SOME EXPERIMENTS ON THE PERCEPTION OF MANDARIN TONES

JOHN M. HOWIE

The principal aim of this study has been to make some general statements about the acoustical properties of Peking Mandarin tones, as they occur in citation forms of monosyllabic morphemes. The generality of the description is attained through the use of speech synthesis, by validating measurements made from a single speaker in perception tests with a number of listeners who are speakers of the language.

Mandarin has four distinctive tones which can be observed in citation forms of monosyllabic morphemes. A set of sixty Mandarin citation syllables was selected so as to include several variants of each tone, occurring in syllables of various segmental structures. This set of citation syllables was constructed from fifteen sets of monosyllabic morphemes minimally distinguished by tone, such as *mā* 'mother', *má* 'hemp', *mǎ* 'horse', *mà* 'to scold'. The caption of Figure 1 presents these Mandarin citation syllables in the *Pinyin* romanization (with tones unmarked).

The prosodic features called tones are implemented mainly by differences in the perceived pitch of the voice. The physical dimension of sound most closely correlated with perceived pitch is fundamental frequency. The course of the fundamental frequency during the entire voiced part of the sixty syllables was measured from spectrograms of the recorded utterances of one male speaker of Peking Mandarin. All the syllables were normalized as to duration, by locating the points chosen for fundamental frequency measurements on the spectrograms on the basis of per cent of duration of the voiced segment. The frequency values for each syllable were then plotted logarithmically against duration, yielding fifteen curves for each tone. Finally, generalized average curves of the four tones were obtained by drawing a smooth curve by eye through the superimposed plotted curves of each tone. All the plotted curves, and the resulting average curves (which may be seen in Figure 1), have, in effect, been 'stretched' or 'shrunk' to the same length through the normalizing procedure.

From the shapes of the generalized average curves, the contours of the four citation tones of Mandarin can be described as quite distinct. By examining Figure 1, one finds the overall range of the average curves to be approximately ten semitones — slightly less than one octave — and one can make the following statements about the contours:
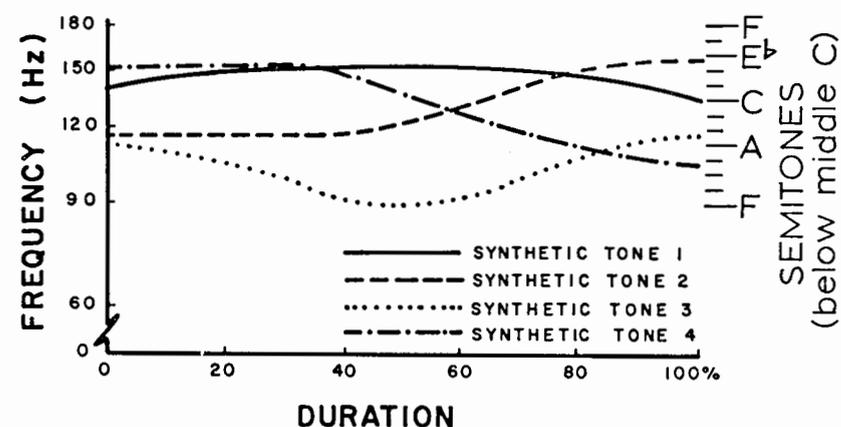
Fig. 1. Average curves of the four citation tones of Peking Mandarin (spoken by one male informant), used as the basis for synthesizing the tones. Fundamental frequency values were plotted for sixty citation syllables representing fifteen sets of monosyllabic morphemes minimally distinguished by tone: *yi, yu, ai, ya, yao, yuan, ying, fa, ma, ke, gɔ, qie, can, jie, zan*. The mean durations of the voiced part of the syllable are: Tone 1, 237 msec; Tone 2, 271 msec; Tone 3, 277 msec; and Tone 4, 251 msec.

Tone 1 is high in the upper five semitones,
Tone 2 rises through the upper five semitones,
Tone 3 remains within the lower five semitones,
Tone 4 falls through the upper five semitones.

