Leslie De'Ath, Associate Editor

Effect of Multifactorial Constraints on Opera-Singing Intelligibility (I)

Nicole Scotto Di Carlo

Nicole Scotto Di Carlo

Journal of Singing, March/April 2007
Volume 63, No. 4, pp. 000–000
Copyright © 2007
National Association of Teachers of Singing

INTRODUCTION

The criticism most often made of opera singers is that they are difficult to understand. Is the lack of intelligibility in lyric art due to the singer's poor diction, as generally believed, or to other causes? What happens when the frequency, intensity, and duration of the sound are considerably greater as they are in singing as opposed to speech? Are the acoustic structures of vowels and consonants preserved? Is syllabic isochrony*[1] maintained? How much can phonemes* be lengthened without losing their identity? What happens when a singer pays particular attention to diction in order to improve the intelligibility of the words being sung? Are linguistic factors the only ones that alter intelligibility?

To answer each of these questions, let us first examine the linguistic factors that lessen the intelligibility of sung vowels, consonants, and syllables, before looking at the extralinguistic factors.

VOWELS

Phonetic Background

Vowels are defined by the fact that the laryngeal sound* passes freely through the vocal tract.* During vowel production, the positions taken on by the vocal organs give the supraglottal cavities the characteristic shape and volume of the vowel in question. By amplifying certain harmonics and dampening others, these cavities modify the spectral makeup of the glottal wave* passing through them. The amplified harmonics are called formants.* The acoustic structure of each vowel is characterized by the frequencies of its formants, which enable listeners to identify it perceptually. However, vowel intelligibility depends more on the relationships between the frequencies of the different formants than on their absolute numerical values. This is why in speech, vowels pronounced in a deep male voice are understood as well as ones pronounced in a child's high-pitched voice. By contrast, in shouting and singing, the greatly altered relationships between the formant frequencies cause different vowel timbres to be perceived.[2] The first two formants (F1* and F2*)—and sometimes the third (F3), which is needed in certain cases to determine the vowel's timbre—are the most important formants in vowel intelligibility, that is, they are the ones that allow us to distinguish one vowel from another. The other formants serve mainly to identify the speaker by providing indications about his/her individual characteristics. In singing, vowels undergo substantial distortion, mainly due to the intensity and frequency at which they are produced.

Effect of Intensity

In spoken French, there are fifteen distinct vowel timbres. In singing, however, when the intensity of the sound is between 90 and 130 dB, only three timbres remain: /i/, /a/, and /u/3 (minimal vocalic system). All other timbres tend to be reduced to one of these basic timbres as follows:[4]

- close anterior vowels /i/ /e/ tend toward /i/;
- open vowels /a/, /ɑ/, /ɔ/, /ɛ/, and /æ/ tend toward /ɑ/;
- close posterior vowels /ø/, /u/, /o/, /y/ tend toward /u/.

Acoustically, this can be explained by the fact that as vocal intensity increases, sound energy is shifted from the lower to the higher frequencies of the spectrum, that is, toward the frequency zone where auditory sensitivity is the greatest[5] (Figure 1). This upward shift makes it difficult to distinguish the formant structures of the different vowels and thus lowers their intelligibility.

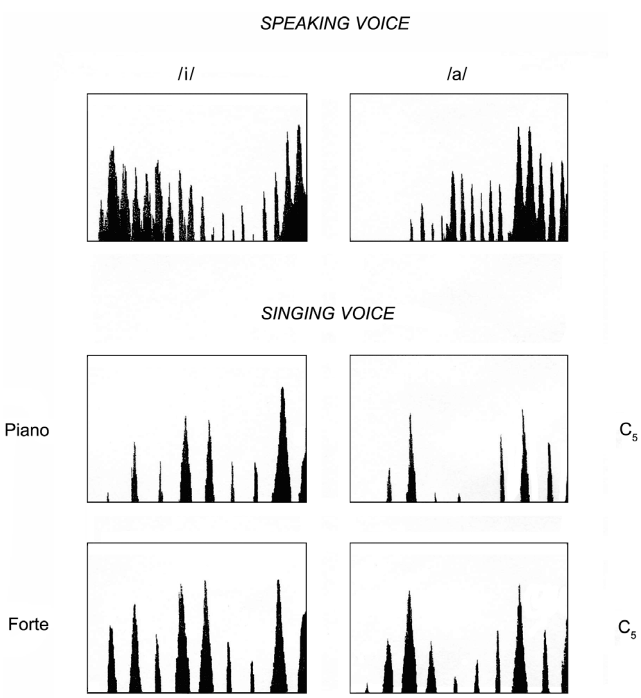


Figure 1. Spectral modifications of the vowels /i/ and /a/ sung by a soprano, as a function of intensity (inverted spectra).

The spectrum of a sound signal is a two-dimensional graph. The intensity (y-axis) of each harmonic is represented by the length of the spectral lines, which can be measured in decibels (dB). On this inverted spectrum, the frequencies (x-axis) are read from right to left. When a sound goes from *piano* to *forte*, energy is transferred to the zone of the spectrum where auditory acuity is the best, i.e., between 2000 and 5000 Hz.

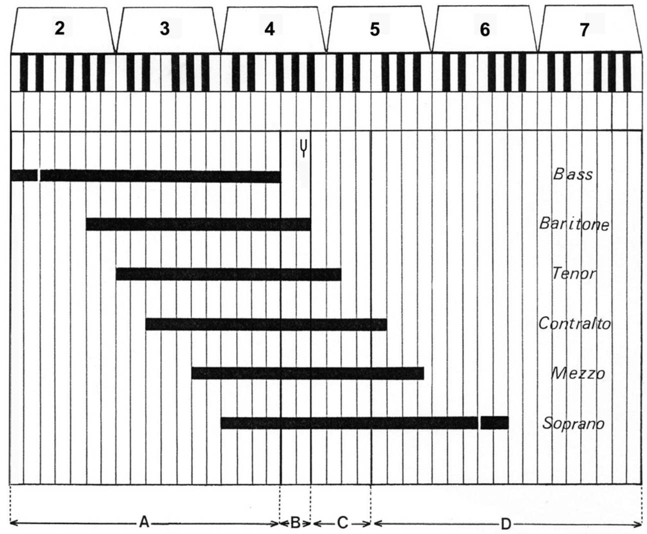


Figure 2. Intelligibility zones of the singing voice, for the tessitura of each vocal category. A. Optimal intelligibility zone; B. Margin of tolerance; C. Elective intelligibility zone; D. Absolute unintelligibility zone.

The diagram shows that the entire tessitura of a bass is situated in the optimal intelligibility zone, which explains why a bass, for example, is understood better than a soprano, only a quarter of whose voice range lies within this zone.

Effect of Frequency

Vowels also undergo distortions due to the frequency at which they are sung (Figure 2).

- Between 65 Hz (C2) and 349 Hz (F4), all vowels are correctly perceived; this is the *optimal intelligibility zone*.
- Between 349 Hz (F4) and 440 Hz (A4), nasal vowels* and /e/, /ɛ/, and /o/ are still clearly distinct: this is the *margin of tolerance*.
- Between 440 Hz (A4) and 659 Hz (E5), only the vowels /i/ and /a/ can be differentiated; this is the *elective intelligibility zone*.

