA) > (DK,GE)	AmE=FR > JA > DK > GE	AmE=FR > (DK,JA) > GE

tip constriction and lip-rounding, which are both lacking in Danish /r/. Danish /l/ is phonetically “light/clear” [l], and thus differs modestly from the AmE “dark” velarized/pharyngealized [ɫ]. Finally, although Danish lacks a phonological /w/, its closest word-initial neighbor /v/ has the same manner as English /w/ but lacks its rounding and secondary velar constriction, and has a different place of articulation (labio-dental rather than bilabial-velar). Also, Danish (unlike German) permits the rounded back vowel /u/ to occur in onset position before /a/ (e.g., *uagtksom*; English: *negligent*), however, it is realized as a two-vowel /u/ + /a/ sequence rather than as a single-syllable approximant-onset /wa/.

## 2.1. Method

### 2.1.1. Participants

Participants were 18 Danish speakers (Aarhus University, Aarhus DK; 15 female, 3 male;  $M_{\text{age}}=23.0$  years,  $s.d.=2.1$ ). A background questionnaire confirmed that each subject met the following selection criteria: no history of hearing loss, native Danish, native Danish-speaking parents, and limited immersion in languages other than Danish (<8 months living in a foreign language environment). Eleven participants had not spent any time in an English-speaking environment; the other seven had been exposed to native English for a mean period of 4.3 months (range: 1–7 months). Twelve participants had lived most of their lives in the East Jutland region around Aarhus; the other six grew up in neighboring regions, but did not differ in any noticeable way from the East Jutlanders. Their mean self-ratings (on a scale of 1=poor to 5=excellent) of English proficiency were 4.1 ( $SD=.9$ ) for *speaking* and 4.2 ( $SD=1.0$ ) for *understanding* spoken English. The subjects were paid 200 DK kroner (ca. 30 USD) for participation in one two-hour session.

### 2.1.2. Stimulus materials

We used the same three 10-step continua of AmE approximant contrasts, /rak/-/lak/, /wak/-/jak/, and /wak/-/rak/, synthesized by Best and Strange (1992) on the basis of careful acoustic measurements of natural productions by an American English phonetician (see original paper for full stimulus details).

### 2.1.3. Procedure

Subjects were tested in three groups of six in one session each in the language laboratory of Aarhus University. The signal from the three audiotapes was routed through Tandberg Educational Media Centre IS-10MM to professional studio-quality circumaural headphones. All other procedural details are as in the previous reports (Best & Strange, 1992; Hallé et al., 1999). The sequence in which the three contrasts were presented was counterbalanced across groups. For each contrast, a two-choice identification test was followed by an AXB discrimination test. For each identification test, stimuli were presented one by one in ten blocks of 20, with an ISI of 3.0 s and an IBI of 5.0 s. For each trial on the

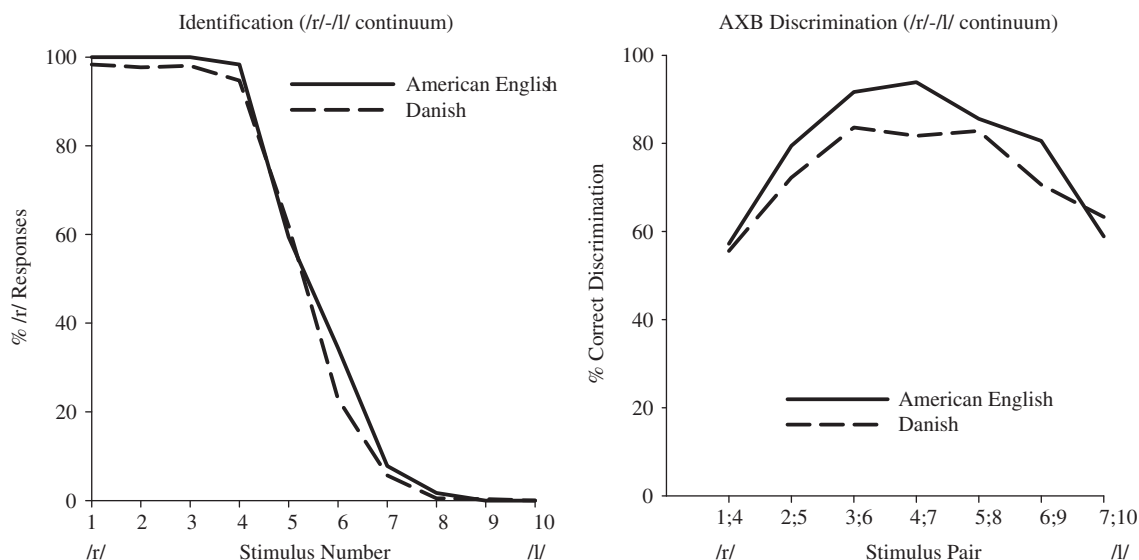
corresponding forced-choice answer sheet, subjects circled the English consonant they heard at the syllable onset (W or Y; W or R; R or L). In each AXB discrimination test, triads of stimuli were presented in which the first and third item were separated by three steps along a 10-step continuum, and the middle item matched the first or third item; all possible 3-step pairings were presented an equal number of times, in all four possible triad orders (AAB, ABB, BAA, BBA). Trials for a given contrast were presented in ten blocks of 14 (inter-stimulus interval=1.0 s; inter-trial interval=3.0 s; inter-block-interval=6.5 s). Subjects used a numbered answer sheet for each contrast, circling “1” or “3” for each trial to indicate whether the second syllable was identical with the first or the third syllable in the triad that they had just heard. After completing all tests, participants completed a post-test questionnaire, giving informal/impressionistic descriptions of the syllable onsets they heard.

## 2.2. Results

For the identification tests in this and each subsequent experiment, we first ran Probit analyses of each subject’s responses to the stimulus continuum items in order to estimate their category boundary (stimulus number at the 50% category crossover where the percept was evenly divided between the two categories) and slope (steepness of the identification function around the 50% crossover) on each contrast (as in Best & Strange, 1992; Hallé et al., 1999; MacKain et al., 1981). For these boundary and slope measures, we conducted one-way analyses of variance (ANOVAs) for each continuum, comparing the Danish results against the AmE listener results of Best and Strange (1992), using listener Language (Danish, AmE) as a between-subject factor.<sup>6</sup> We examined finer-grained within- and between-category differences between the current Danish and previous AmE (from Best & Strange, 1992) listeners with two-way ANOVAs on the full categorization functions for each continuum, using Language (2) as a between-subject factor and Stimulus item (10) as a within-subject factor.

For discrimination analyses, one-way Language group ANOVAs were conducted on three extracted dependent variables: mean percent correct discrimination scores, percent correct discrimination for the stimulus pair that straddled each individual listener’s category boundary (as established by the probit analyses on each individual’s identification test), and a measure of the “flatness/peakiness” of the discrimination function (mean of unsigned difference scores for all adjacent stimulus pairs). We also probed finer-grained within- and between-category differences for each continuum with two-way repeated measure ANOVAs on the full discrimination functions using Language (2) as a between-subject

<sup>6</sup> One-way ANOVAs on 2-level factors are statistically equivalent to paired *t*-tests; for consistent treatment of analyses within and across experiments in this report, some of which were multi-factor and thus required ANOVAs, we report even the one-way paired comparisons as ANOVAs.



**Fig. 1.** Identification functions (left panel) and discrimination functions (right panel) for the /r/-/l/ continuum as perceived by native Danish and American English listeners. (Data for American English listeners in this and all other figures from Best & Strange, 1992.)

factor and percent correct discrimination on each Stimulus Pair (7 levels) as a within-subject factor.

### 2.2.1. The /r/-/l/ continuum

Fig. 1 compares the identification (left panel) and discrimination functions (right panel) for the /r/-/l/ continuum of the Danish listeners in the present study to American English listeners (Best & Strange, 1992). The category boundaries did not differ between the two listener groups, according to the one-way ANOVAs on the probit-derived measures of boundary location (Danish group: 5.3 along the 10-step continuum; English group: 5.4) or slope (Danish: 1.8; English: 2.2),<sup>7</sup> *ns*. In the two-way ANOVA on Language group (2) × Stimulus item (10) for the full categorization functions, the significant Stimulus item main effect confirmed that both groups showed a highly categorical identification function with cross-over boundaries falling between items 5 and 6,  $F(9,225)=254.94$ ,  $p < .0001$ . However, as with the analyses of the two probit boundary measures, this ANOVA failed to find any finer-grained group differences in the pattern of /r/-/l/ identification, i.e., the Language effect and its interaction with Stimulus item were nonsignificant, *ns*.

The one-way ANOVAs on discrimination performance found no significant listener group differences on mean percent correct /r/-/l/ discrimination (Danish: 73%; English: 78%), “peakiness” of the discrimination function (Danish: 12%; English: 12%), or discrimination accuracy at listeners’ cross-category boundaries (Danish mean: 84% correct, English: 93%), both *ns*. The two-way ANOVA on discrimination did, however, find a marginal overall group difference (Danish: 73% correct; English, 78%),  $F(1,25)=3.20$ ,  $p < .09$ . Simple effect ANOVAs on this trend at each stimulus pair found a group difference solely at the *mean* category boundary, pair 4–7,  $F(1,151)=5.49$ ,  $p < .025$ . Specifically, Danish listeners had a lower, flatter between-category discrimination peak (82% correct) than English listeners (94%) (see Fig. 1, right panel). In the two-way ANOVA, the Stimulus Pair effect was also significant,  $F(6,150)=29.92$ ,  $p < .0001$ , indicating a peak in discrimination performance at the category boundary (stimulus pairs 2–5 and 3–6) across the two listener groups.

### 2.2.2. The /w/-/r/ continuum

The groups’ identification and discrimination functions for /w/-/r/ are shown in Fig. 2. The one-way ANOVAs on the probit identification measures again found no reliable group differences in category boundary location (Danish group: 5.1; English group: 4.8) or slope (Danish: 1.5; English: 1.7), both *ns*. The significant Stimulus effect in the two-way ANOVA again confirmed that both groups had highly categorical identification functions,  $F(9,225)=393.66$ ,  $p < .0001$ , with the cross-over boundary between stimulus items 4 and 5 (close to item 5). There was also a marginal Language group difference in mean reporting of “W” across the continuum (Danish: 49%; English: 41%),  $F(1,25)=3.45$ ,  $p < .08$ . Simple effects tests on that trend found a significant group difference for stimulus 5 (Danish: 50%; English: 34%),  $F(1,9)=10.44$ ,  $p < .001$ , and stimulus 6 (Danish: 31%; English: 18%),  $F(1,9)=7.45$ ,  $p < .01$ , the two items just to the /r/ side of the category boundary (see Fig. 2, left panel).

As for AXB discrimination performance, the one-way ANOVAs on the mean percent correct discrimination scores for Danish (72%) and English listeners (75%), “peakiness” of their discrimination functions (Danish: 12%; English: 12%), and their accuracy in individuals’ cross-category discrimination (Danish: 82% correct, English: 86%) all failed to find reliable group differences, *ns*. Moreover, although the two-way ANOVA revealed a significant Stimulus Pair effect, indicating a peak in discrimination at the category boundary (stimulus pair 3–6),  $F(6,150)=17.57$ ,  $p < .0001$ , this did not differ between the listener groups: Neither the Language group effect nor its interaction with Stimulus Pair were significant, *ns*.

### 2.2.3. The /w/-/j/ continuum

The identification and discrimination functions for /w/-/j/ are shown in Fig. 3. Again, the one-way ANOVAs on the probit measures failed to find significant Language group differences for the boundary location (Danish: 5.1; English: 5.4) or slope (Danish: 1.6; English: 1.9), both *ns*. The two-way ANOVA again revealed only that both groups had very similar categorical identification functions, with the cross-over boundary falling between stimulus items 5 and 6 (closer to item 6),  $F(9,225)=226.62$ ,  $p < .0001$ . Neither the Language main effect nor the interaction approached significance, *ns*.

<sup>7</sup> Note: Higher values indicate steeper slopes.

