VOWEL AND CONSONANT IDENTIFICATION AT HIGH PITCH: THE ACOUSTICS OF SOPRANO UNINTELLIGIBILITY

Francis Nolan and Harriet Sykes

Phonetics Laboratory, Department of Theoretical and Applied Linguistics, University of Cambridge
fjn1@cam.ac.uk; hattie141@mac.com

ABSTRACT

Sopranos are notoriously difficult to understand. This study tracks the progressive loss with extreme high f_0 of (a) vowel quality distinctions, and (b) the percept of a syllable-initial lateral. A soprano sang [lV] syllables on the notes of an arpeggio from A_4 to A_5. V ranged over [i, ɛ, a, ɑ, ɔ, u, ə]. She performed as a phonetician, not a trained singer, so that aesthetic adjustments of vowel configuration were avoided to isolate the effect of f_0.

Twenty-seven students with IPA training responded on a forced-choice vowel quadrilateral, reporting also whether [l] was present. At the highest f_0, all vowels sounded open and lateral detection was erratic. Findings are discussed with reference to acoustic analysis. Loss of spectral peak definition is argued to explain the results, but at intermediate pitches there is some recoverability of vowel articulation thanks to differing relative amplitudes in the first three harmonics.

Keywords: soprano, singing pitch, vowel quality, consonant perception, spectral resolution.

1. INTRODUCTION

Soprano singers are notoriously difficult to understand, to the point where it may even be difficult to tell out of context what language is being sung. *A priori*, the primary reason is likely to be that, as fundamental frequency (f_0) rises, wider and wider spacing of the harmonics of the glottal source means progressively poorer manifestation of the supralaryngeal resonance function, on which vowel quality depends. A second reason, related to the harmonic spacing, is that trained singers learn to maximise sound output by articulatorily tuning a resonance of the vocal tract (formants) to a harmonic, at the expense of vowel identity [4–7, 9].

Previous studies [1–9] of high-pitched singers have generally not controlled these factors, the singers being allowed to sing vowels as they would in a musical performance. The present study uses stimuli sung by a phonetically trained soprano who explicitly avoided articulatory modifications, thus isolating as far as possible the contribution of purely acoustic factors to the intelligibility decrement. The study is also innovative in that it includes the audibility of a consonant as well as that of a number of vowels.

2. EXPERIMENT

2.1. Recording of materials

Seven sung syllables [li, lɛ, la, lɑ, lɔ, lu, lə] were recorded by a soprano singer (the second author), the range of vowels being chosen to reveal changes in the vowel space with pitch. Recordings were made in a sound-treated room using a Sennheiser MKH40 P48 microphone, placed about 10 inches from the singer's mouth, and a Marantz PMD670 Professional Solid-State Recorder at a 44khz sampling rate. Each syllable was produced on seven pitches (comprising a two-octave arpeggio), starting on A_3 and finishing on A_5. The syllables were produced in full voice, defined as singing that fully utilises the resources of the vocal mechanism, with normal vibrato for each pitch. Crucially, the vowels were produced as IPA categories, without the quality distortion that singers are often trained to produce in order to maximize acoustic intensity.

2.2. Subjects

Twenty-seven students took part. All had been studying practical phonetics, comprising ear-training and production including the Cardinal Vowels, for at least 5 weeks as part of a phonetics course at the University of Cambridge, and so were familiar with the phonetic symbols used to represent vowels. IPA symbols are unambiguous in their denotation to a phonetician in a way that alphabetic letters or orthographic forms that a naïve listener might use – as in [1, 3] – are not.

2.3. Stimuli

It was decided to use only the highest four pitches A_4, $C^{\sharp}_5$, E_5, and A_5 (≈ 440, 554, 659, and 880 Hz) in the perceptual experiment, as at lower pitches informal listening suggested the full range of vowel qualities could still be discriminated. Overall there were twenty-eight different speech samples: seven syllables over four pitches each. All stimuli began with a lateral.

