Interspeech 2007, Aug. 27-31, Antwerp Belgium,
International Speech Communication Association.

English and French Speakers' Perception of Voicing Distinctions in Non-Native Lateral Consonant Syllable Onsets

Catherine T. Best [1,3], *Pierre A. Hallé* [2, 3] and *Jennifer S. Pardo* [4]

[1] MARCS Auditory Laboratories, University of Western Sydney, Australia
[2] CNRS: Laboratoire de Phonétique et Phonologie, University of Paris-3, France
[3] Haskins Laboratories, U.S.A. [4] Barnard College, Columbia University, U.S.A.
c.best@uws.edu.au, pierre.halle@univ-paris5.fr, jsp2003@columbia.edu

Abstract

English and French listeners were tested on discrimination and open-response categorization of laryngeal contrasts in three non-native syllable onsets differing in gestural complexity, in particular in the phasing between laryngeal and supralaryngeal articulations. All onsets involved a lateral, which was combined with a coronal stop in two contrasts. Results support the view that syllable onsets are perceived as holistic articulatory patterns, in which voicing is more difficult to perceive separately as gestural complexity of the onset increases.

Index Terms: non-native speech perception, laryngeal contrast, articulatory gestures, syllable onsets, cross-language

1. Introduction

Classic cross-language studies of consonant perception revealed that adults may categorize and discriminate non-native minimal consonant contrasts rather poorly [1]. Yet, perceptual difficulty can vary substantially across non-native contrasts and across listener languages [2-6], providing evidence that the perception of unfamiliar consonants is constrained by the listener's native phonology as well as by the consonants' phonological and phonetic properties. The Perceptual Assimilation Model [PAM: 7] offers a principled account of such constraints and of the expected perceptual difficulties according to perceived cross-language similarities.

Much remains to be explained, however, about the basis for perceived cross-language similarities. While they could perhaps derive from universal phonetic features, we will focus here on perceptual similarity from the viewpoint of Articulatory Phonology (AP), where both phonological and phonetic structure are defined in terms of dynamic articulatory gestures [8]. A quick review of prior studies in which poor discrimination of non-native consonant contrasts had been found [2,5,9-11] indicates that the difficult contrasts are often those produced by differentiated actions of a single articulator rather than those involving different articulators. That is, "within-organ" contrasts are usually more difficult than "between-organ" contrasts. The Articulatory Organ Hypothesis [AOH: 12-14] of the AP framework, originally applied to infants' attunement to their native language and concomitant decline in perceiving non-native distinctions, predicts a sharper, earlier decline for within-organ contrasts. This prediction has been successfully supported by empirical data. English-learning 11-month-olds showed a decline, relative to 7-month-olds, in discriminating three non-native Zulu within-organ laryngeal contrasts ([ɮ]-[ɬ], [k^h]-[k'], [p]-[ɓ]) but no decline for the non-native Tigrinya between-organ [p']-[t'] contrast [14]. For adults, discrimination patterns do not seem to systematically follow AOH although, as we have noted, they may encounter more difficulty for within- than between-organ non-native contrasts. However, a recent study [3] demonstrated that while English-speaking adults discriminated the Zulu [p]-[ɓ] laryngeal contrast as poorly as 11-month-olds, they performed quite well on the other two Zulu laryngeal contrasts. The joint categorization and discrimination patterns were consistent with PAM predictions: [ɮ]-[ɬ] and [k^h]-[k'] were assimilated as two-category and category-goodness contrasts with excellent and good discrimination, respectively, whereas [p]-[ɓ] was primarily assimilated as a single-category contrast and was discriminated poorly. AOH failed to predict the differences in adults' performance on these laryngeal contrasts. The simple between- versus within-organ dichotomy appears inadequate to account for the varying degrees of perceived similarity for the contrasts examined in [3]. What other type of perceived similarities may explain listeners' performance on non-native contrasts generally, and the large variation across these within-organ laryngeal contrasts in particular?

