ACOUSTIC CHARACTERISTICS OF THE SHANGHAI-ZHENHAI SYLLABLE TYPES

Philip John Rose

0. INTRODUCTION

This paper presents a quantified description of the main acoustic characteristics of the six contrasting syllable-types in the Shanghai-Zhenhai variety of Chinese (Sh-Zh).[1] Zhenhai is a rural county north-east of the municipality of Ningpo in Zhejiang province, and Sh-Zh can be described as a common type of Zhenhai dialect which exhibits to a greater or lesser extent lexical and phonological influence from the neighbouring prestigious dialect of Shanghai.[2]

Sh-Zh belongs to the Wu dialect group of Chinese, and has typical Wu phonology. This includes complex tone sandhi, restriction of syllable-final consonants to [ŋ] and [ʔ], and respective tripartite and bipartite division, according to manner, of syllable-initial occlusives and fricatives (Yuan 1960:59; Chao 1967).

The acoustic parameters investigated in this study were fundamental frequency (F_0), duration, and oral amplitude (Ao). Fundamental frequency is the acoustic correlate of the rate of vibration of the vocal cords, and is usually assumed to correspond to the perceptual dimension of pitch. Duration, the time dimension of the acoustic signal, corresponds to perceived length, and is determined physiologically by the relative timing of articulatory events. Oral amplitude refers to the integrated time-varying sound pressure transduced at a distance from the speaker's lips. It is important to realise that the oral amplitude thus measured is a function of three factors which vary in an essentially independent manner, and which are consequently difficult to isolate. The first of these, the time-varying amplitude of the glottal source, occurs extrinsically as the result of articulatory gestures which affect the sub-glottal pressure (P_s), such as differing respiratory effort or adjustments in the open quotient of the glottal cycle (Zemlin 1968:198-204). Thus, other things being equal, a change in P_s will be reflected by a change in Ao.

The second factor, the transfer function of the supralaryngeal vocal tract, modifies the amplitude of the glottal source by attenuating the transfer of energy at certain frequencies and passing maximum energy at others (Lieberman 1977:31-36). The transfer function is determined by the shape of the supralaryngeal vocal tract, changes in which - as for example when different vowels are articulated - will then differentially modify the amplitude of the glottal source. Thus, other things being equal, a change in the supralaryngeal vocal tract shape will be reflected by a change in Ao. This is the reason for intrinsic vowel amplitude (Lehiste and Peterson 1959:429).

In addition to the above two factors, changes in Ao will occur as the result of the interaction of harmonic and formant frequencies (House 1959). Thus, if P_s and transfer function are held constant, changes in F_o will be reflected in changes in Ao.

From the articulatory point of view, then, the oral amplitude is of prosodic interest only in so far as it reflects those amplitude features which the speaker is extrinsically controlling, that is, the time-varying amplitude of the glottal source. It is therefore necessary to ensure that fluctuations in Ao due to intrinsic vowel amplitude and the interaction of harmonic and formant frequencies are kept to a minimum. I have done this by 1) selecting examples spoken with as near steady state supralaryngeal configurations as possible, monophthongality being adjudged by reference to formant trajectories on wide band spectrograms, and 2) analysing about equal numbers of open and close vowels in each particular sample. Although it would have been possible to apply a correction factor to the oral amplitude to eliminate the effects of the interaction between formant and harmonic frequencies, I have not done so, because the actual F_o range used by the informant in this study was too narrow to have caused appreciable changes in Ao. The oral amplitude data presented below can therefore be taken to give a reasonable approximation of the time-varying amplitude of the glottal source.

Amplitude is usually assumed to correlate with perceived loudness. In the perception of speech, however, there is evidence that listeners base loudness judgements on features more directly related to the P_s and glottal source amplitude than the oral amplitude (Ladefoged 1967: 35-41).

1. PREVIOUS STUDIES

There have been very few acoustical studies on Wu tones, and there are, to my knowledge, none which pay attention to all three parameters of F_o, Ao, and duration.

Liu (1925) analysed the F_o and duration of two Jiangyin tones kymographically, and there are also some kymographic records of various Ningpo utterances in Tchen (1938). Sokolov (1965) presented data on the F_o and duration of Shanghai citation monosyllables, and the F_o and duration of some Shanghai monosyllabic and polysyllabic utterances have been investigated by Zee and Maddieson (1979). A comparison of their results with those presented below can give an idea of the degree of similarity in F_o shapes between Shanghai and Sh-Zh/Zhenhai dialect.

