



HAL
open science

Contraintes de production et intelligibilité de la voix chantée

Nicole Scotto Di Carlo

► **To cite this version:**

Nicole Scotto Di Carlo. Contraintes de production et intelligibilité de la voix chantée. Travaux interdisciplinaires du Laboratoire Parole et Langage, 2005, 24, pp.159-180. hal-00134207

HAL Id: hal-00134207

<https://hal.science/hal-00134207>

Submitted on 1 Mar 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONTRAINTES DE PRODUCTION ET INTELLIGIBILITE DE LA VOIX CHANTEE

Nicole Scotto Di Carlo

Résumé

Dans le chant, les phonèmes subissent d'importantes distorsions dues à la fréquence et à l'intensité auxquelles ils sont émis. De plus, le vibrato masque les fréquences des formants ainsi que la direction de leurs transitions et rend plus difficile l'identification des voyelles et des consonnes. Enfin, l'allongement considérable que subit la syllabe modifie complètement sa structure temporelle, ce qui perturbe la reconnaissance des formes vocales.

Mots-clés : voix chantée, intelligibilité, contraintes phonatoires, contraintes articulatoires, loi d'intelligibilité, vibrato, singing formant, phénomènes aéro-dynamiques, phénomènes micro-prosodiques.

Abstract

In singing phonemes undergo substantial distortions due to the frequency and intensity at which they are produced. In addition, the vibrato masks the frequencies of the formants and the direction of their transitions, making it all the more difficult to identify vowels and consonants. Finally, the extensive syllabic lengthening involved in singing completely modifies the temporal structure of the syllable and consequently perturbs vocal patterns recognition.

Keywords : singing voice, intelligibility, phonatory constraints, articulatory constraints, intelligibility law, vibrato, singing formant, aerodynamic phenomena, microprosodic phenomena.

SCOTTO DI CARLO, Nicole, (2005), Contraintes de production et intelligibilité de la voix chantée, *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, vol. 24, p. 159-179.

1. Introduction

Qu'il s'agisse de la voix chuchotée, de la voix criée, de la voix projetée, ou de la voix chantée, chaque fois que la parole est produite dans des conditions extrêmes, ses caractéristiques articulatoires et acoustiques se trouvent profondément altérées. De toutes les productions vocales, c'est le chant lyrique qui est réalisé, dans les conditions les plus extrêmes puisqu'il s'étend sur cinq octaves de la₀ (55Hz) à la₅ (1760 Hz) et développe des intensités pouvant atteindre 130 dB.

Les chanteurs d'opéra sont soumis à de multiples contraintes liées à des exigences physiologiques, musicales et esthétiques qui ne sont pas toujours compatibles avec les nécessités linguistiques. Parmi ces contraintes, certaines sont à l'origine même des problèmes d'intelligibilité. C'est le cas, par exemple, des contraintes phonatoires qui résultent des accommodations pharyngo-buccales indispensables à une émission vocale correcte et qui sont en conflit avec les contraintes articulatoires résultant des modifications pharyngo-buccales nécessaires à l'articulation des phonèmes.

Nous allons tenter de savoir comment les chanteurs gèrent ces contradictions au niveau des unités constitutives de la phrase chantée : les voyelles, les consonnes et la syllabe.

2. Voyelles

2.1. Fréquence

Lorsqu'elles sont chantées, les voyelles subissent des distorsions dues à la fréquence à laquelle elles sont émises. On peut distinguer quatre zones d'intelligibilité :

- *la zone d'intelligibilité optimale* située entre 65 Hz (Do₁) et 349 Hz (Fa₃), dans laquelle toutes les voyelles sont correctement perçues,
- *la marge de tolérance*, comprise entre 349 Hz (Fa₃) et 440 Hz (La₃) dans laquelle les voyelles nasales, ainsi que [e], [ɛ], [o] sont encore nettement individualisées,
- *la zone d'intelligibilité élective*, comprise entre 440 Hz (La₃) et 659 Hz (Mi₄), où seules les voyelles [i] et [a] sont différenciées,
- et *la zone d'intelligibilité absolue* située au-dessus de 659 Hz (Mi₄), dans laquelle on ne peut plus distinguer les voyelles les unes des autres (Scotto Di Carlo, 1972), (fig. 1).

Par conséquent, un chanteur a d'autant plus de chance d'être compris du public que la plus grande partie de sa tessiture¹ se trouve dans la zone d'intelligibilité optimale, c.-à-d. au-dessous de 349 Hz. Si l'on comprend mieux une basse qu'un soprano, c'est que la tessiture de la basse est entièrement située dans la zone d'intelligibilité optimale, alors que la tessiture du soprano en occupe le quart et celle du soprano léger colorature le cinquième seulement (Scotto Di Carlo, 1978).

¹ Ensemble des notes qu'un chanteur peut émettre avec le maximum de facilité et de confort.

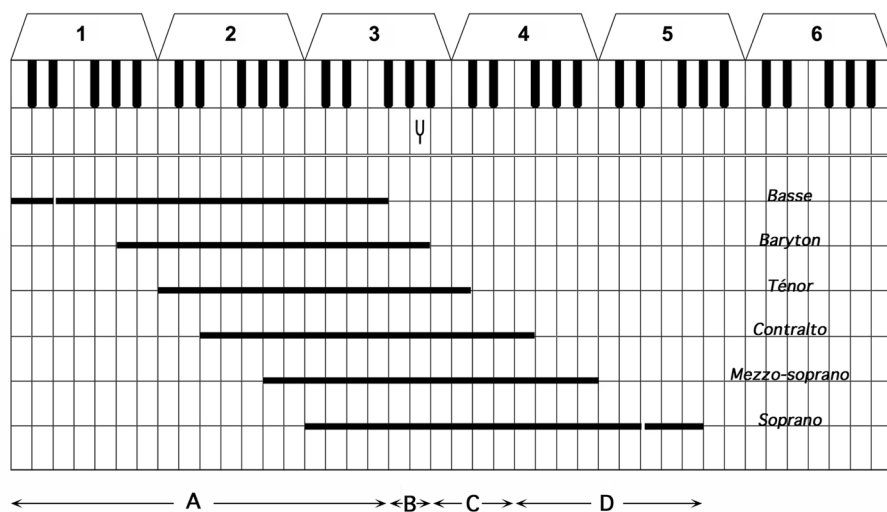


Figure 1
*Zones d'intelligibilité de la voix chantée en fonction de la tessiture
pour chaque catégorie vocale*
A : Zone d'Intelligibilité Optimale, B : Marge de Tolérance
C : Zone d'Intelligibilité Elective, D : Zone d'Inintelligibilité Absolue

[On peut constater sur ce schéma que la tessiture totale de la basse est située dans la Zone d'Intelligibilité Optimale, ce qui explique qu'on la comprend mieux qu'un soprano par exemple, dont la tessiture n'occupe que le quart de cette zone.]

Du point de vue acoustique, ce phénomène s'explique par le fait que plus la fréquence augmente, plus les harmoniques s'espacent et plus leur probabilité de coïncider avec les zones formantiques devient faible. Les formants étant moins bien individualisés, l'intelligibilité des voyelles se trouve altérée.