Perhaps perceived similarity, for adults, is more inclusive than a spotlight on the one critical gesture (or feature, in the standard phonological theory formulation) that distinguishes minimally contrasting non-native phones. Possibly, listeners assimilate and discriminate the *overall* gestural structures of contrasting initial consonants, or indeed of syllable onsets [see 15]. The perceived similarity between onsets that contrast in their laryngeal settings may thus be determined by their overall gestural organization. The Zulu velar and bilabial stop contrasts offer a hint of support for that view. Both these contrasts have a stop that has a close correspondent in English, and another that is deviant from English specifications. In terms of classic phonetic features, both contrasts involve a distinction between pulmonic (English-like [p], [k^h]) and glottalic stops (non-English-like implosive [ɓ], ejective [k']), with an added voicing or aspiration distinction. That is, [p]-[ɓ] and [k^h]-[k'] are each distinguished by two phonetic features. In terms of gestural organization, however, [p] is more similar to [ɓ] than [k^h] is to [k']. Both the velar and the bilabial contrasts distinguish an English-like constriction degree (CD) of the GLOTTIS (the laryngeal articulator in AP), at its default vertical location in the throat, from a non-English glottic constriction location (CL) gesture, coordinated with an English-like supralaryngeal stop closure+release gesture. Specifically, Zulu and English /k/s ([k^h]) correspond in that they phase the release of a velar closure with a glottic opening gesture (CD = WIDE), and their plosive /b/s ([p]) correspond in that they release a bilabial closure coincident with voicing onset (glottic CD = CRITICAL, the default speech setting that yields glottal pulsing in sonorants). Zulu [k'] differs from [k^h] by phasing its velar release to both a glottic *closure* (CD = CLOSED) and a rapid ejective glottic gesture (CL = RAISE), whereas implosive [ɓ] differs from [p] by phasing bilabial release with a rapid ingressive glottic gesture (CL = LOWER), but it maintains CD = CRITICAL so that the descending glottis vibrates, resulting in voicing. English speakers unanimously assimilated [k'] as a "deviant" /k/ and

discriminated it very well from [k^h], yet assimilated [ɓ] primarily as plain /b/ and discriminated it poorly from [p] (also assimilated as /b/). This perceptual difference is compatible with the greater difference in gestural structure for the velar than for the bilabial contrast, suggesting that perceived similarity is based on overall gestural organization.

The Zulu lateral fricative voicing contrast offers further insight about the potential role of gestural organization in perceived similarities. It is similar to the lingual fricative voicing contrasts of English ([s]-[z], [ʃ]-[ʒ]) in that a glottic opening gesture is phased with a frication constriction (CD = CRITICAL) for the voiceless version. There is simply no such glottic gesture for the voiced version, in which voicing occurs throughout the fricative constriction. The lateral constriction in Zulu involves tongue tip (TT) and body (TB) constriction, whereas English [ʃ]-[ʒ] involves a different TT+TB pattern, and [s]-[z] engages TT only. Thus, American listeners could be expected to assimilate Zulu [ɬ]-[ɮ] to contrasting English fricatives, most likely [ʃ]-[ʒ] or [s]-[z], and to discriminate them quite well. While they did discriminate the Zulu contrast very well, however, they gave wildly varied naïve transcriptions in their open-response categorizations, which were often phonotactically impermissible sequences of fricatives, stops or affricates plus /l/ [3]. Intriguingly, not all transcriptions reflected "correct" perception of voicing. There was one constant, though: everyone transcribed the contrast as some sort of phonological distinction (consistent with their excellent discrimination), often with at least one complex syllable onset. Thus, the assimilations appear to be based on gestural organization rather than on a simple fricative voicing contrast.

It was important to follow up the lateral fricative findings for at least two reasons. First, this particular contrast suggests that perceived cross-language similarity depends on overall gestural organization rather than factored-out distinctive features. Based on our analysis of existing data, we speculate that perception of, for example, laryngeal contrasts depends on overall gestural complexity. Second, the organ involved is the GLOTTIS. Glottic (voicing) contrasts in non-native consonants do not always elicit good discrimination and categorization, as was suggested by the first empirical report [1] on cross-language speech perception, which found notable perceptual difficulties for non-native stop voicing contrasts. Despite the impact of that seminal report on cross-language research, relatively few studies have examined voicing contrasts. An additional motivation for looking at voicing contrasts is that phonological accounts of voicing are fraught with difficulties, for example with respect to the minimal set of phonological features required. The phonetic details of voicing contrasts differ widely across languages, consonant types, and syllable positions, leading many phonologists to eschew the notion that phonological voicing can be defined by phonetic properties, because seemingly identical phonetic settings can be perceived as voiced or voiceless depending on the settings employed by the listener's native language.

