ABSTRACT

This paper reports on a study involving the production and perception of 72 consonants and 7 vowels of the Xhosa language by two groups of beginner-learners from disparate linguistic backgrounds. In particular this presentation examines the nature of ejective, implosive and click sounds when produced by 8 adult native English and native Afrikaans speakers learning Xhosa whilst being adjudicated by 10 Xhosa native (L1) listeners. The Xhosa L2 findings are related to findings obtained from 8 Xhosa L1 speakers who also participated in the study in an attempt to make useful comparative analyses between L1 speakers and L2 learners’ intelligibility scores. It is shown that L2 learners experience great difficulties with all three sound classes investigated. The nature and extent of “correctly” identified sounds is discussed cursorily after which a hierarchical ordering of sound classes is introduced demonstrating which classes of sounds seem to be more detrimental to intelligibility than others.

1. INTRODUCTION

An African language such as Xhosa exhibits a wide range of ejective sounds as well ingressive which include an implosive bilabial and a variety of clicks. Due to the fact that Xhosa has acquired official status as one of the official languages of South Africa, a renewed interest in the learning of African languages in general as well as of this language in particular has emerged. Due to the fact that Xhosa utilises different types of air stream mechanisms which do not occur in the respective sound inventories of languages such as Afrikaans or English, learners from these languages are inundated with production problems whilst Xhosa listeners on the other hand experience problems related to intelligibility.

This paper attempts to examine the nature of selected Xhosa L2 consonantal speech productions by adult native English (NEX) and Afrikaans (NAX) learners in view of Xhosa L1 listeners’ perceptual reactions to the respective productions. Findings discussed below are taken from Lewis [8]. More specifically this paper examines:

- Xhosa L2 learners’ productions of clicks, ejectives and the implosive bilabial which are identified as correct realisations of the intended sound;
- Xhosa L2 learners’ intelligibility scores for clicks in relation to the other sound classes.

The problem of acquiring native like accent in L2 learning is well known. Foreign-accented speech may be defined as non-pathological speech which deviates in some respects from first language (L1) speaker pronunciation norms. The speech deviations that L1 listeners perceive may include segmental deviations related to phone substitutions, phonetic distortions, and nonnative suprasegmental patterns [3, 4, 10].

Although L2 learners’ pronunciation is characterised by deviances in several areas, one of the dimensions on which it is judged is in terms of the speaker’s “overall intelligibility”, defined by Schiavetti [13] as “…the match between the intention of the speaker and the response of the listener through the speech passed through the transmission system. When all the words in the listener’s response list match all the words intended to be produced by the speaker, speech intelligibility is perfect. When none of the words in the listener’s response list match the words intended to be produced by the speaker, speech intelligibility is zero. In between these extremes of perfect and zero speech intelligibility lies a continuum on which we may quantify the degree to which the response list of the listener matches the intended productions of the speaker”.

Furthermore the debate on whether segmental phenomena and/or suprasegmentals play a larger or lesser role in this process also still continues [12].

Empirical research on L2 speech learning of South Africa’s indigenous languages in general, and Xhosa in particular is extremely limited. A generally acknowledged difficulty for Xhosa L2 learners is the pronunciation of clicks. Dowling [2], for instance, highlights the tongue-twister, “iqaqa liqhawuk’ uqhoqhoqho” (The polecat split its windpipe, transcribed phonetically as [il’a:li’u:k’ u:qa:qu:qu]) which Xhosa L1 speakers often enjoy presenting L2 learners in order to see if they can pronounce palatal click phonemes. In addition, little is really known about which consonants and vowels of the Xhosa sound system are typically produced as intelligible/unintelligible renditions of an intended sound. Impressionistic observations suggest that Xhosa L2 learners experience greater difficulty with consonants than vowels. Unfortunately, there is no research to support nor to refute such a claim.

2. METHOD

2.1. Stimuli

72 consonants and 7 vowels of the Xhosa language were elicited in this study.

2.2. Subjects

Two groups of subjects whom were studying at the Cape Town College of Education (CTCE), a teachers’ training college, participated in the study. The first group comprised of 8 NEX (native English speakers learning Xhosa), 8 NAX (native Afrikaans speakers learning Xhosa) learners and 8 NX (native Xhosa speakers). The NEX and NAX learners were considered to be inexperienced learners. The task of this group was to produce the consonantal and vocalic target sounds of the Xhosa
language. The second group of subjects comprised of 5 male and 5 female Xhosa L1 listeners. Their task was to identify the speech tokens produced by the NEX and NAX learners and NX speakers.