Above 659 Hz, vowels can no longer be discriminated; this is the *absolute unintelligibility zone.*[6]

Consequently, singers are more likely to be understood by the audience if most of their tessitura is situated within the optimal intelligibility zone, or below 349 Hz. If we understand a bass better than a soprano, it is because his entire voice range falls within the optimal zone. In contrast, only a quarter of a soprano's range is in this zone, and the fraction drops to only a fifth for

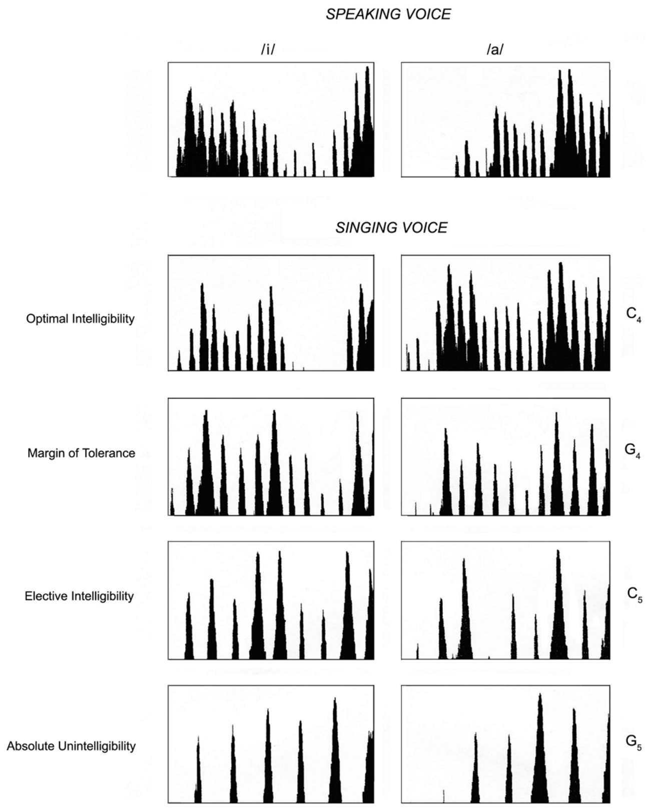


Figure 3. Spectral modifications of the vowels /i/ and /a/ sung by a soprano, as a function of frequency (inverted spectra).

On C4, situated in the optimal intelligibility zone, the spectra of the vowels /i/ and /a/ are very close to what they are in speech. But as the frequency rises, the harmonics move increasingly farther apart and coincide less and less often with the characteristic formant zones of the different vowel timbres. This considerably decreases intelligibility in the upper-middle and upper registers.

a light coloratura soprano.[7] One can account for this phenomenon acoustically: As frequency rises, the harmonics move farther and farther apart and become less and less likely to coincide with the formant zones, making the formants less distinct and adversely affecting vowel intelligibility (Figure 3).

Another factor that affects vowel intelligibility is the frequency relationship between the fundamental and the first formant.

- When the fundamental frequency of a given vowel is much lower than its first formant, as it is in speech, all formants are clearly distinguishable and intelligibility is optimal.
- When a vowel's fundamental frequency is approximately equal to its first formant, intelligibility is almost fully preserved since the fundamental frequency acts as the first formant.
- When a vowel's fundamental frequency is much higher than its first formant, intelligibility is very poor for all vowels except /a/, where the fundamental acts as an average formant located between this vowel's theoretical first and second formants (725 Hz, 1300 Hz). It is not surprising, then, that all vowels sung in the upper register are perceived as /a/.[8]

Vowel intelligibility is also a question of the relationship between the frequencies of the first two formants. At high frequencies, the second formant varies little across vowels.[9] The fact of having smaller F2 differences accounts for a number of confusions between vowels of equal aperture.* This is the case for the close vowels /i/, /y/, and /u/, all of which have their first formant at 250 Hz, or the open vowels* /ɛ/, /œ/, and /ɔ/, whose first formant is 510 Hz. In speech, listeners distinguish between vowels with identical F1s by relying on F2, but this is impossible in singing because in addition to having the same F1, the F2 values are too close together to permit proper discrimination (Figure 4).

Effect of Vibrato

The vocal vibrato is characterized by periodic frequency variations that may or may not be accompanied by periodic synchronous or asynchronous variations in intensity or timbre (Figure 5). As its name suggests, the frequency vibrato is characterized by periodic frequency

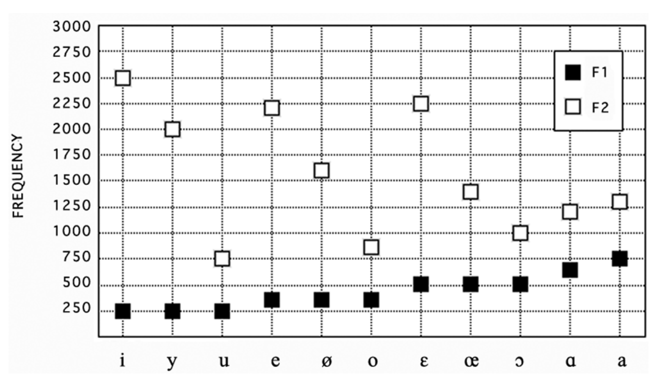


Figure 4. French vowel formant values.

The second formant (F2) differentiates the French vowels /i/, /y/, and /u/ (whose first formant, F1, is 250 Hz), the vowels /e/, /ø/, and /o/ (whose F1 is 350 Hz), and the vowels /ɛ/, /œ/, and /ɔ/ (whose F1 is 510 Hz). While F2 takes on clearly distinct values in speech, it varies little across vowels in singing. This makes it difficult to recognize the individual vowel timbres.

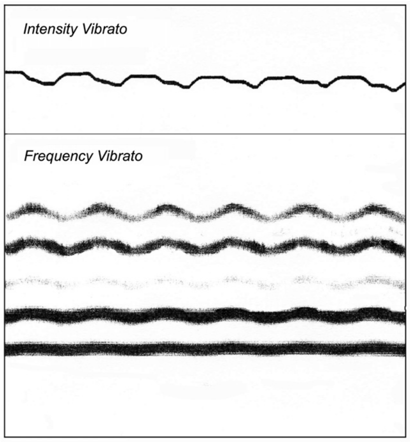


Figure 5. **Straight sounds and vibrated sounds.**

When opera singers are about to sing a note, say C_3 at 262 Hz, they have two choices: either they produce what is called a *straight sound*, in which case the fundamental frequency stays virtually the same for the entire duration of the sound, or they produce a *vibrated sound*, in which case the fundamental frequency oscillates in a periodic fashion around a target frequency (262 Hz), with variations that range between a quarter tone and a whole tone. The vocal vibrato is characterized by periodic frequency variations that may or may not be accompanied by periodic variations in intensity and timbre. On this sonagram, the frequency vibrato and the intensity vibrato are in phase opposition (the frequency-vibrato maxima correspond to the intensity-vibrato minima). The timbre vibrato is also visible here, especially on the last harmonic, whose peaks show an intensity increase. The sinusoidal shape of the vibrato curve makes consonant identification difficult because it masks the direction of the formant transitions.

