NAÏVE IMITATION OF SECOND-LANGUAGE STIMULI: DURATION AND F0

Duncan Markham
Deakin University, Australia

ABSTRACT

The ability of language learners and naïve experimental subjects to reproduce auditory stimuli is an important issue in understanding the causes of foreign accent. Data from imitations of auditory stimuli from Mandarin are presented. Eight native speakers of Swedish, without prior knowledge of Mandarin, imitated three three-morpheme phrases. The paper examines objective measures of articulatory reproduction, focusing on durational and F0 characteristics of the imitations, rather than presenting accent-ratings. Durational measurements revealed considerable individual fluctuation in the closeness of the imitations, although there were some clearly identifiable differences between subjects. Some subjects were also found to be more consistent in reproducing the same pattern. Fundamental frequency data show that some subjects treated contour tones as level tones. Again, some subjects are more successful than others at matching the target, or reproducing the same pattern. These observations agree with a previous study of durational and intonational imitation.

1. INTRODUCTION

The attainment of native-like pronunciation constitutes a great challenge for any person engaged in learning or acquiring a second language (L2). The wealth of studies of L2 pronunciation achievement attests to both the conspicuous nature of non-native accent and the difficulty in explaining the difficulties which learners aspiring to native-like pronunciation face.

Although there is a large body of pronunciation studies, only a small number of published studies have directly focused on the maximum degree of achievement actually possible for non-native speakers. The results of these indicate that a small number of speakers may be able to achieve native-like pronunciation. These findings are based on perceptual measures of nativeness—accent judgements by native speakers of the L2.

It would be desirable to gain insight into actual physical similarities in phonetic behaviour between native speakers and high-achieving non-native speakers, in order to understand the degree to which we can manipulate our articulatory planning and execution behaviours. Some research, notably, e.g., [1, 2], has concentrated on voice onset time (VOT), as it is a comparatively easy physical dimension to measure. The conclusions have usually been that groups of non-natives do not obtain native-like physical patterns, although a smattering of individual subjects can often be observed to fall within native ranges. (For discussion of this, see [3].) The potential for some learners to achieve native-like levels of performance in various linguistic dimensions has been reported by, amongst others, [4,5,6,7].

The processing load imposed on non-native speakers when performing in the L2 varies both in terms of communicative task, and regarding what parts of linguistic performance may be affected. This makes any investigation of the maximum potential of a non-native speaker quite difficult to interpret, due to the variety of experimental tasks and contexts. One way of avoiding problems associated with processing load is by removing linguistic information other than the phonetic signal. By asking subjects to reproduce stimuli from a language of which they have no knowledge it might be possible to test various aspects of speech perception and productional control.

2. PHYSICAL MEASURES — THE PRESENT STUDY

2.1. Background. A large study of achievement in phonetic imitation and learning for second languages ([3]) provides a source of phonetic data for both perceptual and physical examination. Eight native speakers of Swedish imitated stimuli and read texts for other languages they were familiar with and for a number of unfamiliar languages. This paper presents some data from the subjects' imitations of Mandarin.

The examination of these data concentrates on the identification and possible explanation of individual behaviours and patterns which may highlight individual aptitudes.

2.2. Method. The eight subjects, five females, three males, were selected on the basis of their conspicuously better than average performance in L2-pronunciation. Each subject heard stimuli from Mandarin presented via headphones and the subject's imitations were recorded using a head-mounted microphone. Recordings were made to digital audio tape at a sampling rate of 44.1kHz.

Two readings of a complete short text, read by a native speaker of Mandarin, were presented in order to familiarise the subject with the sounds of the language. A short phrase was then played to the subject at two reading speeds ("normal" and "fast") so that the subject was aware of the immediate material for imitation.

Individual morphemes were imitated first, followed by the full phrase, first at the slower reading speed, and then at the faster speed. Each stimulus was presented twice, after which the subject imitated the model. This procedure was repeated twice more for each stimulus, yielding three imitations per stimulus per speaker. As there were 15 stimuli, this yielded a total of 45 utterances per subject. The stimuli are listed in Table 1.

