

**Analyse Acoustique**  
**et Catégorisation**

**d'un Ensemble de Qualités Vocales**  
**Pertinent pour la Description**

**de Voix**  
**Lyriques Masculines**

**Mémoire de Stage du DEA ATIAM**  
**Année 2003-2004**  
**8 Juillet 2004**

**Laboratoire d'Acoustique Musicale de Paris**



**Responsable du stage : Maëva Garnier**

**Co-encadrants : Nathalie Henrich – Michèle Castellengo**

**Etudiant : David Sotiropoulos**

**Université des Sciences Pierre et Marie Curie – Paris 6**  
**Mars – Juillet 2004**

# SOMMAIRE

<b>REMERCIEMENTS.....</b>	<b>4</b>
<b>INTRODUCTION.....</b>	<b>5</b>
<b><u>PARTIE A – CADRE THEORIQUE DU STAGE.....</u></b>	<b>6</b>
<b><u>Chapitre I : Qualité Vocale .....</u></b>	<b>7</b>
I.1)Problématiques de la Qualité Vocale.....	7
I.1.1)Problématique de définition.....	7
I.1.2)Aspect multi-paramétrique.....	8
I.2)Contexte.....	9
<b><u>Chapitre II : Chant Lyrique .....</u></b>	<b>11</b>
II.1)Eléments de Physiologie - Fonctionnement de l'Appareil Vocal.....	11
II.1.1)Source, Conduit vocal et Cavités de résonance.....	11
II.1.2)Formants vocaliques – principe.....	12
II.2)Eléments de technique vocale.....	13
II.2.1)Formant du chanteur.....	13
II.2.2)« Ouverture » et « couverture ».....	13
II.2.3)Prononciation – Triangle vocalique.....	14
<b><u>Chapitre III :Diverses Approches de l'Etude de la Qualité Vocale.....</u></b>	<b>15</b>
III.1)Approche Psychophysique.....	15
III.2)Approche Psycholinguistique.....	16
III.2.1)Perception et catégorisation de la qualité sonore.....	16
III.2.2)Une étude antérieure de la qualité vocale.....	17
III.2.3)Résultats de l'étude acoustique.....	17
<b><u>PARTIE B – PREMIERE ETAPE : MATERIEL ET EXPERIMENTATION DU STAGE....</u></b>	<b>19</b>
<b><u>Chapitre IV : Protocole et Base de Données.....</u></b>	<b>20</b>
IV.1)Objectifs du stage.....	20
IV.2)Utilisation de l'ancienne base de données.....	20
IV.2.1)Test exploratoire de catégorisation libre .....	20
IV.2.2)Résultats et conclusions.....	21
IV.3)Elaboration d'un protocole pour notre étude.....	22
IV.3.1)Protocole.....	22
IV.3.2)Choix de la mélodie et du texte.....	25
IV.4)Enregistrement et Matériel.....	25
IV.4.1)Matériel d'enregistrement.....	25
IV.4.2)Chanteurs.....	26
<b><u>PARTIE C – ANALYSES ACOUSTIQUES ET RESULTATS.....</u></b>	<b>27</b>
<b><u>Chapitre V : Analyses acoustiques des voyelles – déplacement des formants vocaliques suivant les qualités vocales .....</u></b>	<b>28</b>
V.1)Méthode de mesure.....	28
V.2)Comparaison de la voix fry avec la voix normale.....	29
V.3)Comparaison des formants vocaliques pour les voyelles /i/, /u/, /é/.....	32
V.4)Comparaison des formants vocaliques pour les voyelles /a/, /o/, /ou/.....	38

V.5)Quelques conclusions générales.....	42
V.6)Remarques.....	43
<b>Chapitre VI : Analyses de la répartition spectrale de l'énergie pour quelques qualités.....</b>	<b>44</b>
VI.1)Qualités vocales « détimbrée » et « timbrée ».....	44
VI.2)Qualités vocales « brillante» et « sourde ».....	50
<b>CONCLUSION.....</b>	<b>53</b>
<b>PERSPECTIVES.....</b>	<b>54</b>
<b>BIBLIOGRAPHIE.....</b>	<b>55</b>
<b>ANNEXE.....</b>	<b>56</b>

# REMERCIEMENTS

Je tiens en tout premier lieu à remercier très chaleureusement M<sup>elle</sup> Maëva Garnier, ma responsable de stage, qui a apporté tout son soutien à son premier stagiaire pour mener à bien ce projet, prenant énormément sur le temps de sa thèse pour m'épauler dans mon apprentissage, et aussi pour sa bonne humeur,

mes sincères remerciements s'adressent aussi à M<sup>me</sup> Nathalie Bernardoni-Henrich pour son aide, ses précieux conseils et son efficacité, en plus d'une grande gentillesse,

ainsi qu'à M<sup>me</sup> Michèle Castellengo pour son savoir, l'intérêt porté à nos travaux durant ces quatre mois et ses riches enseignements sur l'écoute musicale,

toute ma gratitude se tourne vers M<sup>me</sup> Danièle Dubois pour son aide concernant la psycholinguistique qui fait du L.A.M. un laboratoire original, pour ses corrections, son dynamisme à toute épreuve et pour m'avoir permis d'utiliser son instrument de travail,

je ne peux que dire un grand merci aux chanteurs qui se sont prêtés gracieusement aux enregistrements, Pascal Bézard, Rémy Escarré, Robert Expert, Vincent Hédon et Michel Sotiropoulos, en leur souhaitant bonne chance,

merci à mes deux camarades du DEA ATIAM : Pierre Leveau pour son aide indispensable sur le logiciel Matlab et pour s'être prêté à mes tests ainsi qu'Adrien Mamou-Mani pour sa participation au test ; merci également à Julien Bensa pour sa participation au test,

ma sympathie va à tous les membres du L.A.M., et tout particulièrement mes amis que je n'ai pas encore cités de LAM'Ô Choeur, Fabienne, Boris, Nicolas, Pascal et Sébastien, pour leur grande patience et leur application lors des répétitions de chant choral.

Enfin, un grand merci à Rachel pour son soutien quotidien.