The nearest site to Zhenhai for which descriptions of citation tone pitch values are available is Ningpo town. Ningpo town was one of the sites visited by Y.R. Chao in 1927 when collecting material for his pioneering monograph on the Wu dialects (Chao 1928). Chao's pitch descriptions compare visually very well with the F_o data obtained in this study.

The pitch of the citation tones of Ningpo town has also been recorded in a recent description (Shi 1979) with the five-point system devised by Chao (1930). The pitch values given, however, do not agree well with those in Chao (1928) or the present study, and the transcriptions must, therefore, be treated with caution.

2. SYLLABLE-TYPES

In Shanghai-Zhenhai, any citation monosyllable or monosyllabic word belongs unambiguously to one of six contrasting types. Below are listed examples of these six syllable-types, together with their main auditory characteristics.

1. The pitch of type 1 syllables falls from high in the speaker's pitch-range to low, with a short initial level component. (In terms of Chao's (1930) five-point pitch scale, its value would be 42 or 442.) The fall often starts higher, and sounds more abrupt, in syllables with nasal codas. The first half of the Final[3] seems the loudest, and the length of the Final is shorter than average for the syllable types. Voice quality is normal, and there is a gradual offset to voicing. Examples[4] are:

[sɿ][5] : *'poem'*; [tĩ] : *'shop'*; [ʔlœʏ] : *'to pick'*, e.g. *'the nose'*;
[tɕɪ̃ŋ] : *'needle'*.

2. Syllables of type 2 have a concave pitch contour, starting in

mid pitch-range, falling, and then rising just into the upper third of the pitch-range, where voicing is terminated by a glottal-stop. The pitch shape would be best transcribed by 324 in the Chao notation: this adequately expresses the relative onset and offset heights, and concave shape, but implies too low a dip in pitch. Other possibilities could therefore be 334 or 434.

Loudness is concentrated at the beginning and end of the Final, with the end usually the loudest. Length is above average, and the voice often has a somewhat tense quality. Some examples are:

[tɕiʔ] : *'chicken'*; [sɿʔ] : *'water'*;
[kãʔ] : *'river'*; [tĕŋʔ] : *'to wait'*;
[tɕɪ̆ŋʔ] : *'well (of water)'*.

3. Type 3 syllables have striking auditory characteristics, resulting primarily from the combination of a convex pitch contour which starts in the low part of the pitch-range - 232 or 242 - with loudness energetically concentrated on the first half of the syllable. In addition, voice quality over approximately the first third of the Final is whispery [_̳]. There is gradual offset to voicing. Syllables of type 3 have average length. Examples are:

[nɛ̳ɩ] : *'south'*; [mɑ̳̃] : *'busy'*;
[mə̳̆ŋ] : *'door'*; [d̥ʑ̊ɪ̳̆ŋ] : *'town'*;
[ʑ̊ɿ̳] : *'to be'*; [d̥ɪ̳̃] : *'sweet'*.

4. Syllables of type 4 have a concave pitch contour, starting in the low pitch-range, falling to the bottom of the range and then rising again to mid pitch-range, where voicing is optionally terminated in a glottal-stop. The pitch shape would be best transcribed by 213 or 214. Loudness is concentrated at both beginning and end of the Final, but the end sounds considerably louder, with a typical burst-like quality. Type 4 syllables are the longest, and have whispery voice over the first half to two-thirds of their length.

[ʑ̊ɿ̳] : *'word'*; [d̥i̳ʔ] : *'ground'*;
[d̥ʑ̊ɪ̳̆ŋ] : *'near'*; [mɑ̳̃ʔ] : *'dream'*

5. Syllables of type 5 give an auditory impression of very short, high level pitch, which ends in a glottal-stop - 5. There is a single burst of loudness, and voice quality sounds somewhat tense. Type 5 syllables sound the shortest of the six types. Examples are:

[fĕʔ] : *'law'*; [tɪ̆ʔ] : *'target'*;
[tɕjĕʔ] : *'foot'*; [tɕhɥɔ̆ʔ] : *'to eat'*.

6. Syllables of type 6 have a pitch contour which rises abruptly from low in the speaker's pitch-range to upper mid, where it ends in a glottal-stop - 24 or 23. There is a single burst of loudness. They