Phrase 1			**translation**
stimulus	1a	sān	
	1b	gè	
	1c	fáng jiān	
	1d	sān gè fáng jiān (*slow*)	*three rooms*
	1e	sān gè fáng jiān (*fast*)	
Phrase 2			
stimulus	2a	dōu	
	2b	yǒu	
	2c	qì chē	
	2d	dōu yǒu qì chē (*slow*)	*(they) all have cars*
	2e	dōu yǒu qì chē (*fast*)	
Phrase 3			
stimulus	3a	shì jiè	
	3b	gè	
	3c	dı	
	3d	shì jiè gè dì (*slow*)	*all over the world*
	3e	shì jiè gè dì (*fast*)	

Table 1. Mandarin stimuli (in pinyin romanisation) for imitation.

2.3. Analysis procedure. The recorded imitations were transferred to an Apple Macintosh and analysed using the SoundScope signal analysis program. Measurements of utterance duration were made, with a mean accuracy of ±5ms. Fundamental frequency plots were also obtained and examined with reference to the general contours and relative movements or tonal events in the original stimuli.

2.4. Durational results. The most striking initial observation was that there was considerable variation between speakers with regard to how closely they imitated the target and how consistent their behaviour was across the three attempts for each stimulus. Figure 1 illustrates the degree of variation observed to occur for one stimulus (1a: sān [san˥]).

For all stimuli, discrepancies between the target duration and the imitation durations were both positive and negative, although overshoot (positive differences) was more frequent, as can be seen in Table 2. There was no correlation between length or complexity of the target and the incidence of overshoot. Indeed, no explaining factor could be found for the amount of observed and unpredictable deviation from the target value. Stimulus 2a, for instance, was imitated with durational undershoot in 17 of 24 attempts. This stimulus (dōu [təʊ˥]) was spoken on a high tone, but is otherwise structurally similar to a number of other short stimuli (3b,3c), yet the latter stimuli showed very little undershoot.

stimulus	**duration**	***n* undershoot**	***n* overshoot**
1a	584	8	15
1b	178	3	21
1c	731	3	21
1d	1328	9	14
1e	872	2	22
2a	375	17	7
2b	514	8	16
2c	720	8	16
2d	1169	8	15
2e	775	2	22
3a	764	4	20
3b	361	1	23
3c	261	3	21
3d	1241	1	20
3e	831	1	21

Table 2. Target (stimulus) duration and durational undershoot/ overshoot for imitations of each stimulus.

Individual speakers' deviations from the target showed some speaker-specific patterns, as can be seen in Table 3. For Phrase 1 stimuli, speaker JM undershoots the target in 10 cases, and overshoots in only 5, in contrast to all other speakers who show very few instances of undershoot. More variation is observable for Phrase 2 stimuli, while Phrase 3 stimuli show very little undershoot for any speaker.

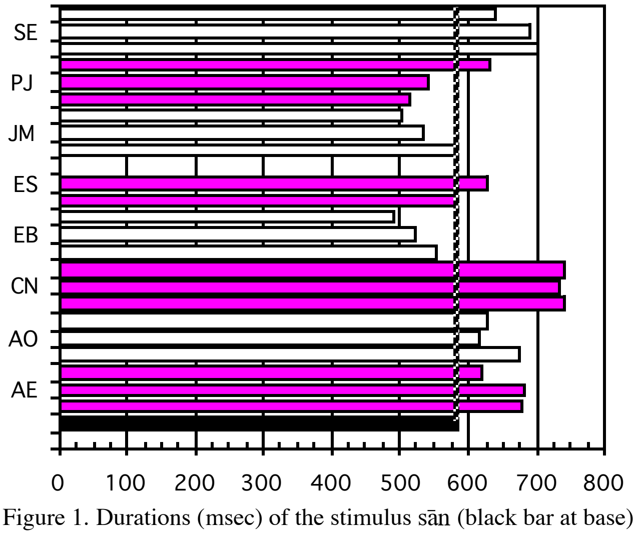

Figure 1. Durations (msec) of the stimulus sān (black bar at base) and imitations by each speaker. The vertical dotted line shows the target duration across all imitations.

