AN ACOUSTIC STUDY OF FRENCH VOWELS IN SPEECH AND SINGING VOICE

E. Florig
Institut de Phonétique de Strasbourg - USHS
22, rue Descartes. 67084 Strasbourg. France.
monpiou@ushs.u-strasbg.fr

ABSTRACT

The major question addressed in this study is, to find out to what extent certain French vowels undergo acoustic tranformations as a function of frequency variation from soprano voices. The vowels investigated are /i, e, ɛ, a, y/ and the main thrust of this research is to verify if the nature of spectral modification is comparable for all vowels.

INTRODUCTION

Singing voice, because of its specific demands, its complex articulatory movements, represents a particular object of analysis. In speech, the main objective is to communicate, *i.e.* to render the phonetic message comprehensible. In singing voice, however, this intelligibility does not seem to be the foremost desired goal. The search for precision, the aesthetic aspect, voice compass and quality play a fundamental role, with an aim to conserve a given homogenous quality along the entire voice range. It is obvious that it is more difficult to understand a sung text than a spoken text; the further the singer gets into the high notes, the more problematic phoneme distinction becomes.

The purpose of this study is thus to examine to what extent certain French vowels undergo acoustic transformations as a function of frequency variation.

METHOD

Corpus

As the purpose of the investigation is to look at acoustic differences between vowels in speech and in singing voice, the corpus is composed of two parts:

- the first part for singing voice that is comprised of vocal exercises: a string of the same vowel at different frequencies satisfied the conditions necessary for our study. Ascending vocal exercises were obtained for each of the vowels cited above. Vocal exercises carried out on the same vowel has the advantage of excluding intervocalic consonants, thus avoiding potential consonantal effects on vowel spectra. Each of the vocal exercises started with the cluster /ts/ thus avoiding the problem of sound attack and also optimizing respiration strategies during the vocal exercises (Figure 1).

- the second part for speech: to the vocal exercises are added sentences that contain the target vowels. The vowels appeared in French words, embedded in the carrier sentences. Vowel context was varied using one of the following consonants /b, v, z, l, ʒ, ʁ/ that vary constriction location. Unvoiced consonants were deliberately excluded to avoid possible vowel devoicing, so also were nasals to avoid vowel nasalizing. The sentences were constructed as follows:

The carrier sentence "Je vais chanter sur /i/, comme dans bise, vie, Suzie, lit, magie et Paris." means "I am going to sing on /i/, as in kiss, life, Suzie, etc.".
"Je vais chanter sur /e/, comme dans bébé, privé, rusé, blé, léger et paré."
"Je vais chanter sur /ɛ/ comme dans bête, vert, zèbre, laide, geste et raide."
"Je vais chanter sur /a/, comme dans bas, vase, visage, lame, jade et rat."
"Je vais chanter sur /y/, comme dans butte, vue, zut, lutte, juste et rustre."

Measurement zones were restricted to the mid-portion of the vowels, thus avoiding transitions due to adjacent contexts.

Recordings

Recordings were carried out using a DAT TDC-D3 recorder and a Neumann K54 microphone. The recordings were made in a sound proof anechoic room for two lyrique sopranos from the Conservatoire National of Strasbourg, digitized and analyzed by software. Analyses were carried out on spectrogrammes, using wideband and narrow band filters. Wideband spectra were used for formant measurements whilst narrow band spectra served for measurements of sound pitch.

Extracts of the corpus were submitted to a dozen subjects for identification. This was not a perception test, rather it served as a control test where subjects had to note the quality perceived and also to say if the vowel was recognizable or not. Such a task would served as an indicator that would confirm or invalidate our results.

RESULTS

Speech

Measurements were carried out on 31 /i/, 28 /e/, 17 /ɛ/, 24 /a/ and 17 /y/ for each speaker. Two values were obtained from two points on each vowel. Typical values were then calculated as the means of measured values (*cf.* Table 1.). These values served as reference values in our analyses of singing voice, since the influence of frequency variation on the acoustic composition of vowels in singing voice, will be determined in relation to the acoustic properties of the same vowels in speech.