En outre, l'intelligibilité des voyelles est liée aux rapports de fréquence existant entre le fondamental et le premier formant :

- Lorsque la fréquence fondamentale d'une voyelle donnée est beaucoup plus basse que son premier formant, ce qui correspond aux conditions rencontrées dans la parole, tous les formants sont bien individualisés et l'intelligibilité est optimale.
- Lorsque la valeur de la fréquence fondamentale est à peu près égale au premier formant de la voyelle, l'intelligibilité est pratiquement conservée dans la mesure où la fréquence fondamentale joue le rôle du premier formant.
- Lorsque la fréquence fondamentale est beaucoup plus élevée que celle du premier formant d'une voyelle donnée, l'intelligibilité est très mauvaise pour toutes les voyelles sauf pour la voyelle [a] car dans ce cas, la fréquence fondamentale joue le rôle d'un formant moyen qui se situe entre le premier et le deuxième formant virtuels de [a]

(725 Hz/1300 Hz), ce qui explique que dans l'aigu toutes les voyelles soient perçues comme des [a] (Howie & Delattre, 1962).

L'intelligibilité des voyelles dépend également des rapports de fréquence entre les deux premiers formants. Dans le chant, lorsque la fréquence augmente, le deuxième formant varie peu d'une voyelle à l'autre (Sundberg, 1975). Ce phénomène résulte de l'importante ouverture buccale nécessaire à l'émission. Le fait que les F2 des voyelles soient moins différenciés explique un certain nombre de confusions entre les voyelles d'aperture égale comme par exemple les voyelles fermées [i], [y], [u] qui ont leur premier formant à 250 Hz, ou les voyelles ouvertes [ɛ], [œ], [ɔ] qui ont leur premier formant à 510 Hz. C'est donc le F2 qui permet de les distinguer les unes des autres dans la parole, distinction rendue impossible dans le chant où non seulement les F1 sont identiques mais les F2 sont très proches (fig. 2).

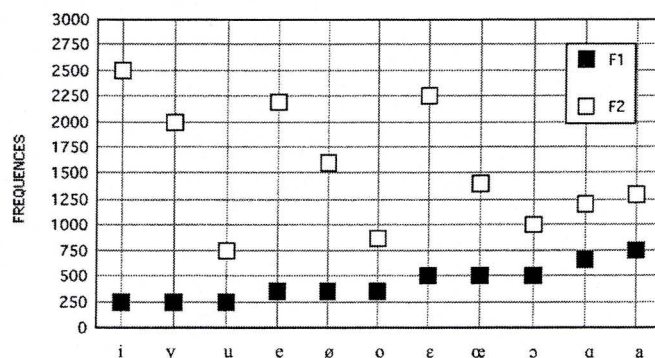


Figure 2
Valeurs formantiques des voyelles du français

[C'est le deuxième formant (F2) qui permet de distinguer les voyelles [i], [y], [u] dont le premier formant (F1) est à 250 Hz, les voyelles [e], [ø], [o] dont le F1 est à 350 Hz ou les voyelles [ɛ], [œ], [ɔ] dont le F1 est à 510 Hz. Or, le F2 qui a des valeurs bien individualisées dans la parole, varie peu d'une voyelle à l'autre dans le chant, ce qui rend difficile la reconnaissance des timbres vocaliques.]

2.2. Intensité

Dans le chant, dès que l'intensité sonore est située entre 90 et 130 dB, le timbre des voyelles est réduit au système vocalique minimal [i], [a], [u].

- les voyelles fermées antérieures [i], [e] tendent vers [i]
- les voyelles ouvertes [ɛ], [œ], [ɔ], [a], [a] tendent vers [a]

• les voyelles fermées postérieures [y], [ø], [o], [u], tendent vers [u], (Husson, 1960 ; Scotto Di Carlo, 1972).

Sur le plan acoustique, ceci s'explique par le fait que lorsque l'intensité vocale augmente, l'énergie sonore se déplace des basses vers les hautes fréquences du spectre, c.à.d. vers la zone du champ auditif où la sensibilité est optimale. Ce transfert d'énergie vers l'aigu rend difficile la différenciation des structures formantiques des voyelles et nuit par conséquent à leur intelligibilité.

2.3. Vibrato

Le vibrato vocal est caractérisé par des variations périodiques de fréquence pouvant être accompagnées de variations périodiques d'intensité synchrones ou asynchrones et des variations de timbre (fig. 3). Le vibrato de fréquence est, comme son nom l'indique, caractérisé par des variations périodiques de fréquence autour de la fréquence-cible qui correspond à la note à émettre. Si ces variations affectent peu l'intelligibilité des voyelles synthétiques (Sundberg, 1977), leur influence sur l'intelligibilité des voyelles naturelles est plus marquée. Une étude de Dalcastello (1994) portant sur des comparaisons entre sons blancs² et sons vibrés montre que le taux d'identification des voyelles est supérieur de 28% lorsqu'elles sont émises sans vibrato et ce, quelle que soit la fréquence à laquelle elles sont chantées. Selon l'auteur, ce phénomène serait dû au fait que les modulations du vibrato de fréquence qui sont de l'ordre du demi-ton mais peuvent atteindre le ton complet, ont pour effet d'occulter les fréquences formantiques et de rendre plus difficile l'identification des voyelles chantées.

2.4. Singing Formant

En plus des formants vocaliques, la voix des chanteurs d'opéra présente un formant extra-vocalique découvert par Bartholomew en 1934 dont les travaux ont été repris par Winckel en 1953, puis par Vennard en 1967, et enfin par Sundberg en 1972. Ce « formant du chanteur » baptisé « *Shimmer* » par Bartholomew, « *Ring* » par Winckel et « *Singing Formant* » par Sundberg, est situé entre 2000 et 3000 Hz pour les voix masculines et entre 3000 et 4000 Hz pour les voix féminines. Le Singing Formant est responsable de la portée de la voix. On ne trouve d'ailleurs ce renforcement des harmoniques dans la zone privilégiée de la sensibilité auditive que dans les voix travaillées : celle des chanteurs d'opéra et celles des comédiens lorsqu'ils utilisent la voix projetée (Scotto Di Carlo, 1991), (fig. 4).

² Sons émis sans vibrato que les chanteurs appellent aussi parfois *son droit*. Dans le chant classique, les sons blancs sont utilisés à des fins expressives.