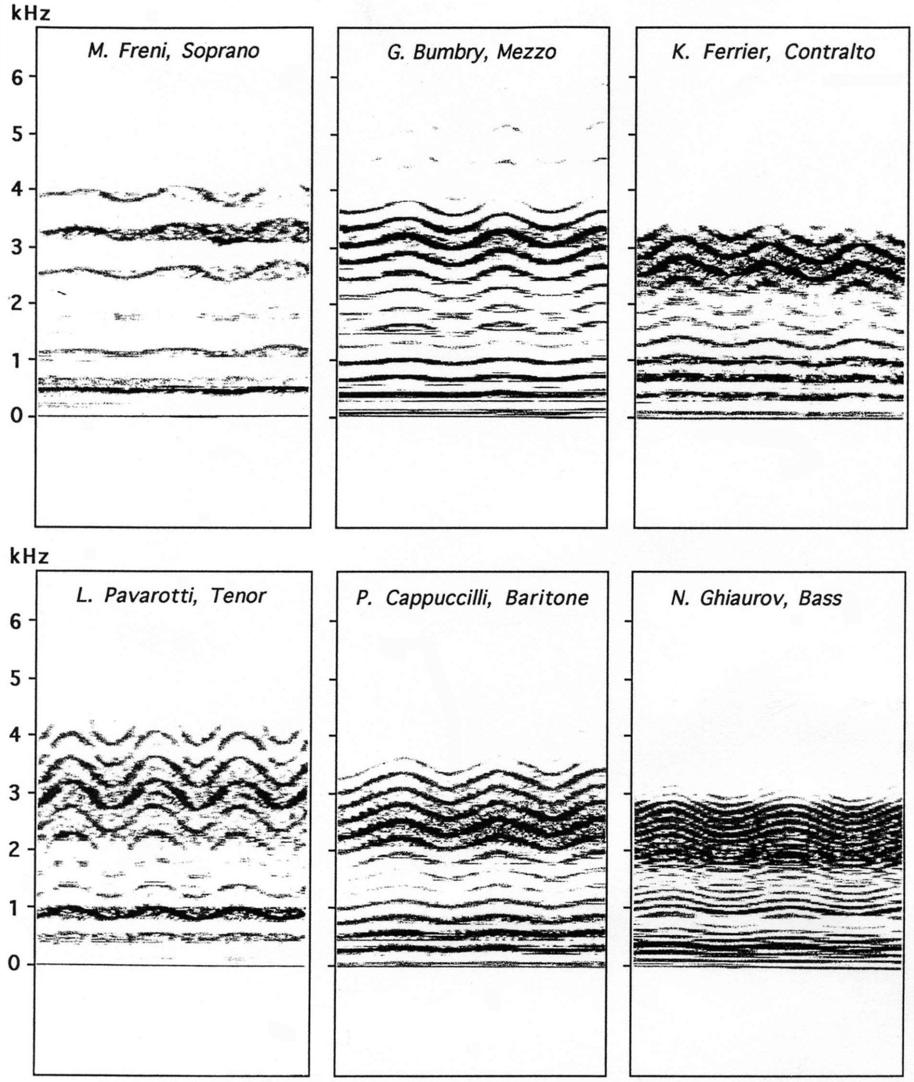


Figure 6. **Frequency of the singing formant for the six main vocal categories.**

The singing formant, which enables the singer's voice to carry, is characterized acoustically by the reinforcement of harmonics situated around 3000 Hz, i.e., in the zone of the spectrum where auditory acuity is optimal. The pitch of the singing formant varies slightly across vocal categories. According to Sataloff, it is about 3100 Hz for sopranos, 3000 Hz for mezzo-sopranos, 2900 Hz for contraltos, 2800 Hz for tenors, 2600 Hz for baritones, and 2400 Hz for basses. We can see from these sonagrams representing each of the six main vocal categories that the pitch differences between the singing formants of male voices are greater than those of female voices.

fluctuations around a target frequency corresponding to the note being sung. While this type of fluctuation has little impact on the intelligibility of synthesized vowels,[10] their effect on natural-vowel intelligibility is much greater. A study by Dalcastello comparing nonvibrated and vibrated sounds* showed that the overall identification rate of vowels sung without vibrato was 28% higher, regardless of frequency.[11] According to the author, this phenomenon is due to the fact that frequency vibrato modulations which range from a semitone to as much as a whole tone mask the formant frequencies of sung vowels and thereby make their identification more difficult.

Effect of the Singing Formant

In addition to the vowel formants, the voice of opera singers has an extravocalic formant. It was discovered by Bartholomew in 1934, studied again by Winckel in 1953, then by Vennard in 1967, and finally by Sundberg in 1972.[12] This formant, called "shimmer" by Bartholomew, "ring" by Winckel, and "singing formant" by Sundberg, is situated between 2000 and 3000 Hz in male voices and between 3000 and 4000 Hz in female voices (Figure 6). The singing formant is what enables the singer's voice to carry. In fact, the reinforcement of harmonics in this auditory sensitivity zone is found only in trained voices, namely, those of opera singers and actors when they use what they call "voice projection."[13]

The singing formant may be partly responsible for the high rate of confusion among diffuse vowels,* particularly /i/ and /y/.[14] Because of its location between 2000 and 4000 Hz, the singing formant masks the true pitch of the second formants of vowels like /i/, /e/, /ɛ/, and /y/, all of which are located in that same frequency range; and as noted above, it is the second formants of /i/ and /y/ (2500 Hz and 2000 Hz, respectively) that distinguish these two vowels, since both have an F1 of 250 Hz.

Effect of Phonatory Constraints*

Singers have a clear-cut tendency to centralize vowels in all registers. This can be explained by the necessity in singing of keeping the vocal tract totally unobstructed and maintaining a high degree of flexibility in the tongue and bucco-facial muscles. Yet the French vowel system is characterized both by vocal tract constriction due to its high proportion of close vowels* (63%) and by extensive muscular tension due to its high percentage of anterior vowels* (73%).[15] Given that constriction and muscular tension are incompatible with the demands of lyrical art, singers attempt to reduce these phenomena by centralizing vowels. This leads to underarticulation.

Furthermore, the *buccal opening*, which increases with pitch, is always wider in singing than in speech, for all registers. The same holds true for *aperture* which, being greater in singing, considerably perturbs the production of close vowels like /i/, /y/, and /e/ (Figure 7).

Finally, in order to free up the pharynx, singers move the *tongue mass* forward in the middle, upper-middle, and upper registers, which obviously hinders the production of posterior vowels* like /o/ and /u/ (Figure 7).

Concerning the *lips*, many current vocal techniques call for substantial labial projection in the lower register and slightly less lip projection in the upper regis-

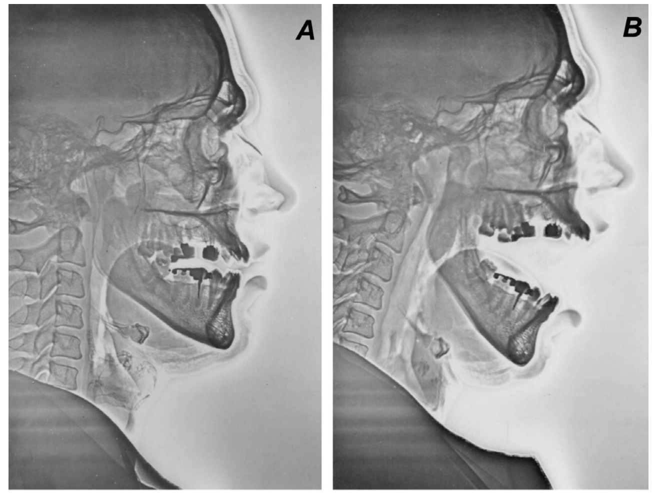


Figure 7. **X-Rays of a tenor singing the vowel /a/ on C3 and C5.**

When singers from all vocal categories go from the lower (A) to the upper (B) register, the anteriorization of the tongue mass hinders the articulation of posterior vowels, and the widened buccal opening and increased aperture (distance between the tongue and the palate) hinder the articulation of close vowels, labial and dental plosives, and certain fricatives, all of which require substantial vocal tract narrowing.