# INTRODUCTION

L'étude de la voix intéresse depuis longtemps de nombreuses disciplines rattachées à des domaines très variés allant des arts (musique, théâtre, cinéma ...) à la politique (art d'orateur), en passant par les sciences physiques (acoustique, phonétique, synthèse, mécanique ...), les sciences humaines (phonologie, (psycho)linguistique...) et la médecine (oto-rhino-laryngologie, phoniatrie, orthophonie...). Chacun de ces domaines permet d'appréhender la voix sous un aspect particulier. Il ne faut pas perdre de vue que la voix est un système de production sonore qui se distingue des autres. En premier lieu, elle dote l'homme d'un moyen de communication tout particulier, lui permettant d'exprimer ses émotions, de partager avec autrui. En cela la voix est porteuse de sens. Cependant nous remarquons que le message n'est pas véhiculé uniquement par les seuls mots employés, mais qu'une grande part d'information est présente dans la qualité de la voix. On peut en effet distinguer deux phrases au même contenu sémantique par un changement dans la qualité vocale (intonation, articulation, émotions...). Dans un deuxième temps, la voix permet également de caractériser un individu par rapport à un autre, ce qui mobilise et passionne depuis longtemps les chercheurs du domaine de la reconnaissance vocale. Là encore, la qualité vocale est à prendre en compte pour optimiser les systèmes.

Enfin, la voix peut être utilisée comme un instrument de musique à la fois riche et complexe. Dans ce cas, la perception des qualités vocales se fera sur des paramètres différents que ceux utilisés pour la parole. L'importance du contexte dans l'étude de la perception de la qualité vocale nous a ainsi amené à restreindre le cadre de notre étude à la qualité vocale dans le chant lyrique.

Le travail proposé s'est inscrit dans une méthodologie originale développée au Laboratoire d'Acoustique Musicale, qui se base sur le point de vue d'auditeurs experts et sur les critères qui ont un sens perceptif pour eux. Deux études préalables réalisées au LAM ont mis en place des bases solides sur lesquelles nous nous sommes appuyés. La première étude, faite par N. Henrich pendant sa thèse, a constitué une première base de données destinée à une étude des mécanismes laryngés et

de la qualité vocale. La seconde étude, de M. Garnier, s'est portée d'une part sur l'étude des représentations mentales de la qualité vocale d'experts du chant lyrique à partir de l'analyse psycholinguistique de leur verbalisation et d'autre part sur des analyses acoustiques de la qualité vocale à partir de cette base de données. Elle a permis de dégager les termes consensuels employés dans le chant lyrique, et d'établir des corrélations entre les critères verbaux et acoustiques de caractérisation de la qualité vocale.

L'objet du stage a été d'approfondir ses études, notamment l'analyse acoustique d'un ensemble de qualités vocales sélectionnées pour leur pertinence vis à vis du chant lyrique.

Nous décrivons dans une première partie les fondements de notre étude, en expliquant les problématiques de la qualité vocale (notamment de sa définition) qui nous ont conduit à mener cette étude, en faisant le bilan de l'état actuel des connaissances utiles à notre démarche, et en délimitant le cadre théorique du stage. Dans une deuxième partie, nous expliquerons le protocole que nous avons mis au point pour enregistrer une nouvelle base de données dédiée à l'étude de la qualité vocale. Enfin, nous livrons nos résultats tirés des analyses acoustiques réalisées sur notre corpus d'exemples de qualités vocales.

# **PARTIE A**

## **CADRE THEORIQUE**

**DU**

**STAGE**

# Chapitre I : Qualité Vocale

## I.1) Problématiques de la Qualité Vocale :

### I.1.1) Problématique de définition

La « voix » désigne, de la même manière que le « son », ce qui est émis et ce qui est reçu. C'est pourquoi, à l'instar du concept de « timbre », nous constatons que celui de « qualité vocale » pose encore de nombreux problèmes et questions quant à sa définition. Comme le font remarquer Jody Kreiman, Diana Vanlancker-Sidtis et Bruce R. Gerratt [16] : « *What is voice? The definitional dilemma* »<sup>1</sup> : une définition univoque est non-envisageable du seul fait de la variété des domaines concernés par la voix (comme nous le remarquons en introduction), et surtout de la variété d'utilisations qu'offre celle-ci.

En tout premier lieu, la voix nous offre la possibilité de parler donc de communiquer de manière directe : elle donne un son à la pensée. Ensuite c'est sur elle que l'on peut se baser pour identifier une personne, ou enfin, elle peut être appréhendée comme un instrument de musique à part entière. Il n'est pas lieu de porter de jugement de valeur mais on conviendra que dans cette dernière fonction, la voix occupe une place importante dans toute l'histoire de la musique de par sa richesse, sa portée et sa complexité, qui la rendent si attrayante. En effet, appréhender cet instrument est difficile – ce qui le rend intéressant – et bien que l'on puisse l'aborder de différentes sortes (point de vue de la physiologie, de la technique vocale, ...), nous sommes toujours confrontés au même problème : comment voir, ressentir, comprendre et perfectionner cet « organe » logé dans notre intérieur ? De plus, c'est un instrument très riche et complet qui s'offre à nous : – riche, car par imitation la voix peut reproduire bien des sons d'autres instruments ou autres sources sonores (elle peut aussi en imaginer !) – complet, car outre son ambitus de plusieurs octaves et le fait qu'il puisse tenir des sons très longs, il prend parfois un usage polyphonique (chant diphonique du Kazakhstan, de Mongolie). En outre une des principales fonctions qui le différencie de loin des autres instruments est le fait que l'on puisse prononcer un texte en musique.

On comprend qu'avec la multitude de fonctions que peut prendre la voix, celle-ci ne peut pas être étudiée seulement au regard du message sémantique qu'elle véhicule. De plus la qualité vocale ne se restreint pas seulement à la qualification du timbre de l'individu, elle considère bien souvent la posture, la technique et d'autres événements ou situations relatifs au corps ([4], [9-11]). Bien-sûr la qualité vocale est très liée au timbre musical – avec toute la problématique qui en résulte –, et certains ne la conçoivent d'ailleurs qu'au travers de cet aspect. Il faut dissocier deux modalités du timbre ([3]) : d'une part, le timbre « causal », caractérisé par les événements localisés que sont les transitoires d'attaque et d'extinction pour une identification de locuteur/chanteur (écoute de type reconnaissance) ; d'autre part, le timbre « global » ou « couleur sonore », caractérisé plutôt par le contenu spectral pour une écoute plus qualitative (balance spectrale, formants, degré d'harmonicité ...).