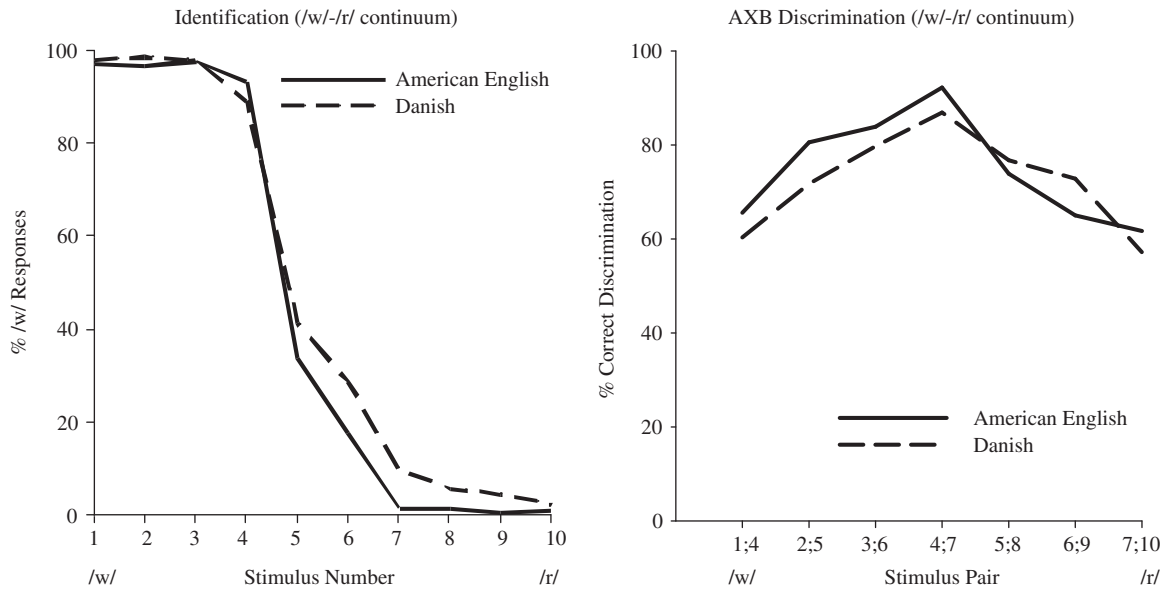


Fig. 2. Identification functions (left panel) and discrimination functions (right panel) for the /w/-/r/ continuum as perceived by Danish and American English listeners.

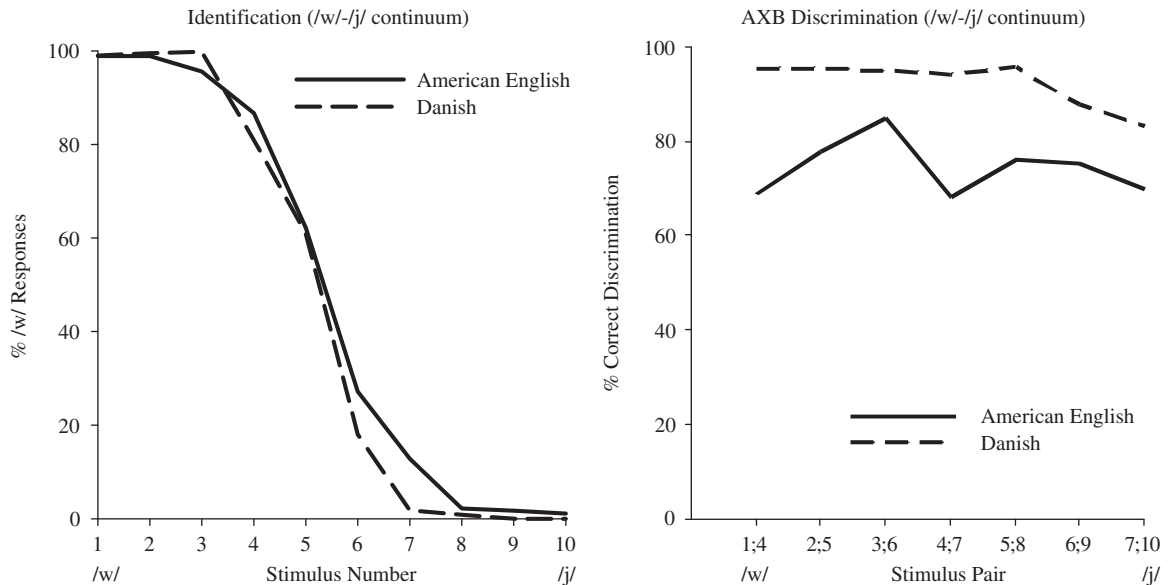


Fig. 3. Identification functions (left panel) and discrimination functions (right panel) for the /w/-/j/ continuum as perceived by native Danish and American English listeners.

In the discrimination task, however, the Danish listeners were substantially and significantly more accurate than native English listeners on /w/-/j/ in mean discrimination level (Danish: 93% correct; English: 75%),  $F(1,25)=48.42$ ,  $p < .0001$ , cross-category discrimination (Danish: 94% correct; English: 78%),  $F(1,25)=23.7$ ,  $p > .001$ , and differed in “peakiness” (Danish: 6%; English: 9%),  $F(1,25)=10.57$ ,  $p < .004$ , which indicated a significantly flatter discrimination function for Danish than English listeners. This picture was further supported by the two-way discrimination ANOVA, which confirmed the mean discrimination advantage of Danish over native English listeners,  $F(1,25)=48.41$ ,  $p < .0001$ . The Stimulus Pair main effect was also significant,  $F(6,150)=7.85$ ,  $p < .0001$ , indicating the average discrimination function was non-flat, showing shallow peaks at pairs 2–5 and 4–7. Importantly, the Language  $\times$  Stimulus Pair interaction indicated those peaks were attributable to the English listeners,  $F(6,150)=4.51$ ,  $p < .0005$  (see Fig. 3, right panel). Simple effects tests found that

Danish listeners outperformed native English listeners on each of the seven stimulus pairs;  $F(1,6)$  values ranged between 6.61–47.37,  $p$  values between .01 and .0001.

### 2.3. Discussion

Danish listeners categorized the /r/-/l/ and /w/-/j/ contrasts in a manner that was statistically indistinguishable from English listeners in every way. On the other hand, although the Danish /w/-/r/ categorization boundary and slope did not differ from English listeners, their categorizations did deviate significantly for two items just to the /r/ side of the /w/-/r/ boundary. By comparison, they failed to show any difference from native English listeners in their discrimination of /w/-/r/, whereas they did show differences from them in discrimination of the /r/-/l/ and /w/-/j/ contrasts. Danish listeners' discrimination of /r/-/l/ differed from English listeners only on the discrimination pair

that straddled the mean category boundary: The Danish peak was significantly lower and flatter than the English peak. In stark contrast, discrimination of /w/-/j/ was remarkably better for Danish than English listeners, across the board. This performance pattern is not fully consistent with any of the three sets of predictions delineated for Danish versus English listeners (see Table 2).

Nativelike categorization and discrimination of AmE /r/-/l/ was predicted by the phonological perspective, based on the reasoning that both languages have a phonological contrast between /r/-/l/. Excellent categorization and discrimination of /r/-/l/ was also predicted by SLM and PAM, for different reasons. According to SLM, Danish listeners should equivalence-classify the velarized AmE realization of /l/ to their own “light” realization of /l/, and the AmE realization of /r/ as (somewhat less) “similar” to the Danish realization of /r/. Danish listeners therefore should easily categorize and discriminate AmE /r/-/l/ as a difference between two native-equivalent phonetic categories. PAM similarly predicted Two Category (TC) assimilation of AmE /r/-/l/ to the native Danish phonological contrast /r/-/l/, but also suggested the possibility of Uncategorized–Categorized (UC) assimilation. Both TC and UC assimilation are expected to yield excellent performance on both perceptual tasks.

However, unlike the phonological approach, both SLM and PAM also predict finer-grained listener group differences for /r/-/l/, based on differences in their phonetic realizations in Danish versus English. Specifically, SLM expectations are that Danish listeners would identify and discriminate the /r/ and /l/ edges of the continuum somewhat less consistently than English listeners, because both phonetic categories are less well-defined for them than for native listeners. PAM posits that the cross-language phonetic realization differences will affect finer-grained aspects of the Danish categorization (boundary location, steepness, localized differences along the continuum), and/or discrimination functions (localized within- or cross-category performance), as compared to native listeners. The highly categorical identification of AmE /r/-/l/ by Danish listeners, which was equivalent with that of English listeners, is compatible with predictions of all three theoretical viewpoints. However, the lower, flatter peak for Danish than for English listeners’ discrimination at the category boundary is consistent with PAM’s assumptions about sensitivity to between-language phonetic realization differences. This finding is less consistent with SLM phonetic-level predictions, and is inconsistent with the phonological model.

Danish perception of /w/-/r/ is also inconsistent with the phonological prediction that because their native phonology lacks /w/ they should have notable difficulties with this contrast. They instead showed highly categorical performance on both /w/-/r/ perceptual tasks, which was quite similar to native English listeners. PAM and SLM had predicted that Danish listeners would categorize and discriminate /w/-/r/ excellently, as was observed. PAM predicted this based on expectations of a TC or UC assimilation pattern, whereas SLM predicted it on the reasoning that both /w/ and /r/ should be equivalence-classified as “similar” to two different Danish phonetic categories. SLM also predicted that Danish performance would differ somewhat from native English listeners at both endpoints. PAM instead predicted modest differences from native English listeners in Danish listeners’ categorization and/or discrimination of /w/-/r/, on the phonetic-level reasoning that cross-language differences in realization of /r/, and the similar but non-identical realizations of Danish /v/ ([v]) and AmE /w/, would impinge on Danish listeners’ within- and between-category perception of AmE /w/-/r/. The categorization analyses found that the Danish listeners reported “W” more often than native English listeners for the two continuum items just on the /r/ side of the /w/-/r/ boundary, showing modest sensitivity near the category

boundary to the phonetic differences between AmE /w/-/r/ and their closest-corresponding native approximants. Thus, SLM and PAM predictions that Danish listeners would show excellent categorization and discrimination of AmE /w/-/r/ were upheld, while the phonological prediction of perceptual difficulties was not. In addition, the finer-grained group differences in “W” categorizations on the /r/ side of the boundary were somewhat more compatible with PAM’s than SLM’s finer-grained phonetic-level predictions.

Danish listeners categorized the /w/-/j/ continuum essentially the same way as native English listeners. The phonological view had predicted poorer-than-native performance on categorization and discrimination of /w/-/j/ because Danish lacks /w/ and therefore lacks this phonological contrast. PAM and SLM had instead both predicted that Danish listeners would perform as well as native English listeners because the phonetic realization of /j/ is phonetically identical in the two languages and the realizations of AmE /w/ and Danish /v/ are quite similar. Thus, PAM reasoned that they would perceptually assimilate AmE /w/-/j/ as a TC contrast, whereas SLM expected them to equivalence-classify AmE /j/ as “identical” to native /j/ and AmE /w/ as “similar” to native /v/ ([v]). The /w/-/j/ categorization results, then, are inconsistent with the phonological viewpoint, but compatible with both PAM and SLM.

The most intriguing finding, however, was unexpected all around: Danish listeners discriminated AmE /w/-/j/ much better than native listeners, performing near ceiling except slightly lower at the /j/ end of the continuum. This unusually high nonnative performance was foreshadowed by French listeners’ similar performance in Hallé et al. (1999), where it was also unexpected. This pattern is quite inconsistent with the predictions of all three models.

Before attempting a comprehensive interpretation of this most surprising finding, or comparing it against native German listeners and other previously-reported L1 groups (Experiment 3), we felt it important to address the possibility that the excellent performance of the Danish listeners in general, and especially on /w/-/j/ discrimination, may have been due to familiarity with spoken English. Our second experiment, therefore, compared two new groups of native Danish listeners who differed substantially in experience with native spoken English. This comparison also permitted us to further compare PAM and SLM, which had each received support from the /r/-/l/ and w/-/r/ findings, as well as the /w/-/j/ categorization results. Specifically, comparing Danish speakers who have minimal versus more extensive experience with natively spoken English allowed examination of the two models’ additional predictions about the effects of second language (L2) experience on perception of nonnative L2 consonants and contrasts.

### 3. Experiment 2

Although the participants of Experiment 1 were relatively inexperienced with English by comparison to native listeners, they had nonetheless had years of classroom instruction and speaking/reading experience with English. This is virtually impossible to avoid given modern education in Denmark, where English is taught as a required school subject from 10 years of age or younger. Therefore, we systematically probed the effects of minimal versus substantial experience with *natively spoken* English, by comparing perception of the AmE approximant contrasts in two additional groups of native Danish listeners who differed in experience with spoken English in native-English environments (i.e., no such experience, versus months/years of native-English immersion experience). We note, nonetheless, that



even high levels of spoken English experience are unlikely to explain unexpectedly good discrimination of /w/-/j/.

SLM hypotheses (e.g., Flege, 1995) about the influence of L2 experience are that little to no perceptual learning should occur for L2 phones perceived as “similar” to L1 phonetic categories, which was hypothesized for AmE /l/, /r/ and /w/, because equivalence classification will block learning of the L2-specific phonetic properties. And of course, no perceptual adjustment need occur for “identical” L2 phones (AmE /j/). Thus, Danish listeners with substantial natively spoken English experience should perform no differently from inexperienced Danish listeners on these three contrasts. That is, both groups should behave as had been predicted for Experiment 1: excellent categorization for all three continua, but somewhat less consistent than English listeners near the /w/, /r/ and /l/ ends of the continua.

