PRODUCTION AND PERCEPTION OF THE SINGING FORMANT

C. Pillot
Institut de phonétique de Paris, France

ABSTRACT

Microphonic and glottographic records of 7 male opera singers confirmed the acoustic properties of the singing formant (SF). MRI measurements and the use of a vocal-tract acoustic simulation program allowed to quantify the effect of the lowering of the larynx. Perceptual experiments reveal that SF influences a number of perceptual dimensions of voice, the phonetic quality of the vowels and the personal vocal quality.

INTRODUCTION

The acoustic characteristics of the singing formant SF are well known [1, 2, 4]. The purpose of this study is to investigate further the acoustic (spectrographic measurements and use of Maeda's vocal tract acoustic simulation program), physiological (MRI) and perceptual aspects of SF for French vocal productions.

I. ACOUSTIC DESCRIPTION

We perform simultaneous microphonic and glottographic records of 4 male professional and 3 untrained singers singing isolated sustained vowels [a], [i], [o] and [u], a given sentence extracted from a familiar piece of music and a melody they have freely chosen.

In accordance with Bartholomew [1] and Sundberg [2], our results show that there is a significant difference of amplitude of the region around SF between the untrained and trained singers ($t_7=7.57$; $p<0.02$) in favour of the latter.

Furthermore, for a given singer, SF frequency doesn't vary whatever the production sung. Finally, SF frequency varies significantly according to the vocal category of the trained singers: for example: about 2620 Hz for the bass singer, 2800 Hz for the baritone and 3406 Hz for the tenor (Fig. 1).

The comparison of the source spectra on the one hand, and of the acoustic spectra on the other hand allowed us to formulate a few hypotheses about the origin of SF:

The intensities of the source spectrum harmonics at the frequencies of the SF are higher for the professional singers.

Furthermore, the bandwidth of SF is about equal to the double ($p<0.0001$) of the theoretical bandwidth of the formants at the same frequencies, Fant [3]: SF doesn't come presumably from one, but from at least 2 formants, as asserted by Sundberg [2].

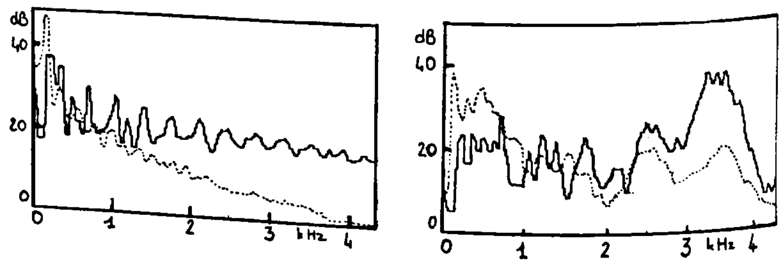

Figure 1. Average spectra of the sentence sung by a professional tenor singer (solid lines) and by an untrained one (dotted lines) at the same tempo and at the same frequency (spectra averaged on 8 seconds). Left: source spectra; Right: acoustic spectra.

II. PRODUCTION

Wave study

We have just noticed that the source produces more energy around 3KHz among the singers who have SF. The influence of the glottal flow parameters on the higher formants in the French vowels [i], [a] and [u] was studied using the Klatt's synthetizer type named Compost (Bailly).

The open quotient (OQ) is the ratio of the opening and closing times to the total duration of the cordal vibratory cycle. The disymmetry quotient (DQ) is the ratio of the opening time to the vocal cords closing time. Our results show that the reduction of OQ and the increase of DQ allow the spectra of the resulting sounds to have more intense high harmonics whatever the vowel sung.

Articulatory study

The preceding source phenomena are not sufficient to explain the emergence of a *peak* like SF.

Maeda's Model

What sort of articulation can generate SF? The hypothesis of a sinus Morgagni's resonance combined with a laryngeal lowering, Sundberg [4], is checked with the Maeda's vocal tract acoustic simulation program [5]. Laryngeal lowering simulation with the 2 tube model entails an emergence of F3 and F4 which come closer, as the F5 amplitude decreases. The variation of the ratio of the laryngeal to the pharyngeal section by modification of the laryngeal section confirms Sundberg's theory according to which the larynx is active as a resonator only if the pharyngeal section is at least six times wider than the laryngeal one.