La qualité vocale est liée au pouvoir de changer de timbre, de modifier certains de ses aspects mais aussi à d'autres, comme ceux relatifs à la production vocale, les vibrations des cordes vocales (mécanismes laryngés).

L'identité, les émotions ou les intentions seront aussi véhiculées par la qualité vocale et ses variations. La qualité vocale serait-elle alors ce qui distingue deux messages au contenu identique, au même titre que le timbre différencierait deux sons de mêmes hauteur, intensité et durée ? L'articulation, le timbre de la voix ou la prosodie, caractérisée par le rythme, l'intensité, la hauteur et l'intonation qui à elle seule peut changer la perception de la qualité d'une voix, feraient alors partie de la qualité vocale. Pour exemple, la simple interjection « ça va » peut prendre

---

1 « Qu'est-ce que la voix? Le dilemme de la définition »

plusieurs sens véhiculés plus par la qualité de la voix que par la phrase elle-même : un ton sec, un ton las ou une interrogation ...

En outre, les questions sur la globalité et la localité se posent aussi pour la qualité vocale (B. Payri, 2000). En effet, la qualité sonore est dépendante de nombreux facteurs propres à l'instrument : attaques, voyelles, les différents modes de jeu, les changements de tessiture qui requièrent une partie plus ou moins ponctuelle du son, et d'autres paramètres « externes » tels que l'acoustique du lieu, les aléas du jeu musical (interprétation, spontanéité ...), etc. qui s'appliquent de manière plus globale. Ainsi la longueur de l'extrait musical choisi pour étudier la qualité est problématique. A cette question, peuvent s'en adjoindre d'autres toutes aussi problématiques : quel type d'extrait musical doit-on prendre, un exercice (tel qu'une gamme) ou un fragment de musique ? Très ou peu connu, ou bien inconnu voire inventé ? Nous y reviendrons dans la partie B.

Enfin, un dernier point important dans la qualification de qualité vocale, aussi bien en parole qu'en chant : cette qualification d'une qualité peut se fonder sur une comparaison entre la voix considérée comme « normale » d'un locuteur/chanteur et sa voix « modifiée », mais elle peut aussi être employée indifféremment pour désigner la voix en émission « normale » d'un chanteur par rapport à un autre extrait précédemment entendu, ou bien par rapport à l'idée que l'on a en mémoire d'une voix normale, ou encore par rapport à d'autres qualités (voir Partie Résultats). Ce ne sont alors pas sur les mêmes échelles que nous nous appuyons et il convient de bien séparer ces modalités de la qualité vocale, significatives de la multitude d'interactions.

En résumé, une qualité peut s'appliquer à une voix :

- « normale » → par rapport à une autre qualité → du même chanteur  
→ d'un autre chanteur  
→ d'une référence mémorisée
- par rapport voix « normale » → d'un autre chanteur  
→ d'une référence mémorisée
- « modifiée » → par rapport à une autre qualité → du même chanteur  
→ d'un autre chanteur  
→ d'une référence mémorisée
- par rapport voix « normale » → du même chanteur  
→ d'un autre chanteur  
→ d'une référence mémorisée

### I.1.2) Aspect multi-paramétrique

Un point complexe dans l'évaluation de la qualité vocale est qu'elle fait intervenir de multiples paramètres, tant au niveau verbal qu'acoustique. Leur interdépendance ou non, est importante à noter et à prendre en compte pour chercher des descripteurs valides de la qualité vocale.

C'est à dire, tout d'abord d'un point de vue verbal, que des critères verbaux utilisés pour caractériser la qualité vocale, sont liés par des relations de plusieurs natures :

- relations de synonymies
- relations de causalités, i.e. d'implications ou d'imputations dues à des raisons **physiologiques**, car la configuration de la physiologie peut impliquer la production de plusieurs qualités en même temps ; ou à des raisons esthétiques qui font aller de paire ces qualités (ou une certaine théorie de l'interprétation) comme le timbrage et le *vibrato*

Les qualités sont donc parfois dépendantes, par contre, on peut juger une voix sur plusieurs qualités qui sont cette fois indépendantes, comme le *vibrato* et les voyelles antérieures / postérieures.

De plus, d'un point de vue acoustique :

- une variation d'un paramètre acoustique peut modifier plusieurs qualités vocales
- inversement, la modification d'une qualité est toujours accompagnée d'une variation de plusieurs paramètres acoustiques dont certains sont plus caractéristiques que d'autres, et certains n'ont aucune influence sur la perception de cette qualité

Donc, du point de vue acoustique, les paramètres les plus saillants vont être recherchés.

## I.2) Contexte :

La prise en compte de la situation, déterminée par le cadre de l'étude, est primordiale pour le choix de la méthode à suivre et l'interprétation des résultats observés. Dans ce cas, le contexte joue sur la stratégie d'écoute à adopter par l'auditeur, impliquant l'utilisation d'un vocabulaire précis pour l'observation de critères qui diffèrent selon les cas. Un contexte de voix parlée est bien différent de celui de voix chantée.

Suivant le type de chant auquel on a affaire – « variété », jazz, chant lyrique, ... – et suivant les cultures – asiatique, africaine, européenne, ... – on ne va pas adopter la même écoute. Dans certains cas, par exemple, on parlera de « vibrato » dans d'autres cela n'aura aucun sens, ou la justesse et le timbre dépendront de l'échelle musicale adaptée ou non au contexte, par exemple la musique orientale ou celle des Balkans utilisent des échelles composées de quarts de ton, celles-ci ont moins de sens pour les cultures européennes ...

Nous nous fixons donc un cadre pour notre étude : celui du chant lyrique (« chant savant occidental de l'adulte »), car dans ce domaine un savant contrôle de la qualité vocale (conscient ou non) est nécessaire, voire indispensable pour répondre aux exigences imposées par la musique. La composition de la palette des qualités vocales, dans une situation concrète musicale, est dépendante de la volonté du compositeur : celui-ci joue, par exemple, sur le partage de ses rôles suivant les tessitures et à l'intérieur peuvent apparaître des décompositions suivant les différentes aptitudes de chaque voix à jouer de certaines qualités vocales. Ces qualités vocales sont parfois bien plus déterminantes pour choisir un chanteur que la technique vocale en elle-même : l'audition d'un chanteur sert avant tout à connaître ses aptitudes à « jouer » de sa voix, autrement dit à faire varier la qualité de la voix. La qualité vocale habituelle d'un chanteur (comprenant aussi sa capacité à véhiculer des émotions) peut l'amener à être sélectionné en particulier pour certains rôles de toute époque : on appréciera une « grosse » voix dans *Falstaff* de G. Verdi ; la musique contemporaine s'attache plus consciemment à ces considérations : dans *Le Grand Macabre*, G. Ligeti préconise des voix « éthérées » ou « ambiguës » pour le couple Amando – Amanda par exemple ; bien des rôles (si ce n'est tous) de Mozart demandent une « clarté » dans la voix (la célèbre « couleur vocale Mozart ») etc.