PAM hypothesizes instead that L2 experience will result in perceptual learning (“attunement”) mainly for nonnative consonants that are assimilated to a native consonant (Categorized) but are notably deviant from it (poor goodness-of-fit). For contrasting nonnative phones that differ in goodness of fit to native categories (CG assimilation), learning is especially likely for the phone that shows the poorer fit to the native category (Best & Tyler, 2007). Given that AmE /j/ is phonetically identical to Danish /j/, and that /l/ is a better fit to Danish /l/ than AmE /r/ and /w/ are to Danish /r/ and /v/ ([v]), by PAM reasoning English experience is most likely to impact on finer-grained phonetic-level aspects of categorization and/or discrimination for /w/-/r/, on both sides of the contrast. Perceptual effects of experience should be smaller to nonexistent for /r/-/l/ and /w/-/j/; if any experience effect emerges, it should be seen more on the /r/ and /w/ sides of those continua, respectively.

### 3.1. Methods

#### 3.1.1. Participants

Thirty L1 Danish speakers participated in the experiments. A background questionnaire confirmed that each subject met the following selection criteria: no history of hearing loss, native Danish speaker, and native Danish speaking parents. Fifteen speakers (6 females; 9 males;  $M_{\text{age}}=27.1$  years,  $s.d.=4.0$ ) had limited exposure to languages other than Danish (i.e., < 3 months in any foreign language environment), and had always worked/studied in a Danish speaking environment. These were assigned to the inexperienced listener group (DK<sub>inexp</sub>). The remaining 15 speakers (10 females, 5 males;  $M_{\text{age}}=24.8$  years,  $s.d.=2.0$ ) were experienced speakers of English (i.e., had spent > 10 months in an English speaking country) and had studied at the English Department of Aarhus University where all teaching is conducted in English. They were assigned to the experienced listener group (DK<sub>exp</sub>). All 30 Danish participants had lived most of their lives in Jutland; 17 had grown up in East Jutland (area surrounding and including Aarhus), the other 13 in neighboring regions. These differences in experience were reflected in the two groups’ self-ratings of their proficiency in speaking (DK<sub>inexp</sub>  $M=3.5$ ,  $SD=.7$ ; DK<sub>exp</sub>  $M=4.5$ ,  $SD=.5$ ) and understanding spoken English (DK<sub>inexp</sub>  $M=4.0$ ,  $SD=.6$ ; DK<sub>exp</sub>  $M=4.9$ ,  $SD=.4$ ). The subgroups did not differ in any other noticeable ways. They received 200 DK kroner (ca. 30 USD) for participation in one two-hour session.

#### 3.1.2. Stimulus materials

We used the same three continua of AmE approximant contrasts, /rak/-/lak/, /wak/-/jak/, and /wak/-/rak/ as in Experiment 1 and in previous research (Best & Strange, 1992; Hallé et al., 1999; MacKain et al., 1981).

#### 3.1.3. Procedure

Participants were tested individually in an IAC sound booth at the English Department of Aarhus University. The three audiotapes were played on a Marantz audiocassette recorder (model CP 430) and presented over professional studio-quality circumaural headphones. All procedural details were as in Experiment 1, except that we gave the listeners free choice of writing in whichever Danish consonant they felt they had heard for the items played in the identification test, in an attempt to minimize imposing an English bias on categorizations, and to determine instead whether the listeners may hear Danish consonants (or vowels) that differ from the forced-choice English response alternatives (“W”, “Y”, “L”, “R”) that were used in Experiment 1.

### 3.2. Results

We report the results from the identification task by describing the native (Danish) phonetic category(s) that each group of listeners used in their open responses, and by statistically comparing the slopes and boundaries between the response categories that each listener actually used. For the two-way ANOVAs on the categorization data, we analyzed percent use of “R” for the /r/-/l/ continuum items, and percent use of either the Danish labiodental approximant “V” or the English letter “W” for the /w/-/r/ and /w/-/j/ continuum items. “W” does not exist in Danish orthography but the great majority of participants used it in their open responses (as described below). The results of the discrimination task were analyzed by comparing the two Danish groups using ANOVAs on the same dependent variables as for Experiment 1.

#### 3.2.1. The /r/-/l/ continuum

Eleven of the DK<sub>inexp</sub> and 12 of the DK<sub>exp</sub> listeners responded with “R” versus “L” in the open response identification task. Of the minority of listeners who did not divide the /r/-/l/ continuum into “R” and “L”, one DK<sub>inexp</sub> subject responded only with “M,” that is, failed to hear any phonetic distinction along the continuum. Three listeners in each DK subgroup instead indicated that they divided the continuum into three, rather than only two, categories: “R” at the /r/ end of the continuum, a labial approximant middle category (“W” for AmE /w/ or “V” for DK /v/), and “L” at the /l/ end of the continuum, which is reminiscent of the prior findings with French listeners (Hallé et al., 1999). The percentage of “R” responses to each continuum item is shown in Fig. 4 (left panel) for the 14 DK<sub>inexp</sub> and all 15 DK<sub>exp</sub> listeners who did report hearing /r/ at the lefthand end of the continuum. One-way ANOVAs on the probit-derived identification measures for the participants who reported hearing an /r/-/l/ contrast found no group difference in boundary (DK<sub>inexp</sub>: 5.1; DK<sub>exp</sub>: 5.2) or slope values (DK<sub>inexp</sub>: 1.3; DK<sub>exp</sub>: 1.5), both *ns*. A two-way ANOVA on percentage of “R” responses across the continuum for the factors of Experience group (2, between-subject) × Stimulus item (10, within-subject) found only a significant Stimulus effect,  $F(9,234)=273.67$ ,  $p<.0001$ , indicating comparably categorical identification functions across the experience groups. Neither the Experience effect nor the interaction was significant, *ns*.

Fig. 4 (right panel) also shows the /r/-/l/ discrimination functions for the full set of 15 listeners in each experience subgroup. ANOVAs comparing the DK<sub>inexp</sub> and DK<sub>exp</sub> listeners’ discrimination performance found no significant experience differences in mean percent correct (DK<sub>inexp</sub>: 76%; DK<sub>exp</sub>: 78%), “peakiness” (10% for both groups), or cross-category discrimination (79% and 82%, respectively). In the two-way ANOVA there was only a significant main effect of Stimulus item,  $F(6,168)=19.36$ ,  $p<.0001$ , indicating a significantly non-flat discrimination function

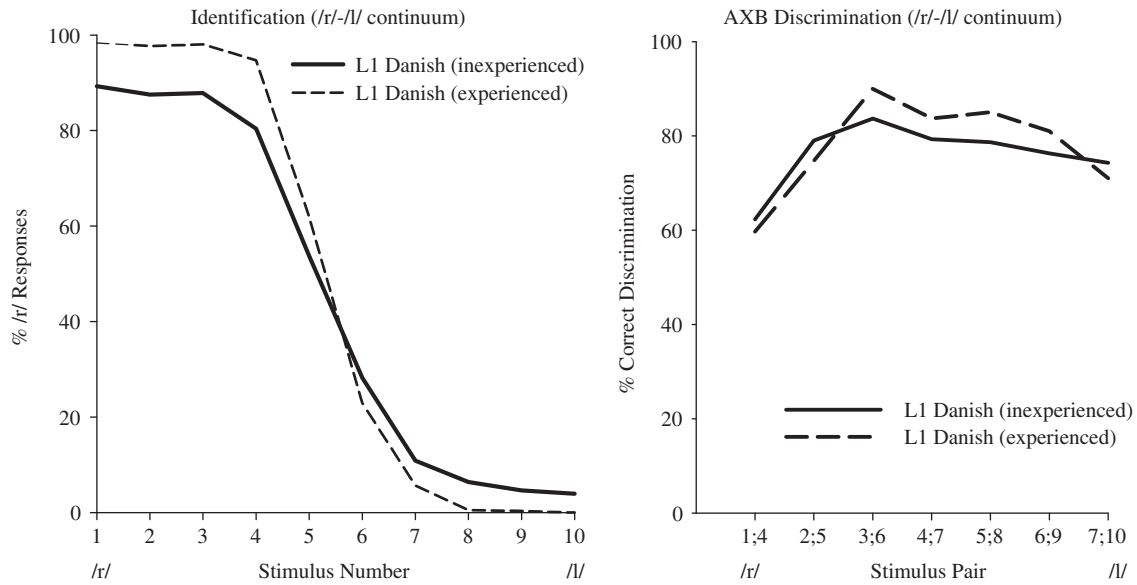


Fig. 4. Identification (left panel) and discrimination functions (right panel) for the /r/-/l/ continuum as perceived by English-inexperienced ( $DK_{inexp}$ ) and English-experienced ( $DK_{exp}$ ) Danish listeners.

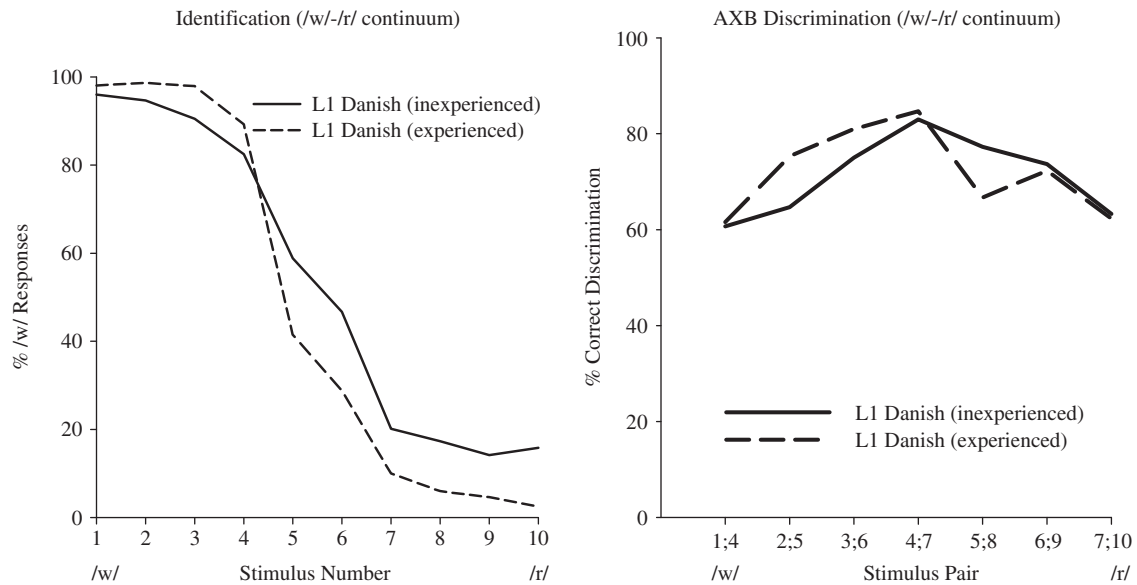


Fig. 5. Identification (left panel) and discrimination functions for the /w/-/r/ continuum as perceived by English-inexperienced ( $DK_{inexp}$ ) and English-experienced ( $DK_{exp}$ ) Danish listeners.

with a peak near the category boundary, at stimulus pair 3–6. Neither the Experience effect nor the interaction was significant, *ns*.

### 3.2.2. The /w/-/r/ continuum

Thirteen  $DK_{inexp}$  and 14  $DK_{exp}$  listeners responded with “W” (or Danish /v/) versus “R” in the open response identification task; another  $DK_{inexp}$  participant divided the continuum into “W” versus “M.” Of the two listeners who did not divide the /w/-/r/ continuum into “W” versus “R/M”, one  $DK_{inexp}$  subject responded only with “V,” and one  $DK_{exp}$  listener divided it into “L” versus “R.” The one-way ANOVAs on the probit-derived categorization measures for the 28 participants who reported hearing “W” versus “R” (or “W” vs. “M”) found no significant group differences in category boundary location ( $DK_{inexp}$ : 5.1;  $DK_{exp}$ : 4.9) or slope (.9 and 1.3, respectively). The two-way ANOVA on these same participants’ percentage of “W” (or /v/) responses revealed a

significant Stimulus item effect,  $F(9,225)=195.72$ ,  $p < .0001$ , indicating categorical identification functions overall (see Fig. 5, left panel). Although Experience did not yield a significant main effect, the Experience  $\times$  Stimulus item interaction was significant,  $F(9,225)=3.35$ ,  $p < .001$ . Simple effects tests found that  $DK_{inexp}$  listeners gave more “W” responses than  $DK_{exp}$  listeners on the /r/ side of the continuum, specifically for Stimulus 5 ( $DK_{inexp}$ : 59%;  $DK_{exp}$ : 42%),  $F(1,9)=7.2$ ,  $p < .01$ , Stimulus 6 ( $DK_{inexp}$ : 47%;  $DK_{exp}$ : 29%),  $F(1,9)=7.85$ ,  $p < .01$ , Stimulus 8 ( $DK_{inexp}$ : 18%;  $DK_{exp}$ : 6%),  $F(1,9)=3.5$ ,  $p = .06$ , and Stimulus 10 ( $DK_{inexp}$ : 17%;  $DK_{exp}$ : 3%),  $F(1,9)=4.88$ ,  $p < .05$ .