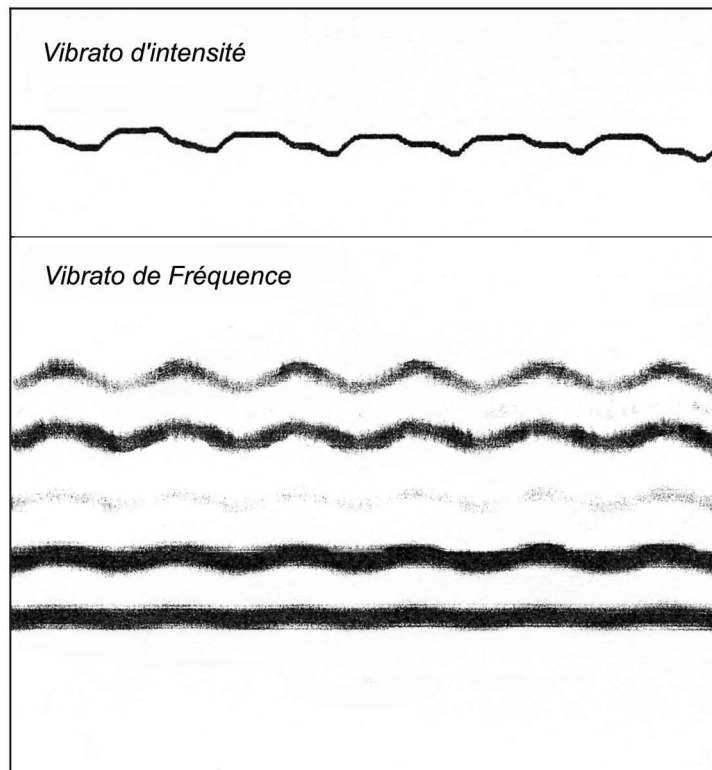
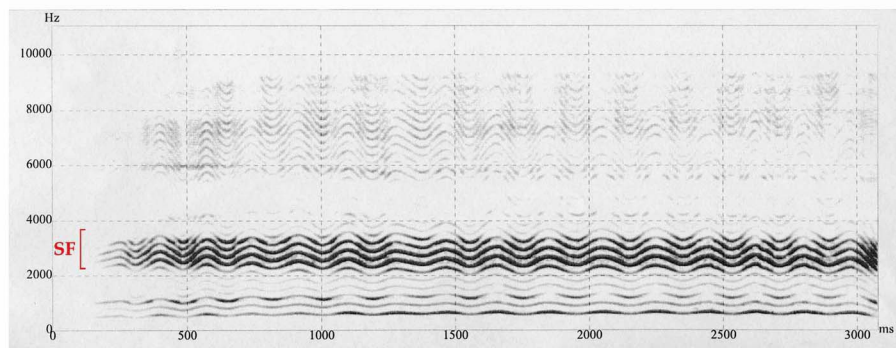


Figure 3
Caractéristiques acoustiques du vibrato vocal

Le vibrato vocal est caractérisé par des variations périodiques de fréquence qui peuvent être accompagnées de variations périodiques d'intensité et de timbre. Sur ce sonagramme, le vibrato de fréquence et le vibrato d'intensité sont en opposition de phase. On peut également observer un vibrato de timbre, mieux visible sur le dernier harmonique dont tous les pics présentent une augmentation d'intensité. Par sa forme sinusoïdale, le vibrato rend l'identification des consonnes difficile en masquant la direction des transitions formantiques.



Caractéristiques acoustiques du Singing Formant

[Le Singing Formant, qui est responsable de la portée de la voix, est caractérisé sur le plan acoustique par un renforcement des harmoniques situés aux alentours de 3000 Hz, c.-à-d. dans la zone du spectre où l'acuité auditive est optimale. La hauteur du Singing Formant varie en fonction de la catégorie vocale. Pour ce soprano lyrique, il est situé entre 2000 et 3500 Hz avec un épïcentre à 2750 Hz.]

Il est probable que le Singing Formant soit en partie responsable du taux élevé de confusions entre les voyelles diffuses et plus précisément entre [i] et [y] (Scotto Di Carlo & Germain, 1985). En effet, on peut penser que le Singing Formant, situé entre 2000 et 4000 Hz, masque la hauteur réelle du deuxième formant des voyelles situées dans cette zone de fréquence comme [i], [e], [ɛ] ou [y]. Or, ainsi qu'on vient de le voir, c'est la hauteur du deuxième formant qui permet de distinguer [i] de [y] par exemple puisque leur F1 est à 250 Hz alors que leur F2 est respectivement à 2500 Hz pour [i] et à 2000 Hz pour [y]³.

2.5. Contraintes phonatoires

Pour l'ensemble des registres, on observe une tendance très nette à la centralisation des voyelles. Cela s'explique par le fait que le chant nécessite un tractus vocal totalement dégagé et une grande souplesse des muscles linguaux et bucco-faciaux. Or, le système vocalique du français est caractérisé à la fois par une constriction du conduit vocal due à une forte proportion de voyelles fermées (63%) et par une tension musculaire importante due à un pourcentage élevé de voyelles

³ Les valeurs mentionnées ici ont été relevées par Lafon, Rossi, Landercy et Renard, sur des voix masculines et confirment les valeurs optimisées, utilisées par Delattre pour la parole synthétique. Actuellement, ce sont les données de Calliope (1989) qui servent de référence, bien que, selon les auteurs "ces valeurs ne puissent, en aucun cas, être considérées comme la norme du français". Elles sont fondées sur l'analyse formantique des dix voyelles orales du français produites par dix locuteurs et neuf locutrices dans un contexte CV ou CVC. En ce qui concerne la voyelle [i], pour les voix d'hommes, Calliope donne comme valeur médiane des deux premiers formants, F1 = 308 Hz, F2 = 2064 Hz et pour la voyelle [y], F1 = 300 Hz, F2 = 1750 Hz.

antérieures (73%) (Delattre, 1953 ; Scotto Di Carlo, 1967). On peut donc penser que le chanteur tente de réduire cette constriction et cette tension musculaire incompatibles avec les impératifs du chant en centralisant l'articulation des voyelles, ce qui revient à sous-articuler.

D'autre part, dans le chant, l'*ouverture buccale* qui augmente du grave à l'aigu est toujours plus importante qu'en voix parlée, quel que soit le registre. Il en est de même pour l'*aperture*, ce qui va gêner considérablement l'émission des voyelles fermées comme [i], [y] ou [e].

Enfin, la *masse linguale*, plus antériorisée dans le médium, le haut médium et l'aigu afin de dégager le pharynx, entrave la réalisation des voyelles postérieures comme [o] et [u] par exemple (fig. 5).

En ce qui concerne les *lèvres*, les techniques vocales qui ont cours actuellement nécessitent une projection labiale importante pour l'émission du grave et un peu moins importante pour l'émission de l'aigu, ce qui va entraîner des difficultés pour la réalisation des voyelles étirées comme [i] ou [e]. En revanche, l'étirement latéral des commissures labiales utilisé dans l'aigu par certaines catégories vocales (voix légères⁴), va gêner la réalisation des voyelles arrondies comme [y], [ø], [œ] et [u].

L'*ouverture vélo-pharyngée*, toujours présente dans le chant, augmente avec la fréquence, jusqu'à une certaine limite au-delà de laquelle le son hypernasalisé devient inesthétique (Scotto Di Carlo & Autesserre, 1987). Dans le chant, les voyelles orales étant déjà nasalisées, lorsque le chanteur veut émettre une voyelle nasale, au lieu d'abaisser davantage son voile, ce qui compromettrait l'esthétique du son, il va élever le dos de sa langue vers le voile du palais : la voyelle émise de la sorte est moins buccalisée et paraît de ce fait plus nasalisée (Husson, 1962).

En conclusion, on peut dire que pour les voix légères, toutes les voyelles ouvertes et apertes, ainsi que les voyelles étirées seront conservées dans l'aigu. En revanche, pour les voix puissantes, les voyelles fermées et les voyelles labiales seront menacées dès le haut médium et l'aigu.

Ces phénomènes ne sont cependant pas les seuls à nuire à l'intelligibilité du texte chanté. Les consonnes subissent elles aussi d'importantes distorsions qui nuisent à leur intelligibilité. De plus, elles ont une incidence sur la justesse des voyelles subséquentes.

⁴ Par opposition aux voix puissantes, les voix légères sont des voix de faible intensité sonore qui brillent par leur agilité et la longueur de leur tessiture (trois octaves au lieu de deux octaves et demi pour les voix puissantes).