Speaker	***n* undershoot/*n* overshoot**				
	Phrase1	**Phrase2**	**Phrase3**	**Total**	**(±20ms)**
AE(f)	2/13	9/6	2/11	**13/30**	6
AO(f)	2/13	5/9	1/14	**8/36**	7
CN(f)	0/15	3/12	0/13	**3/40**	7
EB(f)	5/10	5/10	0/15	**10/35**	4
ES(f)	1/13	3/12	0/15	**4/40**	3
JM(m)	10/5	9/6	3/11	**22/22**	16
PJ(m)	4/11	7/8	3/12	**14/31**	6
SE(m)	1/13	2/13	1/14	**4/40**	7
Total	25/93	43/76	10/105	**78/274**	56

Table 3. Durational undershoot/overshoot by each speaker. Number of imitations falling within ±20ms of the target is shown on the right. (m=male, f=female)

Summed across all stimuli, speaker JM is clearly different from the other speakers, with 50% of imitations showing undershoot, whereas no other speaker exceeds 31% (PJ). The table also shows the number of "close" imitations—those which came within ±20ms of the stimulus. This grouping is arbitrary, but based on frequency distributions, seemed a reasonable range to choose. Once again, JM is noticeably different from the other speakers, with more than twice the number of close imitations.

The undershoot/overshoot values were examined for statistically significant differences between speaker samples. The results of a two-tailed unpaired t-test are shown in Table 4. The test confirms that the observed differences between JM and the other speakers are significant (against all speakers). Speakers CN and ES are significantly different from all other speakers, and this would appear to relate to their strong overshoot of target durations. This is observable in the right part of Figure 2, which shows the means and standard deviations for undershoot/ overshoot for each speaker across all stimuli. Furthermore, the magnitude of CN's overshoot is reflected in the larger standard deviation for this speaker (sd=125ms, compared to 59-80ms for other speakers).

Speaker	AO	CN	EB	ES	JM	PJ	SE
AE	-	*	-	*	*	-	*
AO		*	-	*	*	-	-
CN			*	-	*	*	*
EB				*	*	-	-
ES					*	*	*
JM						*	*
PJ							*

Table 4. Two-tailed unpaired t-test for all imitation/stimulus discrepancies by each speaker. *: significant at p≤0.05.

The consistency of imitations by each speaker—the amount of internal variation, rather than deviation from the target—was examined in terms of the largest difference between the three imitations by each speaker for each stimulus. For example, it can be seen from Figure 3 that speaker CN's imitations varied considerably in duration, over a range of 333ms (the lowest bar is at 1056ms, the second at 1389ms), whereas the imitations by AO and JM differ only by 28ms and 31ms respectively (they are also very close to the target duration, shown by the black bar at the bottom of the graph, and the dotted vertical line).

Averaged across all stimuli, the mean differences between imitations for all speakers were quite similar, lying between 57ms and 66ms (sd :37-46ms), with the exception of speaker CN, whose mean difference was 96ms (sd:101ms), significantly different from the other speakers. A similar pattern obtained for

