IS FOREIGN ACCENT V!SIBLE?

Duncan Markham

Department of Linguistics and Phonetics
Lund University, Sweden

ABSTRACT - The paper presents a pilot experiment testing speakers' ability to identify language and native/non-native status of other speakers from visual stimuli.

INTRODUCTION

The present experiment is motivated by Honikman's (Honikman, 1964) thesis that there exist language-specific 'articulatory settings' which must be adopted in order to attain native-speaker-like phonetic competence. The pilot experiment in this paper tests a method for assessing the role of peripheral articulation — primarily lip and jaw movements/aperture — in the perception of foreign accent and the identification of languages. This research method and the resultant data may prove useful in the development of visual speaker/language recognition systems and synthesis which deal with foreign/non-standard accents.

Most speakers utilise visual cues as well as the acoustic signal in the perception of speech. Research has indicated that speech perception is degraded where listeners do not have access to visual information for the speaker (eg Hashimoto and Seki, 1994) and it is a common experience amongst polyglots that a familiar language can be recognised at a distance just from lip movements, without word identification playing a role. I seek here to gain some measure of the degree to which an unconscious knowledge of 'typical' visible behaviour is accessible. Of particular interest is the transfer of the peripheral articulatory setting from L1 to L2, whether this is visible, and whether near-native pronunciation precludes such transfer. There exists a modest body of literature relating to the transfer of phonetic characteristics (as distinct from phonological transfer) from L1 to L2, whereof only a very small part is directly concerned with phonetic and imitative processes (the emphases in the existing research lie clearly within sociolinguistics and second language acquisition, especially with regard to the question of a critical period of acquisition).

PROCEDURE

Six speakers (4 females, 2 males) with some experience in a number of languages were selected for the experiment. They were chosen so as to represent a mixture of native, non-native, and near-native language abilities for a variety of languages. The speakers were between 20 and 35 years of age. Four of the six speakers were native speakers of Swedish, one of German and one of Thai. The two non-native speakers of Swedish had lived in Sweden for more than 8 years. All speakers had learnt English and German in school at least to a level which allowed them to read a simple text aloud. Two of the informants also had sufficient knowledge of French to read a French text. All informants had normal hearing and no speech disorders according to self assessment.

—Speakers M(ale)1 and M2 speak English and German with marked accent, especially M2. M1 had spent a few weeks in English and German speaking countries, but was only confident in English.
—F(emale)1 has L1 Thai, and moved to Sweden at the age of nine.She speaks at least near-native Swedish, and is a confident reader of English, French, and German. She began learning English at approximately six years of age and speaks easily. She has been to language courses in France.
—F2 speaks English confidently, having spent time in the USA, but with marked accent. She reads German hesitantly, and has trouble speaking it.
—F3 has L1 German. She speaks Swedish confidently, having resided in Sweden for more than eight years. She speaks English with marked accent. Her German phonology shows some effects of Swedish L2 influence.
—F4 speaks German, French and English very confidently, especially the latter. She shows mild foreign accent for all L2s.

The speakers read in random order a simple text in Swedish, English, German and French (where L2 competence was sufficient). The texts were selected from stories aimed at non-adult audiences on the basis of simplicity/familiarity of vocabulary, as it was essential that the informants should be able to read the texts aloud without undue slowness or hesitation. Informants were instructed to read each text silently prior to recording and were given the option of doing a practice reading for each text. Each reading lasted approximately 60 seconds.

Recordings were made in a sound-proof recording studio at the Department of Linguistics and Phonetics at Lund University. Audio recordings were made on Digital Audio Tape at a sampling rate of 44.1 kHz, using a headset microphone with pop-shield. Video recordings were made simultaneously. The video frame encompassed each speaker's chin, mouth and a small part of the nose (see Figure 1). The camera was placed at an angle of approx 30 degrees horizontally from a direct frontal view of the speaker's face and approximately 10 degrees below to maximise visibility of lip rounding and protrusion, and interdental aperture. The speaker was seated. A lamp was placed above the text, both to illuminate the text and the face of the speaker. Speakers were provided with a visor to shade their eyes when reading.