Enfin, nous remarquons qu'une tessiture seule n'est pas suffisante pour qualifier la voix d'un chanteur, c'est pourquoi on lui appose fréquemment une dénomination supplémentaire dont la liste des qualificatifs est longue mais représentative de notre propos : aigu, léger, lyrique, dramatique, héroïque, chantant, bouffe, -Verdi, -spinto, noble ou profond, -Martin, colorature, demi-carctère etc. Par exemple, les termes *colorature* et *dramatique* différencient les *soprani* de telle sorte que la première a une qualité de voix à prédominance « claire », « légère » (cf. rôle de *Lakmé* de L. Delibes avec « l'air des clochettes » ou celui de la « Reine de la Nuit » dans *Die Zauberflöte* de W. Mozart), et la seconde aura une voix plus « sombre » mais toujours de *soprano* (*Norma*, de G. Puccini). Dans le cas d'une voix de *contralto*, *colorature* désigne un timbre rarement « clair », moins éloigné de la *contralto dramatique* (cf. G. Léothaud). On retrouverait là, la dichotomie entre les différentes modalités de qualifications des qualités

vocales, que l'on a rencontrée plus haut (récapitulatif fin de p. 2) : à savoir que colorature ne désigne pas la même particularité suivant le chanteur considéré.

Il serait intéressant d'étudier plus profondément la question mais ce n'est pas le sujet principal de notre stage, qui néanmoins va dans le sens d'une compréhension des qualificatifs utilisés dans le domaine lyrique.

# Chapitre II : Chant Lyrique

## II.1) Éléments de Physiologie – Fonctionnement de l'Appareil Vocal :

Nous donnons ici simplement quelques rappels du fonctionnement de certaines parties de l'appareil phonatoire.

### **II.1.1) Source, Conduit vocal et Cavités de résonance**

La voix est souvent comparée à un modèle source-filtre, où la source est constituée de la soufflerie (poumons) et de l'excitateur (cordes vocales), et le filtre est composé du conduit vocal et de ses cavités.

Le conduit vocal est constitué du larynx, du pharynx, de la bouche et des fosses nasales. Ainsi, les principales cavités en jeu dans la résonance sont le pharynx, la cavité buccale et les cavités nasales. Les articulateurs (langue, lèvres, maxillaires, dents, palais, voile du palais, larynx) modifient la configuration du conduit vocal et ses propriétés acoustiques.

*a – la source – base du larynx*

L'air envoyé par les poumons ne produit en lui-même pas de son. Ce sont les cordes vocales qui, au passage de l'air, s'accolent et se décollent, produisant ainsi l'ouverture et la fermeture de la glotte de manière cyclique. C'est le larynx, constitué de muscles et de cartilages, qui est donc le lieu de la production vocale. Le son est dit « glottique » ou de « fourniture laryngée » et il est donc émis par l'évacuation des ondes aéroacoustiques. Les muscles servent à régler la fréquence fondamentale de ce son. Il existe quatre mécanismes de vibration des cordes vocales (fry, de poitrine, de tête et de sifflet ; mais nous ne rentrons pas dans le détail dans le cadre de ce stage), le son de la source est riche en harmoniques, et l'étude de la source ([14]) nous renseigne sur la façon dont se ferment les cordes vocales et comment le son est enrichi spectralement. C'est pourquoi, la glotte joue un rôle dans la qualité de la voix.

*b – principales cavités*

Le **pharynx** est la cavité verticale de l'arrière-gorge située entre le larynx et la bouche. Ses parois étant musculaires le pharynx peut considérablement varier de volume et de forme donc faire varier les fréquences propres de résonance de la cavité, jouant ainsi un rôle dans les sons de voix « poitrinée ». La cavité pharyngale gère aussi la communication avec les fosses nasales par abaissement ou relèvement du voile du palais. Elle permet ainsi un couplage du conduit vocal avec la **cavité nasale**. Celle-ci participera à la production de sons plus ou moins nasalisés. Cette participation dépend aussi du texte prononcé. Les consonnes nasales ([m], [n], /gn/ de agneau) sont souvent associées aux voyelles (notamment antérieures [i] et [e]) dans les exercices vocaux pour sentir la résonance ([19]).

Enfin, la **cavité buccale** est de loin la cavité la plus importante au niveau des variations volumiques et de forme, comprenant l'intérieur de la bouche (dont la disposition peut varier suivant la place de la langue) et le vestibule labial (espace entre dents et lèvres/joues). Elle va jouer le rôle principal dans l'allure du spectre en positionnant les formants vocaliques.

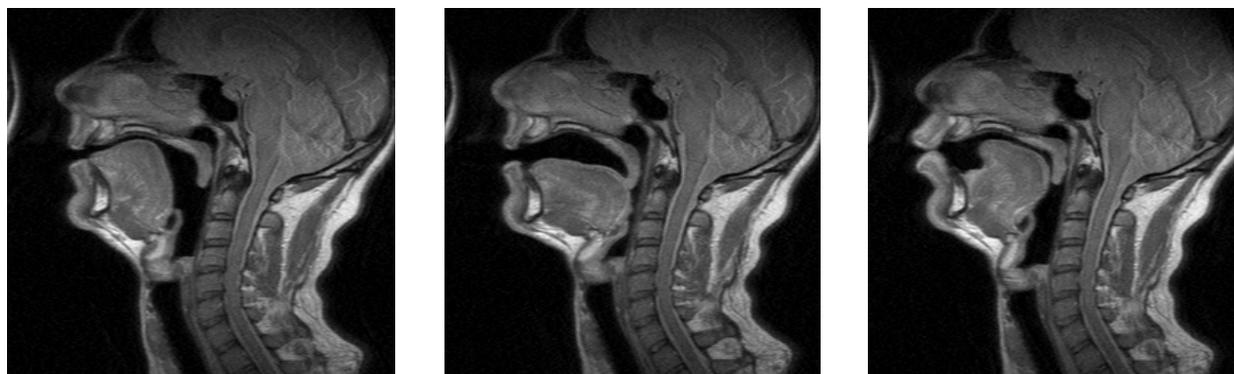
## II.1.2) Formants vocaliques – principe