These statements, as well as the shapes of the average curves, are in good agreement with Professor Y.R. Chao's well-known descriptions of the contours of Mandarin citation tones (Chao 1948:25, 1968:26):

Tone 1 has a high-level contour,
Tone 2 has a high-rising contour,
Tone 3 has a low-dipping contour,
Tone 4 has a high falling contour.

Perception tests were run using Mandarin citation syllables as stimuli and Mandarin monosyllabic morphemes minimally distinguished by tone as inventories of possible responses. Speakers of the language were asked to identify a set of citation syllables with the four tones synthesized on the Vocoder-Digital Spectrum Manipulator, built by the Haskins Laboratories in New York City. The listeners were also asked to identify a set of citation syllables in which the pitch was first made static and then removed entirely, simulating the effects of monotone and whisper. As expected, the synthetic tones were quite acceptable to the Mandarin speakers and easily discriminated by them, while the tones transformed into monotone or whisper were almost impossible to identify.

The Digital Spectrum Manipulator (DSM) is a speech synthesizer that uses a channel Vocoder as one of its principal components. For this reason, and because the Vocoder is a relatively familiar instrument, the DSM will be referred to in this paper

as the Vocoder-DSM. The DSM has been described in some detail by Dr. Franklin S. Cooper, President of Haskins Laboratories, at the Fifth International Congress of Phonetic Sciences (Cooper 1965).

Fundamental frequency values taken from the generalized average curves were used to control the pitch channel of the Vocoder-DSM, imposing new pitch patterns on recordings of real speech citation syllables. The synthetic tones thus have frequency/time patterns corresponding to the average curves of Figure 1. The set of tonally differentiated morphemes used in the perception test of synthetic tones was *bāo*, *báo*, *bǎo*, *bào* (which were not included in the sixty citation syllables measured). The four synthetic tones were imposed on the Tone 1 syllable *bāo*, which has a relatively level pitch pattern. A recording of this syllable, as spoken by the same informant who made the recordings for the spectrographic measurements, was processed with the Vocoder-DSM, manipulating the course of the fundamental frequency to synthesize the four tones. [Demonstrations of the Vocoder output were given here; for each tone, the syllables *ma* and *bao* were presented with the original pitch patterns regenerated, followed by the syllable *bao* with the synthetic tone.]

Five recordings of the Vocoder output were used to obtain five tokens each of the four stimuli for perception test A. The resulting twenty test items were randomized and presented to twelve listeners for identification by means of multiple-choice answer forms with Chinese characters. Intelligibility for the synthetic tones in Test A was 95 per cent. (Eight of the 12 listeners scored 100 per cent in this test.) The results of Test A support the assumption that the fundamental frequency provides effective cues for the perception of the tonal distinctions in Mandarin. It therefore seems reasonable to conclude that the generalized average curves of Figure 1 represent a valid acoustic description of Mandarin tones in citation syllables.

Another kind of experiment on the perception of the tones was undertaken briefly, to complement earlier experiments with real whispered speech and vocoder output in various tone languages. The tonal contours in a set of real speech citation syllables were suppressed in one test by imposing on them a perfectly level pitch, and in another test by substituting an aperiodic sound source, that is making the syllables voiceless. In Test B, a set of tonally differentiated morphemes, *yīng*, *yíng*, *yǐng*, *yìng*, was rendered monotone, at a pitch chosen to be as neutral as possible (which was the point on the frequency scale at which the Tone 2 and Tone 4 contours cross each other). Recordings of these syllables, spoken by the same informant as before, were processed by the Vocoder-DSM, changing the fundamental frequency to a constant 128 Hz. [Demonstrations of the Vocoder output were given here; these four syllables were presented as monotone.] In Test C, the same set of morphemes was reduced to whisper. The same recordings were processed further, with the buzz/hiss control of the Vocoder-DSM set at Hiss. [Demonstrations of the Vocoder output were given here; these four syllables were presented as 'whisper'.]