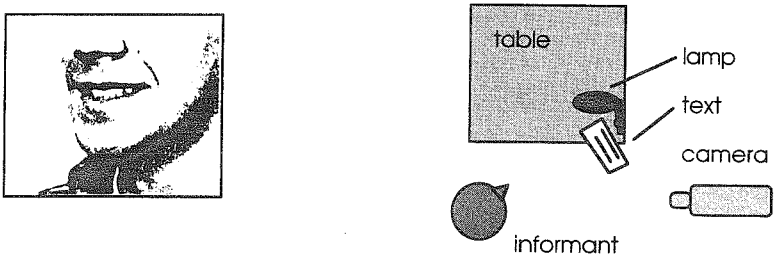


Figure 1. Image of facial area recorded on video, and schematic drawing of experimental environment.

A second group of informants — 'listeners'/observers — consisting of six linguists (three native speakers of Swedish, and native speakers of French, American English, and German with 23, 20 and 2 years residence in Sweden respectively), were (I) shown the video recordings for each speaker, without the acoustic signal, and were first asked to identify the language being spoken. They were given the four possibilities, plus one dummy possibility, Finnish, in the hope that this might reveal whether listeners were just answering randomly. The listeners were told that any given language could appear more than once for each speaker. All listeners had some familiarity with these languages, from previous experience, and because Swedish, French, English and German are regularly heard in the Department, and foreign language material shown on television in Sweden is not dubbed (there are also Finnish news programmes on Swedish television).

The listeners were then (II) given the four texts used, and were asked to judge whether the speaker reading the respective text was a native speaker of the relevant language, and if not, to give a judgement of degree of foreign accent on a four point scale *none-weak-medium-strong*. The listeners were then (III) presented with the audio signal only for all speakers for all texts and were asked to rate speakers for foreign accent and to identify each speaker's L1. In sections II and III, listeners were also asked to comment on the reasons for their judgements, where possible.

RESULTS

The speakers were generally able to read the texts without hesitations, although tempo for the L2 texts was often below a modal reading rate. The listeners all stated that they found the identification tasks (I and II) very difficult and claimed that they for the most part just wrote down guesses, especially for the first, language identification, task.

I - Language identification

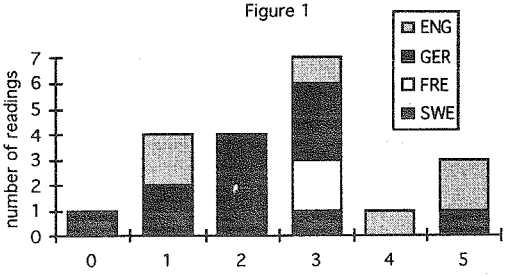
The results of the language identification test are presented in Tables 1 and 2. It is intuitively more likely that a native speaker of a given language will be better at recognising his own language than a non-native speaker. Table 1 shows the performance of the listeners in identifying their own L1. All languages other than French occurred six times in the material (once for each speaker). The probability of a chance correct identification is maximally 6% (2% for French). The probability of two chance correct identifications is maximally 0.3% (0.06% for French). This means that only two correct identifications for Swedish, English or Germany are necessary for the result to be above chance level. This is observable for the German and English L1 speakers and for two of the Swedish L1 speakers (GB, GH).

	GER L1 (MT)	FRE L1 (PT)	ENG L1 (DH)	SWE L1 (GB)	SWE L1 (GH)	SWE L1 (JO)
positive identification L1	**2** / 6	**0** / 2	**4** / 6	**2** / 6	**4** / 6	**1** / 6
false positive L1	**2** / 14	**5** / 18	**na**	**4** / 14	**0** / 14	**4** / 14

Table 1. Positive and false positive identifications of own L1 for the six listeners. Each column represents one listener (in brackets) and his or her respective L1. Identifications are given in bold type. Total possible correct identifications is given after the slash. See text for explanation of DH's scores.

Considerable variability in performance both for the identification of L1 and, as can be seen in Table 2, for all languages. The total number of correct identification for each listener has to exceed 20% of stimuli to be significant. This figure is achieved by listeners MT, PT, GB, and GH. Identifications by listener JO do not exceed this limit, and I can offer no external explanation for this, whilst the score for informant DH (ENG L1) is accounted for by the fact that he only respnded for English stimuli, where he recognised interdental fricatives, and felt entirely unable to give any response for any non-English stimuli.

