Experiments on the Perception of Mandarin Vowels and Tones*

Mandarin vowels and tones were synthesized on the basis of measurements made of the vowel formants and fundamental frequency contours of citation forms of monosyllabic morphemes pronounced by one male speaker of the Peking dialect. Monosyllables containing representative vowels were produced with the Pattern Playback, and the pitch curves of monosyllabic utterances were manipulated with the Digital Spectrum Manipulator to (1) synthesize the four citation tones and (2) to eliminate variation of the fundamental frequency. Recordings of these synthesized vowels and tones were randomized and presented to several speakers of Standard Mandarin for identification.

To date, responses have been obtained from at least six listeners for each test, with the following intelligibility scores:

Synthesized Vowels			Manipulated Tones		
Test 5	CV	75.5%	Test 13	bāo (isol.)	95.4%
Test 6	V	83.3	Test 14	bāo (sent.)	81.3
Test 7	mid:low	98	Test 15	ying (120 Hz)	32.5
Test 8	ɪə:VV	100	Test 16	ying (Hiss)	39.2

The results of Test 5 (just listed) are especially gratifying because of the improvement over the results of Test 4. That test, using stimuli similar to those of Test 5, was described in an extended report--under the same title--in SR 7/8 (1966). In

*See SR-7/8 (1966), pp. 7.1 - 7.11.

Test 4 (referred to in the earlier report as "the second pilot test of CV syllables") the listeners identified both the synthesized vowels [i] (as found in xi) and [ɯ] (shi) overwhelmingly as [ü] (xu). New patterns were made for these two vowels for Test 5, and appropriate identifications now predominate. The listeners' responses in Test 5 are shown in the form of a confusion matrix in Figure 1. While there is still considerable confusion between [i] and [u], it will be seen that [ɯ] is now quite distinct.

The remaining confusion of [ɨ] (si) for [ɯ] (shi) in Test 5 is interesting, since in natural speech they occur only in mutually exclusive environments--which furnish important cues for their perception. The purpose of this test using "neutral" initial consonants was to test the distinguishability of these two vowels (which occur only in CV syllables), from other vowels as well as from each other. Because [ɯ] occurs only after retroflex consonants and [ɨ] occurs only after alveolar consonants, the ambiguous friction noise in these stimuli eliminates an expected contextual distinctive feature and forces listeners to make an unnatural discrimination. Thus it is not surprising that they should hear synthesized [ɨ] as [ɯ] more often than as [ɨ]. What is perhaps surprising is that they should recognize [ɯ] so well.

The synthesized vowels [i] (as found in yi) and [ü] (yu) of Test 6 have the same formant patterns as those of Test 5. The greater intelligibility of [i] in Test 6 is no doubt due to the absence of other high unrounded vowels in this test. The listeners' responses in Test 6 are shown in Figure 2.

Test 7 is described in the earlier report. Test 8 requires a binary choice between [ɨə] (as found in e) and each one of the following syllables [üɛ iɛ id uɔ ud] (yue ye ya wo wa). Because of the very high intelligibility scores in these tests, little would be gained by presenting confusion matrices for them.

Tests 9 through 12 were preliminary tests of tone per-

TEST 5

VOWELS SYNTHESIZED

VOWELS HEARD	Xi	Xü	Xu	Xɨə	Xɯ	Xɨ	Xɑ	TOTALS
Xi	10	4						14
Xü	7	15			3			25
Xu	2		21					23
Xɨə	1			20		1		22
Xɯ	1	1			16	11		29
Xɨ		1		1	2	8		12
Xɑ						1	21	22
TOTALS	21	21	21	21	21	21	21	147

INTELLIGIBILITY: 75.5%

Confusion matrix of vowels preceded by an ambiguous friction noise, as identified by 7 subjects (3rd series of stimuli).

FIGURE 1

TEST 6

VOWELS SYNTHESIZED

VOWELS HEARD	i	ü	u	e	o	ɑ	ɑəʳ	TOTALS
i	16	5						21
ü	2	13	1	1				17
u			14		2			16
e				17	1			18
o			3		10			13
ɑ					2	18	1	21
ɑəʳ					3		17	20
TOTALS	18	18	18	18	18	18	18	126

INTELLIGIBILITY: 83.3%

Confusion matrix of vowels without an initial consonant, as identified by 6 subjects.