We therefore undertook to expand on the lateral fricative findings. In order to probe how overall gestural organization may contribute to the perception of within-organ laryngeal contrasts, we added two more non-native syllable onsets that also involve lateral constriction but differ in overall gestural complexity. We added a native French listener group of native speakers of French, because word-initial stop voicing differs in French (prevoiced vs. voiceless unaspirated) versus English (voiceless unaspirated vs. voiceless aspirated).

For one of the new onsets, we chose Hebrew coronal stop + /l/ clusters with a stop voicing contrast /tl/-/dl/. Their gestural organization is, on the one hand, more complex than for Zulu lateral fricatives since it involves two segments. On the other hand, the phasing between the two segments is not strictly time-locked [16] so that the critical laryngeal-supralaryngeal phasing occurs in the stop component. In our logic, the Hebrew /dl/ and /tl/ clusters are of particular interest because they might be treated as holistic onsets, as has been suggested in [15]: French and American listeners assimilate Hebrew /tl/, which is disallowed in both French and English, *as a whole* to the closest onset cluster /kl/, that is, despite the fact that /tl/ could be parsed into /t/ plus /l/, each of which exist in the listeners' languages. According to [15], this pattern is less clear-cut for Hebrew /dl/ which is less often assimilated to a velar-initial onset. In this regard, French listeners differ from American listeners in more often perceiving Hebrew /dl/ "faithfully" as /d/ + /l/. We thus expect better discrimination of /dl, tl/ by French than American listeners.

The other new onset was a Tlingit lateral affricate voicing contrast /tɬ/-/dɮ/ (phonetically [t͡ɬ]-[t͡ɮ]) [17-18]. English has voiced-voiceless affricates, though not laterals, but French phonology lacks affricates (except in a few loan words, e.g., *jeep*, *Czech*, where onsets may be pronounced as fricatives instead of affricates). On the holistic view of onset perception, then, French listeners may encounter more difficulty than Americans with Tlingit /tɬ/-/dɮ/. As affricates, the gestural structure of the Tlingit onsets' is intermediate between lateral fricatives and coronal stop + /l/ clusters. Like lateral fricatives, /tɬ/-/dɮ/ are monosegmental and contrast voicing of a CRITICAL lateral constriction. Like the clusters, they involve a coronal stop and a lateral, but their gestural structure differs: stop closure is phased to both the lateral and glottic gestures, possibly decreasing detection of the voicing contrast.

2. Method

Participants completed discrimination and categorization tests on each contrast. Discrimination was tested first, to minimize effects of stimulus categorization on performance. A categorial AXB task was used, given its fairly low memory demands and response bias. Open-response categorization was used to reveal how listeners' perception of voicing interacted with the differing gestural structures of the three onsets.

2.1. Participants

Native speakers of English (N=19) and of French (N=16) participated. None had been exposed to the languages or contrasts investigated. None reported hearing, speech or language impairments. Twelve additional American subjects were tested but their data was not retained due to failure to complete all test sessions, inappropriate linguistic background, and/or a high rate of missing responses (> 2.5 s.d.).

2.2. Stimulus Materials

Male native speakers of Hebrew, Tlingit, and Zulu recorded 20 repetitions of open syllables with the targeted onsets followed by /a/. For the stimulus set, 5 tokens each of the two members of each contrast (e.g., Hebrew /dla, tla/) were chosen to be similar in duration and F0. Detailed acoustic measurements were conducted on these, including F0 and formant frequencies at various points in the lateral and vocalic portions of the stimuli, energy and durations. We limit description here to characteristics that differentiate voiced and voiceless onsets in each language. In Hebrew, /d/ in /dla/ was prevoiced (146 ms prerelease voicing lead) and /t/ in /tla/ had a medium voice onset lag (61 ms); integrated energy over stop release was greater for /t/ than /d/ (3.8 dB s vs. 1.3 dB s); F0