Figure 1



Number of listeners who correctly identified each language for each reading (visual)

There were three readings for which five of the six listeners identified the language in question (English, English, Swedish) (Figure 1), four readings where only one listener identified the language correctly (English, Swedish, English, German), and one reading where no listener identified the language (Swedish). There was no concentration of these extremes for any particular speaker.

	MT	PT	DH	GB	GH	JO
CORRECT identification	**12**	**11**	**4**	**9**	**12**	**4**
INCORRECT identification	**8** (2 FIN)	**9** (1 FIN)	**na**	**11**	**8** (4 FIN)	**16** (1 FIN)

Table 2. Correct and incorrect identifications of languages for the six listeners. Each column represents one listener. Total identifications are given in bold type. The number of identifications as Finnish of the total incorrect is given in brackets.

II - Identification of native/non-native status and foreign accent from visual cues

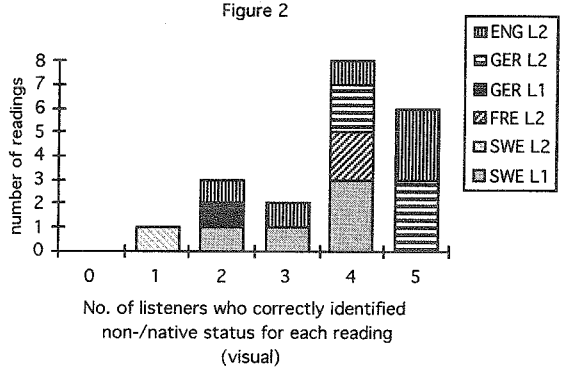
Table 3 shows the number of correct and incorrect identifications of L1 status. Two listeners (MT, PT) attained very high correct scores, with 18 and 17 correct respectively. Six readings were correctly identified unanimously by five listeners (Figure 2), whilst only one reading was misidentified by a large number (4) of listeners.

	MT	PT	DH	GB	GH	JO
CORRECT non-/native	**18** (13/5)	**17** (13/4)	**1**	**15** (12/3)	**12** (9/3)	**13** (10/3)
INCORRECT non-/native	**2** (1/1)	**3** (1/2)	**na**	**5** (2/3)	**8** (3/5)	**7** (3/4)

Table 3. Correct and incorrect identifications of native/non-native status for each language and speaker by the six listeners. Each column represents one listener. Figures in brackets give a breakdown of each total (non-native/native). Listener DH was unable to give judgements in this section.

Judgements of degree of foreign accent showed considerable variation for each reading, although there were only few instances (3) of both high *strong* accent and high native/*weak* accent scores for any single reading. In most cases (10), there was a central strong score with one or two deviating judgements on either side on the scale. It is important to note that listeners were critical to different degrees, with certain listeners tending more towards negative accent scores more than others. If integer values are assigned to each of the accent scores (such that 1=weak, 3=strong), and an average is taken for each listener the following values are obtained: MT [x=**1.77**, sd=0.83], PT [x=**2.33**, sd=0.62], GB [x=**1.57**, sd=0.51], GH [x=**1.17**, sd=0.58], JO [x=**2.08**, sd=0.95]. Clearly there is considerable variation, requiring some form of normalisation to enable proper comparison. Such normalisation would however be inappropriate in this context, not only due to the modest number of data, but also on account of demonstrated inconsistency in some listener judgements (Markham, 1993; 1994). Furthermore, there is no reason to assume that accent judgements for visual stimuli are made according to uniform criteria across listeners.