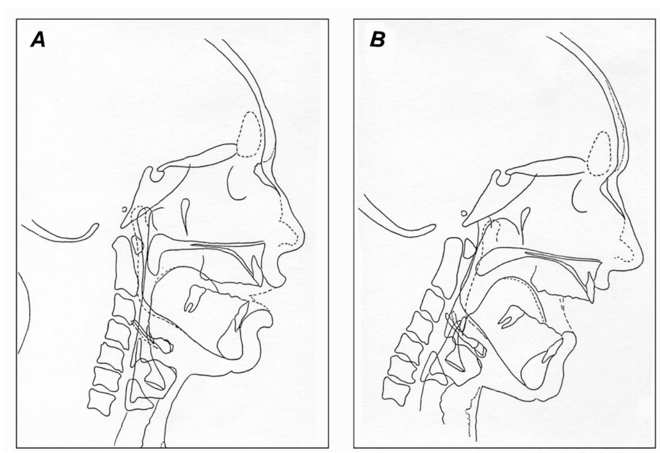


Figure 8. **Role of the lips in singing.**

The space between the teeth and the lips forms a resonance cavity, the labial cavity (A), that can be eliminated by spreading the lips (B). The labial cavity provides a way of increasing the volume of the buccal cavity (by extending it) so as to reinforce low-frequency harmonics and darken the timbre of the voice. Inversely, to decrease the volume of the buccal cavity and thereby reinforce high-frequency harmonics, light sopranos and light coloratura sopranos use lateral lip spreading, so-called "singing with a smile," to produce notes located in the upper part of their voice range. This physiological maneuver clarifies the timbre of the voice. However, because of today's demand for mellow voices, light coloratura sopranos are tending more and more to reduce lip spreading in order to achieve a timbre that fits with current aesthetic trends.

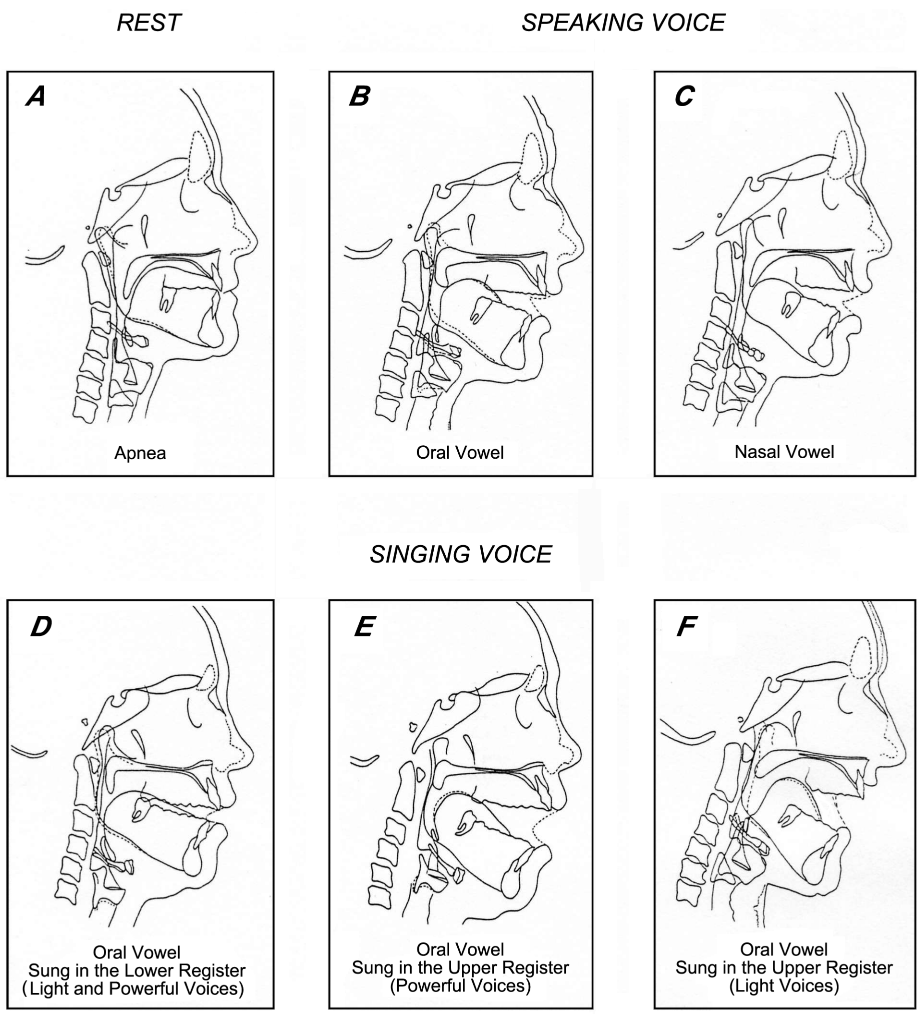


Figure 9. Positions of the velum in speaking and in singing.

At rest, the velum touches the dorsal surface of the tongue so that air passes into the nasal fossae (A). In speech, the velum takes on two extreme positions. When it is raised (B), it is held against the posterior wall of the pharynx to prevent air from entering the nasal fossae; sounds produced in this way are called *oral*. When it is lowered (C), air from the lungs flows through both the nose and the mouth; the resulting sound is called *nasal*. In lower-register singing, the velum occupies (D) the same raised position as in spoken oral vowels, the only difference being that it is moved slightly away from the posterior pharyngeal wall in a way that allows some of the air from the lungs to escape through the nasal passages. For powerful voices, the velopharyngeal opening increases slightly in the upper register (E); for light voices, the velum occupies a partly-lowered, taut position that is never used in speech (F)—which is why this position is so hard to learn—and the velopharyngeal opening is a little wider than in powerful voices. The balance between an accurate sound and a hypernasalized sound is a question of only a few millimeters in this case, which explains why the defective emission mode that singing teachers call "placing the sound in the nose" is most often found in singers with light voices.

ter. This leads to difficulty producing unrounded vowels* like /i/ or /e/. On the other hand, the lateral lip spreading used on high notes by singers in certain vocal categories (e.g., light and coloratura sopranos) hinders the realization of rounded vowels* like /y/, /ø/, /œ/, and /u/ (Figure 8).

The *velopharyngeal opening*, which always exists in singing, increases in size as frequency rises. Beyond a certain point, the sounds become overly nasalized and unpleasant to the ear (Figure 9).[16] In singing, where oral vowels are already slightly nasalized, singers preparing to produce a nasal vowel do not lower the velum, but shift the back of the tongue upward toward the velum, in an attempt to preserve the beauty of the sound. Vowels produced in this manner are less buccalized and therefore sound more nasalized.[17]

In summary, all open vowels and all unrounded vowels sung by high voices will retain their timbre in the upper register. By contrast, close vowels and rounded vowels are threatened as soon as the singer reaches the upper-middle or upper register.

These phenomena are not the only ones that lessen the intelligibility of opera singing. As we have just seen, vowels are subjected to extensive deformation due to frequency, intensity, vibrato, and phonatory constraints, but they also undergo contextual distortions brought about by the consonants that surround them.

CONSONANTS

Phonetic Background

Unlike vowels, which are characterized by the free flow of pulmonary air through the vocal tract, consonants result from airflow obstruction by the articulatory organs (principally the tongue and the lips). The obstacle can be created by complete closure* (occlusion) or narrowing (constriction) of the vocal tract.

Consonants called *plosives*,* which include /p/, /b/, /t/, /d/, /k/, and /g/, are produced by total vocal-tract obstruction. The air that accumulates behind the blockage point of articulation*[18] is released with a bursting noise when the articulatory organs separate. Consonants called *fricative*,* which include /f/, /v/, /s/, /z/, /ʃ/, and /ʒ/, are articulated by narrowing of the vocal tract.[19] The resulting disturbance in the airflow causes friction noises.

Each of these two major classes of consonants is subdivided into two categories: *voiced*, in which case the vocal folds vibrate, as in /b/, /d/, /g/, /v/, /z/, and /ʒ/, and *voiceless*, in which case the vocal folds do not vibrate, as in /p/, /t/, /k/, /f/, /s/, and /ʃ/.

In the speech stream, the rapid succession of vowels and consonants is produced by modifying the shape and volume of the resonance cavities.* Acoustically, the modifications show up as large spectral variations, particularly in the formant transitions,* which reflect the changes in the resonators during the transition from a vowel to the following consonant, or from a consonant to the following vowel (Figure 10). Delattre (1955, 1967) showed that the transitions between the second and third vowel formants supply acoustic cues to the upcoming consonant's place of articulation.*[20] According to this author, each consonant has a virtual point, called its locus,* toward which the transitions of the second and third vowel formants converge. This explains why the formant transitions and loci are so important to consonant recognition (Figure 11).