at the start of /l/ was higher for /tla/ than /dla/ (134 vs. 114 Hz); these values are similar to [14]. In Tlingit, the stop burst was voiceless for both /tɬ/ and /dɮ/; the lateral fricative release was longer for /tɬ/ than /dɮ/ (202 vs. 139 ms) and the initial voiceless portion of the release was longer for /tɬ/ (147 vs. 38 ms), thus 26% vs. 73% of the release was voiced in /tɬ/ vs. /dɮ/; integrated energy over the voiceless portion was greater for /tɬ/ than for /dɮ/ (9.3 vs. 2.5 dB); F0 at the start of /a/ was higher for /tɬa/ than /dɮa/ (142 vs. 128 Hz). The Zulu lateral fricatives were fully voiced for /ɮ/ and almost totally voiceless for /ɬ/; they were slightly longer for /ɮ/ than /ɬ/ (221 vs. 203 ms); F0 was higher at the start of /a/ in /ɬa/ than /ɮa/ (112 vs. 96 Hz); energy in the Zulu fricatives otherwise was equivalent. Thus, the glottic contrasts in these stimuli are best described as +/-[CRITICAL] (+/-glottal pulsing) for Hebrew and Zulu, but as +/-[WIDE] (+/-aspirated) for Tlingit ([t͡ɬ]-[t͡ɮ])[18].

2.3. AXB Discrimination Task

For each contrast, there were 60 AXB triplets: 15 for each of the 4 possible triad types, with the constraint that each syllable token appeared equiprobably. Trials were presented in random order, blocked by language; block order was counterbalanced across subjects (ISI = 1 s, ITI = 3 s, IBI = 5 s). There was a 5 trial training phase. On each AXB trial, participants indicated whether X matched A or to B by pressing buttons labeled '1' and '3'. They were instructed to respond on each trial, even if guessing, and to respond as quickly as possible after hearing all three items. Missed trials were "recycled" so that each subject completed all 180 trials.

2.4. Categorization Task

In the test phase, each of the 30 stimulus tokens was presented 3 times, totaling 90 trials. The test was preceded by a training phase of 6 trials. On each trial, participants were presented with a syllable, which they had to transcribe using the keyboard. If they hesitated among several transcriptions, they could report all of them by using '/' to separate them.

3. Results

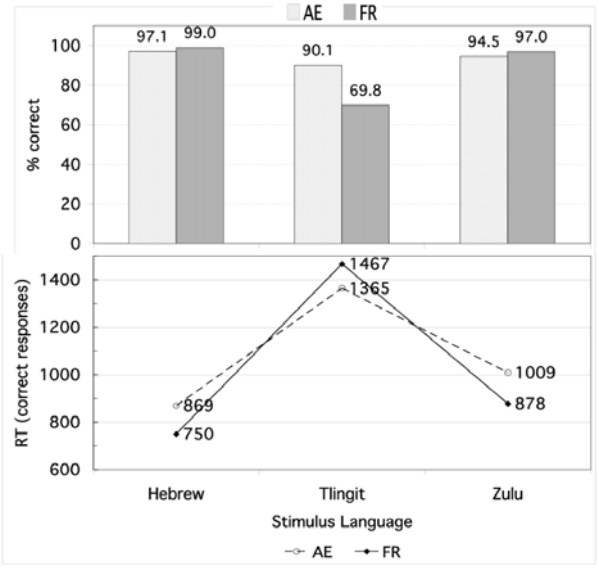

Figure 1: *Percent correct discrimination and RTs.*

3.1. Discrimination Performance

The correct discrimination and response time data are summarized in Fig. 1. Analyses of variance were conducted for both, with stimulus *Language* (Hebrew, Tlingit, Zulu) and listener *Group* (American vs. French) as within- and between-subject factors, respectively. For the accuracy data, American participants performed better overall than French participants (94 vs. 89%), $F(1,33)=13.81$, $p<.001$, but this advantage held only for Tlingit (90 vs. 70%), $F(1,33)=43.02$, $p<.0001$, with the opposite trend (marginally significant) for Hebrew and Zulu. *Language* was highly significant, $p<.0001$; in each *Group*, performance was lowest for Tlingit, highest for Hebrew ($p<.01$ for all pair-wise comparisons). The RT data correlated (negatively) highly with accuracy ($p<.0001$), closely paralleling in reverse the patterns found for accuracy: French participants performed worse than Americans only for Tlingit; they exceeded Americans on Hebrew and Zulu. Both groups responded much more slowly to Tlingit than the other two, $p<.0001$, and were faster for Hebrew than Zulu, $p<.0001$.