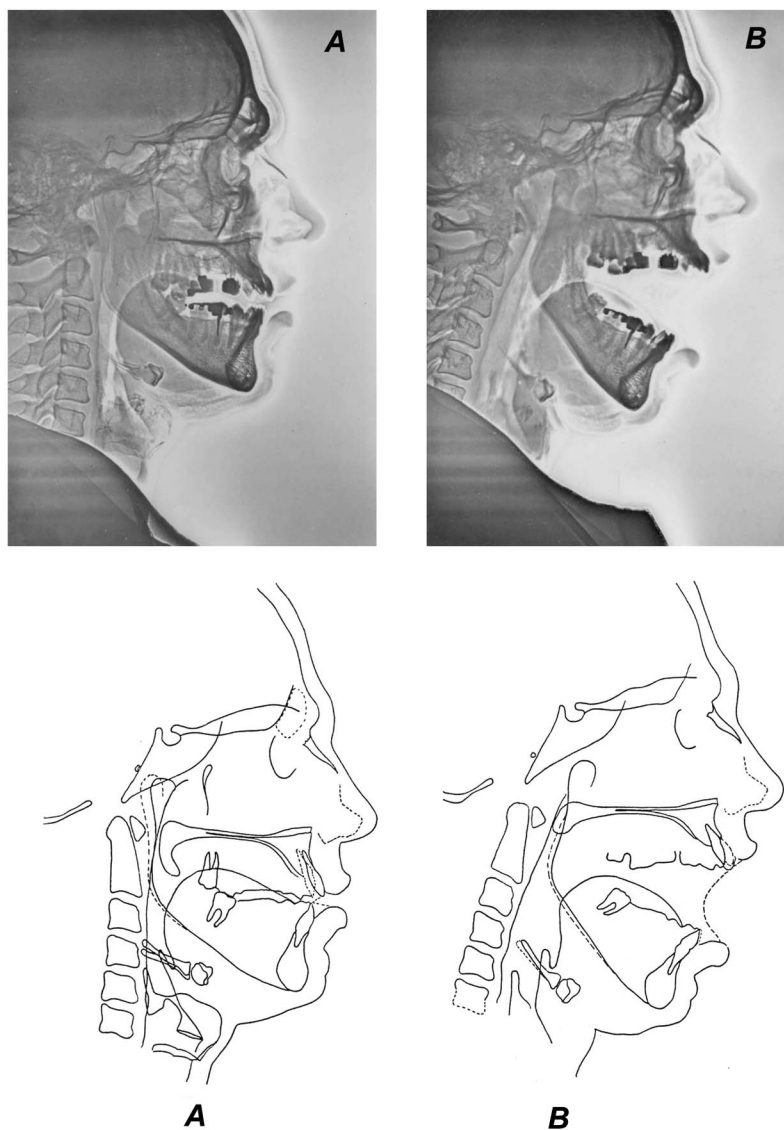


Figure 5 et 5bis
Radiographies de la voyelle [a]
émise sur do2 et do4 par un ténor lirico-spinto

[Lors du passage du grave (A) à l'aigu (B), on peut observer, pour toutes les catégories vocales, une antériorisation de la masse linguale qui entrave la réalisation des voyelles postérieures, ainsi qu'une augmentation importante de l'ouverture buccale et de l'aperture qui gêne la réalisation des voyelles fermées et la production des occlusives labiales et dentales ainsi que celle de certaines constrictives dont l'articulation nécessite un resserrement important du tractus vocal.]

3. Consonnes

3.1. Fréquence

L'intelligibilité des consonnes, comme celle des voyelles, est inversement proportionnelle à la fréquence à laquelle elles sont chantées. Chez les sopranos, on observe dès le médium une baisse de l'intelligibilité de toutes les voyelles et de toutes les consonnes qui chute rapidement jusqu'à l'aigu. Les consonnes restent toutefois mieux compréhensibles que les voyelles puisque leur taux moyen d'identification, tous registres confondus, est de 40,3% contre 31,8% pour les voyelles (Germain & Seassau, 1982 ; Scotto Di Carlo & Rutherford, 1990).

Certaines consonnes résistent mieux que d'autres à l'élévation de la fréquence fondamentale. La constrictive palatale [ʒ] arrive en tête, puis ce sont les consonnes nasales [ɲ], [m], [n], suivies de la consonne liquide [l] et des constrictives palatales [ʃ] et [z]. Viennent ensuite les occlusives palatales [g], [k] puis les occlusives labiales [b], [p] puis [r] et enfin les occlusives et constrictives dentales [d], [v], [t], [f], [s] (Germain & Seassau, 1982).

La meilleure résistance des palatales dans les registres élevés par rapport aux labiales et aux dentales, s'explique par le fait que, comme nous l'avons déjà vu, dans le chant, l'ouverture buccale est beaucoup plus importante que dans la parole⁵. Or, une grande ouverture buccale est davantage compatible avec une articulation palatale relativement ouverte qu'avec une articulation labiale ou dentale très fermées. Quant aux consonnes nasales et liquides, on sait que leur structure acoustique, très proche de celles des voyelles, ne constitue pas une gêne dans le chant et qu'elle facilite même la cantillation⁶ ce qui explique que ce soient elles que l'on utilise spontanément lorsqu'on chante un air dont on ne connaît pas les paroles ("na, na, na, nère" ou "la, la, la, lère") (Scotto Di Carlo, 1993).

L'analyse acoustique permet en outre d'observer un phénomène spécifique à la voix chantée appelé « déstructuration syllabique » par Germain & Seassau (1982) et qui, en provoquant la disparition ou l'atténuation des pentes de transitions formantiques, rend extrêmement difficile la détection des loci et entraîne une perte importante de l'intelligibilité des consonnes (fig. 6).

⁵ Dans le grave, la bouche est environ 2 fois plus ouverte que dans la parole, dans le médium 3 à 4 fois et dans l'aigu 9 à 11 fois.

⁶ Du latin populaire *cantillare* = fredonner). Action de chanter à mi-voix sans articuler les paroles mais en utilisant des syllabes de substitution.

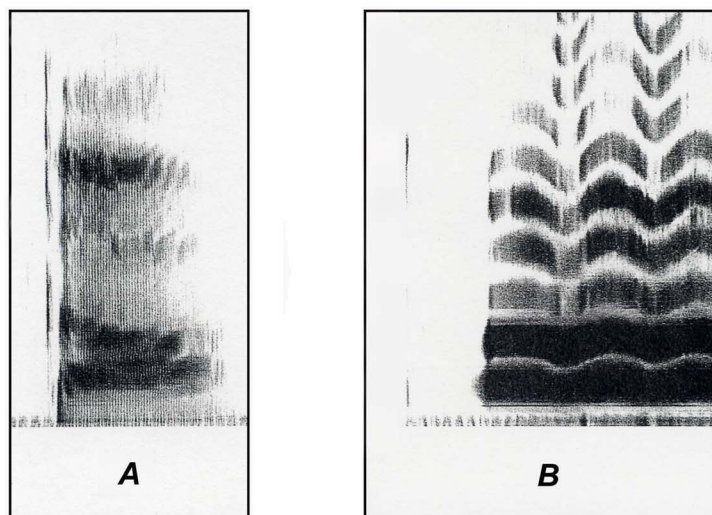


Figure 6
*Sonagrammes de la séquence [ta] parlée
 et chantée sur fa#4 par un soprano léger colorature*

[En A, l'analyse sonographique de la séquence [ta] en voix parlée permet d'observer l'orientation des transitions des trois premiers formants : transition négative pour F1 et positive pour F2 ainsi que F3. En B, la même séquence chantée par le même sujet sur fa#4 (740 Hz) met en évidence la disparition des pentes des transitions formantiques qui sont neutralisées par le vibrato. La détection du locus devient de ce fait impossible et l'identification de la consonne difficile.]

