Intelligibility of high-pitched vowel sounds in the singing and speaking of a female Cantonese Opera singer

Dieter Maurer [1], *Peggy Mok* [2], *Daniel Friedrichs* [3], *Volker Dellwo* [3]

[1] Zurich University of the Arts, Institute for the Performing Arts and Film, Switzerland
[2] The Chinese University of Hong Kong, Dept. of Linguistics and Modern Languages, Hong Kong
[3] University of Zurich, Phonetics Laboratory, Switzerland

dieter.maurer@zhdk.ch, peggy.mok@cuhk.edu.hk, daniel.friedrichs@uzh.ch, volker.dellwo@uzh.ch

Abstract

The question whether or not vowel quality can be maintained at F0 of the sounds exceeding statistical F1 of „normal" speech is still a matter of debate. The present study investigates the perception of long Cantonese vowels /i, y, œ, a, ɔ, u/, spoken and sung in (C)V and (C)V:S context by a well-known female Cantonese Opera singer in the range of F0 of c. 560–860Hz. 172 high-pitched syllables or isolated vowel sounds were selected and extracted from a live recording. Corresponding vowel perception was tested in a listening test performed by 26 students of linguistics. All six vowels proved to be identifiable > 80% up to F0 of c. 700Hz, and sounds of /i, a, ɔ, u/ proved to be identifiable > 80% up to a range of F0 of c. 820–860Hz. Confusion matrices are provided in the present paper and spectral illustrations of all sounds are presented at http://is2014.phones-and-phonemes.org.

Index Terms: vowel, perception, F0, speaking, singing, Cantonese Opera

1. Introduction

Is vowel perception inevitably compromised once F0 of a sound exceeds F1 of the intended vowel as given in formant statistics for "normal" speech? Hitherto, in the literature, no general consensus regarding the intelligibility of high-pitched vowel sounds could be established. However, three positions can be identified: (1) sounds on F0 above statistical F1 of a vowel lose their intelligibility [1]; (2) a correct vowel identification of 80% can be maintained up to F0 = c. 500Hz, as holds true for a 50% identification up to c. 650 Hz, with the exception of sounds of /a/ which may exhibit correspondingly higher identification rates [2]; (3) vowel intelligibility can be maintained up to c. 880Hz if the sounds are produced with a raised larynx position or in a CVC context; this holds true up to F0 = c. 1050Hz for both conditions combined [3].

However, except for rare studies [4, 5], empirical investigations of vowel sounds on varying F0 relate to "Western" styles of singing, mostly to the "classical" style ("legit"). "Classical" singing is characterized by a particular coloring and harmonization of the vowels (phenomenon of "vowel modification") as well as a specific voice characteristic and timbre that allows the voice to be heard in the context of a big orchestra, a vocal strategy in which sound projection and timbre are often favored over text or vowel intelligibility. Therefore, the question of high-pitched vowels produced in singing/ speaking styles and traditions other than "Western" arises.

In this context, the investigation of singing and speaking in the style of Cantonese Opera is of specific interest for several reasons: there is no strict separation of speaking and singing; there is no superordination of sound projection and timbre over text intelligibility; in some pieces, male and female roles are sung by women; some roles concern longer parts of singing and speaking on pitches in the range of 500–900Hz which are intended to be intelligible according to the style of Cantonese Opera. Thus, the investigation of singing and speaking on high pitches in Cantonese Opera offers a promising possibility to clarify the relationship of vowel intelligibility and F0.

Above all, for singing, the investigation of long vowels allows for more reliable results than for short vowels, both with regard to acoustic analysis (less transitions in the vowel nuclei) and perception (higher constancy in vowel quality). In Cantonese, there are seven long and seven medium-long vowels in (C)V and (C)V:S syllables. With regard to corresponding statistical formant frequencies for women and "normal" speech of Cantonese, F1 for /i, y, u/ is given < 380Hz, and F1 for /ɛ, œ, ɔ/ is given < 720Hz [6, 7]. Thus, for the present investigation, sounds at F0 > 550 Hz are of special interest: for sounds intended as /i, y, u/, the first lower spectral energy maximum of the sound will always be near or equal to statistical F1 of another long vowel. (Note also the statistical frequency distance F1-F2 < 550Hz for /u, o, a/.)