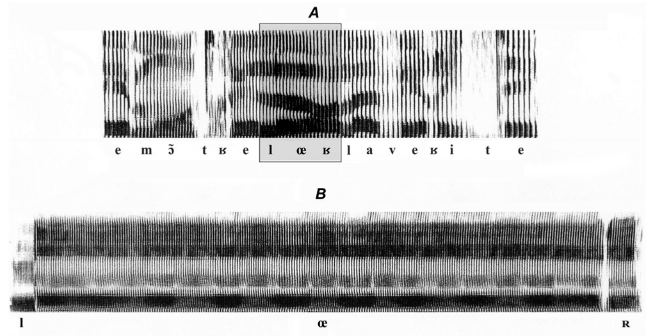


Figure 10. Formant transitions in speaking and in singing. Large spectral variations are clearly visible in this spectrogram of the spoken phrase "et montrez-leur la vérité" ("and show them the truth") from the French version of the opera *Die Zauberflöte*, pronounced by a basso profundo (A). They represent the formant transitions, which are the acoustic counterparts of the resonator modifications that take place during a consonant-vowel or vowel-consonant changeover. The formant transitions supply important acoustic cues for consonant recognition. On the word "leur" in the above phrase sung by the same subject (B), the formant transitions completely disappear, leading to a considerable drop in consonant intelligibility.

Effect of Intensity

As the intensity and fundamental frequency of a sung vowel increase, the intensity of the adjacent consonant decreases, resulting in very weak noise poles (whether a burst or friction). [21] This leads to confusions in the identification of those consonant categories for which noise intensity and frequency are the main acoustic cues (Figure 12).

Effect of Frequency

The intelligibility of consonants, like that of vowels, is inversely proportional to the frequency at which they are sung. For sopranos, the intelligibility of all vowels and all consonants starts to decline by the middle register and continues to drop with increasing rapidity as the sung tone gets higher (Figure 13). Consonants remain a little more understandable than vowels, with an overall mean identification rate (all registers taken together) of 40.3%, as opposed to 31.8% for vowels.[22]

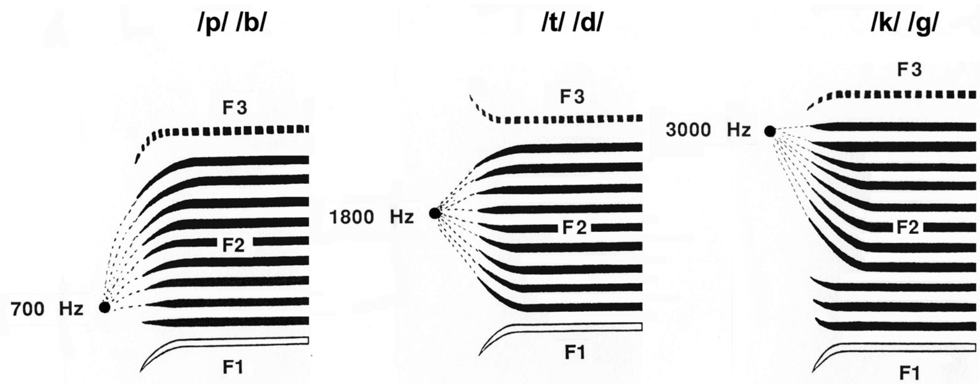


Figure 11. Locus of labial, dental, and palatal consonants (from Delattre, 1967).

According to Delattre, each consonant has a virtual point called a locus toward which the formant transitions of vowels converge. Acoustically, the loci are acoustic place-of-articulation cues. They are situated at different frequencies for each type of consonant. For the labials /p/ and /b/, the locus is at about 700 Hz; the formant transitions have a convex configuration and are referred to as negative. The locus of the dentals /t/ and /d/ is situated at about 1800 Hz, and that of the palatals /k/ and /g/, at about 3000 Hz; their transitions have a concave configuration and are referred to as positive.

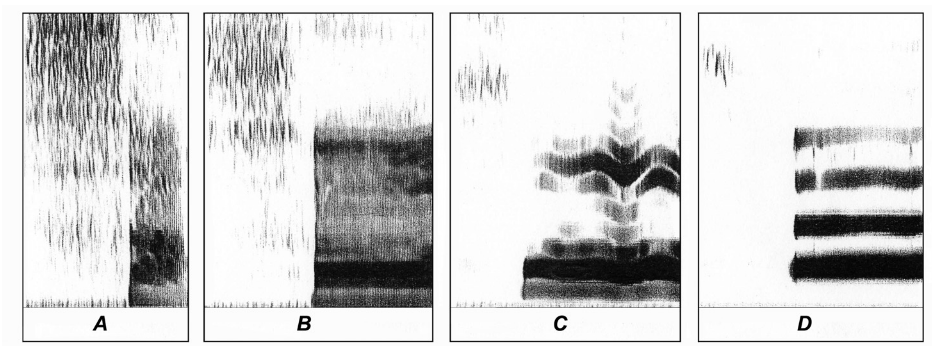


Figure 12. Noise pole attenuation as a function of frequency.

The consonant /s/ is produced with the anterior part of the tongue mass pushed up against the alveolar ridge of the upper incisors. Air from the lungs passing through this very narrow opening makes a high-frequency friction noise in the 3000–16000 Hz frequency range. The consonant /S/, on the other hand, for which buccal-cavity narrowing is not as great, has a lower-frequency timbre, with a 2000–10000 Hz noise band. This accounts for why, when we run a tape recorder at half speed to slow down the speech rate, an /s/ is perceived as a /ʃ/ since its noise band drops from 3000–16000 Hz to 1500–8000 Hz, the approximate frequency range of a /ʃ/ noise band. When /sa/ is spoken (A) and then sung by a soprano in the lower (B), middle (C), and upper (D) registers, the noise band of /s/ in the lower register is less intense and briefer than in speech. As the frequency rises, the /s/ noise band becomes less and less intense and increasingly narrow and brief. The resulting loss of information makes recognition difficult.

Some consonants are more resistant to the rise in fundamental frequency than others (Figure 14). The palatal fricative /ʒ/ holds up the best, followed by the nasal consonants* /ɲ/, /m/, /n/, and then the liquid consonant /l/. Next come the palatal plosives /g/ and /k/, the labial plosives /b/ and /p/, and finally the dental plosives and fricatives /d/, /v/, /t/, /f/, and /s/.[23]

As compared to labials and dentals, the greater resistance of palatal consonants in the upper registers is due to the fact that the buccal opening in singing increases as the note becomes higher—a wide buccal opening is more compatible with a relatively open palatal articulation than with a very close labial or dental articulation. The nasal and liquid consonants,* on the other hand, have an acoustic structure that is similar to that of vowels, so their articulation does not hinder singing.[24]

Acoustic analyses of the singing voice have also brought out a phenomenon specific to singing, which Germain and Séassau called syllabic destructuring.[25] By reducing or eliminating the formant-transition slopes, syllabic destructuring makes it extremely difficult to detect the loci, and this contributes to decreasing consonant intelligibility (Figure 15).