2.3. Elicitation procedure
NEX and NAX learners and NX speakers were presented with aural and written-format cues to facilitate the elicitation procedure. Once hearing the aural cues, NEX and NAX learners and NX speakers were required to produce the target consonants in the medial position of a fixed vowel frame /u_u/, appearing in the carrier phrase, “Ndithi u_u apha” (I say u_u here) and target vowels in a fixed consonant frame /s_s/ appearing in the carrier phrase, “Ndithi s_s apha” (I say s_s here). Once having placed the target sound in the vocalic or consonantal frames, NEX and NAX learners and NX speakers were expected to read the complete sentence (carrier phrase and the specific target sound of interest) aloud. On completion, they were then required to say the sentence aloud. Lip-to-microphone distance was maintained at approximately 15-30 cm for each speaker, and subjects were asked to produce all carrier phrases containing the target sound at a deliberate rate. This method of approach to elicit all target sounds was chosen in preference to spontaneous or extemporaneous speech to ensure a systematic and comparable collection of data. Results obtained here would probably differ from those approaches using other speech styles.

2.4. Stimulus preparation
Both tokens (that is the carrier phrase that was read as well as the carrier phrase that was spoken aloud) were digitized at 16 kHz with 16-bit resolution and saved as an audio file. Only the second token that was spoken aloud was edited. Signal editing was carried out visually and aurally with a Kay CSL (model 4300B). In each case, the target sound (together with its consonant or vowel frame) was edited from the carrier phrase, and this was completed at the start and end of all /C_C/ and /V_V/ frames. To remove any distracting transients resulting from editing at the rise and fall of the edited signal (utterance), each signal was multiplied by a Tukey window with a rise and fall time of 5% of the total signal length. The final stimulus set consisted of 1896 edited tokens (72 consonants x 24 talkers + 7 vowels x 24 talkers).

2.5. Identification procedure
The elicited speech productions of the Xhosa L2 learners and NX speakers where presented in 79 blocks to the 10 Xhosa L1 listeners in a forced-choice identification experiment effected on a personal computer. Only one intended sound was evaluated in each block. The total number of stimuli presented to each listener was 1896 (24 speakers x 79 sounds) tokens. The computer programme employed to facilitate the forced-choice judgements was created locally. Each block contained 24 tokens (8 productions by NEX learners, 8 productions by NAX learners and 8 productions by NX speakers) of an intended sound. Listeners knew beforehand what the intended sound for each block was supposed to be when identifying the stimulus. For each of the intended sounds being assessed, listeners were offered a forced-choice response alternative with which to label tokens, e.g., “ch, q, x, tsh, c”. The 24 tokens in each block were randomly presented for listeners to identify. The response alternatives other than the “correct” sound were in the majority of cases based on results obtained from a transcriptional analysis of NAX and NEX learners’ elicited speech productions. In total 18 960 tokens (79 intended sounds x 24 tokens x 10 listeners) were present in the data corpus [8]. Each score for an intended sound per language group would be out of 80 (8 speakers’ productions per language group x 10 listeners’ judgements).

3. RESULTS
Figures 1-3 illustrate the mean percentage of sounds per sound class produced by NX speakers and Xhosa L2 learners which are judged 10 Xhosa L1 listeners to be correct realisations of an intended Xhosa sound. Each percentage mean value is based on a language group’s total of 80 forced-choice judgements (e.g., 8 NAX learners x 10 listeners) for an intended sound. A 95% confidence interval is fitted around the number correct out of 80 as a percentage for each production and language group. For a given production, if the range plots between different language groups do not overlap this signifies that the language groups differ significantly from each other (p<0.05). For example, with regard to the production of /l/ (Figure 3), NAX and NEX learners score significantly lower than the NX group, however, no significant difference is noted between these two Xhosa L2 groups for this particular sound as the values overlap.
3.1. Ejectives

Ejectives function phonemically in Xhosa. The most important observation regarding ejective stops /p', t', c', k'/ in Figure 1 and 2 respectively, is that in no instance do the NEX and NAX values approximate the values obtained by NX speakers. These differences are large and indicate that Xhosa L2 learners have a problem articulating ejectives, i.e. synchronising glottal closure with oral closure in such a fashion that audible ejectives result. In all of the cases though it seems as if NAX speakers do relatively better than their NEX counterparts. This is probably due to the fact that stops in Afrikaans are radical whilst they are normally aspirated in English. The fact therefore that NEX speakers seem to do better with respect to the aspirated stops /p', t', k'/ is to be expected. As a matter of fact, apart from the pre-palatal aspirated stop /ζ/ (which in any case is not part of the inventory of English or Afrikaans), there seems to be no significant difference between the intelligibility scores of aspirated stops between NEX and NX speakers, or even for that matter between NAX speakers.