### *a – principe*

L'action portée sur les cavités du conduit vocal implique une configuration de forme, de souplesse, de volume et de densité. Chaque configuration engendre à la fois une répartition de l'énergie acoustique et des résonances en fréquence différentes. Ainsi, vont exister dans le spectre des bandes de fréquences renforcées par rapport à d'autres : ce sont les formants. Les formants vocaliques sont donc propres à une configuration permettant de prononcer une voyelle de la langue. Les différences spectrales entre les voyelles chantées, de nature articuloire, s'identifient au niveau des deux premiers formants F1 et F2 (Luchsinger et Arnold, 1965 ; [1]). La place de ceux-ci permet de discriminer des voyelles reconnaissables (Kantner et West, 1968). Bien que la position exacte des formants soit propre à chaque voix, on les retrouve dans des zones bien déterminées, et ce déplacement de zone changera la perception que l'on a de la voyelle.

Dans les modèles sources-filtres ([8] Fant 1964) traditionnellement employés pour décrire le spectre d'un son vocal, on considère que l'atténuation du spectre de la source glottique est en  $-6$  dB/octave. L'action du filtre (conduit vocal) est neutre quand à la dynamique spectrale (pente nulle) et favorise l'émergence de zones spectrales appelées zones formants. La présence des lèvres impose une pente spectrale de  $-6$  dB/octave supplémentaire ( $-12$  dB/octave au total). S'il est vrai que cette modélisation donne des résultats concordant bien avec la réalité, cela n'est valable que pour celle de la parole. Nous savons que la voix chantée sort des limites du modèle ([14]), néanmoins il permet de comprendre qualitativement ce qu'est un formant. Il faudrait considérer les différentes interactions (plus non-linéarité) engendrées par le chant lyrique entre la source et les cavités pour modéliser la place des formants vocaliques dans le spectre.

### *b – évolution des formants vocaliques*



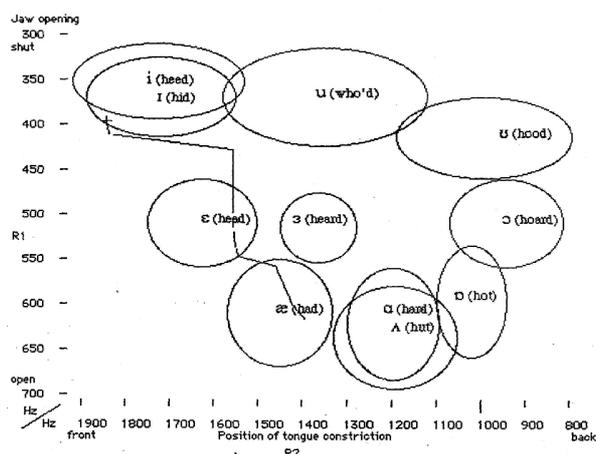
*vue du conduit vocal par IRM pour [i], [a], [y], P. Badin, ICP*

La protrusion des lèvres ou la position de la langue vont agir sur F1 et F2 jouant ainsi sur les voyelles [i] et [y].

Ils dépendent de beaucoup de paramètres dont on connaît quelques implications (Miller, Titze, in [6]). Si le larynx descend et que les lèvres se resserrent, on a un allongement du conduit, qui a pour effet de baisser les fréquences des formants. Inversement, quand les lèvres s'écartent, les formants montent ; le rôle de la mâchoire affecte le premier formant en l'augmentant plus celle-ci s'ouvre. Enfin au niveau des voyelles [i] et [y], la constriction de la langue dans la cavité buccale et la protrusion des lèvres atténuent F1 et augmentent F2, et inversement, pour les voyelles [a] et [o], la constriction de la langue au niveau du pharynx renforce F1 et diminue F2.

En règle générale le premier formant est affecté par le pharynx et le deuxième par la cavité buccale (et le changement de forme de la langue).

On peut observer une relation triangulaire entre les voyelles suivant 3 axes : [i] – [u], [i] – [a] et [u] – [a] matérialisant un déplacement voyelles ouvertes – voyelles fermées et un déplacement antérieur – postérieur (devant, derrière).



Zones où se trouvent formants F1 et F2 pour des voyelles anglaises, in Epps J. [7]

## II.2) Éléments relatifs à la technique vocale :

On postule que c'est par un apprentissage et un entraînement répété que le chanteur mémorise les positions de ses divers organes phonatoires relatives à un certain rendu perceptif. C'est sur une somme d'éléments divers, employés et contrôlés, que le chanteur va pouvoir s'appuyer pour affirmer sa voix. Les choix de stratégies dans le chant lyrique sont gouvernés par des canons esthétiques. Il en est ainsi du *vibrato*, du formant du chanteur (« *singing-formant* »), et de tous les contrôles au niveau physiologique (posture, respiration, articulation ...).

### II.2.1) Formant du chanteur<sup>2</sup>

Parmi les pics de résonance présents dans le spectre de la voix, il existe une région très sollicitée dans le chant lyrique. Elle se situe au-delà de la zone des deux premiers formants définissant les voyelles. Le chanteur lyrique (idem pour les acteurs « classiques »), à la voix cultivée, développe ce formant particulier dans une zone à peu près fixe pour toutes les voyelles et propre à chaque chanteur ([2]). Il s'agit de la bande de fréquence située entre 2000 et 4000 Hz, souvent centrée sur 2500-3000 Hz qui correspond à la zone de plus grande sensibilité de l'oreille (Fletcher). Le but est à la fois esthétique car il joue sur la qualité du timbre (cf. ci-après « brillance », « timbrage ») et pratique car il permet l'émergence de la voix du chanteur au dessus d'un orchestre par exemple (augmentation de la « portée » de la voix intéressante pour les grandes salles). Son effet peut être très important dans la qualité vocale suivant le rapport entre l'énergie maximale dans cette région et l'énergie totale rayonnée ([23]). En témoigne le résultat perceptif lorsqu'on filtre l'énergie dans ce formant ([3]).