Singing voice

10 vocal exercises on /i/, 8 on /e/, 7 on /ɛ/, 9 on /a/ and 10 on /y/ were analyzed for each speaker. This amounts to 90 notes for /i/, 72 for /e/, 63 for /ɛ/, 81 for /a/ and 90 for /y/. Two measurement points were retained, here also, on each vocalic portion.

Determining thresholds

The acoustic composition of the different vowels in singing voice was not clear along the entire frequency range of vocal exercises. Actually, as from certain high notes, it becomes very difficult, let impossible, to distinguish formants in a precise manner. Also, it was observed that in these high frequencies, subjects could no longer recognize the different vowels. Another observation was that the first hamomic and fondamental frequency were rapidly above the first reinforcement zone that had been detected for the same vowel in speech. It is thus Fo that will be reinforced to play the F1 role. However, the Fo value will also rapidly attain a level so high, such that the different harmonics will not coincide with reinforcment zones characteristic of vowels. Formant structure of vowels in the singing voice context will therefore no longer ressemble that of vowels in speech. Thresholds were established, beneath which the formant structure of the vowel is relatively close to a speech-like vowel and above which, formant detection becomes difficult as values no longer correspond to speech reference values (*cf.* Table 2). – [5] refers to this phenomenon as "intelligibility thresholds", since vowel intelligibility depends on the distinctiveness of its formants. A mean formant value was obtained for all vowels on each note of the vocal exercises. These formant values allowed us to establish thresholds, by comparing obtained values with reference values for speech and by controlling our results with observations that had been made during the audition test.

It was also noticed that in zones where acoustic patterns were relatively unclear, formant structures were close to those of speech, on certain notes. This, presumably, is due to Fo frequency that allows harmonics to coincide with reinforcement zones, charactersitic of vowels.

The presence of a supplementary reinforcement zone was also detected, the "Singing Formant", as described by [7]. This formant was located, for the two female speakers, around 360 Hz, regardless of the vowel analyzed

Quality variation

Vowel quallity was also compared across the singing and the speech conditions, using mean values (*cf.*

Figure 2). Two values were determined: the first in an average pitch, not very far from speech and the second, in a higher pitch.

Notice the difference in formant values between speech and singing voice for /i/, regardless of the note on which it was produced: F1 increases slightly — due to the combined effect of pitch and a slight aperture increase — whereas F2 decreases — due certainly to a slight lip rounding (*sic*). Actually, around 20% of listeners hesitated in recognizing /i/ and /y/ in the high frequencies.

For vowel /e/, the strategy is to increase both F1 and F2 thus attaining values for /i/ in speech. The two vowels are confused by 70% of the listeners in the high frequencies.

Vowel /ɛ/ has formant values different from those in speech as from the onset of low frequencies, because F1 is immediately superior while F2 starts descending. It seems that this vowel was slightly rounded since some listeners perceived a timbre close to the rounded vowel /œ/.

F1 for /a/ remains stable for quite a while — up to around D 4 — with an F2 lower than in speech. This vowel is seemingly rounded in singing voice and is perceived, in the high frequencies, as /ɑ/ or /ɔ/ by 86% of listeners.

For /y/ the tendency is to increase F1 — provoked certainly by pitch increase and by jaw lowering —, while F2 starts lowering. Apparently, jaw lowering, that is responsible for frequency increase, also causes the unrounding of the vowel.

CONCLUSION

It is clear that frequency variations provoked modifications in the acoustic structure of sung vowels. It is possible to determine thresholds or progressive steps beyond which formant distinction and vowel recognition becomes difficult. Moreover, it was noticed that vowel quality was somewhat different across conditions. This is a general tendency that is manifest on all vocal exercises. All vowels are characterized by an increase in F1, probably due to the conjugated effect of pitch increase and jaw lowering. Vowels /i, ɛ, a/ seem to undergo slight rounding while vowel /y/ seems to lose some of its rounding feature.