Five recordings of the Vocoder output were used to obtain five tokens each of the four stimuli for each of these tests. The resulting twenty test items in Tests B and C

were randomized and presented to six listeners for identification, as in Test A. The results are largely what might be expected: the percentage of 'correct' responses is not much greater than chance. In Test B, with the monotone syllables, intelligibility was 31 per cent; in Test C, with the 'whisper' syllables, it was 39 per cent. The results of Tests B and C, in which the feature of pitch was suppressed, confirm Arthur Abramson's observations that pitch is the primary feature in the perception of tones for speakers of a tone language (Abramson 1972).

The purpose of this study has been to make some general statements about the acoustical properties of the tones in Mandarin citation syllables.

Experiments on the perception of isolated monosyllables processed by Vocoder have led to the following conclusions.

(1) Since highly intelligible tones can be synthesized using the average curves derived from fundamental frequency measurements of real speech, the average curves represent a valid general description of the acoustical properties of Mandarin tones.

(2) The fundamental provides effective cues for the perception of the tonal distinctions in Mandarin.

(3) When variation of the fundamental is instrumentally suppressed, tonal discrimination in Mandarin citation syllables is nearly impossible, even in the presence of concomitant features which may be observed in the utterances. This result points to the primacy of pitch in the perception of tones.

University of Missouri
Columbia, Missouri

REFERENCES

Abramson, A.
  1972 "Tonal Experiments with Whispered Thai", in *A la mémoire de Pierre Delattre*, A. Valdman, ed. (The Hague, Mouton).
Chao, Y.R.
  1948 *Mandarin Primer* (Cambridge, Harvard University Press).
  1968 *A Grammar of Spoken Chinese* (Berkeley, University of California Press).
Cooper, F.S.
  1965 "Instrumental Methods for Research in Phonetics", in *Proceedings of the Vth International Congress of Phonetic Sciences*, Münster, 1964 (Basel, Karger) pp. 142-171.

DISCUSSION

JÜRGENSEN (Copenhagen)
Do any of the examples quoted as being capable of carrying more than one meaning have one chief sense which, when heard in isolation, would make the listeners choose that particular sense of the word rather than a more obscure one?

HOWIE

To the best of my knowledge, differences in frequency of occurrence and familiarity among the four morphemes of these sets are not especially great, and would not be complicating factors in the listeners' responses. (Actually, the case is not the same as with English homophones, since these Mandarin morphemes do have distinct pronunciations.)

In the tests with both the monotone and the 'whisper' syllables, the distribution of the listeners' responses was, to a very large extent, random — especially with the monotone syllables. Very little patterning of the responses can be seen in these experiments.

ADAMS (Haberfield, Australia)

Do you consider that intensity could be at all significant as a cue for the perception of tone on whispered syllables?

HOWIE

In replaying the 'whisper' syllables, I would like to draw your attention to the falling intensity contour of Tone 4, which parallels the falling pitch contour. To a limited extent, the listeners were able to make use of this concomitant feature of intensity as a supplementary cue for the perception of Tone 4, resulting in a much higher recognition score for Tone 4 than for the other 'whisper' tones, i.e., 60 per cent. This is still very poor recognition, compared with that for the synthetic tones, and suggests that native speakers of tone languages are not easily able to utilize phonetic features other than pitch for the perception of tones.

ROTHMAN (Gainesville, Fla.)

Did the vocoder make any changes in spectral output to produce an aural percept of change for Tone 2?

HOWIE

In the pronunciation of these four syllables, *yīng, yíng, yǐng, yìng*, differences in duration, intensity, and vowel quality may be observed, accompanying the different tones. These concomitant phonetic features are not the result of processing the speech with the Vocoder-DSM, but were always present in the spoken utterances. They were not so readily perceptible when the pitch feature was also present, and are made more prominent when variation of the fundamental is suppressed by the Vocoder-DSM.