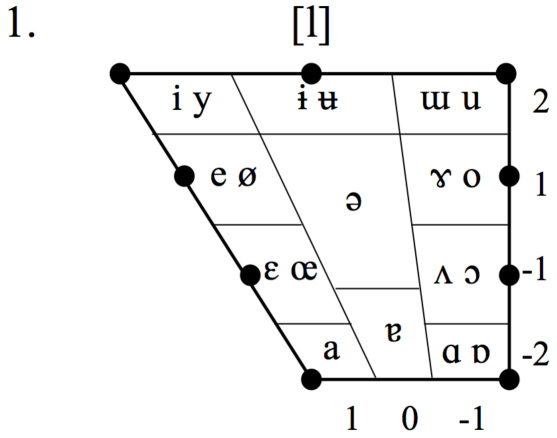
2.4. Presentation

The twenty-eight samples were embedded randomised in a Microsoft PowerPoint slideshow. Three different slideshows were created, so that three responses per subject to each stimulus could be elicited. The randomisation was different for each slideshow to neutralise possible order effects. Each slide lasted for eight seconds, during which the stimulus in question would play automatically twice. Before the experiment commenced, the participants were told that they would hear a voice singing at various pitches, and that they would hear each stimulus twice and have a few seconds to respond. After the samples had played, the participants would have approximately 6 seconds left to select a response. All twenty-seven participants heard all three batches, which lasted 3 minutes and 44 seconds each. This was done so as to have as many results as possible, in order to make any results or patterns identified more reliable.

2.5. Task

Each subject had a forced-choice response sheet that comprised, for each numbered stimulus, a vowel quadrilateral divided into eleven sections based on the phonetic space intuitively occupied by that vowel (Fig. 1). This was to constrain and speed responses, and facilitate their tabulation. Subjects were told to circle the best symbol for the vowel they perceived, and to circle [l] if they heard one. They should not worry if some vowels were used more than others. The slides were numbered consecutively. Subjects were shown a practice slide-show of 4 slides before the experiment proper, so that they could become accustomed to the task, and ask questions if they found the task confusing.

Figure 1: Response template for vowel quality and lateral. Numbers on the x- and y-axes are used in the tabulation of results (section 3.1) and were not seen by subjects. Top left: stimulus number.