Magnetic Resonance Imaging

The area functions corresponding to the French vowels [i], [a] and [u] spoken and sung by a professional bass singer (Fig. 2) were estimated from Magnetic Resonance Imaging (MRI) midsagittal images (Magnetic field: 0,5T; Gyrex V machine; ET: 15ms; RT: 33ms; head coil; acquisition time: 8s). Fig. 2 (bottom) shows for [a] the creation of SF (merging and reinforcement of F3 and F4) and Fig. 2 (top) suggests a significant lowering of the larynx (arrow 1) and the jaw (arrow2), a slight labial protrusion and a lingual posteriorisation (arrow 3).

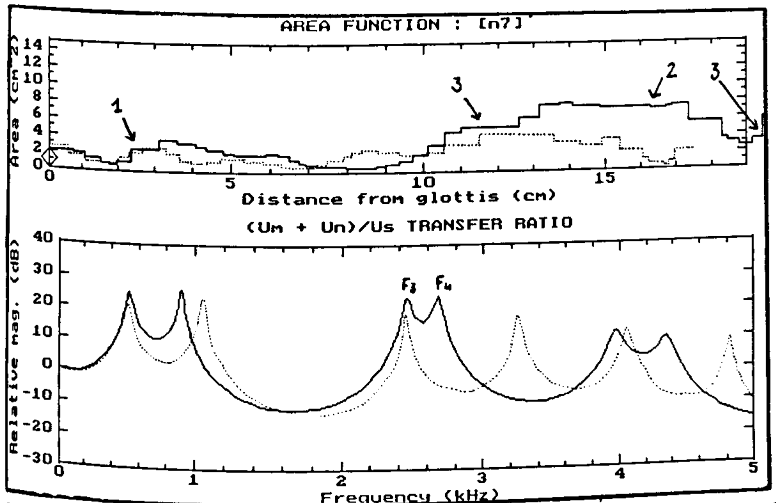

Figure 2. Area function and transfer function of the vowel [a] spoken (dotted lines) and sung (solid lines) by a professional bass singer (same frequency: 100Hz). Data coming from the Maeda's vocal tract acoustic simulation program [5].

III. PERCEPTION

To study SF perceptive significance, 2 comparative tests of pairs of words and orchestral extracts were performed with naive auditors.

Words

The word "solitaire" was sung by 7 singers. The SF of the 4 professional singers was filtered, while the corresponding SF frequency zone of the untrained ones was amplified. In a discrimination task, all of the 22 auditors perceive a difference between words with and without SF. In a forced decision task ("Quel est le mot le plus riche?"), the word considered as "plus riche" have the SF for the professional singers (70%). The artificially created SF for the untrained singers was not as perceptually effective as the original SF (40%).

Musical sentence (orchestral context)

We choose as stimuli an excerpt of a CD-record (The opera Faust by Gounod) during 15 seconds. The SF of the tenor voice was then filtered. 54 auditors were asked to freely judge the perceived difference between the original and the filtered versions.

Table 1. Distribution of the responses for the comparison of 2 musical extracts among 54 auditors.

Sort of responses	Result in %	Sort of responses	Result in %
Timbre	31,9	Pitch	7,34
Aesthetics	13,07	Others	6,62
rticulation	12,4	/orchestra	4,36
Emotion	9,17	Duration	3,44
Intensity	8,49	Distance	3,21

Our results show that the SF is first perceived in terms of timbre ("clair, brillant, riche"). The intensity only intervenes in the fifth position ("forte"). The voice of the singer who has the SF is more "articulée", "belle" (aesthetics), "proche" (distance), "courte" (duration) and "plus aigue" (pitch). The interpretation of the results is complex.

Vocalic identification

22 naive auditors were asked to identify the original vowels of the professional singers, and their filtered versions, and the original vowels of untrained singers and their SF added versions. The singer productions were [i], [a] and [u] vowels, and the listeners had an open choice (15 French vowels).