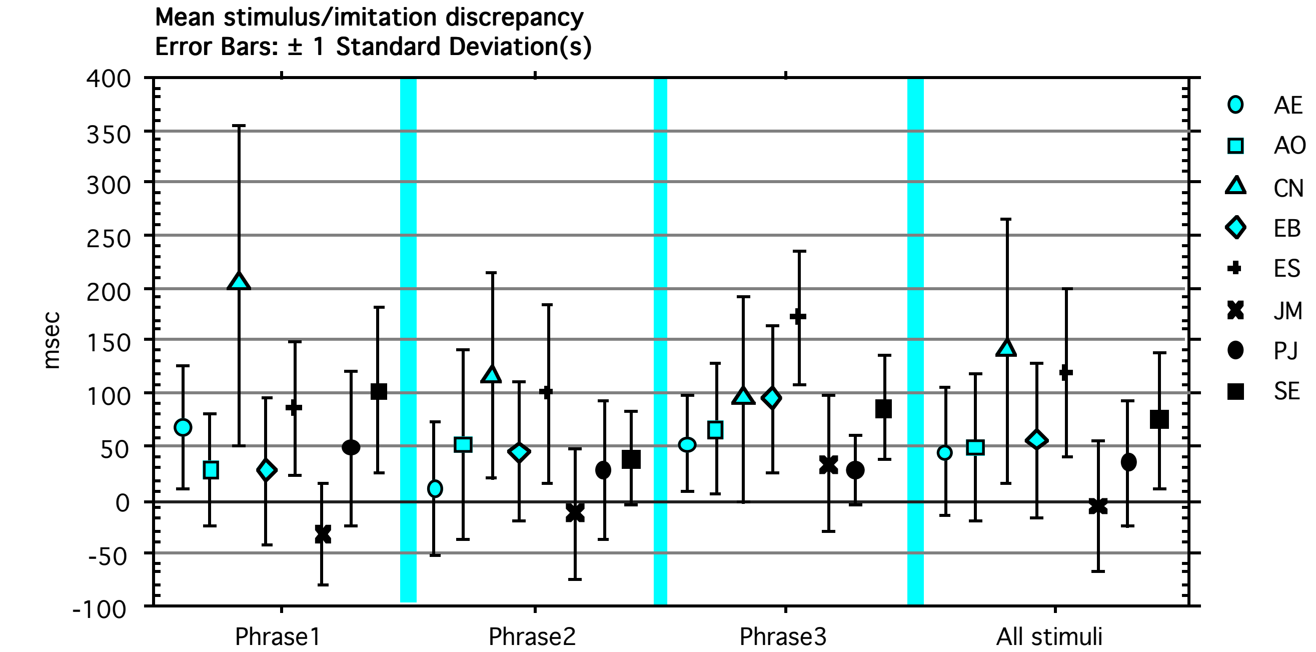

Figure 2: Means and standard deviations of discrepancies between stimulus and imitation durations, by speaker, and by stimulus group.

Phrase1 stimuli as a group, but there was less predictable variation for Phrases2 and 3.

2.5. Intonational results Speakers appeared to behave in more predictable ways in their imitation of the F0 patterns of the stimuli, in that some speakers were clearly better at matching the patterns of the targets and consistently imitated the same stimulus in a similar way.

Speakers AE, AO, JM, and PJ appeared to be more consistent than the remaining speakers, in that multiple imitations of the same stimulus were close to identical in terms of absolute frequency and tonal movement. The best match with the patterns of the target were achieved by JM, PJ, and SE. Frequent inconsistency in the form of multiple imitations was found for ES.

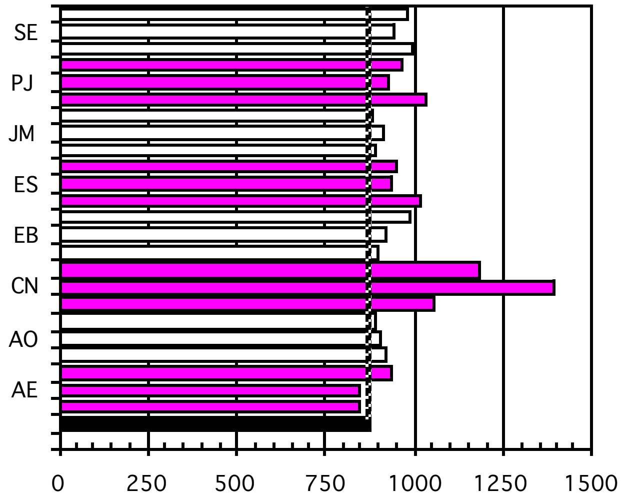

Figure 3. Durations (msec) of the stimulus sān gè fáng jiān (fast) (black bar at base) and imitations by each speaker. The vertical dotted line shows target duration across all imitations.

Of particular interest was the rising tone on the first syllable of stimulus 1c (fáng jiān [faŋ˧˥ tɕɪɛn˥]). This sequence of rising plus high tone was reproduced by four speakers more like a low plus high tone. The rising percept was obtained by F0 movement through the nasal of FANG in the order of ≥+30Hz (females) and ≥+25Hz (males). One speaker consistently produced rising F0 throughout the utterance. Figure 4 shows rising+high imitations by one speaker (AO) and low+high imitations by another (AE) of the stimulus discussed above.

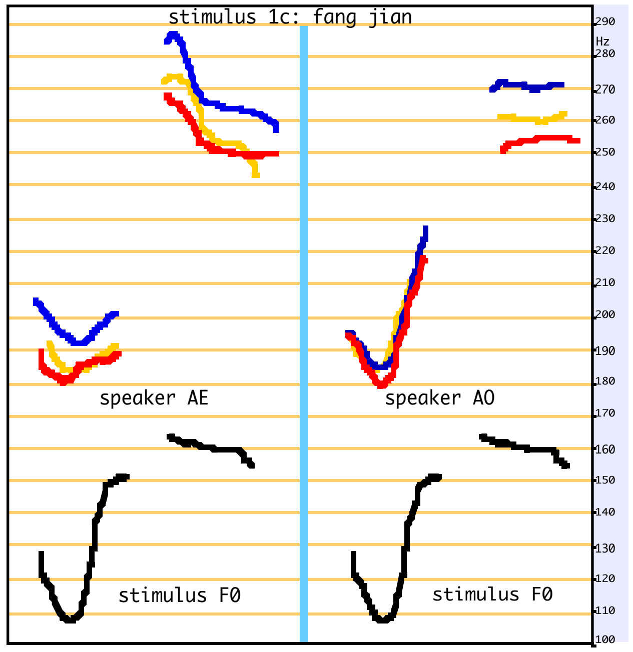

Figure 4. F0 (smoothed) for imitations of stimulus 1c (fáng jiān) by speakers AE and AO.