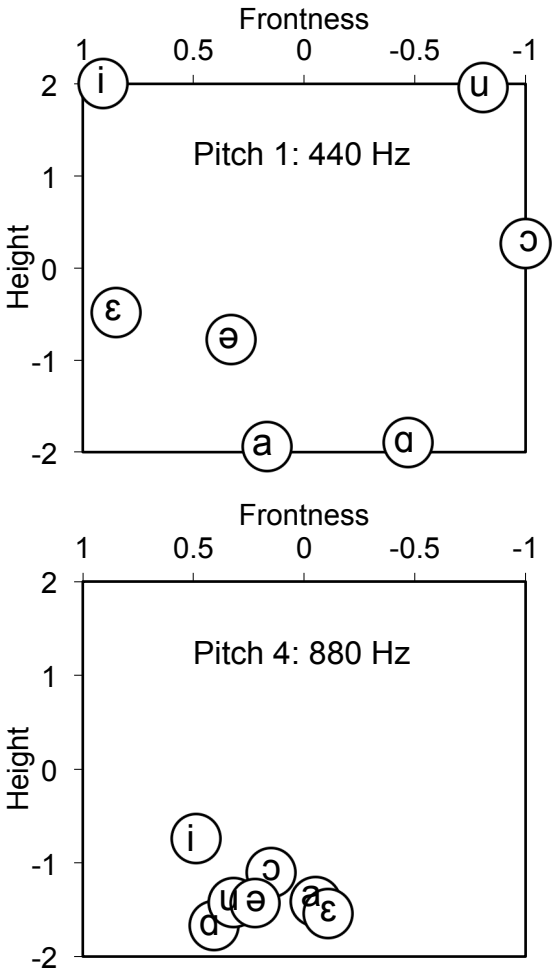


3. RESULTS

3.1. Vowels

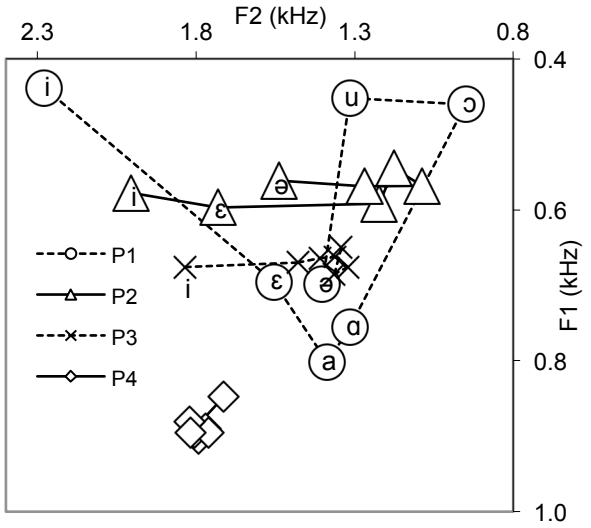
Responses were categorised on the scales for height (2 to -2) and frontness (1 to -1) shown in Figure 1, with schwa being 0 on both scales. The values for each stimulus were averaged over the 27 subjects and three exposures (N=81). Fig. 2 plots the mean values for each vowel in the lowest and highest pitches. At the highest pitch, as can be seen, quality is lost dramatically and all stimuli are judged to have open vowels.

Figure 2: Mean perceptual vowel responses for the lowest and highest of the four pitch conditions.



It is assumed from general principles that the loss of vowel discrimination with increasing f_0 results from reduced spectral definition because of the dearth of harmonics exciting the transfer function of the vocal tract (cf. [3]). As a relatively crude test of this, LPC formant tracking was run on the 28 stimuli. The results plotted in an F2~F1 space can be seen in Fig. 3.

Figure 3: LPC 'formant' tracking for vowels sung at pitches P1 (440 Hz) to P4 (880 Hz).

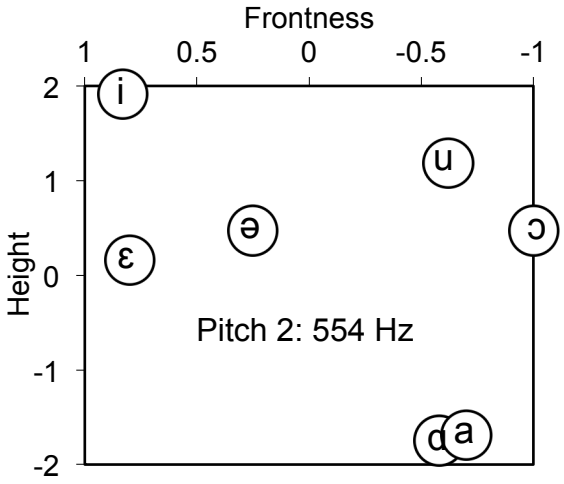


At Pitch 1 (≈440 Hz), the distribution of vowels bears a fair resemblance to that expected in speech, with F2 correlating with frontness and F1 with openness; though there is a clustering of the unrounded mid and open vowels. At Pitch 4 (≈880 Hz), the output of the formant tracker mirrors the lack of perceptual discrimination; it may be inferred that the tracker has locked onto H1 (the first harmonic) at around 880 Hz as its estimate of F1, and H2 (around 1760 Hz) for F2 – regardless of the vowel articulated. At intermediate pitches a mixed picture emerges.

At Pitch 2 (≈554 Hz), the F1 estimate for all vowels roughly corresponds to H1. The F2 estimates for four vowels cluster around 1100–1270 Hz, which might correspond to H2. But [i], [ɛ], and [ə] must have sufficient high frequency emphasis that H3, or in the case of [i] H4, may be attracting the estimate. In the case of Pitch 3 (≈659 Hz) the tracker locks on to H1 as its 'F1' in all vowels, and to H2 for all except [i], where it seems likely that H3 is strongly influencing the estimate.