On peut donc affirmer que l'augmentation de la fréquence fondamentale a une influence sur la perception des consonnes chantées dans la mesure où elle est responsable de la disparition d'importants indices acoustiques comme les transitions formantiques ou les zones de bruit, ce qui va engendrer des erreurs d'identification.

3.2. Intensité

Au fur et à mesure que la fréquence fondamentale et l'intensité de la voyelle chantée augmentent, l'intensité de la consonne diminue⁷. De ce fait, les pôles de bruit (qu'il s'agisse des bruits

⁷ On trouve également ce phénomène dans la voix criée où les voyelles sont plus intenses que les consonnes, alors que c'est l'inverse qui se produit pour la voix chuchotée où ce sont les consonnes qui sont plus intenses que les voyelles (Rostolland, 1979).

d'explosion ou de friction) sont très faibles, ce qui va entraîner des confusions dans l'identification de certaines catégories de consonnes pour lesquelles l'intensité et la fréquence du bruit sont les principaux indices acoustiques qui permettent de les distinguer (fig. 7).

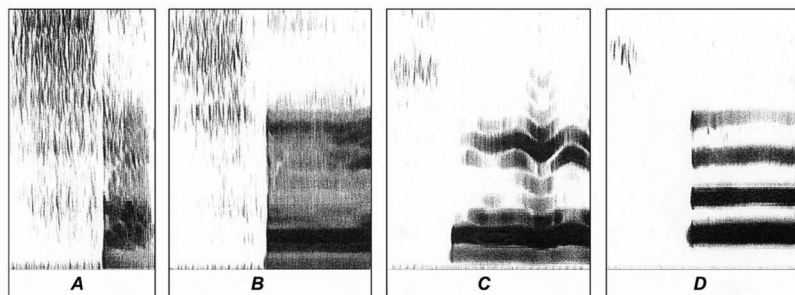


Figure 7
Atténuation des pôles de bruit en fonction de la fréquence

[La séquence [sa] parlée (A) puis chantée par un soprano léger dans le grave (B), le médium (C) et l'aigu (D) montre que dans le grave, la bande de bruit de [s] est moins intense, plus aiguë et plus brève que dans la parole. Au fur et à mesure que la fréquence augmente, la zone de bruit de [s] devient de moins en moins intense, de plus en plus restreinte et de plus en plus brève, d'où une perte considérable d'information qui va rendre sa reconnaissance difficile.]

3.3. Vibrato

Par sa forme sinusoïdale, le vibrato contribue à masquer la direction des transitions formantiques, rendant difficile la distinction entre les différentes catégories de consonnes et notamment entre les labiales qui ont une transition négative (configuration convexe) et les palatales qui ont une transition positive (configuration concave).

3.4. Contraintes phonatoires

Comme pour les voyelles, l'augmentation de l'ouverture buccale en fonction de la fréquence fondamentale va gêner, dans le haut médium et l'aigu, la production des occlusives labiales ou dentales telles que [p], [b], [t], [d], ou des constrictives qui nécessitent un resserrement important du tractus vocal comme c'est le cas de [f], [v], [s]. En revanche, les consonnes palatales seront conservées dans le registre aigu. D'autre part, la production des occlusives sonores telles que [b],

[d], [g] s'accompagne d'un abaissement laryngé peu compatible avec l'émission de l'aigu, en particulier chez les chanteurs qui utilisent des techniques vocales à larynx haut (voix légères).

En résumé, les consonnes dont les accommodations articulatoires sont compatibles avec les contraintes phonatoires, c.-à-d. les consonnes palatales, les consonnes qui se rapprochent le plus des voyelles comme les liquides [l], [r] sont maintenues dans le chant. Les occlusives et les constrictives, quant à elles, vont être menacées dès que l'on atteindra le haut médium et l'aigu, les plus menacées étant celles qui nécessitent un relèvement antérieur de toute la langue, pointe comprise, comme [t] ou [d] par exemple.

3.5. Phénomènes aéro-dynamiques

Les consonnes résultent d'une obstruction totale ou partielle du tractus vocal. Or, pour chanter, le tractus vocal doit être totalement libre. On imagine aisément le problème que pose au chanteur cette incompatibilité fondamentale entre les impératifs esthétiques et les nécessités linguistiques.

Pendant l'émission des consonnes constrictives, les phénomènes de turbulence aérienne provoqués par le resserrement du conduit vocal se manifestent sur le plan auditif par des bruits de frottement que le chanteur essaie d'amortir parce qu'il les juge inesthétiques, ce qui va avoir pour conséquence de détruire en partie les indices acoustiques de bruit si importants pour l'intelligibilité des consonnes. De même, au cours de la production des occlusives où le conduit vocal est momentanément fermé, les silences dus à l'interruption de l'écoulement du flux aérien rompent la continuité de la ligne mélodique en introduisant dans le continuum sonore des zéros acoustiques qui perturbent le legato⁸ de la phrase chantée et que le chanteur essaie instinctivement d'atténuer en réduisant la durée et la force articulatoire des consonnes. Or, la réduction de la durée et de la force articulatoire sont des critères de sous-articulation (fig. 8) (Scotto Di Carlo, 1978, 1981).

3.6. Phénomènes micro-prosodiques

Outre les perturbations qu'elles créent au niveau du continuum sonore, les consonnes ont également une influence sur les voyelles qui les entourent, et en particulier sur celles qui les suivent. Dans la parole, cette influence s'exerce sur l'attaque et la tenue de la voyelle dont elle modifie la hauteur. On peut se demander si ces phénomènes micro-prosodiques sont identiques dans le chant ou si leurs effets se trouvent neutralisés par les contraintes de justesse auxquelles sont soumis les chanteurs.

⁸ Passage d'un son à un autre sans interruption de la ligne mélodique ni altération de l'intensité.

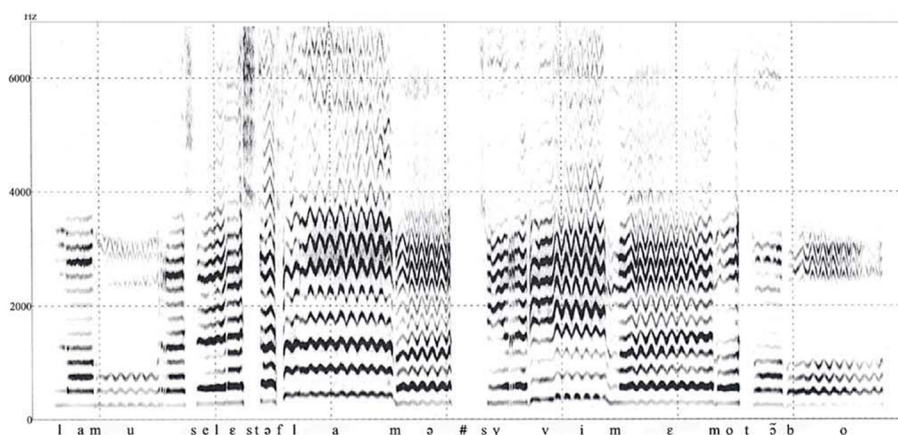


Figure 8

Influence des consonnes sur la qualité du legato

[Sur le sonagramme de la séquence « *L'amour, céleste flamme, survit même au tombeau* » (*Roméo et Juliette* de Gounod) chantée par un ténor lyrique, on peut observer les coupures que font les consonnes dans le continuum sonore. Les consonnes nuisent au legato en rompant la continuité de la ligne mélodique et en altérant la régularité de la courbe d'intensité.]