### II.2.2) « Ouverture » et « couverture »

On envisage ici, pour éviter toute ambiguïté, voix « ouverte » et voix « couverte » comme la traduction des termes italiens *voce aperta* et *voce coperta* employés dans la technique lyrique italienne. On appelle traditionnellement la « couverture » (*copertura*) l'action contrôlée par le chanteur masculin qui monte dans l'aigu de sa voix de poitrine (mécanisme laryngé I). Elle a pour but d'aider le passage dans une région tendue de l'aigu et permet ainsi d'étendre celle-ci. Néanmoins, cette technique essentielle (voire obligatoire) dans l'aigu, peut aussi être utilisée dans le grave ou le médium, à des fins esthétiques ou pour l'interprétation, modifiant ainsi la qualité de la voix. Dans le chant lyrique, la « couverture » apparaît comme une technique incontournable et très difficile pour équilibrer le timbre de la voix sur toute la tessiture vocale. Sa réalisation se fonde sur l'*aggiustamento* ([19]), phénomène de modification vocalique. Sans cette modification,

2 Découverte, dans les années 1930, de W.T. Bartholomew qu'il nomme « *the Ring of the Voice* » (cf. JASA 23 (6)). Depuis, son origine est toujours controversée : laryngée (Husson ...) ou résonnante (Sundberg 1981 cf. Miller ...)

très importante dans l'aigu, la voix reste alors de nature « ouverte » et amène un timbre bien différent caractéristique d'un manque de technique. Le risque est l'apparition de disparités entre les registres que la « couverture » vise à homogénéiser.

Quelques études se sont intéressées à ce phénomène surtout pour expliquer ce « passage » dans l'aigu et pour la compréhension des différents registres de la voix chantée ([6]). Il existe une controverse sur le terme de registre mais elle ne nous concerne pas ici.

Nous livrons quelques résultats (*ibid.*) concernant l'analyse des formants dans les productions de sons « couverts » et « ouverts » :

- la « couverture » tend à abaisser les 2 premiers formants des voyelles ouvertes (/a/ ...), un peu moins pour les voyelles fermées
- au niveau harmonique, un son « ouvert » aura tendance à renforcer H2 (harmonique 2) tandis que pour un son « couvert » on observera un abaissement du formant F1 donc du rapport d'énergie H2/H1, et un déplacement de F2 dépendant de la voyelle : vers H3 (postérieures), H4 (antérieures ouvertes) et H5 (antérieures fermées)

**Conclusions** : au niveau perceptif, cet écart entre les sons « couverts » et « ouverts » va se ressentir sur la prononciation des voyelles. Nous observerons en partie III les réalités acoustiques de cette technique et dans quelle mesure cela affecte les voyelles et la perception de la qualité vocale.

### II.2.3) Prononciation – Triangle vocalique

#### *a – recherche et observation des formants vocaliques*

L'analyse spectrale du **fry**<sup>3</sup> permet d'estimer (marge d'erreur de 6%) la place des deux premiers formants des différentes voyelles. La technique proposée par Miller DG (in Chuberre [6] p. 98) consiste à faire chanter (ou parler) le sujet sur un texte en le faisant passer en voix fry sur chaque voyelle afin de garder la position des cavités comme elles sont et d'observer ainsi les formants générés par celle-ci. Le seul inconvénient est d'ordre pratique : le fry demande un entraînement pour certaines personnes peu habituées. Donc leur demander de basculer sans cesse sur une phrase de quelques secondes peut s'avérer délicat. De plus, cela est fatiguant pour le chanteur et complique un peu une situation qui devrait être la plus naturelle possible, si l'on veut obtenir une homogénéité sur la phrase musicale (cf. Partie B).

#### *b – action des formants sur la perception des voyelles et leur détermination*

Dans le chant lyrique, on conçoit la prononciation d'une manière différente que pour la parole (où l'on réfléchit peu à ce que l'on fait). Elle semble moins précise du fait que le chanteur doit prendre en compte des aspects techniques impliquant de nombreux paramètres. On a vu que les techniciens (Miller) préconisent l'*aggiustamento*, s'opérant dans les aigus. Des études acoustiques ont aussi montré depuis longtemps qu'il y avait des déplacements de formants caractéristiques de la réduction d'articulation des voyelles chantées dans les aigus par rapport aux graves : Bloothoof et Plomp observe des variations spectrales entre les voyelles en fonction de quelques qualités vocales et de la hauteur fondamentale ([1]). Eux-mêmes citent des recherches ayant observé la diminution de l'intelligibilité de voyelles chantées isolément avec l'augmentation de la fréquence fondamentale (Gottfried en 1980, Chew et Gottfried en 1981, Smith and Scott en 1979 et 1980).

---

3 Le fry est le mécanisme 0 des 4 mécanismes laryngés de la voix (au niveau de la vibration des cordes vocales) : il correspond à un relâchement de la tension et émet donc un son très grave (le fondamental est situé entre 20 et 80 Hz avec une moyenne de 50 Hz sans grande différence entre hommes et femmes ([21])). La vibration est presque harmonique (donc les fréquences sont assez rapprochées) et subit le filtrage des cavités. Le fry permet donc d'observer les formants sur un sonagramme.

# **Chapitre III: Diverses Approches de l'Etude de la Qualité Vocale**

Rappelons que nous cherchons à trouver des descripteurs de la qualité vocale au niveau verbal et acoustique, et à étudier les corrélations entre ces deux domaines.

Pour établir ces relations, nous nous proposons de parler des deux grandes stratégies d'approches existantes, qui ne sont pas nécessairement toujours opposées ou contradictoires, mais plutôt complémentaires. Nous donnons ici simplement les grandes lignes de la première approche pour se focaliser ensuite sur celle abordée par certains chercheurs du Laboratoire d'Acoustique Musicale lors d'études antérieures.