Curiously, for the entire phrase sān gè fáng jiān at a slow reading speed, all female speakers produced a rise for the final tone, rather than the target level tone (in fact showing a decline of approx 10Hz). Male speakers did not do this. It is possible that the high tone end of the phrase (in this stimulus approx 35Hz higher than the beginning of the rise) was interpreted as a continuation tone (as in sentence prosody), which would normally rise in the speakers' L1, Swedish. A high tone termination of an isolated declarative utterance in Swedish would be unusual. For the fast reading of the above phrase, four of the five females showed a rise, albeit smaller than in the slow reading imitations. The fast stimulus was similar to the slow one, except that the distance from the beginning of the rise to the onset of the high tone was approx 25 Hz.

3. DISCUSSION

The data presented here illustrate what can be observed in naïve imitators of second language stimuli. Whilst one can state broadly that considerable inconsistency and speaker-specific behaviour is found, the few systematic tendencies which can be observed do not appear to shed much light on what occurs in the imitative process.

Clearly, there is a strong tendency to overshoot, rather than match or undershoot the target, although the degree of overshoot can be both speaker and stimulus-specific. Conditioning factors were not identifiable within the stimuli, at least with regard to durational characteristics. It is possible that segmental difficulties influence overshoot (although these measurememts are of quasi-fluent imitations, irrespective of the actual accuracy of individual segments).

Frequent similarities in the consistency of imitations (irrespective of how close the match to the target) indicate that often speakers will either develop a model of the stimulus and not revise it (fixed production), or interpret the stimulus in the same way at each attempt (fixed perception).

Analysis of the imitation of F0 in the stimuli showed that some speakers are clearly more capable of producing close imitations. Possible interference from the L1 was observed in the conversion of high final tones to high rising tones, akin to a continuation tone in Swedish or English discourse. Given that Swedish (the subjects' L1) is tonal, one might have expected some sensitivity to tonal movement, yet these results and anecdotal evidence from language teachers indicate that a tonal L1 imparts little or no initial advantage on a learner of a tonal L2.

It is interesting to note that this task, designed to remove linguistic processing load and thereby facilitate maximal performance, did not yield consistent close imitation. It is possible that the phonetic unfamiliarity of the stimuli presents an initial hurdle which must be overcome before global perception or planning becomes accurate (assuming this is possible for all speakers).

The existence of stimulus specific and speaker-specific differences or effects in these data draw our attention to how little we know about the perception-to-production process in utterance planning and articulatory control for complex stimuli. The results agree with observations in [5] on two subjects' imitations of their own previously recorded complex phrases (in their L1) showing clear differences beteen subjects, in terms of both durational and intonational reproduction ability.

REFERENCES

1. Flege, J. E. and W. Eefting 1987. Production and perception of English stops by Spanish speakers of English. *Journal of Phonetics*, 15, 67-83.

2. Flege, J. E. 1991. Age of learning affects the authenticity of voice-onset-time (VOT) in stop consonants produced in a second language. *JASA*, 89, 395-411.

3. Markham, D. J. 1997. Phonetic Imitation, Accent, and the Learner. Lund University Press, Lund.

4. Bongaerts, T., B. Planken and E. Schils 1995. Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton and Z. Lengyel (Eds.), *The Age Factor in Second Language Acquisition.* Multilingual Matters, Clevedon.

5. Markham, D. J. 1994. Prosodic imitation: Productional results. In *ICSLP'94,* Vol. 3, 1187-1190, Acoustic Society of Japan, Yokohama.

6. Ioup, G., et al. 1994. Reexamining the critical period hypothesis: A case study of successful SLA in a naturalistic environment. *SSLA*, 16, 73-98.

7. Flege, J. E., M. J. Munro and I. R. A. MacKay 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication*, 16, 1-26.