3.6.1. Incidence sur la tenue

En ce qui concerne la tenue, dans la parole, les consonnes sourdes ([p], [t], [k], [f], [s], [ʃ]) ont tendance à élever la hauteur moyenne de la voyelle subséquente, et les consonnes sonores ([b], [d], [g], [v], [z], [ʒ]) à l'abaisser (Di Cristo, 1982). Dans le chant, on retrouve cette même tendance, bien que moins marquée, pour les sons blancs (Scotto Di Carlo & Raphael, 1977). En revanche, pour les sons vibrés, si les consonnes semblent ne pas affecter la hauteur de la voyelle subséquente, c'est que les variations qui interviennent au niveau de la tenue de la voyelle sont masquées par le vibrato de fréquence dont l'amplitude moyenne est de l'ordre du demi-ton, ce qui permet au chanteur des variations de hauteur non perceptibles à l'intérieur de cette marge de sécurité.

3.6.2. Incidence sur l'attaque

Dans la parole, seules les consonnes sonores ont pour effet d'abaisser la fréquence fondamentale de la voyelle subséquente, alors que les consonnes sourdes correspondantes ont tendance au contraire à l'augmenter (Di Cristo, 1982). Dans le chant, ce phénomène ne se produit que pour les consonnes émises à une fréquence inférieure à 659 Hz (Mi4). À partir de 659 Hz, l'ensemble des consonnes, qu'elles soient sourdes ou sonores, aggravent l'attaque de la voyelle subséquente (Scotto Di Carlo & Raphaël, 1977), (fig. 9). Cela signifie que toutes les consonnes chantées par un soprano dans le haut médium et l'aigu vont entraîner une attaque trop basse de la voyelle qui suit. Cet effet augmentant avec la fréquence, dans le registre aigu du soprano, la chute du fondamental

peut atteindre et même dépasser l'octave. C'est ainsi par exemple que dans l'aigu, pour la consonne [d] on a pu enregistrer chez un soprano léger colorature, une attaque basse de 406 Savarts⁹, soit un intervalle de dixième¹⁰ entre l'attaque et la tenue de la voyelle. Les contraintes de justesse auxquelles sont soumis les chanteurs d'opéra vont les amener à mettre au point des stratégies dans le but d'atténuer ce phénomène : réduction du degré d'occlusion, de la durée et de la force articuloire des consonnes, qui sont les principales caractéristiques de la sous-articulation. Ces stratégies articuloires seront d'autant plus développées que le chanteur possèdera une voix plus aiguë (Scotto Di Carlo, 1978).

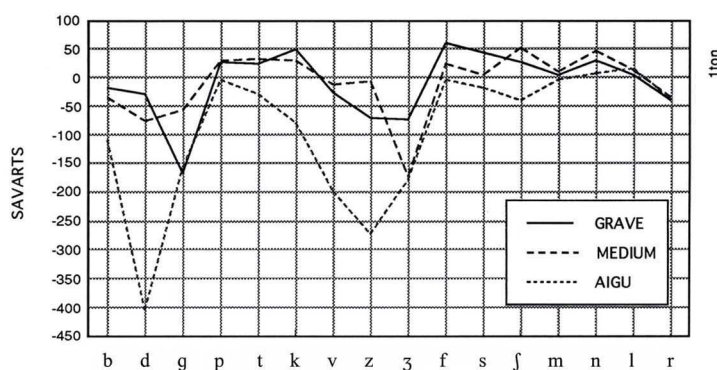


Figure 9
*Justesse des attaques post-consonantiques
en fonction des différentes consonnes chez un soprano léger*

[Sur ce graphique, la ligne zéro représente une attaque vocalique parfaitement juste, les valeurs positives situées au-dessus de cette ligne correspondent à des attaques trop hautes, les valeurs négatives situées au-dessous correspondent à des attaques trop basses. Chaque graduation représente 50 Svt soit un ton. On peut

⁹ Le Savart (Svt) est l'unité d'intervalle musical utilisée par les acousticiens français. Il correspond au plus petit intervalle musical perceptible dans des conditions optimales (1 comma = 5 savarts). Pour mesurer les intervalles musicaux, les anglo-saxons utilisent le *Cent* qui correspond au centième de demi-ton tempéré.

¹⁰ Un intervalle musical peut être constitué de deux ou plusieurs degrés. Le nom de l'intervalle indique le nombre de notes qui sépare la note de départ de celle d'arrivée. Un saut de tierce, par exemple, indique que cet intervalle musical comporte trois degrés, c.-à-d. trois notes successives de la gamme : do (ré) mi. Un saut de quarte comporte quatre degrés (do-fa) ; un saut de quinte, cinq degrés (do-sol) ; un saut de sixte, six degrés (do-la) ; un saut de septième, sept degrés (do-si) ; un saut d'octave, huit degrés (do-do à l'octave). Les intervalles peuvent dépasser les limites de l'octave. On parle alors de neuvième, de dixième, de onzième, etc. Ils peuvent aussi être inférieurs au ton (intervalles d'un demi-ton, d'un quart de ton ou même d'un comma (qui est considéré en musique comme le plus petit intervalle perceptible)).

constater que dans le registre aigu du soprano, exception faite de [m] [n] [l], toutes les consonnes aggravent l'attaque des voyelles qui les suivent. La consonne [d] qui entraîne une attaque vocalique basse de 406 Svt soit 8 tons, a le pouvoir aggravant le plus fort.]

Après l'analyse des divers facteurs qui nuisent à l'intelligibilité des voyelles et des consonnes chantées, il convient d'étudier les problèmes qui se posent au niveau de l'unité phonétique que constitue la syllabe et plus particulièrement au niveau de ses caractéristiques temporelles.

4. Syllabe

4.1. Durée relative des voyelles et des consonnes

Dans le chant, la durée moyenne des voyelles augmente considérablement alors que la durée moyenne des consonnes diminue légèrement par rapport à la parole (fig. 10). L'allongement des voyelles est dû à un important ralentissement du tempo inhérent à toute production musicale, ainsi qu'à une augmentation de la durée articulatoire liée à l'amplitude considérable des mouvements effectués par les organes phonatoires.