REFERENCES

[1] Cornut, G., Lafon, J.-C. (1960) "Etude acoustique comparative des phonèmes vocaliques de la voix parlée et chantée", *Folia Phoniatrica*, vol. 12, n°3, pp. 188-196.

[2] Harmegnies, B., Landercy A. (1993) "Analyse spectrale et voix chantée. Contribution à une métrologie objective de la maîtrise du chant", *Revue de phonétique appliquée*, n°106, pp. 35-59.

[3] Leothaud, P. (1991) "L'acoustique vocale", *La voix dévoilée*, Paris : Romillat, pp. 71-92.

[4] Miller, R. (1990) *La structure du chant*, traduction de Gouëlou J.-M., Paris : IPMC.

[5] Scotto Di Carlo, N. (1976) "Etude acoustique et auditive des facteurs d'intelligibilité de la voix chantée", *Travaux de l'Institut de Phonétique d'Aix*, vol. 1, pp. 115-129.

[6] Scotto Di Carlo, N. (1991) "La voix chantée", *La Recherche*, n°235, vol. 22, pp. 1016-1025.

[7] Sundberg, J. (1973) "The acoustics of the singing voice", *Scientific American*, vol. 236, n°3, pp.

[8] Sundberg, J. (1987) *The science of the singing voice*, Illinois : Northern Illinois University Press.

[9] Sundberg, J. (1991) "Phonatory vibrations in singers. A critical Review", *STL-QPSR*, vol. 1, Stockholm.

[10] Sundberg, J., Johansson, C., Wilbrand, H. (1982) "X-Ray Study of Articulation and Formant frequencies in two female Singers", *STL-QPSR*, vol. 4, Stockholm.

Table 1. Reference values for formants in speech given in Hz.

	F1	F2	F3
[i]	334	2627	3295
[e]	378	2317	3136
[ɛ]	635	2109	3080
[a]	732	1727	2980
[y]	329	2021	2688

Table 2. Frequency thresholds

	distinct acoustic composition	imprecise acoustic composition
[i]	up to A sharp 3	from A sharp 3 to E4
[y]	up to B3 - C4	from B3 - C4 to D4 - D sharp 4
[e]	up to G sharp 3 - A3	from A sharp 3 to D4
[ɛ]	up to A3	from A3 to C sharp 4
[a]	up to C sharp 4 - D4	from C sharp 4 - D4 to E4

Figure 1. Example of vocal exercises

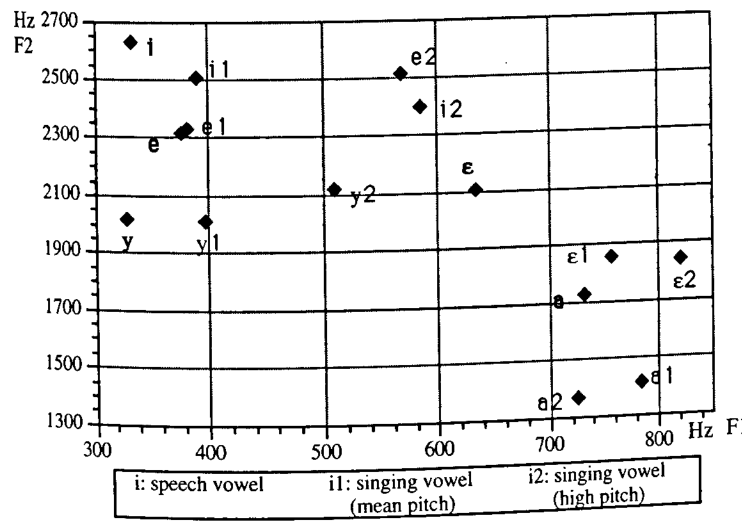

Figure 2. Formant values for speech and for singing voice.