Figure 2



III - Identification of native/non-native status and foreign accent from audio signal

All speakers with an L1 included in the test were identified for their native language (ie, four Swedes, one German). This identification was unanimous for the Swedes, but the German speaker was assigned no L1 by one listener. It was of course possible to assign more than one L1, as any speaker could have been bilingual. The Thai L1 speaker (F1) was identified as an L1 speaker of Swedish by four of the six listeners, with one of the remaining listeners suggesting an L1 German accent for her Swedish reading. The same speaker was assigned native status for her English reading by two listeners. The German L1 speaker (F3) was identified as an L1 German speaker by five listeners, including the German listener (MT), but her slight non-native accent was also commented on by two of the same listeners. Two listeners were unsure as to whether she was an L1 speaker of Swedish (DH-English L1, PT- French L1). Speaker F4 was identified as L1 English for her English reading by listener GB (L1 Swedish).

listener	ρ (n)	p
MT	**0.69 (20)**	**0.0025**
PT	**0.74 (20)**	**0.0013**
DH	0.77 (5)	0.1211
GB	**0.60 (20)**	**0.0085**
GH	0.28 (20)	0.2297
JO	0.04 (20)	0.8492

Table 4. rho and p-values for visual and auditory conditions (Spearman Rank Correlation) for each listener.

Average accent scores assigned by each listener were: MT [x=**1.86**, sd=0.86], PT [x=**2.00**, sd=0.78], DH [x=**2.00**, sd=0.58], GB [x=**1.50**, sd=0.65], GH [x=**1.92**, sd=0.76], JO [x=**1.64**, sd=0.74]. Note that there is considerably more similarity in these means than in the means for visual accent scores. Accent judgements by three of the listeners for the

visual and auditory conditions were found to be correlated at $p<0.01$ by Spearman Rank Correlation tests (Table 4).

Interlistener accent judgements for visual foreign accent (section II) showed correlation for the listener pair MT-GB at the 1% level of probability. A further three pairs were correlated at the 5% level. Interspeaker judgements for auditory foreign accent (section III) showed correlation for the pairs MT-GB, DH-PT, DH-MT, and GB-JO at or below the 0.1% level. Eight pairs were correlated below the 1% level.

		identified (I)	visual accent (II)		auditory accent (III)	
speaker	language		L1 n	mean accent/sd, (n)	L1 n	mean accent/sd, (n)
M1	**SWE**	**3**	**4**	**2.00** / 0.0, (1)	**6**	
	GER	**3**	**0**	**2.00** / 0.7, (5)	**0**	**2.00** / 0.6, (6)
	ENG	**1**	**1**	**1.25** / 0.5, (4)	**0**	**1.50** / 0.5, (6)
F1	FRE	**3**	**1**	**2.50** / 1.0, (4)	**0**	**1.67** / 0.8, (6)
	ENG	**3**	**0**	**2.00** / 1.0, (5)	**2**	**1.50** / 0.6, (4)
	SWE	**0**	**3**	**1.00** / 0.0, (2)	**4**	**1.50** / 0.7, (2)
	GER	**3**	**1**	**1.75** / 0.5, (3)	**1**	**2.20** / 0.4, (5)
F2	GER	**2**	**0**	**2.00** / 1.0, (5)	**0**	**2.33** / 0.8, (6)
	SWE	**2**	**2**	**1.67** / 1.2, (3)	**6**	
	ENG	**5**	**2**	**2.00** / 1.0, (3)	**0**	**1.83** / 0.8, (6)
F3	ENG	**5**	**0**	**2.20** / 0.8, (5)	**0**	**1.67** / 0.5, (6)
	SWE	**1**	**4**	**1.00** / 0.0, (1)	**2**	**1.25** / 0.5, (4)
	GER	**2**	**2**	**2.00** / 1.0, (3)	**5**	**1.00** / 0.0, (3)
M2	GER	**3**	**1**	**2.00** / 0.7, (5)	**0**	**3.00** / 0.0, (6)
	SWE	**2**	**4**	**1.00** / 0.0, (1)	**6**	
	ENG	**1**	**0**	**1.60** / 0.9, (5)	**0**	**2.33** / 0.5, (6)
F4	**SWE**	**5**	**4**	**1.00** / 0.0, (2)	**6**	
	FRE	**3**	**1**	**1.40** / 0.5, (5)	**0**	**1.50** / 0.8, (6)
	ENG	**4**	**3**	**1.67** / 0.6, (3)	**1**	**1.20** / 0.4, (5)
	GER	**1**	**0**	**2.00** / 0.6, (6)	**0**	**1.50** / 0.5, (6)

Table 5. Summary of identifications of language and L1 status for each section, and average accent scores by all listeners for sections II and III.