Fig. 5 (right panel) shows the discrimination functions for /w/-/r/. The three dependent measures of mean percent correct ( $DK_{inexp}$ : 71%;  $DK_{exp}$ : 72%), “peakiness” (12% for both subgroups), and cross-category discrimination (83% and 82%, respectively) all failed to show significant differences between the two Danish listener groups. However, the two-way ANOVA revealed not only

a significant Stimulus item effect,  $F(6,168)=15.12$ ,  $p < .0001$ , but also an Experience x Stimulus item interaction,  $F(6,168)=2.62$ ,  $p < .02$ . Simple effects tests found that on the within-/w/ pair 2–5, the  $DK_{inexp}$  group discriminated more poorly (64.67% correct) than the experienced group (75%),  $F(1,6)=5.36$ ,  $p < .025$ , while the opposite was true for the within-/r/ pair ( $DK_{inexp}$ : 77%;  $DK_{exp}$ : 67%),  $F(1,6)=5.36$ ,  $p < .025$ .

### 3.2.3. The /w/-/j/ continuum

The open responses for identification of /w/-/j/ were much more variable than for the other two continua. Only four listeners divided the continuum into two categories. Just one  $DK_{inexp}$  and one  $DK_{exp}$  listener responded with “W” versus “J” while two other  $DK_{inexp}$  listeners responded with, respectively, “V” versus “J,” and with “V” versus “L.” The open responses from 12  $DK_{inexp}$  and 14  $DK_{exp}$  listeners suggest that they heard three categories: A labial category near the /w/ endpoint that they mostly labeled “W,” an intermediate “L” category, and a category near the /j/ endpoint that they mostly labeled “J.” It is of interest that this particular 3-way pattern is reminiscent of French (Hallé et al., 1999) and American English listeners’ responses (Best & Strange, 1992) to this same continuum.

We analyzed the 3-way categorizations of these listeners, who formed by far the majority. The one-way ANOVAs on the “W”–“L” boundary location ( $DK_{inexp}$ : 3.9,  $DK_{exp}$ : 3.7) and slope ( $DK_{inexp}$ : 2.2,  $DK_{exp}$ : 1.97), or for the “L”–“J” boundary location (6.1 for both subgroups) and slope ( $DK_{inexp}$ : 2.1,  $DK_{exp}$ : 1.9) found no differences between the Danish subgroups. The two-way ANOVA on percent use of “W” or “V” by the 28 participants (14  $DK_{inexp}$ , 14  $DK_{exp}$ ) who used those categories (shown in Fig. 6, left panel) found a significant Stimulus effect,  $F(9,234)=211.91$ ,  $p < .0001$ , but neither Experience nor its interaction with Stimulus were significant, *ns*.

Fig. 6 (right panel) shows the discrimination functions of the two Danish subgroups on the /w/-/j/ continuum. They did not differ significantly in mean percent correct discrimination ( $DK_{inexp}$ : 92%;  $DK_{exp}$ : 95%), “peakiness” (which was relatively flat:  $DK_{inexp}$ : 6%, for  $DK_{exp}$ : 5%), or discrimination of stimuli straddling the category boundary ( $DK_{inexp}$ : 97%,  $DK_{exp}$ : 96%). Moreover, the two-way ANOVA on discrimination found only a

significant Stimulus Pair effect,  $F(6,168)=11.40$ ,  $p < .0001$ ; neither Experience nor its interaction with Stimulus Pair were significant, *ns*.

### 3.3. Discussion

The difference in spoken English experience between the two Danish listener subgroups in Experiment 2 had no significant impact on any measure of identification or discrimination for AmE /r/-/l/ or /w/-/j/, even though the groups differed in having minimal versus fairly extensive exposure to natively spoken English. Moreover, the open responses from the two Danish groups provided no indication of different assimilation patterns of AmE approximants to Danish categories as a function of spoken English experience. This outcome indicates that it is native language experience, rather than L2-English experience, that shapes Danish listeners’ perception of these two AmE approximant contrasts. This conclusion is, in fact, consistent with numerous findings of persisting L1 influences on L2 perception by not only late L2 learners, but also by very early fluent bilinguals (e.g., Pallier, Bosch, & Sebastian-Galles, 1997; Pallier, Colomé, & Sebastian-Galles, 2000). Importantly for the present study, the null effect of L2-English experience on perception of these two AmE contrasts implies that the excellent performance of the Danish listeners in Experiment 1, and especially their unexpected out-performance of native American English listeners on discrimination of /w/-/j/, were not due to English proficiency but instead to the relationships between Danish and AmE approximant realizations.

The lack of English-experience effects on Danish listeners’ categorization and discrimination of /r/-/l/, or on their categorization of /w/-/j/, is compatible with both SLM and PAM predictions of minimal effects of experience for these two contrasts. However, the significant effects of experience on both their categorization and discrimination of /w/-/r/ is consistent with PAM but not SLM predictions. Even more striking, however, is that again neither PAM nor SLM predicted the exceedingly high discrimination performance that Danish listeners displayed on AmE /w/-/j/, regardless of spoken English experience.

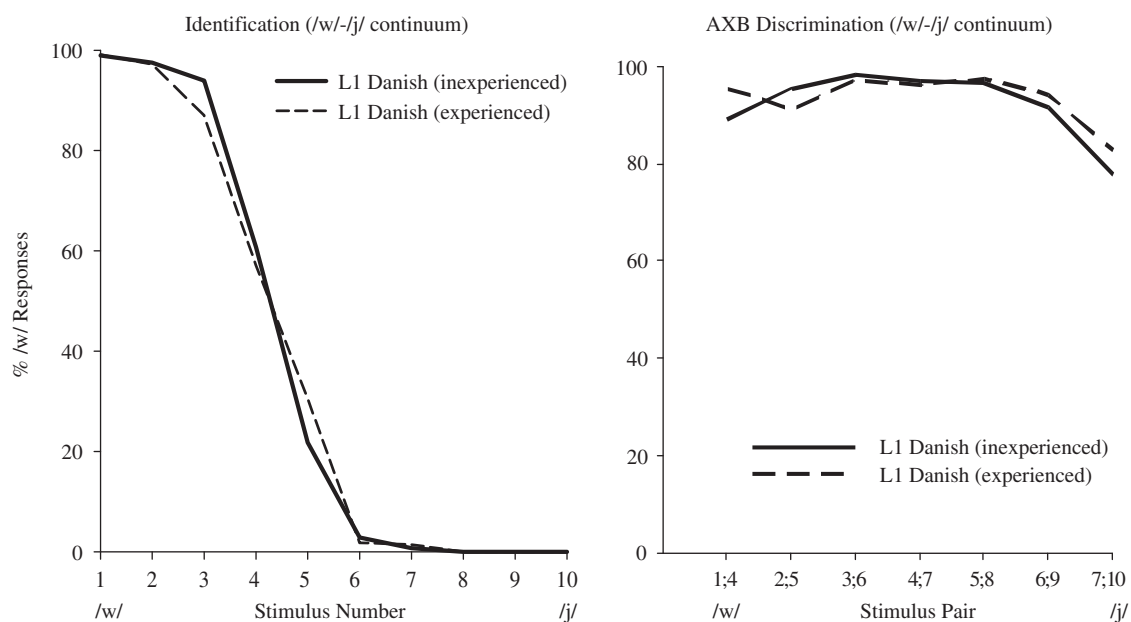


Fig. 6. Identification (left panel) and discrimination functions (right panel) for the /w/-/j/ continuum as perceived by English-inexperienced ( $DK_{inexp}$ ) and English-experienced ( $DK_{exp}$ ) Danish listeners.

We still cannot entirely exclude the possibility that the inexperienced Danish listeners in Experiment 2 had some spoken English experience even though they had not spent time immersed in native English-speaking environments. In Denmark the language of many products of the entertainment industry (e.g., movies, TV shows and documentaries) is presented in spoken English with Danish subtitles, which makes even “inexperienced” Danish speakers still relatively experienced auditorily, as compared to nonnative speakers from countries in which English is used much less frequently in the popular media, such as Germany.

No amount of spoken English experience, however, can explain our most intriguing finding, that Danish listeners discriminate AmE /w/-/j/ much *better* than native English listeners. If anything, English listening experience should have resulted in more English-like performance on this non-Danish contrast. Therefore, Experiment 3 extended the study to native German listeners. Our primary purpose was to examine the role of phonological and phonetic factors in cross-language perception in greater depth, given the similarities in the front-rounded vowel systems but differences in the approximant systems of German and Danish. Also of interest was that German listeners are less experienced than Danes with hearing natively spoken English. Formal English language experience within the German educational system is similar to that of the native Danish listeners in Experiment 1, but they hear much less spoken English in their everyday experience: foreign language media (films, television shows) are almost always dubbed in Germany whereas they are rarely dubbed in Denmark (except for young children’s shows).

## 4. Experiment 3

The German approximant inventory (/r l j/) is simpler than English (/r l j w/) and especially Danish (/r l j/) and approximant realizations of /v/ and intervocalic /b d g/). While German /j/ is typically produced as a palatal approximant identical to /j/ in English, Danish, French and Japanese, German displays key differences from English in its phonetic realization of the other target approximants. German /r/ is typically realized as a voiced uvular fricative/approximant [ʁ/ʀ], as in French, but unlike the realizations of /r/ in English, as well as that of Danish (or of Japanese). German /l/ is “light” [l] like that of French and Danish, thus differing phonetically from the English velarized/pharyngealized alveolar [ɫ]. German, like Danish, lacks /w/, and its closest phonetic neighbor is /v/. However, German /v/ is a labio-dental voiced fricative [v], identical to the /v/ of English and French but differing from the Danish /v/ (labiodental approximant [ʋ]). These phonetic and phonological characteristics of German, relative to English and Danish, make it particularly useful for further tests of the predictions laid out in the general Introduction (Table 2). In particular, it provides a further probe into possible factors that may underlie the unexpectedly high discrimination performance of Danish and French listeners on the AmE /w/-/j/ contrast.

### 4.1. Method

#### 4.1.1. Participants

Eighteen native northern German speakers participated as unpaid volunteers. They were students at Kiel University and were in all other critical ways comparable to the participants in Experiment 1 (10 females, 8 males,  $M_{\text{age}}=21.3$  years,  $s.d.=1.9$ ). A background questionnaire confirmed that each subject met the following selection criteria, corresponding to those used for the Danish participants of Experiment 1: no history of hearing loss, native German speaker, native German speaking parents, and limited experience in language environments other than German

(i.e., < 8 months immersed in a foreign language environment). All of them realized GE /r/ as [ʁ] or [ʀ].<sup>8</sup> Their mean self-ratings (on a scale of 1 = poor to 5 = excellent) of English proficiency were 3.5 (SD=.8) for *speaking* and 3.7 (SD=1.0) for *understanding* spoken English.

#### 4.1.2. Stimulus materials

The same stimuli were used again, as in Experiments 1–2.

#### 4.1.3. Procedure

The procedure was identical to Experiment 1, except that testing took place at the Kiel University language laboratory. The set-up and headphones were comparable in design and quality to those of the Aarhus University language lab (Experiment 1).

## 4.2. Results

Data analysis followed the same approach as in Experiment 1.

### 4.2.1. The /r/-/l/ contrast

Fig. 7 displays /r/-/l/ identification and discrimination by German and English listeners. The one-way ANOVAs on the probit-derived categorization measures found no significant Language difference for boundary location (German: 5.4; English: 5.4) or slope (German: 1.5; English: 2.2). The two-way ANOVA on the full categorization function found only a significant Stimulus item effect,  $F(9,225)=203.25$ ,  $p < .0001$ , but no Language difference or interaction, *ns*, despite apparent discrepancies in German labeling near both ends of the continuum, especially near the /l/ end (Fig. 7, left).

The one-way ANOVAs on mean percent correct discrimination (German: 73%; English: 78%), “peakiness” (German: 11%; English: 12%), and cross-category discrimination (German: 85%; English: 93%) were all nonsignificant, *ns*. However, significant effects were found by the two-way ANOVA for Stimulus Pair,  $F(6,150)=26.66$ ,  $p < .0001$ , and for the Language  $\times$  Stimulus Pair interaction,  $F(6,150)=3.21$ ,  $p = .005$ . Simple effects tests on the interaction indicate that discrimination was significantly lower for German than English listeners on the /l/ side of the category boundary, specifically for stimulus pairs 4–7 (German: 80% correct; English: 94%),  $F(1,25)=6.08$ ,  $p < .02$ ; and 5–8 (German: 74%; English: 86%),  $F(1,25)=4.69$ ,  $p < .04$ ; and 6–9 (German: 68%; English: 81%),  $F(1,25)=4.91$ ,  $p < .03$ .

### 4.2.2. The /w/-/r/ contrast

Fig. 8 shows the groups’ identification and discrimination functions for /w/-/r/. The one-way ANOVAs on boundary location (German: 5.4; English: 4.8) and slope (German: 1.5; English: 1.7) failed to find significant group differences, *ns*. However, the two-way ANOVA revealed not only a significant Stimulus item effect,  $F(9,225)=313.23$ ,  $p < .0001$ , but also a main effect of Language (German mean: 48% “W” responses; English mean: 44% “W”),  $F(1,25)=4.53$ ,  $p < .05$ , and a significant interaction,  $F(9,225)=2.69$ ,  $p = .005$ . Simple effects tests of the interaction found that the Germans gave significantly more “W” responses than the English listeners for stimulus items 5 (German: 50% correct; English: 34%),  $F(1,25)=9.52$ ,  $p < .002$ , and 6 (German: 40% correct; English: 18%),  $F(1,25)=17.94$ ,  $p < .0001$ , that is, just near the /r/ side of the category boundary.