## **III.1) Approche Psychophysique :**

C'est l'approche classique, la plus couramment pratiquée. Ses fondements reposent sur la procédure selon laquelle un stimulus, paramétré dans les sciences physiques, déclenche une réaction, réponse ou jugement d'un sujet : soit, dans une conception contemporaine en sciences cognitives, un processus du bas (stimulus comme élément du monde qui est perçu) vers le haut (l'effet produit sur le sujet) dans le sens stimulus → récepteurs → cerveau → sensation ; la perception est alors considérée passive et non sélective, l'oreille est assimilée à un capteur et l'humain à un type d'instrument de mesure très particulier (cf. travaux de Mc Adams). Ainsi l'étude va s'orienter sur les paramètres physiques du stimulus en question (amplitude, fréquence, durée ...), et, en faisant varier ceux-ci, de manière analytique, l'expérimentateur obtient des valeurs de correspondances entre données perceptives (des sujets) et des paramètres physiques (« Lois de variations », [3]). Pour obtenir des résultats quantifiables et exploitables, le psychophysicien doit se servir de stimuli normalisés en laboratoire sur des dimensions a priori bien définies dans une description physique : rappelons ici les études menées sur les sons purs (Fletcher, Plomp, Stevens, Terhardt, ...) ou sons complexes (Espaces des timbres de Grey, 1977). Dans l'expérience de Grey, les stimuli sont pour la première fois des sons d'instruments réels, traités par l'ordinateur, dont on peut contrôler les paramètres physiques de manière indépendante (ensuite les études porteront sur des sons de synthèse pour un contrôle total) sur une seule hauteur commune à tous les instruments de musique considérés, de même durée et de même intensité. Toutes ces recherches ont apporté une grande connaissance du fonctionnement de l'appareil auditif (comme capteur) en créant des unités et normes de référence (dB SPL, sonie, phonie, tonie ...).

Concernant la qualité vocale, cette approche cherche de la même manière à observer les variations de chaque paramètre recensé comme ayant un rôle dans la perception de qualités différentes. Puis ces paramètres sont évalués indépendamment sur des exemples vocaux, ceux-ci étant soumis ensuite à une évaluation perceptive suivant une grille de questions fermées. Cette méthode peut être satisfaisante lorsqu'on connaît à l'avance (ou que l'on présuppose) les représentations mentales que l'auditeur s'est construit concernant la qualité sonore des objets évalués, et si l'on postule que l'éventuelle différence de ces représentations n'influencera pas la perception. Ce type de méthode cherche donc à établir des mesures subjectives correspondant à des mesures physiques en proposant des choix forcés.

Dans le cas de la voix, surtout chantée, nous n'avions jusqu'à très récemment pas assez de connaissances sur la représentation complexe que l'on s'en fait. C'est pourquoi, d'autres approches pouvant prendre à la fois des orientations différentes et complémentaires se sont mises en place. Elles tiennent compte justement de l'étude de ces représentations qui permettent de trouver les critères que l'on va étudier. Elles ont ainsi amené quelques réponses qui rendent désormais possible l'utilisation de l'approche psychophysique pour avancer en parallèle.

## **III.2) Approche Psycholinguistique :**

Par opposition à la précédente, celle-ci peut être qualifiée de « top-down » (*ibid.*).

Elle présuppose un schéma descendant des représentations préalables à la saisie du stimulus, et se caractérise par une conception active et sélective de la perception. Celle-ci dépend des auditeurs, de leur mémoire, de leur apprentissage et de leur fonctionnement cognitif, de langage et de leurs connaissances et expertises. Cette approche considère en outre que l'individu, au centre de la perception, a construit ses représentations de manière multisensorielle (donc complexe), et qu'ainsi les simplifications des situations monosensorielles de laboratoire suscitent des traitements spécifiques différents des traitements normaux et spontanés. De plus, sur la base de cette expérience sensible, l'individu donne sens à ce qu'il perçoit. Pour accéder à cette signification individuelle, la langue constitue un accès partagé par les mêmes locuteurs. L'approche psycholinguistique semble alors appropriée pour connaître ces représentations à la fois individuelles et partagées. C'est au travers de la description verbale de l'auditeur que l'on peut comprendre, par l'analyse linguistique de son propos, ce qui est important et pertinent dans ce qu'il écoute et comment s'organisent ses représentations mentales (étude linguistique de la verbalisation).

Au Laboratoire d'Acoustique Musicale, cette approche a été adoptée depuis quelques années et bénéficie d'une collaboration entre sciences cognitives et sciences physiques. Cette méthodologie prend en compte l'avis d'experts, dont les conceptions amènent parfois à suivre d'autres pistes et critères que les physiciens n'auraient pas forcément conçus. Tout ceci dans le but de rechercher les paramètres acoustiques les plus saillants dans la perception des phénomènes.

Cette approche a déjà donné des résultats lors de recherches portant sur la qualité sonore dans des environnements bruités (urbains) et en basses fréquences ([12], [18]) et très récemment lors d'une étude de la qualité vocale ([9-11]) à la suite de laquelle notre travail s'inscrit. On en rappellera donc ici les principales grandes lignes.

### **III.2.1) Perception et catégorisation de la qualité sonore (partie rédigée avec D. Dubois)**

La catégorisation est un autre cadre conceptuel emprunté aux sciences cognitives et qui permet de rendre compte de l'organisation en mémoire des représentations. Une des tâches souvent utilisée pour accéder ainsi aux catégories cognitives consiste à demander aux sujets de former des groupes de stimuli (et éventuellement de les commenter). Des classes d'objets sont ainsi identifiées de même que, entre elles, une éventuelle structure hiérarchique, et à l'intérieur de chacune un gradient de typicalité (faisant que certains exemplaires sont plus représentatifs que d'autres) (Rosch, 1978 ; Dubois, 1991).

Cette méthode a déjà été utilisée dans l'étude de l'environnement sonore urbain sera utilisée ici pour formuler des hypothèses sur les catégories pertinentes pour différentes qualités de la voix chantée.

La catégorisation permet de constituer des groupes par similarité entre les objets composant les catégories.

### **Conclusion :**

Toutes ces méthodes et le processus engagé soulignent et témoignent de l'originalité des travaux menés, qui sur le long terme apporteront des résultats importants dans la compréhension des qualités des voix, des mécanismes mis en jeu et de la perception de ceux-ci.

### III.2.2) Une étude antérieure de la qualité vocale

#### *a – déroulement de la précédente étude et principaux résultats*

Cette étude a consisté tout d'abord en la recherche de l'existence de termes consensuels entre les individus, utilisés pour décrire les diverses qualités vocales d'un corpus d'exemples vocaux réalisés par des chanteurs. Le but était d'obtenir des résultats signifiants pour un grand nombre de sujets afin de se baser sur une terminologie consensuelle pour ce groupe de chanteurs/locuteurs. Les analyses ont montré la diversité du langage mais aussi l'ambiguïté ou la polysémie de certains termes qui varient entre les personnes, en l'occurrence les chanteurs.