3.2. Categorization Patterns

Participants' open responses were translated into homogeneous "broad" phonetic transcriptions, then coded for phonetic properties of the syllable onsets: voice, manner, and place of the "primary," and "secondary" consonants transcribed (e.g., /k/ and /l/ for [tl]). Hesitation between several transcriptions ('/') was coded as "ambiguous" for the appropriate property. The categorization data closely predicted discrimination performance, which is computed by assuming that two items will be discriminated only if they are categoryized differently. Thus, many aspects of the categorization data are related to the patterns of discrimination reported above. For example, categorization of the onsets as voiced or voiceless was consistent with the stimulus voicing category, except for the French listeners' transcriptions of the Tlingit affricates. But the categorization data also provide more detailed insights about how the complex gestural structures of the nonnative onsets were perceived. In the following, we report only statistically significant patterns.

For Hebrew, judgments of voicing were quite congruent with stimulus voicing. Unambiguous responses of French participants showed no voicing "errors." They performed better than Americans, who occasionally categorized /tla/ as voiced and /dla/ as voiceless (~9% of errors). Americans reported single-stop onsets for 10% of the /tl/ and 1% of the /dl/ items, whereas French participants always reported clusters. As for place, French listeners only reported velar onset for /tl/ but more often a dental than velar onset for /dl/ (61% vs. 35%). This voicing asymmetry was much weaker for Americans: 93% and 71% velar responses for Hebrew /tl/ and /dl/. Only Americans ever reported labials (2% for /tl/, 10% for /dl/).

For Tlingit, judgments of voicing were almost error-free for /tɬ/ but not for /dɮ/, which induced much more voicing ambivalence (~11% vs. 1%) and was otherwise judged as voiceless 65% and 11% of the time by French and American listeners, respectively. French participants gave "voiceless" more often than "voiced" responses to /dɮ/ (65% vs. 22%). There were more single-stop responses to the Tlingit than the Hebrew stimuli; these were more frequent for /tɬ/ than for /dɮ/ (32% vs. 9%), and for French than American listeners (29% vs. 12%), correlating negatively with "correct" voicing judgments. As for place, the transcriptions reflected a dental-to-velar perceptual shift (97% velar responses on average), for both French and American listeners, and for both /tɬ/ and /dɮ/.

For Zulu, voicing judgments were near ceiling for French listeners (98.4% for both /ɬ/ and /ɮ/). Americans performed

less well (76% and 81% correct voicing for /ɬ/ and /ɮ)/, with more ambiguous responses (12%) than French listeners (4%). Subjects rarely reported a single-stop onset (6% and 2% for French and Americans); they reported many single-fricative onsets but also stop+fricative onsets, homorganic or not. They never reported stop+/l/ onsets. The stop+fricative responses were equally frequent in American and French listeners for /ɮ/ (39% and 36%) but more frequent in French than American listeners for /ɬ/ (54% vs. 20%). There were otherwise more correct coronal responses in the American than French data (69% vs. 40%). Coronal responses were more frequent for /ɬ/ than /ɮ/ in the American data (81% vs. 57%) but equally frequent for /ɬ/ and /ɮ/ in the French data (38 and 43%). Ambiguous-place responses were, for the most part, a place-ambiguous stop plus a coronal fricative (e.g., *bz-ks, ps-psh-ts-tsh*). Ambivalent responses were either between labial and coronal or between labial and velar place. The latter case was more frequent than the former (American data: 21% vs. 5%; French data: 42% vs. 9%), reflecting, perhaps, a trace of dental-to-velar or, less likely, dental-to-labial shift.

4. Discussion and conclusions

The results show that the non-native within-organ laryngeal contrasts we examined are perceived differently depending on (a) the gestural context in which the contrast occurs, and (b) listeners' native laryngeal settings for voicing contrasts. Evidently, they fail to detect the relevant laryngeal (voicing) contrast with equal ease across non-native contrasts that differ in gestural complexity, particularly with respect to the relative phasing between supralaryngeal and laryngeal gestures. Perception of these voicing contrasts seems better explained as a response to the contrasting onsets as *overall gestural structures*. All non-native onsets included a coronal lateral gesture (TT+TB constriction), and perceptual difficulty rose as the phasing between lateral and laryngeal gestures became tighter. In Hebrew /dl/-/tl/ clusters, the /l/ and /d, t/ gestures are out of phase, and listeners' performance was best. In Zulu /ɬ/-/ɮ/, the peak of a laryngeal WIDE gesture is phased with the peak of a lateral TT+TB constriction, and discrimination was lower. Listeners had difficulty transcribing these lateral fricative onsets as monosegmental. Although they were largely correct with respect to voicing, their assimilation of segmental composition was quite inconsistent. In Tlingit /tɬ/-/dɮ/, the onset pattern was perceived more consistently, most often as /kl/-/gl/, but the voicing contrast was perceived the worst. Listeners were biased toward perceiving a velar (*not* coronal) stop + /l/ cluster. Again, responses suggest that they did not readily extract voicing as a separate parameter but, rather, perceived the onsets as complex gestural structures.