As for discrimination, neither of the one-way ANOVAs on “peakiness” (12% for both groups) nor on cross-category discrimination (German: 82%; English: 86%) differed significantly between the two groups, *ns*. However, English listeners’ mean

<sup>8</sup> None realized it as an apical trill [r] or came from an area where /r/ is commonly realized as [r].



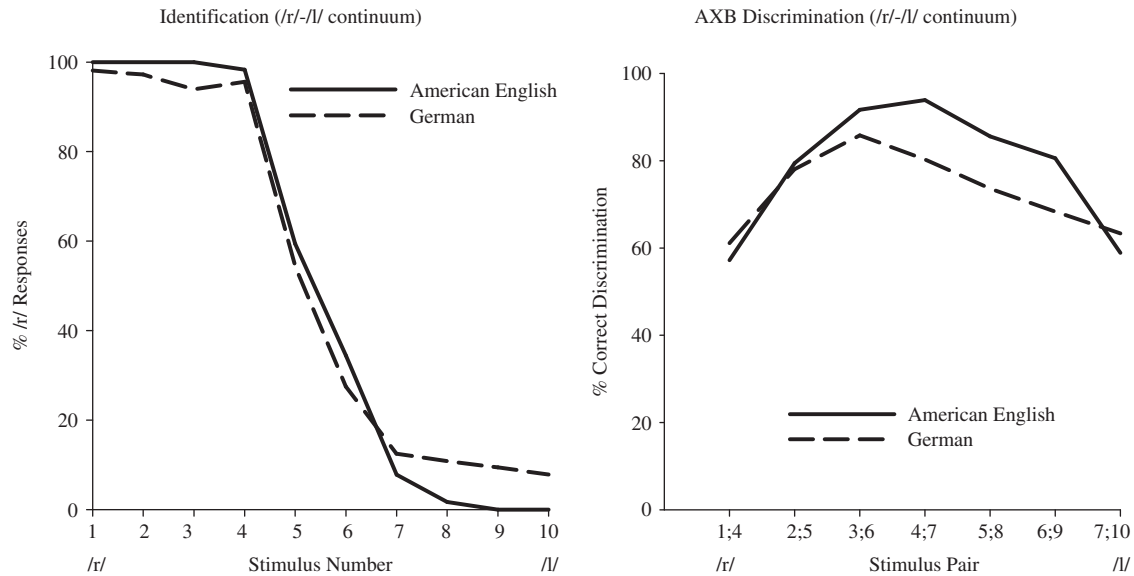


Fig. 7. Identification (left panel) and discrimination functions (right panel) for the /r/-/l/ continuum as perceived by native German and American English listeners.

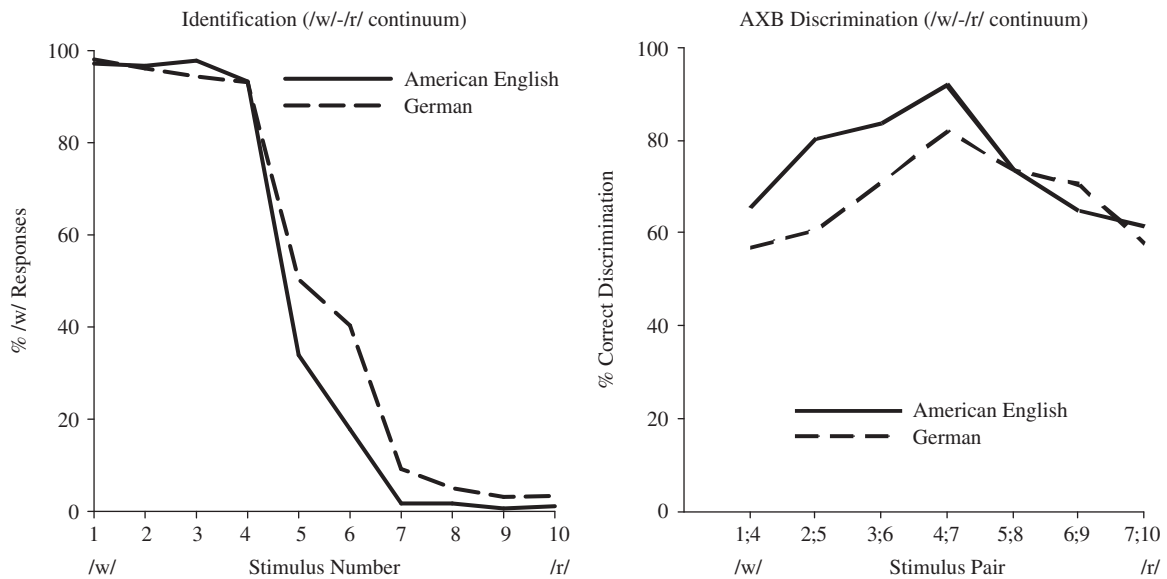


Fig. 8. Identification (left panel) and discrimination functions (right panel) for the /w/-/r/ continuum as perceived by native German and American English listeners.

discrimination was significantly higher (75%) than that of the Germans (68%),  $F(1,25)=3.29$ ,  $p < .05$ . The two-way ANOVA provided further insight into this difference. Both the Language,  $F(1,25)=5.72$ ,  $p < .03$ , and Stimulus Pair,  $F(6,150)=14.44$ ,  $p < .0001$ , main effects were significant, as was their interaction,  $F(6,150)=2.77$ ,  $p < .02$ . Simple effects tests found significantly lower discrimination by German than by English listeners on the /w/ side of the boundary, for stimulus pairs 2–5 (German: 61%; English: 81%),  $F(1,25)=12.79$ ,  $p < .0001$ , and 3–6 (German: 71%; English: 84%),  $F(1,25)=5.40$ ,  $p < .03$ ; the difference was marginally significant for stimulus pair 4–7 (German: 82%; English: 92%),  $F(1,25)=3.29$ ,  $p = .07$ .

#### 4.2.3. The /w/-/j/ contrast

The /w/-/j/ identification and discrimination functions are shown in Fig. 9. The one-way ANOVAs found no reliable differences between German and English listeners for either boundary location (German: 5.6; English: 5.4) or slope (German: 1.8;

English: 1.9). Nor was there any evidence of a listener group difference in the two-way ANOVA, where only the Stimulus item effect was significant,  $F(9,225)=203.79$ ,  $p < .0001$ .

The discrimination analyses, however, revealed large and pervasive group differences. The one-way ANOVAs found that German listeners were much more accurate than native English listeners on mean discrimination (German: 90% correct; English: 75%),  $F(1,25)=22.69$ ,  $p > .001$ , and across the category boundary,  $F(1,25)=9.57$ ,  $p < .001$ , while “peakiness” was greater for English listeners (German: 7%; English: 9%),  $F(1,25)=3.16$ ,  $p = .05$ . The two-way ANOVA revealed significant main effects of Stimulus Pair,  $F(6,150)=5.65$ ,  $p < .0001$ , and Language group,  $F(6,150)=26.13$ ,  $p < .0001$ . Simple effects tests of the significant interaction,  $F(6,150)=2.37$ ,  $p < .05$ , showed that the Germans outperformed the native English listeners on all but one stimulus pair throughout the continuum,  $F_s(1,25)$  ranged from 6.69 to 23.72,  $p < .02$  to  $.0001$ , the exception being pair 3–6, which is near the higher English peak (see Best & Strange, 1992, for an account of their two peaks).

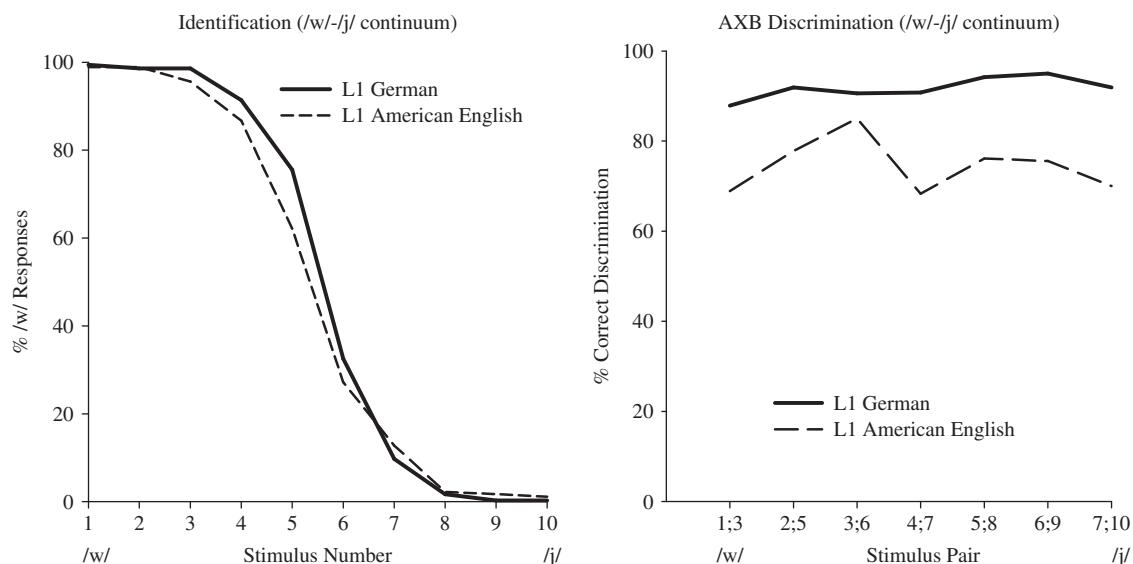


Fig. 9. Identification (left panel) and discrimination functions (right panel) for the /w/-/j/ continuum as perceived by native German and American English listeners.

#### 4.3. Discussion

As with Experiments 1–2, the present findings are not fully consistent with any of the three classes of predictions presented in Section 1; however, they are more in line with the fine-grained phonetic predictions of PAM and/or SLM than with purely phonological predictions. The /r/-/l/ results are superficially consistent with the phonologically based prediction that native German listeners will show excellent categorical perception that does not differ from native English listeners because, like Danish, German has an /r/-/l/ contrast. However, based on differences between English and German phonetic implementations of /r/ and /l/, SLM predicted in addition that German listeners should show less reliable perception of /r/-/l/ than English listeners near both endpoints of the continuum, while PAM predicted they would differ from English listeners on the /r/ side of the category boundary. While there was evidence of within-category differences between German and English listeners, it was not fully in line with either PAM or SLM. The German listeners appeared to show less reliable labeling than English listeners mainly toward the /l/ end of the continuum, but this difference was not statistically significant. They did, however, show significantly poorer discrimination than English listeners, again on the /l/ side of the boundary. PAM, however, had predicted greater effects on the /r/ side of the boundary where the English–German realization difference is greater, and SLM expected differences near both endpoints of the continuum. One might consider extending SLM to account for the lack of an /r/-end difference, i.e., that it results from Germans' perception of AmE /r/ as “new,” i.e., a new L2 category had already formed due to lack of equivalence-classification to German /r/. However, this reasoning deviates from the earlier extrapolations of SLM principles regarding perception of “new” and “similar” nonnative phones; nor would it apply consistently across other findings in the present report.

All three approaches anticipated that German listeners would perceive /w/-/r/ less categorically than English listeners as a nonnative contrast, though the basis for the prediction as well as expectations about finer-grained group differences varied across the models. While the German listeners did differ from English listeners in several ways on both tasks, those differences did not line up with phonological predictions: German listeners' categorization boundary location and slope, and the “peakiness” of their discrimination function, failed to differ from native

English performance. That is, their performance on both tasks was equivalently categorical to that of native listeners. Their points of divergence from native listeners on AmE /w/-/r/ are instead more compatible with phonetically based within-category differences, addressed by PAM and SLM. Specifically, German listeners gave significantly more “W” responses just near the /r/ side of the category boundary, as compared to English listeners, and showed less accurate discrimination for several within-/w/ stimulus pairs but not for the pair that used the /w/-endpoint. This set of German listener differences is somewhat more compatible, though not perfectly so, with PAM than SLM phonetic-level predictions.

The most striking finding, however, is that once again none of the models correctly predicted that the German listeners' categorization of /w/-/j/ would be indistinguishable from that of native English listeners. Most pointedly, they all failed to anticipate that German listeners' discrimination accuracy on this nonnative contrast would be superior to the native listeners. Overall, German listeners' /w/-/j/ performance poses the same problems for the three theoretical perspectives as had that of Danish listeners: not only did both of these nonnative groups show much better discrimination than anticipated by any of the three perspectives, they both far exceeded the native English listeners on /w/-/j/ discrimination.