Against such a background, the utterances of a famous female actress performing as a female character in a Cantonese Opera piece were investigated, focusing on syllables with averaged F0 > 550 Hz of the related vowel nuclei.

2. Method

2.1. Subject, recording, and test material

The speaking and singing of MUI, SHET SZE performing as a female character in three Cantonese Opera pieces, represents the basis of the present investigation (recording of live performance on a special DVD edition, see [9]). The procedure to create the test sample of sounds was as follows: (1) Longer parts of the actress' singing and speaking on high pitches were labeled. (2) Single syllables or isolated sounds of long and medium-long vowels were extracted. The intention of syllables and vowels were taken from the subtitles and were crosschecked acoustically. (3) This first sample was screened by the authors and sounds with generally poor sound quality were excluded. Because no instances of the vowel /ɛ/ with good vowel quality were found, this vowel was also excluded from the investigation. (4) A final test sample of 172 sounds on F0 ≥ 550Hz of the vowels /i, y, œ, a, ɔ, u/, and a training set of 7 phrases on high pitches and 60 syllables on F0 < 550Hz were created. (For /u/, test sounds had to be included in the training set since no sounds with F0 < 550Hz were found).

2.2. Acoustic analysis

For each syllable, the sound nucleus of the long vowel was labeled visually using PRAAT [8]. For the segment selected, average F0 was calculated using the PRAAT standard command. Calculations were acoustically crosschecked relating to the C-major scale. If substantial differences were found, average F0 was estimated acoustically.

2.3. Listening test

26 university students of linguistics (19 women, 7 men, no history of hearing problems), aged 18 to 23 years and native speakers of Cantonese, participated in the listening test. Before the test, they were informed about the purpose of the test, and they listened to the training set (single sounds in random order) without reacting, as to become familiarized with the test sounds.

For the listening test, the sounds of the syllables were presented in random order. The subjects listened to the sounds on a PC using a commercial headphone and were asked to write down the perceived long or medium long vowel quality into a prepared list. Long and medium-long vowels were not differentiated because they do not differ in vowel quality.

3. Results

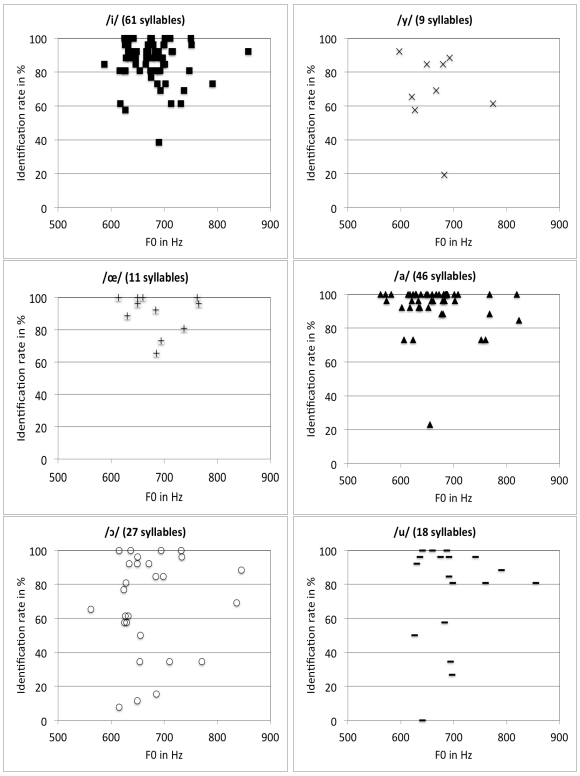

Figure 1: Vowel identification. For each vowel, F0 and the percentage of correspondence between the intended and the perceived vowel quality of single sounds are shown. Symbols: ■ = /i/, ✕ = /y/, ✚ = œ, ▲ = /a/, ○ = /ɔ /, ▬ = /u/.