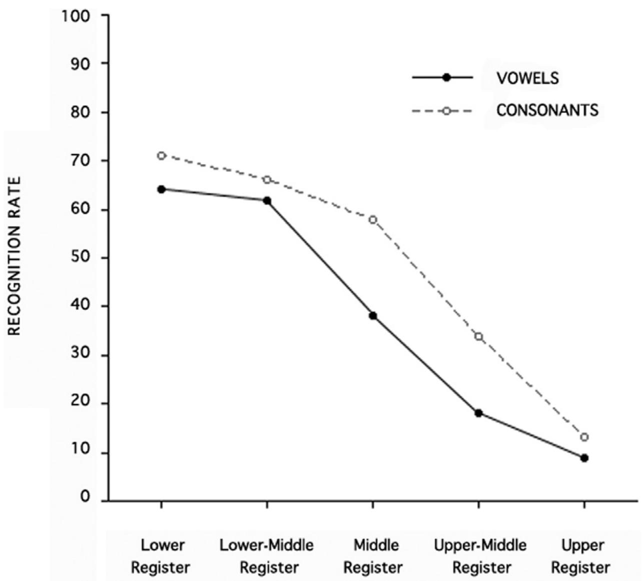


Figure 13. **Identification rate of vowels and consonants, by register.**

The intelligibility of vowels and consonants declines as the frequency at which they are sung rises. As soon as sopranos reach the middle register, intelligibility drops, although more abruptly for vowels than for consonants, which remain fairly understandable on the whole.

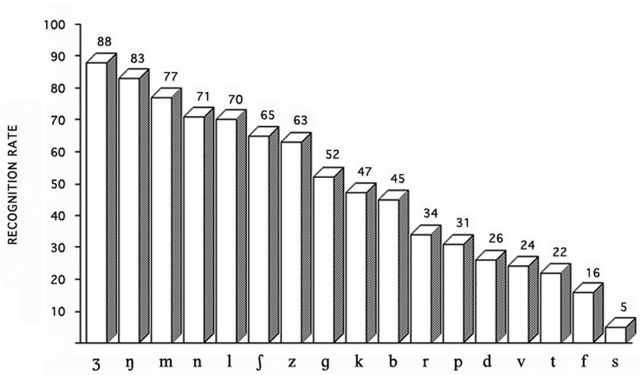


Figure 14. **Mean recognition rate of French consonants.**

Some consonants (palatals, nasals, and liquids) are more resistant than others (labials and dentals) to an increase in frequency. Palatals hold up better in the upper registers because the buccal opening widens as the frequency of the sung note rises. A wide buccal opening is more compatible with a relatively open palatal articulation than it is with a close labial or dental articulation. Vocalic consonants (nasals and liquids) have an acoustic structure that is very similar to that of vowels, so their articulation is not a hindrance in singing.

In sum, an increase in fundamental frequency has an impact on the perception of sung consonants because it eliminates certain acoustic cues, such as the formant transitions and noise-related features, and thereby causes identification errors.

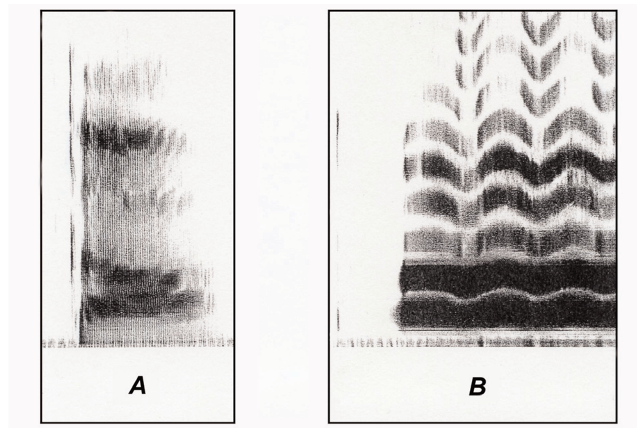


Figure 15. **Sonagrams of /ta/ spoken and sung on $F^{\#}5$ by a soprano.**

The sonagram of a spoken /ta/ (A) reveals the direction of the first three formant transitions: the F1 transition is negative and the F2 and F3 transitions are positive. The same sequence sung by the same subject on $F^{\#}_5$ (B) shows that the formant transition slopes have disappeared and seem to be neutralized by the vibrato. This makes locus detection impossible and consonant identification difficult.

Effect of Vibrato

Because of its sinusoidal wave form, the vibrato masks the direction of the formant transitions. This hinders consonant differentiation, particularly between the labials, which have a negative transition (convex), and the palatals, which have a positive transition (concave) (Figure 11).

Effect of Phonatory Constraints

As above for vowels, because the buccal opening increases as the fundamental frequency rises, singing in the upper-middle and upper registers will hinder the production of labial or dental plosives like /p/, /b/, /t/, and /d/, and fricatives like /f/, /v/, and /s/, which require substantial vocal tract narrowing. On the other hand, palatal consonants are well preserved in the upper register. Voiced plosives like /b/, /d/, and /g/ are produced by laryngeal lowering, which is incompatible with upper-register singing for lyrical artists who use vocal techniques with a high-position larynx (light voices).

In summary, intelligibility is preserved in singing for consonants whose articulatory accommodations are compatible with phonatory constraints, namely, palatal consonants, and vowel-like consonants such as liquids (/l/ and /ʀ/*) and nasals (/m/, /n/, and /ɲ/). By contrast, plosives and fricatives are threatened in the upper-mid-

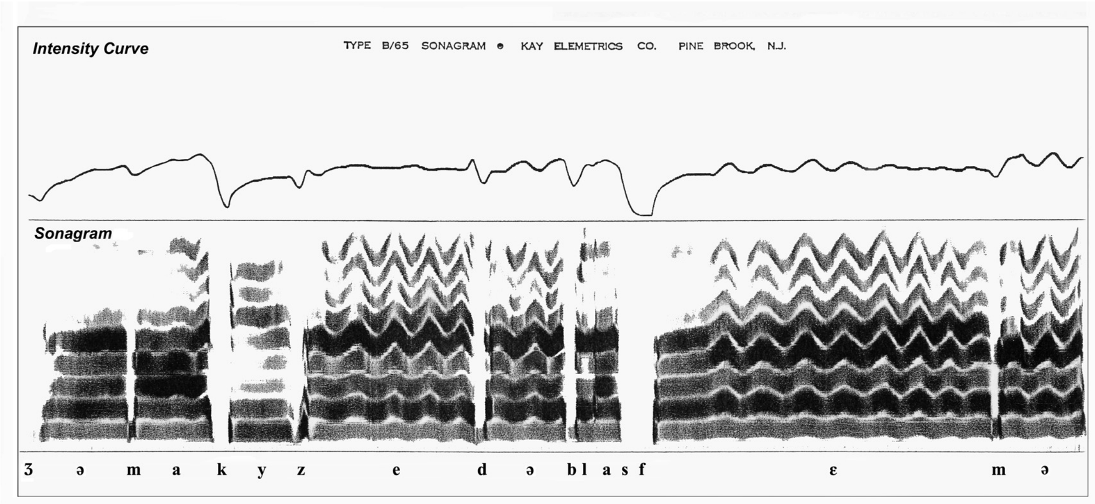


Figure 16. **Effect of consonants on legato quality.**
On this sonagram of the sequence "Je m'accusais de blasphème" ("I was accusing myself of blasphemy") from the opera *Carmen*, sung by a tenor, we can see the consonantal breaks in the sound continuum. The discontinuity consonants generate in the melodic line and the irregularities they cause in the intensity curve are detrimental to the legato.

dle and upper registers, especially ones that require anterior lifting of the entire tongue, including the tip, such as /t/ and /d/.

Aerodynamic Phenomena

Consonants result from the total or partial obstruction of the vocal tract; yet to sing properly, the vocal tract must be completely free. We can easily imagine the singer's problem with this fundamental incompatibility between aesthetic demands and linguistic requirements.