#### 5. Cross-language comparisons of the perception of AmE approximants

In order to have an optimally comprehensive foundation for interpreting the results of Experiments 1–3, and especially for discussing the implications of the /w/-/j/ findings, we conducted systematic, direct comparisons between the results of Experiments 1 and 3 and those of two previously published reports. Those other reports had examined perception of the same AmE approximant contrasts by native listeners of Japanese (Best & Strange, 1992) and French (Hallé et al., 1999). One-way ANOVAs were conducted on the same set of probit-derived variables for identification performance, and the summary variables of the discrimination functions, as we had used to examine the Experiments 1 and 3 data, except that for the present analyses the Language factor had 5-levels: English, Danish, German, French and Japanese. We also conducted two-way ANOVAs on the

identification and discrimination functions, as in Experiments 1 and 3 but again with five levels for the Language factor. Simple effects tests and Tukey pairwise comparisons were used to break down significant Language effects and interactions, to inform the General Discussion (Section 6).

### 5.1. Perception of /r/-/l/

The AmE /r/-/l/ identification and discrimination functions of the five listener groups are shown in Fig. 10. In the one-way ANOVAs on identification of the /r/-/l/ stimuli, no reliable Language differences were found in boundary location, *ns*. However, the Language groups did differ significantly in the slopes of their category boundaries,  $F(4,65)=3.32$ ,  $p < .02$ . Tukey pairwise comparisons found significant differences between the boundary slopes of the native Japanese listeners (.9) and those of the American English (2.2), and Danish (1.8) groups; the latter two groups did not differ from one another. The slope values of the German and French listeners (both 1.5) did not differ significantly from any of the other groups.

By comparison, the two-way ANOVA on the full identification functions revealed not only a significant main effect of Stimulus item,  $F(9,586)=519.15$ ,  $p < .0001$ , but also a significant Language  $\times$  Stimulus item interaction,  $F(9,586)=3.13$ ,  $p < .0001$ . Simple effect tests on that interaction found significant listener group differences for Stimulus items 3, and 5 through 9 of the 10-item continuum,  $F(4,582)$  values ranged between 2.14 and 5.08, with  $p$  values ranging from .05 to .005. We applied Tukey pairwise tests for each stimulus item that showed a significant simple effect, to determine which Language group differences had contributed. Group differences were limited to the /l/ side of the boundary with the exception of stimulus 3 (within-/r/), for which the only significant Language difference was that the Japanese listeners gave fewer “R” responses (80%) than all other language groups (English: 100%; Danish: 98%; German: 94%; French: 97%), all  $ps < .01$ . Reliable Language group differences were found only for three of the stimuli on the /l/ side of the continuum. Japanese gave *more* “R” responses to Stimulus 7 (26%) than Danish (5%),  $p < .01$ , or German (9%), French (7%) or English (8%) listeners,  $ps < .05$ . Japanese listeners also gave more “R” responses for stimulus 8 (JA: 21%; others: 2–5%) and 9 (JA: 17%; others: 0–3%), all  $ps < .01$ . No other language group differences were significant.

In the one-way ANOVAs on the /r/-/l/ discrimination comparisons, there was a significant Language effect for mean percent correct performance (English: 78%; Danish: 73%; German: 73%; French: 73%; Japanese: 64%,  $F(4,65)=3.48$ ,  $p < .02$ ). Pairwise Tukey tests revealed that the Japanese listeners had a significantly lower mean discrimination level than the native English listeners; all other between-group differences were non-significant. The Language effect for “peakiness” was nonsignificant, *ns*. A significant Language effect was also found for discrimination accuracy at the category boundary,  $F(4,65)=4.68$ ,  $p < .01$ . Japanese listeners were significantly less accurate (68%) than English (93%), Danish (84%) or German listeners (85%), who did not differ significantly from each other. The French listeners (79%) did not differ from the Japanese listeners or any of the other groups. No other between-group differences reached significance in the one-way ANOVAs, but the two-way ANOVA yielded significant effects of Language,  $F(4,65)=4.84$ ,  $p < .02$ , Stimulus item,  $F(6,390)=41.31$ ,  $p < .0001$ , and a Language  $\times$  Stimulus item interaction,  $F(24,390)=1.82$ ,  $p < .02$ . In the Tukey tests on the Language effect, the only groups that differed significantly were the English from the Japanese listeners,  $p < .01$ . Simple effects tests on the interaction indicate that the listener group differences were concentrated in the boundary region, specifically for stimulus pair 3–6,  $F(24,343)=4.79$ ,  $p < .001$ , and pair 4–7,  $F(4,343)=5.85$ ,  $p < .0001$ , and 5–8,  $F(4,343)=3.14$ ,  $p < .02$ . Tukey pairwise comparisons on the interaction found that native English listeners discriminated pair 4–7 better than Japanese and French,  $ps < .01$ , or Danish and German listeners,  $ps < .05$ , and they discriminated pair 5–8 better only than the Japanese,  $p < .05$ . No other between-group discrimination differences were found.

### 5.2. Perception of /w/-/r/

Fig. 11 compares the identification and discrimination functions of the five listener groups for /w/-/r/. The one-way ANOVA on the category boundary values revealed significant differences among the five groups,  $F(4,65)=5.57$ ,  $p < .001$ . Tukey pairwise comparisons found that the French boundary (6.7) differed significantly from the other listener groups (English: 4.8; Danish: 5.1; German: 5.4; Japanese: 5.6),  $ps < .05$ . All other between-group differences were nonsignificant, *ns*. The one-way ANOVA

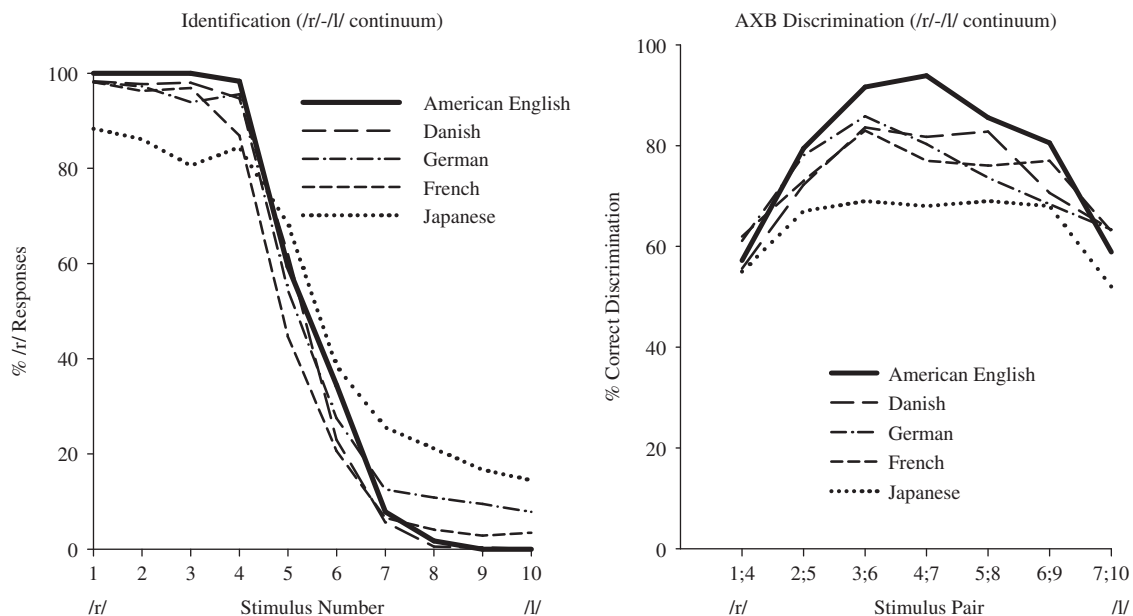


Fig. 10. Identification and discrimination functions for the /r/-/l/ continuum as perceived by native American English, Danish, German, French and Japanese listeners.

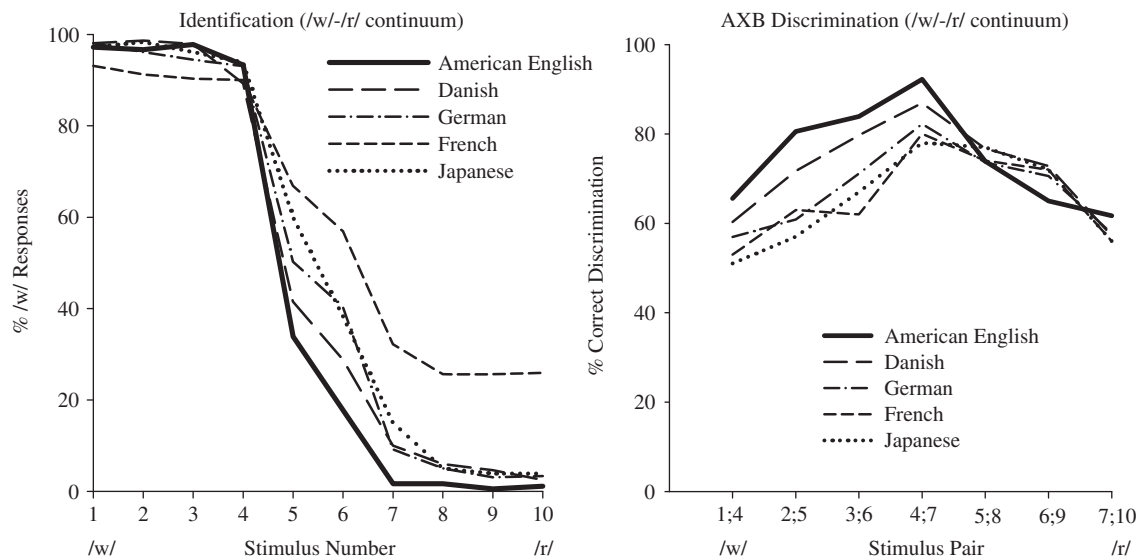


Fig. 11. Identification and discrimination functions for the /w-/r/ continuum as perceived by native American English, Danish, German, French and Japanese listeners.

on slope values was also significant,  $F(4,64)=5.11$ ,  $p > .01$ . Tukey tests indicated that the English and Danish groups had steeper category boundaries (slopes of 1.7 and 1.5, respectively) than the French group (.8),  $ps < .05$ . Although the German and Japanese mean slopes were numerically intermediate to the Danish and French values (both 1.2), neither those nor any other group differences were significant, *ns*. The two-way ANOVA on Language  $\times$  Stimulus provided information about where exactly these group differences occurred within the continuum. In addition to a significant Stimulus item effect, which indicates strongly categorical identification across groups,  $F(9,585)=587.41$ ,  $p < .0001$ , there was a main effect of Language,  $F(4,65)=5.24$ ,  $p < .001$ , for which Tukey tests found the French listeners reported hearing “W” significantly more often (59%) on average across the continuum than English (44%),  $p < .01$ , Danish (49%), German (49%), or Japanese listeners (51%),  $ps < .05$ . Simple effects tests of the significant Language  $\times$  Stimulus item interaction,  $F(36,585)=4.02$ ,  $p < .0001$ , found this difference to be localized to the /r/ side of the continuum, where there were significant Language effects for stimulus items 5–10,  $F(4,416)$  values ranged from 5.06 to 10.47,  $ps$  ranged from .0001 to .001. Tukey pairwise tests of Language group differences on each of the affected /r/ stimuli found that the French listeners gave more “W” responses than the English listeners on all five stimuli 5–10,  $ps < .01$ . The French also gave more “W”s than Danish listeners on stimuli 6 through 8,  $ps < .05$ , more than German and Japanese listeners on stimulus 8,  $ps < .05$ , and more “W”s than all the other groups on stimuli 9–10,  $ps < .01$ . The only other significant group difference was that for stimulus 7 alone, Japanese listeners gave significantly more “W” responses than English, Danish and German listeners, but the French gave more “W”s than the Japanese,  $ps < .05$ . No other group differences were found for any stimulus items, *ns*.

The one-way ANOVA on Language for mean percent correct discrimination was significant,  $F(4,65)=3.83$ ,  $p < .01$ . Tukey tests found that the English listeners (mean 75% correct) differed significantly from the French and Japanese listeners (both 66%),  $p < .05$ , but not from the Danish (72%) or German listeners (68%); no other pairwise comparisons reached significance, *ns*. Neither “peakiness” nor discrimination accuracy at the boundary showed significant Language effects in the remaining one-way ANOVAs.

The two-way ANOVA brought out additional finer-grained differences in the groups’ discrimination functions. Not only were

there significant main effects for Language,  $F(4,65)=4.10$ ,  $p < .005$ , and Stimulus item,  $F(6,390)=38.41$ ,  $p < .0001$ , but their interaction was also significant,  $F(4,65)=1.93$ ,  $p < .006$ . Tukey tests on the Language effect found significantly better discrimination overall by English than French or Japanese listeners,  $ps < .05$ ; no other group differences were significant. Simple effects tests on the interaction found significant Language group differences for the four discrimination pairs on the /w/ side of the continuum (1–4, 2–5, 3–6, 4–7),  $F(6,384)$  values ranged from 2.36 to 6.25,  $ps$  ranged from .05 to .0001. Tukey comparisons found no pairwise group differences on stimulus pair 1–4, but the English listeners significantly outperformed Japanese and French listeners on stimulus pairs 2–5 and 3–6,  $ps < .05$  or .01, and they also outperformed the Japanese on pair 4–7, and the Germans on pair 2–5,  $p < .01$ , while the Danish outperformed the French on pair 3–6,  $p < .05$ . No other group differences were found, *ns*.