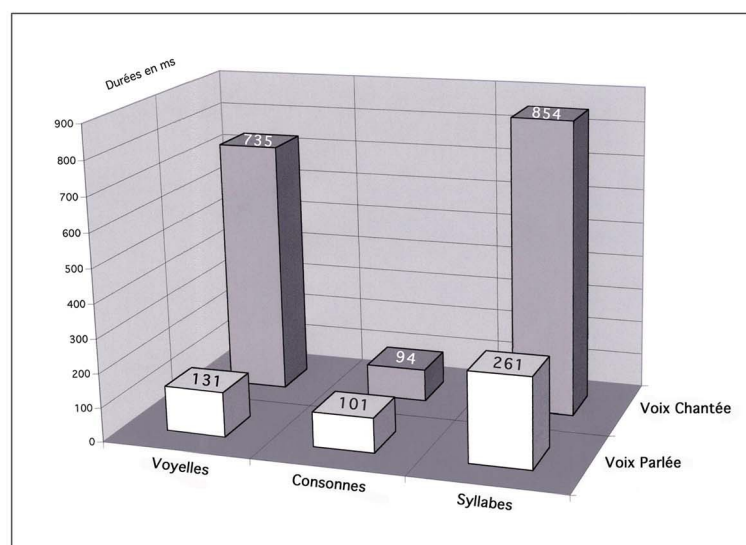


Figure 10
*Durées moyennes des consonnes, voyelles et syllabes
dans la parole et dans le chant*

[En passant de la voix parlée à la voix chantée, la durée moyenne des voyelles augmente considérablement tandis que celle des consonnes diminue légèrement. Ceci est dû au fait que les voyelles constituent un support mélodique idéal alors que les consonnes interrompent la ligne mélodique en formant un obstacle à l'écoulement du flux aérien.]

La durée des consonnes sourdes et des consonnes vocaliques est comparable à celle que l'on trouve en voix parlée. Quant aux consonnes sonores, qu'il s'agisse des occlusives ou des constrictives, elles sont plus brèves dans la parole que les consonnes sourdes, tendance qui se manifeste encore plus nettement dans le chant où les consonnes sonores subissent un abrègement très net, de l'ordre de la moitié ou du quart de la durée qu'elles ont en voix parlée, en raison de contraintes physiologiques. En effet, une durée trop importante aboutirait au dévoisement des consonnes sonores en entraînant une augmentation de la pression intra-orale puis une égalisation des pressions au-dessus et au-dessous de la glotte et enfin une décroissance brutale du débit d'air transglottique, provoquant une diminution de l'effet Bernouilli (fig. 11). Si l'abrègement des consonnes résulte des puissantes contraintes aérodynamiques qu'elles subissent au niveau de leur production, il ne faut pas sous-estimer l'importance des contraintes esthétiques que s'imposent les chanteurs afin de préserver le legato, qui est l'un des paramètres essentiels de la musicalité (cf. 3.5).

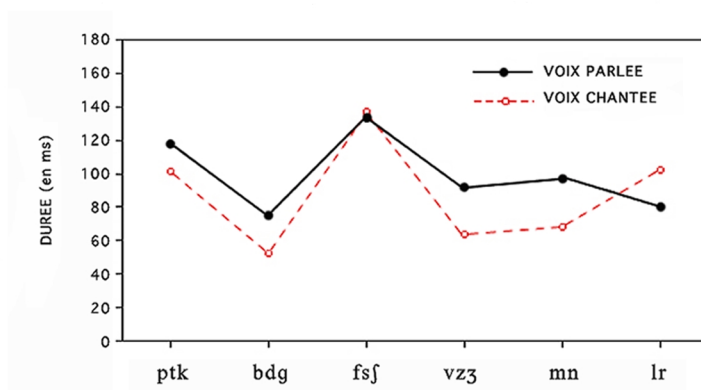


Figure 11

Durée moyenne des classes de consonnes en voix parlée et chantée

[Dans la parole comme dans le chant, on observe un abrègement des consonnes sonores par rapport aux consonnes sourdes correspondantes qui est dû à de puissantes contraintes aérodynamiques de production. Les résultats obtenus pour [l] et [r] chantés s'expliquent par le fait qu'il s'agit de consonnes prolongeables grâce à l'adjonction d'éléments-voyelles (svarabhakti). Ainsi par exemple, le mot « calme » prononcé [kal'm] ou encore les nombreux éléments vocaliques intercalés entre les closions du [r] alvéolaire battu utilisé dans le chant. Autrement dit, même si le chanteur donne l'impression de prolonger les consonnes, en réalité, il ne fait qu'intercaler des éléments vocaliques.]

4.2. Structuration temporelle de la syllabe chantée

Dans le chant, l'allongement très important que subit la syllabe entraîne une modification de sa composition : plus la syllabe s'allonge, plus la proportion de voyelle augmente et plus la proportion de consonne diminue. C'est ainsi par exemple que la durée des consonnes qui représente la moitié de la durée totale de la syllabe parlée ne représente plus que le huitième de la durée totale de la syllabe chantée. L'équilibre temporel de la syllabe se trouve donc détruit (Scotto Di Carlo & Autesserre, 1992 ; Scotto Di Carlo, 2005). Or, notre perception de l'unité phonétique que constitue la syllabe est basée sur la loi d'isochronisme syllabique, selon laquelle toutes les syllabes tendent à être de durées égales. Cet équilibre syllabique est maintenu par les rapports de durée qui s'instaurent entre les voyelles et les consonnes : lorsqu'à l'intérieur d'une syllabe parlée une consonne est très longue, la voyelle sera plus brève et inversement. En raison des contraintes rythmiques qui font que les syllabes chantées ont des durées différentes selon les figures de notes qui leur sont attribuées, il n'y a pas d'isochronisme syllabique dans le chant. De ce fait, dans la mesure où, selon la Gestalttheorie, la perception est syncrétique, les anamorphoses créées sur les images acoustiques de la syllabe par l'allongement vocalique vont perturber globalement la reconnaissance des formes vocales non seulement au niveau de la syllabe mais au niveau du mot.

La durée moyenne des consonnes comme celle des voyelles augmente lorsque le tempo¹¹ ralentit, mais alors que les durées vocaliques varient dans des proportions importantes, la durée des consonnes croît dans une limite réduite, ce qui aggrave le déséquilibre syllabique.

4.3. Limites d'élasticité des voyelles et des consonnes

Les limites d'élasticité des consonnes sont pratiquement identiques dans la parole et dans le chant, la limite de compression étant de 16 ms pour la parole et le chant et la limite d'expansion de 324 ms pour la voix parlée et 340 ms pour la voix chantée (Scotto Di Carlo, 2005).

En ce qui concerne l'*expansion des consonnes*, elle est limitée par des contraintes physiologiques de type aérodynamique. En effet, ainsi que nous l'avons vu, une durée trop importante de leur tenue entraîne une augmentation de la pression intra-orale qui fait chuter le débit d'air transglottique et entrave la vibration des cordes vocales, ce qui aboutit notamment au dévoisement des occlusives sonores. Quant aux limites de *compression des consonnes*, elles sont dues à des contraintes neuro-physiologiques comme l'inertie des effecteurs et à des contraintes perceptives. En effet « *la limite à la compression est impérative, car il faut préserver l'intelligibilité du message qui dépend du maintien de l'identité du*

¹¹ Mouvement ou vitesse à laquelle on exécute une œuvre.

phonème fondée entre autre sur la durée : en-deçà d'une certaine durée, il y a destruction de l'identité du phonème » (Duez, 1987).

Ce phénomène accroît le déséquilibre dans la structure temporelle de la syllabe et perturbe l'intelligibilité non plus au niveau phonémique mais au niveau de l'intégration syllabique (Scotto Di Carlo, 1972).

L'*expansion des voyelles* est très importante dans le chant où physiologiquement elle n'est limitée que par la tenue de souffle du chanteur et musicalement par la partition, à l'exception toutefois des voyelles nasales. En effet, dans le chant, on ne peut tenir une voyelle nasale, car en raison de son inertie, le voile du palais s'affaisse sur le dos de la langue au bout de quelques centièmes de seconde, obstruant ainsi la cavité buccale. Pour éviter que la voyelle nasale devienne un son émis bouche exclue dont le timbre hypernasalisé est particulièrement inesthétique, les chanteurs utilisent un stratagème qui consiste à tenir la voyelle orale correspondante et à ne la nasaliser qu'au dernier moment.