During the production of fricative consonants (like /f/, /v/, /s/, /z/, /ʃ/, /ʒ/, and /ʀ/), airflow disturbances triggered by vocal tract narrowing are manifested auditorily by friction noises that the singer finds unpleasant to the ear and tries to reduce. This results in the partial destruction of acoustic noise cues that are critical to consonant intelligibility. Likewise, during the production of plosives such as /p/, /t/, /k/, /b/, /d/, and /g/, where the vocal tract is momentarily blocked, the silent stretches caused by the interrupted flow of air break the continuity of the melodic line by introducing acoustic gaps in the sound continuum. This results in perturbations of the *legato*. The singer instinctively attempts to "fill in the gaps" by reducing consonantal duration and articulatory force, which amounts to (Figure 16).[26]

Microprosodic* Phenomena

In addition to perturbing the sound continuum, consonants affect the vowels surrounding them, especially the subsequent vowel. In speech, this influence has an impact on the onset and hold of the vowel, whose pitch is thereby modified. A question that arises is whether the same microprosodic phenomena apply to singing, or whether these effects are neutralized by the tonal accuracy constraints imposed on singers.

Effect on Hold

In speech, voiceless consonants (/p/, /t/, /k/, /f/, /s/, /ʃ/) tend to raise the average pitch of the subsequent vowel, while voiced consonants (/b/, /d/, /g/, /v/, /z/, /ʒ/) tend to lower it.[27] The same tendency—although not as pronounced—is found in singing for straight sounds.[28] For vibrated sounds, consonants do not seem to affect the pitch of the next vowel because vowel-hold modifications are masked by the frequency vibrato (whose mean amplitude is about a semitone). This allows the singer to produce pitch variations that fall within the safe range and therefore are unperceived.

Effect on Onset

In speech, only voiced consonants lower the fundamental frequency of the subsequent vowel, unlike their voiceless counterparts, which tend to raise it.[29] In singing, this phenomenon occurs only for consonants produced at a frequency below 659 Hz (E_5). Starting at 659 Hz, all consonants, whether voiced or voiceless, lower the onset frequency of the following vowel (Figure 17).[30] This means that whenever a consonant is sung by a soprano in the upper-middle or upper register, the following

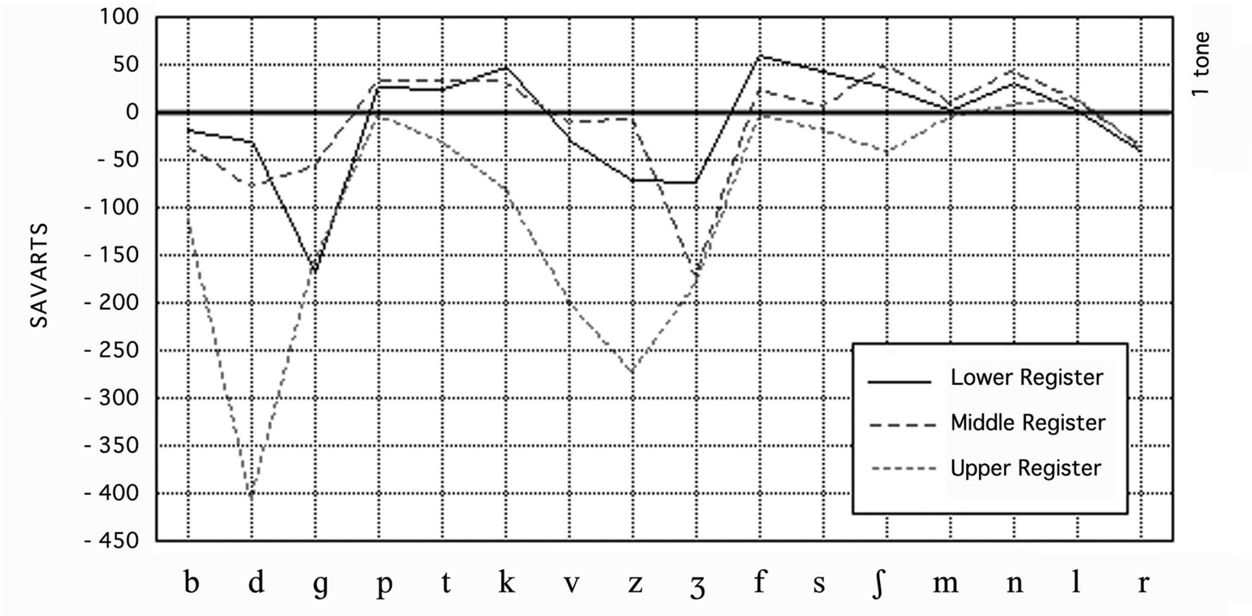
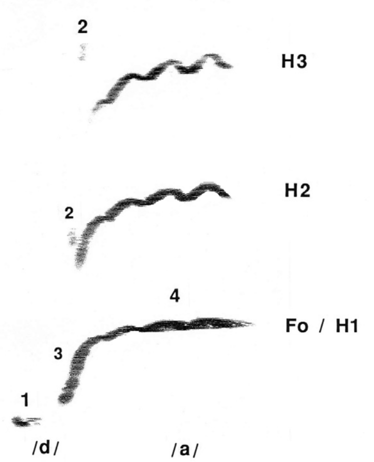


Figure 17. **Post-consonantal onset accuracy of French consonants sung by a soprano.**
On this graph, the zero represents a perfectly accurate vowel onset. Positive values (located above the zero line) correspond to onsets that are too high; negative values (located below it) correspond to onsets that are too low. Each gradation represents 50 Svt* (one tone). We can see that in the upper register of this soprano, except for /m/, /n/, and /l/, all consonants lower the onset of the vowels that follow. Note that the consonant /d/ has the greatest power to decrease the next vowel's onset (by up to 8 tones).



vowel's onset will be too low. Because this effect increases with frequency, the fundamental frequency of a vowel sung by a soprano in the upper register may drop by as much as an octave or more. This phenomenon has been observed, for example, in a subject asked to overarticulate the consonant /d/ while singing in the upper register—the onset was lowered by 406 Savarts (more than eight tones), making an interval of a tenth between the onset and hold of the vowel (Figure 18). In order to sing in key, opera singers devise strategies aimed at reducing this phenomenon by decreasing the degree of occlusion, the duration, and the articulatory force of

Figure 18. **Sonagram of an overarticulated /da/ sung by a light soprano on her highest C.**
This document shows the different phases of /da/ sung on C6 by a soprano asked to overarticulate.
1. **Voicing bar of /d/: laryngeal vibrations during the production of the voiced consonant.**
2. **Burst of /d/: abrupt opening of the vocal tract caused by the pressure build-up behind the point of articulation* (tip of the tongue against the upper incisors) during the consonant hold.**
3. **Transitions between /d/ and /a/: movements of the voice organs during the passage from the consonant to the vowel.**
4. **Hold of /a/: stable part of the vowel on which the pitch of the note is perceived.**

In this overarticulated syllable, there is an interval of a tenth between the onset and the hold of the vowel [a]. Voiced consonants, especially voiced plosives, have the particularity of lowering the fundamental frequency of the subsequent vowel's onset because they are produced with a lowered larynx. When they are sung in the upper register, it takes time for the larynx to move from the lower position it occupies for producing the consonant to the higher one required for singing the vowel, causing the vowel onset to be out of tune. The duration of these postconsonantal onsets is insufficient for the ear to integrate them as off-key, so untrained listeners merely sense them as unpleasant. Trained listeners perceive a parasitic element along with the onset, but they are unable to identify it.