DISCUSSION

The results presented above show that some 'listeners' (observers) can identify languages on the basis of visual information at rates above chance. However, this ability would not seem to be directly related to recognition of an observer's L1, as one of the three best overall identifiers (PT) gave five false positives for his native language, French, although the other two observers performed much better, with GH making no false positive identifications. Interestingly, GH is a former teacher of Swedish as a second language.

It is particularly interesting that these listeners appear to be able to identify languages despite the speakers' non-native accents. There was absolutely no correlation between auditory judgements of foreign accent (III) and the number of positive visual identifications (I), where one might reasonably have expected better identification of readings which were more native-like.

In section III, non-native listeners were found to be less capable of identifying foreign accent where the speaker's accent was near native.

The lack of correlation for interlistener visual accent judgements (II) is indicative of the different criteria used by different listeners in judging visual stimuli. This is also reflected in the difference in performance between certain listeners.

Listener comments

Listeners were also asked to comment on the reasons for their choices and judgements during the experiment. Very few cases of language identification on the basis of word recognition were cited, but some cases of phoneme recognition did occur. This related primarily to recognition of the English interdental fricatives [θ, ð], although some listeners claimed they had recognised the Swedish labiovelar fricative [ɧ]. Listeners also commented on language specific lip movements (rounding, protrusion) and perceived rhythmic patterns. These comments were correct only in approximately 50% of cases.

In judging native/non-native status, listeners were often distracted by pauses or hesitations in the readings, regarding these as cues to L2 status. Whilst this was often valid, it was equally possible for native speakers to hesitate or make corrections in their readings. Overall tempo was certainly the most strongly identified cue to L2 status, with particularly slow readings usually being given strong accent scores.

Experimental design factors

This pilot experiment was structured in such a way as to hopefully reveal areas in its execution which would require careful control in a later, fully fledged investigation. Problems encountered related especially to speaker over-confidence. The speakers often did not feel the need to practise the texts, resulting in avoidable hesitations and resets, later used as cues by the listeners. Familiarity with the texts is therefore essential. This is also true for the listeners, who were reliant on overall facial movements for their judgements, rather than being able to relate expected phoneme-specific movements to the visual stimuli. The author of this paper, being familiar with the texts and the video recordings (having watched them at least six times!) found that he could see clear deviations from the expected norm for certain speakers.

SUMMARY

In this investigation it was observed that three of six observers of visual stimuli, involving the reading of texts in up to four different languages, could identify the language being read for over 50% of the readings. This identification, whilst often being for readings where the speaker had little or no auditory foreign accent, also occurred for some cases of strong auditory foreign accent (Table 5). This neither confirms nor disproves the existence of unconcious knowledge of typical visible behaviour. The significant correlations between judgements of visual and auditory foreign accent for three of the listeners demonstrates that some degree of visual knowledge is accessible. As indicated above, tighter controls of reading behaviour and text familiarity may yield clearer results.

Hashimoto, M. and H. Seki (1994) "Limitations of lip-reading advantage by desynchronizing visual and auditory information in speech". In ICSLP 94 , 3:1155-1158. (Acoustical Society of Japan: Yokohama, Japan)

Honikman, B. (1964) "Articulatory settings." In D. Abercrombie et al. (eds), *In honour of Daniel Jones* (pp 73-84). (Longmans: London)

Markham, D.J. (1993) "Prosodic imitation: perceptual results." In J.S. Pettersson (ed), *Seventh Swedish Phonetics Conference (Fonetik-93).* Reports from Uppsala University, Dept of Linguistics (RUUL), (pp. 94-97). (Dept of Linguistics, Uppsala University: Uppsala)

Markham, D.J. (1994) "Prosodic imitation: 'reliable' listener results." In G. Bruce, D. House and P. Touati (eds), *Eighth Swedish Phonetics Conference (Fonetik-94)* Working Papers, (pp. 96-99). (Dept of Linguistics and Phonetics, Lund University: Lund-Backagården)