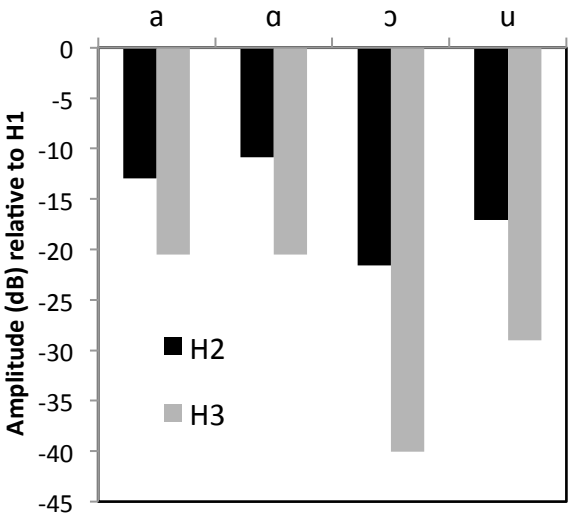
Fig. 4, showing the perceptual responses at Pitch 2, reveals however that the formant tracking for Pitch 2 in Fig. 3 (triangles) underpredicts the discriminability of [a, ɑ] versus [ɔ, u]. Informal listening confirmed that a degree of quality difference was still audible. It was hypothesised that even if the observed spectrum lacks peaks corresponding to the resonances of the vocal tract because of the paucity of harmonics, the resonance function may nonetheless leave its trace in the relative amplitude of those harmonics. To test this, estimates were made of the amplitude of the first three harmonics of [a, ɑ, ɔ, u] in the Pitch 2 stimuli.

Figure 4: Mean perceptual vowel responses for Pitch 2.



Because vibrato made it hard to choose a representative point for spectral analysis, long term average spectra (LTAS) with effective bandwidth 50 Hz were computed over the whole vowel. The amplitude was then taken at the frequencies of H1, H2, and H3. Fig. 5 shows the decrease in amplitude from H1 to H2 and from H1 to H3. It is evident that [ɔ, u] show a steeper spectral slope, which, though different from the absolute values of spoken vowels, retains their relative distinctiveness.

Figure 5: Intensity of harmonics 2 and 3 (relative to H1) of four vowels sung at pitch 2 (≈554 Hz).



3.2. The lateral

Table 1 shows the mean percentage of trials in which an initial lateral was heard, broken down by pitch. It is clear that the higher two pitches affect consonant detection as well as the perception of vowel quality. There is, however, an unexpected upturn in detection at the highest pitch.

Table 1: Mean percentage of [l] detection in the four pitch conditions

Pitch	1	2	3	4
% [l]	99.6	95.4	64.6	77.4

Examination of spectrograms of [lɑ] reveals that at pitch 3 the spectral discontinuity at the release of the lateral becomes less evident. The formant tracker also locks onto H1 and H2, which are of course effectively constant between liquid and vowel because the pitch is determined by the note sung. The improvement in pitch 4 may arise from a lower relative intensity of the lateral, as indicated in Table 2, providing an alternative cue. It should be noted, though, that in three stimuli there was a short pitch glitch at the transition between lateral and vowel, and two of these occurred in pitch 4 stimuli; this may also have contributed to the consonant percept. It might be that the occlusion of the lateral reduces airflow enough to interfere with phonation at extreme pitches. Overall, though, high f_0 clearly impedes perception of the lateral. The main factor obscuring the consonant articulation – as with the erosion of vowel quality – is almost certainly loss of definition of the spectral envelope associated with a given vocal tract configuration, which in turn arises from the paucity of harmonics exciting the vocal tract transfer function.

