Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona, Spain. August 3–9, 2003. Universitat Autònoma de Barcelona.

1292

!Xóõ click perception by English, Isizulu, and Sesotho listeners

Catherine T. Best[1,2], Anthony Traill[3], Allyson Carter[4], K. David Harrison[2,5], and Alice Faber[2]
[1] Wesleyan University (USA), [2] Haskins Laboratories (USA), [3]University of the Witwatersrand (South Africa), [4]University of Arizona (USA), [5]Swarthmore College (USA)

cbest@wesleyan.edu, atraill@icon.co.za, allcarte@yahoo.com, dharris2@swarthmore.edu, faber@haskins.yale.edu

ABSTRACT

Many, though not all, nonnative phonological contrasts pose discrimination difficulties. The Perceptual Assimilation Model attributes discrimination differences to listeners' assimilations of nonnative phones to their native phonologies, which vary across languages. We examined perception of two !Xóõ click contrasts by American English speakers and speakers of Isizulu and Sesotho, African click languages that lack the target contrasts. The Africans should assimilate !Xóõ clicks to native ones and discriminate accordingly; Americans should perceive them as nonspeech and discriminate them well. Isizulu's click system is richer than Sesotho's, so Isizulu speakers should perform better on at least one contrast. Americans should excel on contrasts that Africans assimilate to a single click. As predicted, Isizulu listeners assimilated !Xóõ clicks to native clicks most often, Americans heard nonspeech most often. Sesotho listeners were poorest on one contrast they had difficulty categorizing. Americans excelled on the other, which the Africans assimilated to a single click.

1. INTRODUCTION

Until the late 1980's, research and theory suggested that all unfamiliar nonnative minimal contrasts should be difficult for adults to discriminate and categorize. Subsequently, however, varying perceptual effects have been found for a range of nonnative contrasts. While many are indeed difficult to discriminate, others are moderately easy, and some are discriminated at native-like levels [1, 2]. Several theoretical models have been proposed to account for variations in nonnative speech perception.

The Speech Learning Model (SLM) [3] addresses how second language (L2) learners acquire production and perception of individual L2 phonemes. SLM attributes the relative ease or difficulty of acquiring a given L2 phoneme to its degree of similarity to the closest native (L1) phoneme. An L2 phoneme may be identical to an L1 phoneme, similar to one, or completely new and dissimilar from any L1 phoneme. Forming a separate L2 category is expected to be easy for new phonemes, difficult for similar ones, and moot for identical ones.

Another current model of nonnative speech perception, the Native Language Magnet model (NLM) [4], focuses on listeners' perception of acoustic variation within native versus nonnative phoneme categories. NLM posits that early exposure to the acoustic properties of a native phoneme leads to formation of a category prototype, which acts as a perceptual magnet. The magnet effect makes discrimination more difficult among acoustic near-neighbors of the prototype than among those of a non-prototype. Nonnative categories lack this perceptual prototype structure for naive listeners who are completely unfamiliar with the L2. NLM was not developed to address variations in perceiving different types of nonnative phonemes, but it may be possible to extend it to this issue. By extrapolation, nonnative phonemes that are acoustically similar to a native prototype might act like that prototype, i.e., display a magnet effect, making discrimination among their acoustic near-neighbors difficult. However, those dissimilar to any native prototype should behave like non-prototypes, i.e., near-neighbors should be easy to discriminate.

Another potentially relevant model was proposed primarily as an account of early developmental changes in nonnative speech perception. According to this model, perceptual sensitivity to psychoacoustically fragile phonetic contrasts declines early in development, if the language environment provides no exposure. Fragile contrasts are difficult for adults to distinguish and to re-learn. Conversely, psychoacoustically robust contrasts remain discriminable until later in childhood and are easy for adults to distinguish and to re-learn [5]. Fragile vs. robust contrasts are defined in terms of their perceptual salience, as well as their rarity vs. universality across languages. Like NLM, the fragile-robust hypothesis focuses on naive listeners. However, it addresses perception of nonnative contrasts rather than of individual phonemes.

Finally, the Perceptual Assimilation Model (PAM) [6] is centrally concerned with perceived similarity of nonnative and native phonological elements, like SLM. But its emphasis is on naive listeners' perception of nonnative phones, like NLM and the fragile-robust hypothesis, and focuses on perception of nonnative phonological contrasts, like the fragile-robust hypothesis, because these conditions should tap the maximal level of L1 phonological influence on perception. PAM's core hypothesis is that discriminability of a given contrast depends on how its members are perceptually assimilated relative to the L1 phonological system, e.g., if assimilated equally to a single L1 phoneme, they will be difficult to discriminate. But if each nonnative phone is assimilated to a different native phoneme, they will be discriminated well, like an L1 contrast.