The differences between American and French subjects, also, were in line with our predictions. American listeners outperformed French listeners on Tlingit /tɬ/-/dɮ/, conceivably for two reasons. First, the voicing contrast in Tlingit /tɬ/-/dɮ/ is closer in laryngeal settings to those used in English than in French. Second, affricates are essentially foreign to French. This latter fact suggests, again, that the primary difficulty French listeners have with these onsets relates to their overall gestural organization. The slight advantage of French listeners on Zulu /ɬ/-/ɮ/ may seem puzzling, but could relate to phonotactic differences [see 15]: initial /ʒ/ occurs in French but not English. Further research could clarify this possibility.

5. Acknowledgements

Supported by grants to the 1st (NIH: DC00403) and 2nd authors (French Ministére de Recherche: Cognitique LACO 1).

6. References

[1] Abramson, A. S. and Lisker, L. "Discriminability along the voicing continuum: Cross-language tests", Proc. Internat. Congr. Phon. Sciences, Vol. 6, 1970, p 569-573.

[2] Best, C. T., McRoberts, G. W. and Sithole, N. "Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants", J. Exper. Psych: Human Perc. and Perform., Vol. 14, 1988, p 45-60.

[3] Best, C. T., McRoberts, G. W. and Goodell, E. "American listeners' perception of non-native consonant contrasts varying in perceptual assimilation to English phonology," J. Acous. Soc. Am., Vol. 1097, 2001, p 775-794.

[4] Polka, L. "Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions". J. Acous. Soc. Am., Vol. 89, 1991, p. 2961-2977.

[5] Harnsberger J. "The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis". J. Phon., Vol. 29, 2001, p 303-327.

[6] Calderón, J. & Best, C. T. "Perceptual assimilation of non-native vowel contrasts to the American English vowel system", Meeting of Acous. Soc. Am, May 1996.

[7] Best, C. T. "A direct realist perspective on cross-language speech perception", In W. Strange (ed.) Speech perception and linguistic experience, p 167-200. Timonium MD: York Press, 1995.

[8] Browman, C. and Goldstein, L. "Articulatory phonology: An overview". Phonetica, Vol. 49, 1992, p 155-180.

[9] Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. "Developmental aspects of cross-language speech perception", Child Dev., Vol. 52, 1981, p 349-355.

[10] Polka, L. "Characterizing the influence of native experience on adult speech perception", Perc. & Psychophys., Vol. 52, 1992, p 37-52.

[11] Polka, L., Sundara, M. and Colantonio, B. "A cross-language comparison of /d/-/ð/ perception: Evidence for a new developmental pattern", J. Acous. Soc. Am., Vol. 109, 2001, p 2190-2201.

[12] Goldstein, L. and Fowler, C. "Articulatory phonology: A phonology for public language use". In N. Schiller and A. Meyer (eds.) Phonetics and phonology in language comprehension and production, p 159–208. Berlin: Mouton, 2003.

[13] Studdert-Kennedy, M. and Goldstein, L. "Launching language: The gestural origin of discrete infinity". In M. Christiansen and S. Kirby (eds.) Language evolution: The states of the art, p 235–254. Oxford U. Press, 2003.

[14] Best, C. T. and McRoberts, G. W. "Infant perception of non-native consonant contrasts that adults assimilate in different ways", Lang. Speech, Vol. 46, 2003, p 183-216.

[15] Hallé, P. A. & Best, C. T. "Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters", J. Acoust. Soc. Am, in press.

[16] Nam, H., Goldstein, L. G. and Saltzman, E. "Self-organization of syllable structure: A Coupled oscillator model. In T. Pellegrino, E. Marisco, and I. Chitoran, (eds.), "Approaches to phonological complexity." Berlin: Mouton, in press.

[17] Ladefoged, P. and Maddieson, I. "The sounds of the world's languages." 1996, Oxford: Blackwell.

[18] Maddieson, I., Smith, C., Bessell, N. "Aspects of phonetics of Tlingit," Anthro. Ling., Vol. 43, 2001, p 135-176.