FIGURE 2

TEST 13

TONES IMPOSED

TONES HEARD	1¯	2/	3v	4\	TOTALS
1¯	58			1	59
2/	1	54	2		57
3v		6	58		64
4\	1			59	60
TOTALS	60	60	60	60	240

INTELLIGIBILITY: 95.4%

Confusion matrix of synthetic tones on bāo, as identified by 12 subjects, with the carrier sentence deleted.

FIGURE 3

TEST 14

TONES IMPOSED

TONES HEARD	1¯	2/	3v	4\	TOTALS
1¯	58	9			67
2/		44	26		70
3v		7	33		40
4\	2		1	60	63
TOTALS	60	60	60	60	240

INTELLIGIBILITY: 81.3%

Confusion matrix of synthetic tones on bāo, as identified by 12 subjects, with the carrier sentence retained.

FIGURE 4

ception, undertaken before experimenting with synthesized tones. The stimuli for these four tests retained the original tone contours, successfully verifying the hypothesis that real-speech tones in citation form can be recognized in isolation. These pilot tests, as well as Tests 13 and 14, are described in the earlier report.

The synthesized tones were highly intelligible in Test 13, where they are isolated from the context in which they were originally spoken. Indeed, half of the listeners scored 100 percent in this test. A confusion matrix of all responses in Test 13 is shown in Figure 3.

The same synthesized tones were less intelligible--especially Tone 3--in Test 14, where the phonological environment of the carrier sentence is retained. The implications of these results are discussed in the earlier report. The listeners' responses in Test 14 are shown in Figure 4.

In Tests 15 and 16, the effects of monotone and whisper, respectively, are imposed upon the original pitch curves. The fundamental frequency was set at a constant 120 Hz for monotone, and the Vocoder sound source was set at hiss instead of buzz for "whisper". The listeners' responses in Tests 15 and 16 are shown in Figures 5 and 6. The results are largely what might be expected: the percentage of correct responses is not much greater than chance.

The results of Tests 15 and 16 seem to confirm earlier observations (Abramson, 1962) that pitch is the primary feature in the perception of tones for speakers of a tone language. This conclusion seems to be further supported by the slightly greater intelligibility of the "whispered" syllables produced with hiss excitation as compared with the monotone syllables. When the distracting effect of the monotone pitch is absent, the concomitant features of intensity and duration can, in the "whispered" syllables, play a greater role in the perception of the tones.

TEST 15

TONES SPOKEN

TONES HEARD	1¯	2/	3v	4\	TOTALS
1¯	8	3	15	13	39
2/	6	15	4	4	29
3v	6	8	5	5	24
4\	10	3	4	11	28
TOTALS	30	29	28	33	120

INTELLIGIBILITY: 32.5%

Confusion matrix of original tones for ying, with fundamental constant at 125 Hz, as identified by 6 subjects (carrier sentence deleted)

FIGURE 5

TEST 16

TONES SPOKEN

TONES HEARD	1¯	2/	3v	4\	TOTALS
1¯	7	4	8	4	23
2/	6	12	11	1	30
3v	11	13	10	7	41
4\	6	1	1	18	26
TALS	30	30	30	30	120

INTELLIGIBILITY: 39.2%

Confusion matrix of original tones for ying, with the Vocoder set at hiss source, as identified by 6 subjects (carrier sentence deleted).

FIGURE 6

In the most striking case of features other than fundamental frequency serving as supplemental cues for the perception of tonal distinctions, the sharply falling pitch curve of Tone 4 can be rather readily heard as a sharp drop in the intensity of the hiss. Thus it will be seen in Figure 6 that Tone 4 had by far the greatest intelligibility in Test 16. It is interesting to note, however, that no very well defined pattern of tone discrimination emerges from this confusion matrix, such as was obtained for at least some subjects by Abramson (1958) in similar experiments with Thai tones and Vocoder hiss.

John M. Howie
University of Missouri

References

Abramson, A. S. Vocoder output and whispered speech in a tonal language: Thai. Paper presented at the New York meeting of the Modern Language Association, December 28, 1958.

Abramson, A.S. The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments. Indiana U. Res. Cent. in Anthro., Folkl., and Ling., 20 (Bloomington, 1962).