One way to begin teasing apart the factors contributing to variations in perception of nonnative speech would be to compare perception of a given, mutually-nonnative stim-

ISBN 1-876346-48-5

ulus set by naive listeners from diverse languages. Given the differences among the phonological systems of the world's languages, perception of the same phonemes may vary, perhaps even considerably, among listeners of different L1s. Such cross-language differences would be expected according to PAM and SLM, which stress the phonological relationships between native and nonnative speech. On the other hand, NLM and the fragile-robust hypothesis do not make clear predictions about cross-language perceptual differences for the same nonnative stimuli. NLM posits that if a phoneme is nonnative and exposure to it is minimal, it should fail to show prototype effects regardless of listener L1. However, as suggested above, it may be that nonnative phonemes that are acoustically similar to native prototypes would display magnet-like perceptual effects, while acoustically dissimilar ones would not. The fragile-robust hypothesis seems to predict no L1 differences in perception of nonnative speech. If the stimulus contrasts are nonnative with respect to all listeners, then discrimination of a given contrast should not differ by listener L1. Only fragility-robustness of the contrast should affect discrimination, and equally so across L1s.

To explore the differences in theoretical accounts, we examined cross-language differences in nonnative speech perception. The listener groups were selected to differ markedly in the relationship between their L1 phonological systems and the nonnative target contrasts. To maximize the effects of L1 phonology on nonnative speech perception, our listeners were completely naive of the target language and contrasts. We used nonnative contrasts, rather than single phonemes, to maximize the influence of the contrastive structure of listeners' L1 phonologies.

The stimuli were two click consonant contrasts from !Xóõ, a Khoisan language of Botswana that uses five places of articulation for the primary (anterior) click release (bilabial [ʘ], dental [|], alveolar [!], palatal [ǂ], lateral [‖]). At each place, clicks are produced with one of eleven accompaniments [7] -- the phonetic features of the mandatory secondary (posterior) consonantal release. Thus, !Xóõ has 55 distinct clicks, in addition to click+/C/ clusters [8]. Notably, clicks are rare across languages, occurring only in a subset of African languages. Bilabial clicks are particularly rare, being used only in certain southern Khoisan languages [8-10]. Nonetheless, clicks are perceptually very salient, with all places of articulation substantially easier to identify in noise than other consonant types [8]. The contrasts used for this study were [ʘx]-[|x] and [!x]-[ǂ], produced with the velar fricative accompaniment [x].

The L1s of the three listener groups differ with respect to presence and types of clicks. One group was American English speakers who lacked any exposure to click languages. The English phonological system contains no clicks, of course. The other groups were speakers of Isizulu and Sesotho, two southern African languages that use clicks, but lack both of the target click place distinctions. Both are of the Bantu language family, rather than the Khoisan family, as !Xóõ is. Moreover, neither includes clicks with the velar fricative accompaniment. Importantly, the representation of clicks differs substantially between the phonological systems of these two languages. Isizulu uses three places of articulation ([|], [‖]), [!]) and five accompaniments for clicks, yielding 15 distinct clicks. By comparison, Sesotho has only dental [!], produced with three accompaniments, and thus has only three distinct clicks, which are not distinguished by place.

What predictions can be generated for these listener groups and contrasts by the four theoretical models? In SLM [3], all four !Xóõ clicks would presumably be new phonemes for Americans. Thus, although SLM doesn't explicitly predict discrimination levels for nonnative contrasts, it is reasonable to expect that Americans would discriminate both new-new contrasts well. For the Sesotho listeners, Xóõ [ʘx] and [|x] should be new, while [!x] would be similar but not identical to their L1 [!], which cannot have the velar fricative accompaniment. Acoustic and articulatory measures of alveolar and palatal clicks [8] suggest that [ǂx] would, likewise, be similar to the native [!]. Thus, the [ʘx]- [|x] contrast involves two new phonemes, and so should be easily discriminated by this group. But [!x]-[ǂx] involves two clicks that are similar to the same L1 click, so it should be more difficult for Sesotho listeners. For the Isizulu listeners, only [ʘx] should be new. Xóõ [|x] and [!x] would be similar but not identical to the corresponding Isizulu clicks, and [ǂx] should again be similar to the L1 [!]. The Isizulu group, then, would be expected to perform like the Sesotho group, i.e., to discriminate [ʘx]-[|x] contrast well, but not [!x]-[ǂx]. Thus, although for different reasons in the [ʘx]-[|x] case, the African groups should show the same discrimination patterns, but Americans should outperform them on [!x]-[ǂx].