Table 2: Mean increase in intensity between lateral and vowels

Pitch	1	2	3	4
Δ dB	1.49	0.19	2.68	8.16

4. DISCUSSION

The loss of vowel distinctiveness and the dominance of open vowel percepts at high singing pitches observed in previous studies have been confirmed. Innovatively, articulatory adjustments of the kind used by singers for volume and aesthetics were avoided in the present experiment, in order that the effect of high f_0 could be isolated. It is clear that very high f_0 alone can cause appreciable perceptual migration of vowels towards the open area of the vowel quadrilateral, and ultimately merger. The loss of phonetic quality is progressive, and at pitches intermediate between 440 and 880 Hz some vowel distinctions survive, even when no true formant peaks are evident. This persistence of quality is hypothesised to be due to observable differences in amplitude between the first few harmonics, from which some perceptual reconstruction of the overall spectral shape, and hence articulation, is possible. Whether this would be equally true of a trained singer employing aesthetically-driven modifications to articulation is questionable; it is a matter that would require the same singer to repeat the recording with and (as here) without such modifications.

The study has also made a first foray into the effect of high f_0 on consonant perception. Even though all syllables began with a lateral, the consonant failed to be identified at the higher pitches up to a third of the time. As with vowel quality, this can be attributed to a reduction in spectral definition when harmonics are in short supply. Laterals, though vowel-like in having clear formant structure, are normally characterised by a 'fault line' where they abut the following vowel, with often a step up of F1 and resumption of F3 which may be cancelled by an antiresonance during the lateral. These cues will be greatly weakened by reduced spectral definition. There was some evidence, however, that the lower intensity of the lateral, which seems to be more marked as pitch increases, serves as a useable cue at the highest f_0.

5. CONCLUSION

The present study touches on basic questions of speech perception and articulation. For instance, assuming that vowel quality depends on the hearer being able to reconstruct the transfer function of the vocal tract, and arguably the articulatory configuration potentially underlying it, it is impressive that at intermediate pitches (pitch 2 in particular) where the fundamental harmonic is too high to excite the true F1 of some vowels, those vowels can still be inferred with some, albeit limited, degree of success.

This brings into question any assumption that we rely on one strategy, for instance formant detection, to hear vowel quality. In addition, given that received wisdom about consonant perception has traditionally placed great reliance on transitional effects in formants, maybe the next step should be to extend work on consonant perception at high singing fundamental frequencies to include a range of consonants.

In more practical terms, there is a very reassuring take-home message from this study. No concert-goer or hi-fi enthusiast should feel the slightest embarrassment at not having a clue what the soprano is singing about.

6. REFERENCES

[1] Benolken, M. S., Swanson, C. E. 1990. The effect of pitch-related changes on the perception of sung vowels. *J. Acoust. Soc. Am*. 87(4), 1781–1785.

[2] Hollien, H., Mendes-Schwartz, A. P., Nielsen, K. 2000. Perceptual confusions of high-pitched sung vowels. *Journal of Voice* 14(2), 287–298.

[3] Howie, J., Delattre, P. 1962. An experimental study of the effect of pitch on the intelligibility of vowels. *The NATS Bulletin* 18, 6–9.

[4] Scotto di Carlo, N., Germain, A. 1985. A perceptual study of the influence of pitch on the intelligibility of sung vowels. *Phonetica* 42(4), 188–197.

[5] Sundberg, J. 1975. Formant Technique in a Professional Female Singer. *Acustica*, 32(2), 89–96.

[6] Sundberg, J. 1977. The acoustics of the singing voice. *Scientific American* 236(3), 82–91.

[7] Sundberg, J. 1994. Perceptual aspects of singing. *Journal of Voice* 8(2), 106–122.

[8] Thorpe, C. W., Watson, C. I. 2000. Vowel identification in singing at high pitch. The Eighth Australian International Conference on Speech Science and Technology – Conference Proceedings, Australian Speech Science and Technology Association, Canberra. 280–286.

[9] Westerman Gregg, J., Scherer, R. C. 2006. Vowel Intelligibility in Classical Singing. *Journal of Voice* 20(2), 198–210.