PHONETIC SYMBOLS FOR THE FRENCH LANGUAGE

The phonetic symbols used to transcribe language sounds are codified by the International Phonetic Association (IPA) and constitute the international phonetic alphabet

VOWELS			CONSONANTS			SEMI-VOWELS		
i	=	s**i**x, l**y**s	p	=	**p**ont	j	=	p**i**ed
e	=	**é**mu, **ai**mer, p**ei**ner	t	=	**t**oile	õ	=	p**u**is
o	=	p**o**t, b**eau**	k	=	**c**ar, **qu**and, é**ch**o	w	=	**ou**i
ɛ	=	m**è**re, **ê**tre, p**ai**re	b	=	**b**eau			
a	=	p**a**tte	d	=	**d**ans			
ɑ	=	p**â**te	g	=	**g**ain, **gu**érir			
ɔ	=	p**o**rt	f	=	**f**orêt, **ph**iltre			
o	=	p**o**t, b**eau**	s	=	**s**ourire			
u	=	**fou**	ʃ	=	**ch**ance, **sch**éma			
y	=	él**u**	v	=	**v**oix			
ø	=	p**eu**	z	=	**z**ébre, ro**s**e			
œ	=	fl**eu**r, **oeu**f	ʒ	=	**j**aune, **g**ens			
ə	=	p**e**tit	l	=	**l**oin			
ɛ̃	=	p**in**, b**ain**, s**ein**	ʁ	=	**r**ond			
ɑ̃	=	b**an**c, m**en**the	m	=	**m**er			
ɔ̃	=	b**on**	n	=	**n**om			
œ̃	=	br**un**	ɲ	=	a**gn**eau			
			?	=	smoki**ng**			

consonants, that is, they underarticulate. Articulatory strategies of this type are more marked in singers with a high voice.[31]

Now that the various factors that alter the intelligibility of sung vowels and consonants have been analyzed, Part II will examine how intelligibility can be affected by problems related to the syllable, particularly its temporal characteristics.

NOTES

1. Glossary entries are identified with an asterisk. It will appear with Part II.
2. D. Rostolland, "L'audition de la parole en présence de bruit" (Doctoral thesis, Paris, 1979).
3. See list of phonetic symbols.
4. N. Scotto Di Carlo, "Etude acoustique et auditive des facteurs d'intelligibilité de la voix chantée," *Proceedings of the VIIth International Congress of Phonetic Sciences* (The Hague: Mouton, 1972), 1017–1023.
5. Auditory acuity is optimal between 2000 and 5000 Hz. The lowest absolute threshold of audition is 3150 Hz.
6. Scotto Di Carlo.
7. N. Scotto Di Carlo, "Pourquoi ne comprend-on pas les chanteurs d'opéras?" *La Recherche* IX, no. 89 (1978): 495–497.
8. J. Howie and P. Delattre, "An Experimental Study of the Effect of Pitch on the Intelligibility of Vowels," *The NATS Bulletin* 18 (1962): 6–9.
9. J. Sundberg, "Formant Technique in a Professional Female Singer," *Acustica* 32, no. 2 (1975): 89–96.
10. J. Sundberg, "Vibrato and Vowel Identification," *Archives of Acoustics* 2, no. 4 (1977): 257–266.
11. M. Dalcastello, "Vibrato e intelligibilità dei vocali," *Bolletino di Psicologia* I, no. 3 (1994): 360–376.
12. W. Bartholomew, "A Physical Definition of Good Voice Quality in Male Voice," *Journal of the Acoustical Society of America* 6 (1934): 25–33; F. Winckel, "Physikalische Kriterien für objecktive Stimmbeurteilung," *Folia Phoniatrica* 5 (1953): 232-252; W. Vennard, *Singing, the Mechanism and the Technique* (NY: Carl Fischer, 1967); J. Sundberg, "Production and Function of the 'Singing Formant'," *Report of the XIth Congress of Acoustics* (Copenhagen, 1972), 679–686.
13. N. Scotto Di Carlo, "La voix chantée," *La Recherche* XXIII, no. 235 (1991): 1016–1025.
14. N. Scotto Di Carlo and A. Germain, "A Perceptual Study of the Influence of Pitch on the Intelligibility of Sung Vowels," *Phonetica* 42, no. 4 (1985): 188–197.
15. P. Delattre, "Les modes du français," *French Review* 27 (1953): 45–58; N. Scotto Di Carlo, "Contribution à l'étude statistique

du français contemporain" (Master's thesis, Université de Provence, 1967).

16. N. Scotto Di Carlo and D. Autesserre, "Movements of the Velum in Singing," *Journal of Research in Singing* XI, no. 1 (1987): 3–13.
17. R. Husson, *Physiologie de la phonation* (Paris: Masson, 1962).
18. The point of articulation is the point where the lower lip touches the upper lip for the bilabial consonants /p/ and /b/, where the tip of the tongue touches the upper incisors for the apico-dental consonants /t/ and /d/, and where the dorsum of the tongue touches the hard palate or velum for the dorso-palatal and dorso-velar consonants /k/ and /g/.
19. Narrowing is achieved by bringing the lower lip closer to the upper incisors for the labio-dental consonants /f/ and /v/, the anterior part of the tongue closer to the upper alveolar ridge for the predorso-alveolar consonants /s/ and /z/, and the tip of the tongue closer to the anterior part of the hard palate for the prepalatal consonants /ʃ/ and /ʒ/.
20. P. Delattre, A. M. Liberman, and F. S. Cooper, "Acoustical Loci and Transitional Cues for Consonants," *Journal of the Acoustical Society of America* 27 (1955): 769–773; P. Delattre, "Acoustic or Articulatory Invariance?" *Glossa* 1, no. 1 (1967): 3–25.
21. This phenomenon also is found in shouting, where vowels are more intense than consonants. The opposite occurs in whispering, where consonants are more intense than vowels (see Rostolland).
22. N. Scotto Di Carlo and A. Rutherford, "The Effect of Pitch on the Perception of a Coloratura Soprano's Vocalis System," *Journal of Research in Singing* XIII, no. 2 (1990): 11–24; A. Germain and H. Séassau, "Influence de la fréquence sur l'intelligibilité et les indices acoustiques des consonnes du français en voix chantée" (Master's thesis, Université de Provence, 1982).
23. Germain and Séassau.
24. N. Scotto Di Carlo, "Diction et Musicalité," *Médecine des Arts* 5 (1993): 4–11.
25. Germain and Seassau. Additional studies are needed to confirm the occurrence of syllabic destructuring in singing, i.e., to determine whether the adjacent consonant and vowel are produced successively rather than simultaneously as they are in speech.
26. Scotto Di Carlo, "Pourquoi . . ."
27. A. Di Cristo, *Prolégomènes à l'étude de l'intonation. La microprosodie* (Paris: Collection Sons et Parole, Editions du C.N.R.S., 1982).
28. N. Scotto Di Carlo and A. Raphaël, "Etude acoustique et statistique de l'influence des consonnes sur la justesse des voyelles subséquentes en voix chantée" (Paper presented at the 2nd International Symposium of Research in Singing, I.R.C.A.M., Paris, July, 1977).
29. Di Cristo.
30. Scotto Di Carlo and Raphaël.

Nicole Scotto Di Carlo is a research director at the French National Scientific Research Center (CNRS). She specializes in odology, a scientific discipline she created in 1970, which deals with the singing voice at the physiological, acoustic, and perceptual levels. The entry of the term odology (from the Greek *ôdê* = singing and *logos* = science) into dictionaries has publicly officialized this new science.

Together with her research on singing, N. Scotto Di Carlo teaches her speciality at the University of Provence, in addition to carrying out vocal control for national music conservatory students and France's opera company choristers, and acting as a consultant both for speech and singing therapists and for professional opera singers with vocal problems. In 2004, she received the Manuel Garcia Award, granted to scientists or physicians whose research has made a notable contribution to advancing the science of singing.

With about a hundred publications on the functioning of the singing voice, Dr. Scotto Di Carlo has put the field into a new and original light. Among her recent research topics are internal phonatory sensitivities, opera singers' facial expressions and gestures, cervical spine deformations in professional singers, effects of the tuning fork rise on soprano voices, and the importance of proprioceptive memory in opera singers' absolute pitch.