NLM's [4] predictions about discrimination by each group are less clear, given its emphasis on within-category perceptual differentiation. Certainly, all clicks fall completely outside the language experience of American English listeners, and should function as highly discriminable non-prototypes. It is less clear how !Xóõ clicks would be perceived by the Africans, whose languages have clicks but lack two/three of the target places and the velar fricative accompaniment. Lack of experience should make [ʘx] a non-prototype for both African groups. [|x] would also be a non-prototype for Sesotho listeners, so they should easily discriminate [ʘx]-[|x] as non-prototypes and show no perceptual magnet effects. But would !Xóõ [!x] satisfy the L1 alveolar click protoypes of Sesotho and Isizulu, or would its nonnative velar fricative accompaniment push it to non-prototype status? The same question must be asked regarding !Xóõ and Isizulu [|] clicks, as well as for the alveolar clicks of !Xóõ and both languages, given reported acoustic similarities between alveolar and palatal clicks [8]. If !Xóõ clicks fit L1 prototypes for our African listeners, then its [!x]-[ǂx] contrast should be difficult for both groups to discriminate, but [ʘx]-[|x] should be easier for Isizulu listeners as a prototype vs. a non-prototype, and for Sesotho listeners as two non-prototypes. However, if all !Xóõ clicks are non-prototypes for these African languages, then they should discriminate both contrasts well.

As for the fragile-robust hypothesis [5], studies have found that clicks form a unique perceptual class: they are

ISBN 1-876346-48-5

perceptually robust, psychoacoustically salient, and easier for L1 listeners to identify correctly than non-click consonants [8, 11]. This suggests that both !Xóõ contrasts may be robust, and easily discriminated by all three listener groups. However, click contrasts could vary in psychoacoustic salience, with possible effects on discrimination. This may be the case for the two target contrasts. The primary release bursts (clicks) of alveolar and palatal clicks are abrupt and very intense (higher amplitude than the vowel), and differ strikingly in their spectra. Conversely, the primary release bursts of bilabial and dental clicks are longer-duration, noisier, and much lower in amplitude, and their spectra are more similar (though still distinct). These differences suggest that [!x]-[ǂx] may be more salient, and thus discriminated better, than [ʘx]-[|x]. However, the three groups should still not differ.

The PAM [6] hypotheses that are most relevant to the target stimuli and listener groups are that:

1) if listeners assimilate contrasting nonnative phonemes equally to a single L1 phoneme (SC: Single Category assimilation), discrimination is relatively poor;
2) if they assimilate them to contrasting L1 phonemes (TC: Two Category), to an L1 phoneme vs. an uncategorized consonant (UC: Uncategorized-Categorized), or to a goodness of fit difference in one L1 phoneme (CG: Category Goodness), discrimination is better;
3) if they fail to even hear them as phonological elements (NA: Non-Assimilable), instead hearing them as nonspeech, then discrimination can range from fair to excellent, depending on the perceptual salience of the nonlinguistic auditory difference between them.

English speakers tend to perceive clicks as nonspeech sounds, i.e., they show the NA pattern, with discrimination ranging from modest to excellent [1, 12]. Thus, Americans should discriminate both !Xóõ click contrasts. However, the possible difference in nonlinguistic auditory salience suggests that they may discriminate [!x]-[ǂx] better than [ʘx]-[|x]. Click language speakers should be more likely to assimilate nonnative clicks to L1 clicks. Perceptual patterns should differ between listeners of languages with richer click systems (Isizulu) vs. reduced systems (Sesotho). Thus, Isizulu listeners should assimilate !Xóõ clicks to L1 clicks more often than Sesotho listeners, who should be more likely to hear some as nonspeech. However, nonspeech percepts should be most frequent in American listeners. Isizulu listeners should be more likely than Sesotho listeners to assimilate [ʘx]-[|x] to either an L1 click contrast (TC assimilation) or to a goodness difference (CG) within L1 [|], and thus to discriminate it better. Conversely, given the similar properties of alveolar and palatal clicks, and the fact that both African languages have alveolar but not palatal clicks, both African listener groups may show single category (SC) assimilation to L1 alveolar clicks. Several other predictions follow from PAM hypotheses: English listeners should do no worse overall than Isizulu and Sesotho speakers in discriminating nonnative clicks, and should outperform the Africans on [!x]-[ǂx].