For 160 of the 172 sounds tested, the identification rate is ≥ 50% according to the intended vowel of the singer. Moreover, an upper limit of F0 of possible vowel identification > 80% was found at 692Hz for /y/, 765Hz for /œ/, 823Hz for /a/, 845Hz for /o/, 855Hz for /u/, and 859 for /i/ (see Figure 1). Remarkably, sounds of /u/ and /i/ prove to be intelligible on F0 > 800Hz comparable to sounds of /a/, although F1 of /i/ and /u/ are given as the lowest values and F1 of /a/ is given as the highest value in formant statistics [6, 7].

In a parallel online presentation, additional acoustic analysis and entire confusion matrices of all sounds investigated are presented; see **http://is2014.phones-and-phonemes.org**.

4. Discussion

Considering that these findings relate to utterances of a live performance and not to utterances under experimental conditions, the results strongly indicate that vowel discrimination in CV or CVC context can be maintained up to a frequency range of F0 of 800–900Hz, confirming the third position mentioned in the introduction.

Several questions arise from such a statement. Above all, possible vowel discrimination is found for F0 of sounds exceeding statistical F1 of all vowels except /a/. For /u/, F0 of high-pitched sounds can even match statistical F2. Concerning vowel production, it is questionable whether a possible raised larynx position alone can explain such a phenomenon. Concerning vowel perception, vowel discrimination obviously cannot rely directly on F1-F2. Concerning the vowel context, discrimination of vowel sounds on high pitches produced in isolation and in logatomes should be investigated in order to assess the influence of transitions and meaning. Concerning the investigation of artistic singing and speaking styles, vowel identification proves to be style-specific. Above all, for "Western" Opera singing, no corresponding F0 level with comparable vowel discrimination is reported.

Further studies on vowel production and perception at different F0 and in different contexts of the vowels are needed to clarify the relationship between sounds, F0, and vowel qualities. Such future investigations concern both artistic as well as everyday utterances and should address documentations of spontaneous everyday utterances, live performances as well as systematic vocalizations under experimental conditions.

5. Acknowledgements

We would like to thank Christian d'Heureuse, Casper Hui and Heidy Suter for assisting in the investigation and publication.

6. References

[1] E. Joliveau, J. Smith, and J. Wolfe, "Vocal tract resonances in singing: The soprano voice," *J. Acoust. Soc. Am.*, vol. 116, no. 4, pp. 2434–2439, 2004.

[2] J. Sundberg, "Perception of Singing," in *The Psychology of Music*, 3rd ed. London: Academic Press, pp. 69–106, 2012.

[3] L. A. Smith and B. L. Scott, "Increasing the intelligibility of sung vowels," *J. Acoust. Soc. Am.*, vol. 67, no. 5, pp. 1795–1797, 1980.

[4] D. Maurer and T. Landis, "F0-dependence, number alteration, and non-systematic behaviour of the formants in German vowels.," *Int. J. Neurosci.*, vol. 83, pp. 25–44, 1995.

[5] D. Maurer, C. D'Heureuse, and T. Landis, "Formant pattern ambiguity of vowel sounds," *Int. J. Neurosci.*, vol. 100, pp. 39–76, Jan., 2000.

[6] E. Zee, "Frequency analysis of the vowels in Cantonese from 50 male and 50 female speakers," *Proceedings of the 15th ICPhS*, pp. 1117-1120, 2003.

[7] W.-S. Lee, "Vowel formant frequency characteristics of adult and preadolescent males and females," *Chinese Journal of Phonetics*, vol. 2, pp. 90-97, 2009.

[8] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International,* vol. 5, pp. 341-345.

[9] 龍情詩意半世紀, 龍劍笙：梅雪詩