### 5.3. Perception of /w-/j/

The five groups’ identification and discrimination functions for /w-/j/ are displayed in Fig. 12. The one-way ANOVAs found significant group differences in category boundary location,  $F(4,65)=4.34$ ,  $p < .01$ . Tukey tests indicated that the Japanese listeners’ boundary (6.5) was shifted significantly rightward, toward the /j/ end of the continuum, relative to the Danish (5.1) and French listeners (4.9). No other group differences were found, *ns*. The language group effect for the slope values was nonsignificant, *ns*. The two-way ANOVA provided more detail about the region of the continuum that showed group differences. Both main effects were significant: Stimulus item,  $F(9,585)=467.21$ ,  $p < .0001$ , and Language group,  $F(4,65)=4.79$ ,  $p < .002$ . Tukey pairwise comparisons of the Language effect indicated that the Japanese used more “W” responses on average across the continuum than the English listeners,  $p < .05$ , as well as more than the Danish and French listeners,  $ps < .01$ . More detail is offered by the interaction of Language  $\times$  Stimulus item,  $F(36,585)=2.90$ ,  $p < .0001$ . Simple effects tests on this interaction indicated significant Language differences for stimulus items 4–7, near the boundary and to the /j/ side of it,  $F(4,473)$  values ranged between 3.00 and 15.76,  $ps$  ranged between .02 and .0001. Tukey pairwise comparisons of the Language effect indicated that Japanese listeners gave significantly more “W” responses than all other groups on stimuli 6 and 7, on the /j/ side of the continuum,



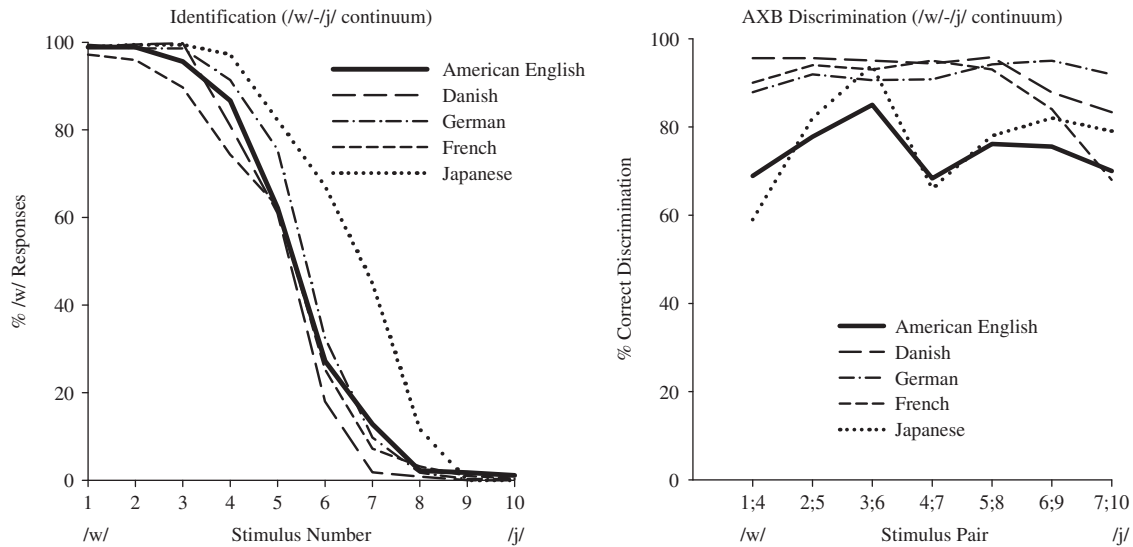


Fig. 12. Identification and discrimination functions for the /w/-/j/ continuum as perceived by native American English, Danish, German, French, and Japanese listeners.

$p < .01$  ( $< .05$  for Germans on item 6), but there were no reliable group differences on stimuli 4 and 5, *ns*.

The one-way ANOVA on overall percent correct discrimination scores differed significantly across the language groups,  $F(4,65) = 16.51$ ,  $p < .001$ . Tukey tests found that this was due to the Danish, German and French groups (Danish: 93%; German: 90%; French: 88%), which did not differ from one another, significantly outperforming the English group (75%),  $p < .05$ . The Danish and German listeners also had significantly higher scores than the Japanese group (77.1%),  $p < .05$ , whose scores did not differ significantly from either the English or the French group. The one-way ANOVA on language effect on discrimination accuracy at the category boundary was significant,  $F(4,65) = 9.95$ ,  $p > .001$ ; Tukey comparisons found that the Danish (94%), German (93%), and French (96%) listeners, who did not differ from one another, were significantly more accurate than the English (78%) and Japanese (81%) listeners, who did not differ from one another. The one-way ANOVA on “peakiness” scores was also significant,  $F(4,65) = 6.56$ ,  $p < .001$ , due to Japanese listeners having a significantly higher “peakiness” score (13%) than Danish (6%), German or French (both 7%) listeners. The English listeners (9%) did not differ significantly from any other group, and there were no further group differences on this measure.

The two-way ANOVA provided additional detail on the locations and degrees of group differences in discrimination throughout the continuum. Both the Language effect,  $F(4,65) = 18.24$ ,  $p < .0001$ , and the Stimulus pair effect,  $F(6,390) = 21.51$ ,  $p < .0001$ , were significant, as was their interaction,  $F(36,390) = 7.21$ ,  $p < .0001$ . Tukey tests of the Language effect found that the English and Japanese groups, which did not differ from each other, showed significantly poorer discrimination overall than the Danish, French and German groups,  $p < .01$ ; the latter three groups did not differ from one another. Simple effects tests on the interaction revealed significant Language group differences at each stimulus pair except for the stimulus 3–6 comparison,  $F(4,65) = 5.64$  and 27.25,  $ps < .0001$ . Tukey tests on the significant simple effects stimulus pairs indicated that the Japanese and English groups, which did not differ from each other, showed significantly poorer discrimination for pairs 1–4, 2–5, 4–7 and 5–8 than did the Danish, German and French listeners,  $ps < .01$ . The latter groups did not differ from one another, *ns*. For stimulus pair 6–9, only the American and German listener groups differed,  $p < .01$ ; no reliable group differences emerged for pair 7–10 despite the fact that it had shown a significant simple Language effect.

## 6. General discussion

The experiments reported here were designed to test the influence of phonetic and phonological properties of two native languages, Danish and German, on the perception of the American English (AmE) approximants /r l w j/. We found that neither phonological considerations nor the phonetically-based predictions of Flege’s Speech Learning Model (SLM) and of Best’s Perceptual Assimilation Model (PAM) were completely successful in predicting the nonnative listeners’ discrimination and identification of stimuli drawn from three AmE approximant continua, /r/-/l/, /w/-/r/, and /w/-/j/. Overall, however, phonetic similarities were better predictors of cross-language perception of AmE approximants than phonological correspondences. Below, we summarize the findings for the three continua with respect to the predictions of the theoretical perspectives laid out in Section 1, and we focus our discussion on those aspects of the results that are problematic for the predictions generated by PAM and by SLM. In this discussion we will also draw on our comparisons in Section 5 with the previous studies that examined the perception of the AmE approximants by native speakers of Japanese (Best & Strange, 1992) and French (Hallé et al., 1999).

The phonological predictions fared best for the perception of the /r/-/l/ contrast. Both native Danish and native German listeners, whose L1s have an /r/ and an /l/ phoneme, perceived this contrast much like AmE listeners. Their identification of /r/-/l/ was statistically indistinguishable from AmE listeners, and their discrimination was quite similar to the native listeners. While both PAM and SLM predicted categorical perception of AmE /r/-/l/ by the nonnative listeners, these models also predicted that the considerable phonetic differences between the realizations of /r/ and /l/ in AmE as opposed to Danish and German would result in clear differences between the three listener groups. However, phonetic sensitivity to differences between native approximants and AmE [ɹ] and [ɫ] was only revealed by a flatter between-category discrimination (for the native Danish listeners), which was correctly predicted by PAM, and lower discrimination on the /l/ side of category boundary (for the native German listeners), which was the reverse of predictions by SLM that perception would be more compromised on the /r/ side of the continuum.

Why were the native Danish and the native German not (much) affected by the large phonetic differences between native approximants and AmE /r/-/l/? One could surmise that the Danish

and German listeners of Experiments 1 and 3, respectively, were too experienced with English to serve as subjects in tests of predictions that were generated for inexperienced listeners. (Recall that PAM was originally designed for naïve cross-language perception and that the SLM predictions used in the present study were extrapolated from core SLM principles for inexperienced L2 learners.) To address this question, we conducted Experiment 2 with two groups of native Danish listeners differing in English language experience. We found no differences in either identification or discrimination of the /r/-l/ continuum between the experienced and the inexperienced native Danish listeners, suggesting that the native-like performance of the Danish listeners in Experiment 1 (and the German listeners in Experiment 3) was probably due to factors other than English language experience. However, given that our subjects – university students in Denmark and in Germany – have had English as an obligatory first foreign language for several years, and that they cannot avoid exposure to English in everyday life, we cannot exclude the possibility that English language experience contributed to the near-native perception of AmE /r/-l/. This question should be addressed in future experiments, perhaps with young children who have much less English language experience than the young adults of our experiments.

Assuming that the results of Experiment 2 do indeed indicate that the near-native perception of /r/-l/ was not primarily related to English language experience, why were the Danish and German listeners almost unaffected by the phonetic differences between their native /r/s and /l/s and AmE /r-l/? Both PAM and SLM predict (in PAM terms) Two Category assimilation, which should lead to categorical perception, but both models assume that the phonetic differences between native and non-native realizations of /r/ and /l/ should compromise within-category perception. We suggest that several factors may conspire so that non-native listeners who TC-assimilate AmE /r-l/ may perceive it in a native-like fashion despite phonetic differences. First of all, the unique phonetic properties of AmE [ɹ] (a “bunched” dorsal or retroflex approximant with additional labial and pharyngeal constrictions, resulting in a very low F3) may make it easy to keep it perceptually distinct from any other consonant in Danish or German. Secondly, higher-level phonological properties of /r/, such as its quite similar phonotactic patterning across AmE, Danish, and German (see footnote 2), could contribute to the ease with which the non-native listeners identified and discriminated AmE /r/-l/ (see also Best & Tyler, 2007).

Our additional cross-language analyses of the perception of AmE /r/-l/ (Section 5) revealed that the French listeners' perception of AmE /r/-l/ was very similar to the native Danes' and the native Germans', and that only the native Japanese listeners differed from the other language groups. This could lead one to expect that if a language has an /r/-l/ contrast (irrespective of how it is phonetically realized), then speakers of that language will perceive AmE /r/-l/ very well, not just because of TC assimilation (PAM) or equivalence classification (SLM), but also because of the unique phonetic properties of AmE [ɹ] and of the native-language-like higher level phonological characteristics of AmE /r/. To test this assumption, and to isolate the factors that may contribute to the very good perception of AmE /r/-l/, it would be instructive to examine the perception of AmE /r/-l/ in listeners whose native language has (a) an /r/ phoneme that is realized quite differently from the /r/s in Danish, German, or French (e.g., apical trill [r] in Spanish), and/or (b) an /r/ that behaves differently from theirs in (some) phonological processes (e.g., syllabic /r/ in Slavonic languages).

The other two approximant contrasts examined in this study involved /w/, which occurs in neither Danish nor German. Accordingly, the phonological prediction was that the non-native

listeners should have considerable difficulties with AmE /w/-r/ and /w/-j/. This prediction was clearly wrong: Both Danish and German listeners showed highly categorical identification and discrimination of /w/-r/, and they identified /w/-j/ categorically while discriminating this contrast in a continuous fashion and at much higher levels than the native AmE listeners. These results are clearly more in line with phonetically based PAM and SLM predictions, which both anticipated categorical perception of /w/-r/ and /w/-j/, though they differed in some predicted details. Interestingly, however, both PAM and SLM failed to predict the near-ceiling discrimination of /w/-j/ by the non-native listeners.

Identification of /w/-r/ by the non-native listeners was almost indistinguishable from the AmE listeners, except that both native German and native Danish listeners gave more /w/ responses to two stimuli near the /r/ side of the category boundary. This indicates some modest sensitivity to the phonetic differences between AmE [w-ɹ] and their corresponding Danish ([v-ʁ]) and German ([v-ʁ]) contrasts, which is somewhat more consistent with PAM than SLM predictions. According to SLM, differences between native AmE and non-native listeners should be evident primarily near the endpoints of the continuum, whereas PAM anticipated sensitivity to cross-language differences in within- and/or between-category perception. This sensitivity was not evident in the native Danish listeners' discrimination of /w/-r/, which was indistinguishable from the native AmE listeners'. The native German listeners, however, discriminated /w/-r/ less well on the /w/ side of the continuum, showing a sensitivity to cross-language differences between German [v] and AmE [w]. The fact that the native Danish listeners were not affected by cross-language phonetic differences in their discrimination of /w-r/, while the native German listeners were, could be due to the relatively large difference between the German labiodental fricative [v] and English [w] relative to the small phonetic difference between the Danish labiodental approximant [v] and English [w].