2. METHOD

AT recorded several native speakers of !Xóõ in Lokalane, southern Botswana, as they produced multiple /Ca/ tokens of !Xóõ [ʘx], [|x], [!x], [ǂx], and [‖x]) clicks (velar fricative accompaniment). One female's tokens were chosen for use. CB and AC selected four tokens of each click (except laterals), matched as closely as possible on noncriterial acoustic properties (vowel formants, F0, duration, intensity). These were waveform-edited to further equate syllable intensities and to narrow the range of the syllable durations ($M_{modified}$ = 903 ms, range = 871-937 ms). Identity and quality of all final tokens were verified by AT. Perceptual tests included four contrasts: [ʘx]-[|x], [!x]-[ǂx], [|x]-[!x], and [|x]-[ǂx]. Results are discussed only for the first two contrasts in this report.

There were 16 American English listeners and 13 listeners each of Isizulu and Sesotho[1]. All were college students tested in their native language and country; all had normal hearing and speech/language/reading abilities. None had experience with !Xóõ or the target clicks. Americans had no experience with any click languages.

Listeners completed categorial AXB discrimination tests involving the multiple tokens of each click [for procedural details: 2], for each !Xóõ contrast. They then completed a categorization task in which they provided speech and/or nonspeech labels for the individual stimulus tokens (presented four times in random order). The test was open-response for the Africans, who were asked to use their (phonologically transparent) L1 orthographies to indicate which consonants they heard, or describe any nonspeech sounds they heard, at the onset of each syllable (transcribed/translated by AT). Preliminary work with Americans indicated that closed-set choices would be needed for them; the final set of English consonants and nonspeech descriptors was developed through extensive pilot testing. African participants took the tests via DAT tape, and gave paper and pencil responses. Americans were tested with PsyScope programs on a Macintosh computer. All instructions and test forms were presented in the listener's L1.

3. RESULTS

A language (3) x click (4) x response type (speech, nonspeech, mixed) analysis of variance (ANOVA) on the categorization data yielded a significant language x response interaction, $F(4, 78) = 14.09$, $p < .0001$. Isizulu listeners were more likely to report hearing !Xóõ clicks as speech (87% overall) than Sesotho (79%) or American listeners (28%), who gave nonspeech (34%) and mixed responses (36%) more often than Isizulu (5% and 1%, respectively)[2] or Sesotho listeners (19% and 1%). The 3-way interaction, $F(12, 234) = 4.30$, $p < .0001$, indicated that Isizulu listeners heard [ǂx] as nonspeech more frequently (14%) than the other clicks (0-4%). Sesotho listeners did not vary across the clicks (all 18-20% nonspeech). Americans reported pure speech more frequently for [ʘx] (35%) and [ǂx] (43%) than for [|x] or [!x] (18%).

ISBN 1-876346-48-5

Isizulu listeners most often heard both [!x] and [ǂx] as their L1 [!h], displaying a predominant SC assimilation pattern. They assimilated [|x] almost universally to Isizulu [|h], but they assimilated [ʘx] more variably to [g|], [‖h], and [|h], as well as to the clusters [|hx] and [‖hx], suggesting that [ʘx]-[|x] constituted a CG or UC assimilation for them. When Sesotho listeners indicated hearing speech rather than nonspeech sounds, they also showed SC assimilation of both [!x] and [ǂx] to their L1 [!], [!h], or [!x]. They likewise showed SC or UU (Uncategorized[2]) assimilation of !Xóõ [ʘx] to [!hx], [!hgx], [xh], [t] and of !Xóõ [|x] to L1 [!hx], [!h], [xh] (or nonnative [|x], [‖], [‖ x], [‖hx]). Americans showed NA or TC assimilation for both contrasts, hearing [!x] most often as "pop" or as [k]; [ǂx] as "snap," "click," "whack," "pop" or [t]; [ʘx] as "fizzle" or [θ] or [f]; [|x] as "tsk" or "kiss" or [θ] or [t].