Another difference in behaviour with respect to ejectives and aspirated stops in Xhosa has also been detected in a previous study of Dogil and Roux [1]. It was shown that ejective stops in Xhosa effectively block vowel-to-vowel co-articulation in a /N_N/ sequence, whereas it is not the case with aspirated stops. This encoding into a CV syllable (or lack of it) may probably have an influence on the perception of the C and as such may explain some of the variation above.

3.2. Implosive

From Figure 2 it becomes quite clear that the implosive bilabial stop /b/ is probably one of most difficult for Xhosa L2 learners to produce correctly. Egressive voiced stops seem to be less of a problem as they obtain relatively high scores whilst the voiced bilabial plosive is consistently used as a substitute for the implosive. This is nevertheless to be expected as both English and Afrikaans entertain these voiced plosives. Acoustic analyses have shown [11] that an increase in voicing intensity in the latter portion of the closure can be considered as a leading cue for implosives, irrespective of continued voicing or total closure of the vocal folds during the closure phase. Acoustic analyses of L2 renditions of the intended implosive did not display any increase in voicing intensity. It is clear that learners will consciously have to learn to articulate this sound, especially with respect to the downward movement of the larynx which in turn will create a greater intensity of voicing.

3.3. Clicks

As expected, this class of sounds indicates the largest differences between NX speakers and L2 learners in the identification test (see Figure 3). As mentioned earlier, it is generally accepted that clicks are difficult for Xhosa L2 learners to articulate. The question of how intelligible Xhosa L2 click renditions are has not been established clearly until now. Figure 3 indicates that the only click from a total of 15 Xhosa click phonemes to indicate no significant difference between NX and Xhosa L2 learners is the voiced alveo-lateral click /Ɂ/. This, interestingly enough is observed only among the NEX learners.

In eleven cases there are no significant differences between the NEX and NAX learners although there are considerable differences between their scores and those of NX speakers. In the four remaining cases where significant differences are attested between NEX and NAX learners (i.e. with /ʃ, Ɂ, /Ɂ, /Ɂ/ and /ʃ, /Ɂ/), the NEX learners consistently score a higher mark. This seems to suggest that English learners of Xhosa seem to have less problems with the production of clicks than their Afrikaans counterparts.

One specific feature may be pointed out with respect to the nasalised clicks /ʃ, /Ɂ/ as a group they seem to be more easily produced than clicks with other accompaniments. This group also proved to be the only group of clicks not to block vowel-to-vowel co-articulation in /N_N/ sequences [1] whilst co-articulation is effectively blocked by all other click types. From an articulatory point of view it is easy to explain this relatively good performance of nasalised clicks. Insufficient timing of velo-pharyngeal closure and necessary dorso-velar contact for the click causes the velar closure to take precedence resulting in an open nasal port. In the acquisition of clicks by mother tongue speakers of Xhosa as well as by L2 learners the nasal clicks more than often are acquired first [9].

3.4. Intersound class differences: the position of clicks

Results of intersound class differences for each of the respective language groups when judged by Xhosa L1 listeners are
illustrated in Figure 4 above. These results indicate that for both NEX and NAX learners, vowels (mean score out of a total of 80 per language group) appear to be the class of sounds identified correctly in the largest number of cases. On the other hand, clicks appear the most difficult for the L2 learners to produce.

Results of this nature may suggest that there are certain classes of Xhosa sounds which probably deserve more emphasis than others in Xhosa L1 curricula. Based on the evidence provided by Lewis [8] it seems as if the following type of hierarchy exists with respect to the identification of degrees of difficulty:

<table>
<thead>
<tr>
<th>Xhosa L2 learners</th>
<th>Xhosa L1 speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clicks</td>
<td>Vowels</td>
</tr>
<tr>
<td>Affricates</td>
<td>Resonants</td>
</tr>
<tr>
<td>Stops</td>
<td>Affricates</td>
</tr>
<tr>
<td>Resonants</td>
<td>Clicks</td>
</tr>
<tr>
<td>Continuants</td>
<td>Continuants</td>
</tr>
</tbody>
</table>

Table 1. A hierarchical ordering (from top to bottom) indicating which particular sound classes need be addressed more than others in pronunciation teaching.

4. CONCLUSION

From this study it has become clear that the control of velaric and glottalic airstream mechanisms is extremely problematic for mother tongue speakers of English and Afrikaans. Considering clicks in particular, it is difficult to support some earlier remarks of Ladefoged and Traill [6] which imply that clicks are quite ordinary consonants which are not only fairly easy to produce, but also perceptually salient to a degree which should make them “highly favoured consonants in the worlds languages.” This is clearly not the case with the data analysed in this study.

REFERENCES