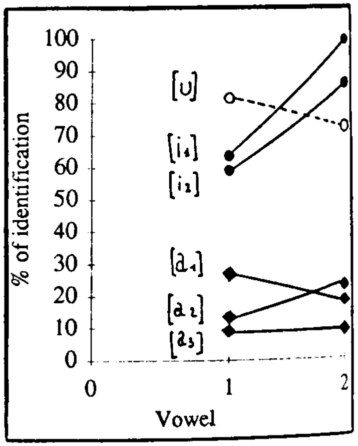

Figure 3. Average percentages of correct identification of vowels sung by the professional singers. 1: without SF; 2: with SF.

As shown in Fig. 3, the percentages of identification for the professional singers are statistically smaller for the vowel [a] than for [i] and [u] ($t_{22}=-9.7$; $p<0.0001$). Furthermore, SF has more influence on the identification scores of [i] and [u] than [a] ($t_3=-9.07$, $p<0.005$). The suppression of SF (on the left) worsens the identification of [i], and allows a better recognition of [u]. [i] and [u] of the untrained subjects are better identified (100 %), without (original) or with added SF.

The filtered vowel [i] of the professional singers is mainly misperceived as [y] (77.2 % of the errors), while original [u] (with SF) is often perceived like [o] (91.67 % of the errors). [a] without SF is perceived like [ɔ].

DISCUSSION

The acoustic analysis of sung productions allows to formulate a few hypotheses about the origin of the SF, which have to been confronted with physiological data.

The study of Magnetic Resonance Imaging (MRI) of sustained vowels confirms the *compensatory articulation* hypothesis of Sundberg & al [6]: MRI shows us "supershapes" of the sung vowels; for example, a lowered larynx and jaw position in spite of a raised tongue can be interpreted as "superpalatalization" for the vowel [i], according to Sundberg [6]. Singing requires indeed a vocal tract free of any constriction and, in the same time, a high degree of flexibility of phonatory muscles. This investigation entails methodological problems due to the supine position of the subject or to the width-to-area conversion.

The study of SF shows also the importance of its perceptive relevance: it's located in a frequency region where the auditory sensitivity is maximal. (The threshold of audibility is minimal at this frequency) [7] This allows the singer's voice not to be masked by an orchestral accompaniment and this without pathogenic vocal effort. Trumpets have the same peak of intensity at 3 KHz, and this instrument merges in an orchestra. The spontaneous qualification of two musical extracts (identical except presence or absence of SF) showed us a great variability in the sort of obtained responses: indeed, a physical value like the SF affects a multiplicity of perceptual values, including the timbre.

Musicians have started to be interested by the timbre of their vocalizations only in the nineteenth century, and expansion of orchestra have obliged singers to develop new vocal strategies to make their voices more audible, at a time where electric amplification didn't exist.

CONCLUSION

SF, which is the real signature of the occidental operating singing, probably exists in other vocal and sound productions. Its origin is multifactorial : it's produced by a singer who have learned to face the orchestral accompaniment by means of phonatory and articulatory modifications.

SF has an effect non only on the vocalic quality, but also on the personal vocal quality of the singer.

ACKNOWLEDGEMENTS

This work was greatly aided by the careful experimental assistance of D. Rostolland and M. Elmaleh. We want to express our gratitude to J. Vaissière and S. Maeda for their support and comment on an earlier draft of the paper.

REFERENCES

[1] Bartholomew, W.T. (1934), "A physical definition of good voice quality in the male voice", *JASA*, vol.VI; pp 25-33.

[2] Sundberg, J. (1987), *The science of the singing voice*, Dekalb Illinois, Northern Illinois University Press, 216p.

[3] Fant, G. (1970), *Acoustic theory of speech production*, The Hague, Mouton & Co., 2nd edition, 328p.

[4] Sundberg, J. (1974), "Articulatory interpretation of the singing-formant", *JASA*, vol.55,n°4, pp. 838-844.

[5] Maeda, S. (1992), "Modélisation articulatoire du conduit vocal", *Journal de Physique IV*, vol.2, pp. 307-314.

[6] Sundberg, J. & Lindblom, B. (1971), "Acoustical consequences of lip, tongue, jaw and larynx movement", *JASA*, vol.4, pp. 1166-1179.

[7] Castellengo, M. (1973), "Considérations sur la voix des chanteurs professionnels", *Bulletin d'acoustique musicale*, n°67, pp 9-14.