The discrimination data were subjected to a language (3) x contrast (4) ANOVA (recall that only two of the contrasts are reported in this paper). The language x contrast interaction, $F(6, 117) = 2.43$, $p < .03$, reveals that the Sesotho listeners were somewhat less accurate (75%) on [ʘx]-[|x] than Isizulu (81%) or American listeners (80%). In contrast, Americans discriminated [!x]-[ǂx], which the Africans had assimilated to L1 alveolar clicks (SC assimilation), much better (89%) than the Isizulu (77%) and Sesotho listeners (78%). The Americans also discriminated it better themselves than the less perceptually salient [ʘx]-[|x]; Isizulu listeners showed the opposite pattern.

4. CONCLUSIONS

In summary, the results indicate that Isizulu listeners showed SC assimilation of !Xóõ [!x]-[ǂx], and CG or UC assimilation of [ʘx]-[|x]. In contrast, Sesotho listeners appear to have shown SC or UU assimilation of both contrasts. Correspondingly, Isizulu listeners discriminated [ʘx]-[|x] but not [!x]-[ǂx] somewhat better than Sesotho listeners. By comparison, American English listeners were most likely to show NA assimilation of both contrasts to distinct nonspeech sounds, or (less likely) TC assimilation to different English consonants. The Americans discriminated [ʘx]-[|x] as well as the Isizulu group, and thus somewhat better than the Sesotho group. Finally, the Americans performed substantially better than the two African groups on the presumably more salient [!x]-[ǂx] contrast. This pattern of results across the groups is most consistent with PAM predictions regarding assimilation patterns for each group on each contrast, and corresponding group differences. The predictions generated from SLM, NLM and the fragile-robust hypothesis, on the other hand, each fail on one or more of the present findings. Further work will be needed to confirm these nonnative perceptual patterns, and especially to determine what stimulus information listeners of different languages rely on (e.g., articulatory, phonetic, phonological) as they categorize and discriminate nonnative phonological contrasts.

Acknowledgment. Supported by NIH (U.S.A.) grants DC00403 (C. Best) and HD01994 (Haskins Labs).

REFERENCES

[1] C.T. Best, G.W. McRoberts and N.M. Sithole, "Examination of perceptual reorganization for nonnative speech constrasts: Zulu click discrimination by English-speaking adults and infants," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 14, p. 345-360, 1988.

[2] C.T. Best, G.W. McRoberts and E. Goodell, "American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology," *Journal of the Acoustical Society of America*, vol. 109, p. 775-794, 2001.

[3] J.E. Flege, "Second-language speech learning: Theory, findings, and problems", in *Speech perception and linguistic experience*, W. Strange, Ed. Timonium MD: York Press, 1995.

[4] P. Kuhl and P. Iverson, "Linguistic experience and the "perceptual magnet effect"", in *Speech perception and linguistic experience: Issues in cross-linguistic research*, W. Strange, Ed., p. 121-154. Baltimore MD: York Press, 1995.

[5] D.K. Burnham, "Developmental loss of speech perception: Exposure to and experience with a first language," *Applied Psycholinguistics*, vol. 7, p. 207-240, 1986.

[6] C.T. Best, "A direct realist perspective on cross-language speech perception", in *Cross-language speech perception*, W. Strange and J.J. Jenkins, Eds., p. 171-204. Timonium, MD: York Press, 1995.

[7] D.M. Beach, "The phonetics of the Hottentot language," Cambridge U.K.: W. Heffer & Sons, Ltd., 1938.

[8] P. Ladefoged and A. Traill, "Clicks and their accompaniments," *Journal of Phonetics*, vol. 22, p. 33-64, 1994.

[9] A. Traill, "Another click accompaniment in !Xóõ," *Khoisan Linguistic Studies*, vol. 5, p. 22-29, 1978.

[10] A. Traill, "The phonological status of !Xõo clicks", in *Khoisan Linguistic Studies 3*, A. Traill, Ed., p. 107-131. Johannesburg: University of Witwatersrand African Studies Institute, 1977.

[11] A. Traill, "The perception of clicks in !Xoo," *JALL*, vol. 15, p. 161-174, 1994.

[12] C.T. Best and R.A. Avery, "Left hemisphere advantage for click consonants is determined by linguistic significance," *Psychological Science*, vol. 10, p. 65-69, 1999.

[1] Xhosa listeners were also tested, but excluded from these analyses.
[2] Missing answers account for the remaining percentage of this group.

ISBN 1-876346-48-5