Les limites de *compression des voyelles* sont également très proches en voix parlée et en voix chantée (respectivement 38 et 40 ms). Les raisons de ces limites de compressibilité sont d'ordre perceptif, une dilatation excessive des phonèmes étant incompatible avec leur intelligibilité (Scotto Di Carlo, 2005).

Conclusion

Dans le chant, les phonèmes subissent d'importantes distorsions dues à la fréquence et à l'intensité auxquelles ils sont émis. Le vibrato perturbe l'intelligibilité de la parole chantée en masquant les fréquences des formants ainsi que la direction de leurs transitions, ce qui rend difficile l'identification des voyelles et des consonnes. Quant au Singing Formant, il nuit à la reconnaissance des voyelles antérieures, fermées en masquant la hauteur réelle de leur deuxième formant. En outre, la composition syllabique, totalement modifiée en raison de l'allongement considérable qu'elle subit, entrave la reconnaissance des formes vocales.

La voix chantée nécessite, par ailleurs, un tractus vocal totalement dégagé. Toutes les manœuvres physiologiques destinées à diminuer la constriction du tractus vocal et à réduire la tension musculaire due à l'antériorisation vocalique du français ou à la force articuloire des consonnes reviennent à sous-articuler. Cette sous-articulation nuit à la richesse et à la pureté des timbres vocaliques et entraîne la disparition des pentes de transitions consonantiques qui sont des éléments importants de l'intelligibilité.

Enfin, les accommodations pharyngo-buccales indispensables à une émission vocale correcte ne sont pas toujours compatibles avec les accommodations articuloires nécessaires à l'articulation

phonémique. Les différentes observations et constatations que nous avons faites tout au long de cette étude nous amènent à énoncer la loi d'intelligibilité de la voix chantée : *Chaque fois que les accommodations phonatoires coïncident avec les accommodations articulatoires, l'intelligibilité des phonèmes est conservée.* C'est le cas, par exemple, des voyelles labiales et des voyelles fermées dans le grave des voyelles ouvertes ou des consonnes palatales dans l'aigu. *Dans les autres cas, l'intelligibilité est d'autant plus perturbée que les accommodations phonatoires et les accommodations articulatoires sont antagonistes.*

Bibliographie

- BARTHOLOMEW, W. (1934). A physical definition of good voice quality in male voice, *Journal of the Acoustical Society of America*, 6, p. 25-33.
- CALLIOPE (1989). *La parole et son traitement automatique*, Collection technique et scientifique des Télécommunications, Masson, Paris.
- DALCASTELLO, M. (1994). Vibrato e intelligibilità dei vocali, *Bolletino di Psicologia*, I, 3, p. 360-376.
- DELATTRE, P. (1953). Les modes du français, *French Review*, 27, p. 45-58.
- DI CRISTO, A. (1982). *Prolégomènes à l'étude de l'intonation. La microprosodie*, Collection Sons et Parole, éd. du CNRS, Paris, 232 p.
- DUEZ, D. (1987). *Contribution à l'étude de la structuration temporelle de la parole en français*, Thèse de Doctorat d'Etat, Aix-en-Provence, 575 p.
- GERMAIN, A. ; SEASSAU, H. (1982). *Influence de la fréquence sur l'intelligibilité et les indices acoustiques des consonnes du français en voix chantée*, DEA, Université de Provence, 115 p.
- HOWIE, J. ; DELATTRE, P. (1962). An experimental study of the effect of pitch on the intelligibility of vowels, *National Association of Teachers in Singing Bulletin*, 18, p. 6-9.
- HUSSON, R. (1960). *La voix chantée*, Collection Science et Techniques d'aujourd'hui, éd. Gauthier-Villars, Paris, 205 p.
- HUSSON, R. (1962). *Physiologie de la phonation*, Masson, Paris, 590 p.
- ROSTOLLAND, D. (1979). *L'audition de la parole en présence de bruit*, Thèse de doctorat d'état, Paris, 281 p.
- SCOTTO DI CARLO, N. (1967). *Contribution à l'étude statistique du français contemporain*, DES, Université de Provence, 37 p.
- SCOTTO DI CARLO, N. (1972). Étude acoustique et auditive des facteurs d'intelligibilité de la voix chantée, *Proceedings of the VIth International Congress of Phonetic Sciences*, Mouton, La Haye, p. 1017-1023.

- SCOTTO DI CARLO, N. ; RAPHAEL, A. (1977). Etude acoustique et statistique de l'influence des consonnes sur la justesse des voyelles subséquentes en voix chantée, *Communication*, II^e Symposium International de Recherches sur le Chant, IRCAM, Paris, juillet. Texte intégral dans *Travaux de l'Institut de Phonétique d'Aix*, 1977, 4, p. 237-260.
- SCOTTO DI CARLO, N. (1978). Pourquoi ne comprend-on pas les chanteurs d'opéras ? *La Recherche*, IX, 89, p. 495-497.
- SCOTTO DI CARLO, N. (1981). Perturbing effects of Overarticulation in Singing, *Zeitschrift für Phonetik*, 34, 2, p. 197-210.
- SCOTTO DI CARLO, N. ; GERMAIN, A. (1985). A perceptual study of the influence of pitch on the intelligibility of sung vowels, *Phonetica*, 42, 4, p. 188-197.
- SCOTTO DI CARLO, N. ; AUTESSERRE, D. (1987). Movements of the velum in Singing, *Journal of Research in Singing*, XI, 1, p. 3-13.
- SCOTTO DI CARLO, N. ; RUTHERFORD, A. (1990). The effect of pitch on the perception of a coloratura soprano's vocalic system, *Journal of Research in Singing*, XIII, 2, p. 1-24.
- SCOTTO DI CARLO, N. (1991). La voix chantée, *La Recherche*, XXIII, 235, p. 1016-1025.
- SCOTTO DI CARLO, N. ; AUTESSERRE, D. (1992). L'organisation temporelle de la syllabe dans la parole et dans le chant, *Communication*, XII^e Congrès International des Sciences Phonétiques, Aix-en-Provence, août 1991. Texte intégral dans *Travaux de l'Institut de Phonétique d'Aix*, 1992, 14, p. 189-204.
- SCOTTO DI CARLO, N. ; AUTESSERRE, D. (1993). Sons et sens dans l'opéra, *Médecine des Arts*, 4, p. 4-11.
- SCOTTO DI CARLO, N. (1993). Diction et Musicalité, *Médecine des Arts*, 5, p. 4-11.
- SCOTTO DI CARLO, N. (2005). Structure temporelle de la syllabe parlée et chantée, *Médecine des Arts*, 52, p. 2-11.
- SUNDBERG, J. (1972). Production and function of the "Singing Formant", *Report of the XIth Congress of Acoustics*, Copenhagen, 2, p. 679-686.
- SUNDBERG, J. (1977). Vibrato and vowel identification, *Archives of Acoustics*, 2, 4, p. 257-266.
- VENNARD, W. (1967). *Singing, the mechanism and the technique*, Carl Fischer ed., New York, 275 p.
- WINCKEL, F. (1953). Physikalische kriterien für objektive Stimmbeurteilung, *Folia Phoniatica*, 5, p. 232-252.