Again, the results of Experiment 2 with two native Danish listener groups differing in English language experience are helpful in interpreting these results. Experiment 2 revealed that the difference in experience with natively spoken English did not affect any of the measures of categorical perception (slope, boundary location, “peakiness”, cross-category discrimination). This could suggest that English language experience is unlikely to be the reason for the very good perception of /w-r/ by the Danish listeners in Experiment 1 and by the German listeners in Experiment 3, and thus that the one major difference between the two language groups – German listeners' lower discrimination levels on the /w/ side of the continuum – is a genuine language effect. Specifically, the language group difference seems more likely to be due to the mismatch between German [v] and English [w] than to potential differences in spoken English experience between the German and Danish listeners.

The cross-language comparisons with earlier studies further confirmed the importance of fine phonetic detail in the perception of AmE approximants. The most striking result of this comparison was that the groups which deviated most from the AmE listeners were the native French and the native Japanese listeners, even though both languages have a /w-r/ contrast. On phonological grounds, perception of AmE /w/-r/ by native French and native Japanese listeners should be perfect. However, unlike the other language groups, the French listeners did not consistently label stimuli near the /r/ end of this continuum as /r/, and both the French and Japanese listeners' discrimination was less accurate than that of the other groups.

However, neither PAM nor SLM were completely successful in predicting how fine phonetic detail would affect perception of /w/-r/. Both models predicted that native French listeners would be more accurate in their perception of stimuli from the /w/ side

of the continuum because English and French /w/ is realized as [w], whereas the other listener groups should show some sensitivity to the mismatches of AmE [w] with Danish [v], German [ʋ], and Japanese [ɰ]. That was not the case; all non-native listeners identified the /w/ side of the continuum perfectly. Moreover, only the Japanese, and more surprisingly the French, listeners discriminated the /w/ side of the continuum less accurately than the other groups.

It appears that neither phonological considerations nor differences in fine phonetic detail can account for the native Danish and the native German listeners' unexpectedly good perception of /w/-/r/. Even though Experiment 2 showed that, for Danish listeners, spoken English experience was unrelated to perception, it could be that the differences between the French and Japanese listeners, on the one hand, and the Danish and German listeners on the other, result from the status of English (in the educational systems and the public sphere at large) in the native communities. For relatively well educated Danes and Germans, auditory exposure to English is to be expected even for relatively inexperienced speakers, whereas this was not the case for the native speakers of Japanese and of French in the prior studies (Best & Strange, 1992; Hallé et al., 1999). The very good perception of /w/-/r/ by native speakers of Danish and of German may be due to their relatively high level of experience with English. It would be instructive to examine the perception of AmE /w/-/r/ with relatively young native speakers of Danish and German whose spoken English experience is much more limited.

For the /w/-/j/ continuum, the non-native listeners' identification was indistinguishable from the AmE listeners, which is incompatible with the phonologically based prediction that the lack of /w/ in Danish and in German should cause perceptual problems on the /w/ side of the continuum. The identification results were more in line with the PAM and SLM predictions, which anticipated categorical perception. However, these models also predicted that differences in fine phonetic detail would cause less correct identification on the /w/ side of the continuum, which was not the case.

Again, the results from Experiment 2, which compared two native Danish listener groups differing in English language experience, are useful in interpreting this surprising result. Specifically, these results indicate that the nativelike identification of AmE /w/-/j/ by native Danish and native German listeners in Experiment 1 and Experiment 3 is unlikely to be due to spoken English language experience.

The most striking result of the present study is the unpredicted near-ceiling and continuous discrimination of /w-j/ by the Danish and German listeners. Indeed, on logical grounds alone, better-than-native discrimination of /w/-/j/ by these two groups, and especially by French listeners (Hallé et al., 1999), clearly could not have derived from spoken English experience. Increasing amounts of spoken experience should have resulted in discrimination performance increasingly more like American listeners (i.e.,  $DK_{exp} > DK_{inexp}$  and  $GE \gg FR$ ). It did not. The models presented in Section 1 all failed to predict that the discrimination of these three non-native listener groups would be superior to AmE listeners.

The cross-language comparisons in Section 5 may help understand why the pattern of results for the Danish and German listeners' discrimination of /w/-/j/ is inconsistent with the predictions of all three models. One way to address this puzzle is to consider what French, Danish, and German have in common, as well as what distinguishes them all from English and Japanese, which enables listeners from these languages to discriminate a /w/-/j/ continuum at very high levels, well above the performance of American English listeners. Hallé et al. (1999) suggested that the superior performance of the native French listeners was due

to a richer inventory of approximants in French (with /r l j w ɥ/) than AmE (with /r l j w/) and Japanese (with /r j w/), which could result in greater sensitivity to approximant contrasts. However, the present results make it clear that this cannot be the only reason for the French listeners' superior performance. Both Danish and German have only three syllable-initial approximant phonemes (/r l j/), while on the other hand Danish has notably more allophonic approximants in medial position than either French or German. Despite these phonological and allophonic differences among the three languages with respect to approximants, Danish and German listeners discriminated AmE /w/-/j/ much like French listeners, i.e., exceedingly well and much better than native English listeners.

But there is another systemic factor that French, Danish and German do have in common, and that distinguishes them from English and Japanese: a set of front rounded vowels which they systematically distinguish from corresponding front unrounded vowels. That is, lip rounding is distinctive for high front vowels in all three languages whose listeners discriminated /w/-/j/ at very high levels, but rounding is not phonologically distinctive, nor does it occur phonetically, for high front vowels in the other two languages whose listeners showed overall lower and more "peaky" discrimination of /w/-/j/. We suggest that the highly overlearned sensitivity to lip rounding distinctions in vowels enables native listeners of languages with such distinctions to discriminate an approximant contrast near ceiling, if this approximant contrast is importantly differentiated through lip rounding, as is /w/-/j/ but not, e.g., AmE /w/-/r/. This reasoning is consistent with the semivowel status of /w/ and /j/, i.e., they are short, nonsyllabic versions of a rounded (/u/) versus an unrounded vowel (/i/).

Our proposal that the design of the listener's native vowel system affects cross-language approximant perception generates testable predictions, and it calls for amendments of both PAM and SLM. We predict that native speakers of other languages with vowel systems in which lip rounding contrasts are important would discriminate AmE /w-j/ at similarly high levels as the native Danish, German, and French listeners. This should be the case for native Swedish and native Norwegian listeners (both languages have /i y ɥ u/), or for native Turkish listeners (Turkish has /i y u u/), with these three languages having /j/ but not /w/, just like Danish and German. It would also be informative to explore how native speakers of languages with a rounding contrast only for back vowels, such as Korean or Portuguese (both with /i u u/ but lacking /y/), would perceive the AmE /w/-/j/ contrast, especially because Korean lacks /w j v/, and Portuguese has only /v/ but not /w j/. Conversely, we predict that speakers of languages in which lip rounding is not contrastive but yet have counterparts to which AmE /w-j/ could be assimilated (like Farsi, Hebrew, or Czech with /i/-/u/ and /v/-/j/) would not discriminate /w/-/j/ at continuously high levels like Danish, German and French listeners, but rather would perform similarly to English listeners or perhaps even show phonetic-level differences from English listeners at the /w/ end of the continuum.

In conclusion, our interpretation of the present results for the discrimination of AmE /w-j/ by the native French, native German, and native French listeners makes it necessary to extend both PAM and SLM to address not only the contribution of native contrasts and their phonetic characteristics to cross-language speech perception, but also the impact of more general characteristics of the native inventory. Our results suggest that a basic phonological principle of the native vowel system, specifically whether or not it uses contrastive rounding, may affect listeners' sensitivity to phonetic evidence of rounding differences in non-native approximant consonant contrasts. We have suggested several ways in which this novel hypothesis could be critically

tested. If it finds wider support, certain fundamental assumptions about the major class distinction between consonants and vowels may need to be reconsidered.<sup>9</sup> We welcome alternative hypotheses and further examination of the factors contributing to better-than-native discrimination of nonnative contrasts such as /w-/j/.

## Acknowledgments

Supported by NIH grant DC00403 to C. Best and by a grant from the Danish National Research Foundation to O.-S. Bohn. We thank Rikke Bundgaard-Nielsen, Eva Sophia Myers Laris, Signe Meinicke, and Meike Bohn for collecting, entering, and reducing the Danish and German data, and Anja Steinlen and Thorsten Piske for recruiting subjects in Germany.

## References

- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issue in cross-language research* (pp. 171–204). Baltimore, MD: York Press.
- Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic features on cross-language perception of approximants. *Journal of Phonetics*, 20, 305–330.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. Munro, & O.-S. Bohn (Eds.), *Second language speech learning* (pp. 13–34). Amsterdam: John Benjamins Publishing.
- Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language*, 26, 191–211.
- Boyce, S., & Espy-Wilson, C. (1997). Coarticulatory stability in American English /r/. *Journal of the Acoustical Society of America*, 101, 3741–3753.
- Delattre, P., & Freeman, D. C. (1968). A dialect study of American R's by x-ray motion picture. *Linguistics*, 44, 29–68.
- Dupoux, E., & Peperkamp, S. (2002). Fossil markers of language development: phonological deafnesses in adult speech processing. In B. Laks, & J. Durand (Eds.), *Phonetics, phonology, and cognition* (pp. 168–190). Oxford: Oxford University Press.
- Flege, J. E. (1995). Second language speech learning: Theory, problems, findings. In W. Strange (Ed.), *Speech perception and linguistic experience: issue in cross-language research* (pp. 233–277). Baltimore, MD: York Press.
- Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer, & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production, differences and similarities* (pp. 319–355). Berlin: Mouton de Gruyter.
- Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, 40, 467–491.
- Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5, 171–189.
- Fujisaki, H., & Kawashima, T. (1969). *On the modes and mechanisms of speech perception* (Annual Report of the Engineering Research Institute No. 28): University of Tokyo Faculty of Engineering.
- Gillette, S. (1980). Contextual variation in the perception of r and l by Japanese and Korean speakers. *Minnesota Papers in Linguistics and the Philosophy of Language*, 6, 59–72.
- Grønnum, N. (1998). Danish. *Journal of the International Phonetic Association*, 28, 99–105.
- Grønnum, N. (2003). Why are the Danes so hard to understand? In H. G. Jacobsen, D. Bleses, T. O. Madsen, & P. Thomsen (Eds.), *Take Danish for instance: Linguistic studies in honour of Hans Basbøll, presented on the occasion of his 60th birthday* (pp. 119–130). Odense, Denmark: U Press Southern Denmark.
- Hallé, P. A., Best, C. T., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, 27, 281–306.
- Ingram, J. C., & Park, S.-G. (1998). Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America*, 103, 1161–1174.
- Lindau, M. (1985). The story of /r/. In V. A. Fromkin (Ed.), *Phonetic linguistics. Essays in honor of Peter Ladefoged* (pp. 157–168). Orlando, FL: Academic Press.
- MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2, 369–390.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, 18, 331–340.
- Pallier, C., Bosch, L., & Sebastian-Galles, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64, B9–B17.
- Pallier, C., Colomé, A., & Sebastian-Galles, N. (2000). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, 6, 445–449.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, 13, 253–260.
- Polivanov, I. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* 4, 79–96. English translation (The subjective nature of the perceptions of language sounds). In A. A. Leont'ev (Ed.), *E. D. Polivanov: Selected works* (pp. 223–237). The Hague: Mouton/A. A. Leont'ev (Ed.), *E. D. Polivanov: Selected works* (pp. 223–237). The Hague: Mouton.
- 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, 129–145.
- Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3, 243–261.
- Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., & Öhman, S. E. G. (1969). Cross-language study of vowel perception. *Language and Speech*, 12, 1–23.
- Takagi, N., & Mann, V. (1995). The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics*, 16, 379–405.
- Trubetzkoy, N. S. (1969). *Principles of phonology* (C. A. M. Baltaxe, Trans.). Berkeley, CA: University of California Press. (Original work published 1958).
- Vance, T. (1987). *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.
- Wiese, Richard (2001). The phonology of /r/. In T. Alan Hall (Ed.), *Distinctive feature theory*. Berlin: Mouton de Gruyter.
- Westbury, J. R., Hashi, M., & Lindstrom, M. L. (1998). Differences among speakers in lingual articulation for American English /r/. *Speech Communication*, 26, 203–226.
- Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics*, 52, 376–392.
- Zawadzki, P. A., & Kuehn, D. P. (1980). A cineradiographic study of static and dynamic aspects of American English /r/. *Phonetica*, 37, 253–266.

<sup>9</sup> We thank Rikke Bundgaard-Nielsen for suggesting this possibility.