Des experts (11 professeurs de chant) ont participé à de séances de tests d'écoute afin de commenter les voix lyriques précédemment produites. Dans un premier temps, les experts ont commenté les extraits en répondant à des questions ouvertes, ce qui a permis de caractériser les qualités vocales. Dans un deuxième temps, ils ont évalué les différentes émissions vocales sur des échelles bipolaires à 5 points fondées sur des critères identifiés par une étude de psycholinguistique.

#### **Conclusion :**

cette analyse a permis de mieux comprendre comment les auditeurs perçoivent et conçoivent les qualités vocales, et de dégager les termes pertinents et consensuels constituant le point de départ de notre exploitation de la qualité vocale.

#### *b – identification de corrélations verbales entre qualités*

Différents points très importants sont à relever, à commencer par les multiples natures des corrélations, rendant problématique les interprétations sur une seule « dimension » :

- corrélation linguistique : au niveau verbal entre les termes
- corrélation statistique entre échelles bipolaires au niveau des termes qui obtiennent des résultats liés

Parmi les principaux résultats, nous retiendrons pour notre étude tout d'abord le consensus important sur les termes « timbré/détimbré », « brillant/sourd », « clair/sombre » et présence ou absence de « vibrato ». On observe une relation entre les qualités « clair/sombre », les notions techniques vues au chapitre II) d'« ouverture/couverture » et la prononciation des voyelles de manière « antérieure/postérieure ». Les évaluations établissent un lien entre l'« air sur la voix », la « brillance », le « timbrage » et le « vibrato » remettant parfois en question certaines associations concernant des qualificatifs couplés qui avaient été considérés comme contraires ; c'est le cas de « timbré » et « détimbré » où le premier semble porter sur la « brillance » et le « vibrato », tandis que le second plutôt sur le caractère « sourd » et « soufflé » de la voix. Les corrélations se font aussi pour désigner des voix déséquilibrées par les caractères « sombre », « sourd » ou avec l'apparition de « souffle » ou de « chevrottement ».

Enfin d'autres notions, par contre, paraissent moins partagées, comme le « trémolo », « l'air sur la voix » (voix soufflée) ou le « chevrottement », ou d'autres semblent moins faciles à évaluer dans les exemples sonores, bien que reconnues de manière consensuelle : c'est le cas de la « nasalité ».

### III.2.3) Résultats de l'étude acoustique

#### *a – procédure analytique*

Comme nous l'avons vu, un grand nombre d'études cherche à résumer une qualité vocale par la combinaison pondérée de tous les paramètres « dimensionnels » du son. Pourtant tous les paramètres ne jouent pas le même rôle et ne sont pas aussi saillants pour notre perception les uns que les autres. L'approche décrite plus haut cherche au contraire à mettre en évidence les critères les plus pertinents dans le jeu de leurs interactions. Toujours dans l'esprit de continuité entre

l'étude verbale et l'étude acoustique, les premières analyses ont été entreprises sur ce qui avait été identifié comme ayant un sens dans l'étude perceptive pour voir s'il en était de même d'un point de vue acoustique.

*b – identification de corrélations entre paramètres acoustiques et qualités vocales*

Les corrélations peuvent, ici aussi, être de plusieurs types :

- corrélation, entre critères verbaux, liée au partage de propriétés physiques, entre qualités vocales : un paramètre acoustique se trouve être un facteur principal intervenant dans plusieurs qualités
- la recherche de corrélats entre l'étude acoustique et l'étude verbale (i.e. si un caractère augmente sensiblement le paramètre corrélé doit augmenter également de manière sensible)

Nous retiendrons les principaux résultats observés. ([10])

Au niveau des qualités de voix « **Brillante et sourde** », la différence entre les deux semble se situer principalement sur l'émergence du formant du chanteur (2000-4000 Hz) par rapport aux formants vocaliques. La « brillance » se caractérise par une plus forte participation de cette bande de fréquences, contrairement à la « surdité ». La corrélation se situerait donc avec la différence d'amplitudes entre zone 0-2000 Hz et 2000-4000 Hz (participation de l'aigu du spectre).

Le timbrage est proche de la notion de brillance : on observe sur les qualités « **Timbrée / détimbrée** » une modification du spectre entier. L'observation d'une atténuation dans le spectre entre 2000-4000 Hz témoigne d'une corrélation possible entre « détimbré » et « sourd », cependant l'atténuation (en niveaux réels) se remarque aussi au-delà de 4000 Hz pour « détimbré ». Au niveau de la répartition spectrale en pourcentages, un changement considérable apparaît entre une voix « timbrée » et une voix « détimbrée » : les premiers formants sont renforcés au détriment du formant du chanteur pour une voix « détimbrée », alors que la zone 2000-4000 Hz est assez importante en général pour les voix « timbrées », et pour la région supérieure à 4000 Hz, une voix « détimbrée » est très faible par rapport aux voix « timbrées ». Le caractère « détimbré » affirme le fondamental mais au détriment du reste.

L'étude des qualités « **Claire et sombre** » par l'analyse acoustique de deux voyelles (/a/ et /i/) confirme le discours institué dans les leçons de chant ainsi que les remarques des experts ([19]) : on observe le déplacement des maximums en amplitude sur les harmoniques correspondant aux deux premières régions formantiques témoignant du déplacement des formants vocaliques et indiquant une modification dans l'ouverture et la couverture des sons de voyelles ([6]).

**Conclusion :**

Les bases et problématiques de notre stage sont ainsi maintenant définies. Nous avons pu relever les critères importants à mettre en évidence dans les analyses.

Cependant, il faut aussi prendre garde au fait que cette étude n'avait pas été réalisée sur des exemples conçus spécialement à cet effet (c'est un point essentiel de notre stage). Certaines conclusions sont donc à relativiser, du fait notamment que chaque chanteur peut adopter une stratégie différente pour réaliser une même qualité. Si certaines conclusions ont été établies, ce qui est le cas pour les caractères « timbré / détimbré », il reste encore à approfondir les recherches quant à certaines qualités qui n'ont pas donné de résultats aussi probants ; c'est le cas de « clair / sombre » dans le lien avec « l'ouverture » et la « couverture » (cf. Partie B (sur ce qui a motivé nos choix) et C (résultats)).

De plus, la notion d'opposition de certaines qualités vocales, réelle dans l'étude verbale, est peut-être à relativiser au niveau de la corrélation acoustique (et perceptive).

**PARTIE B**

**PREMIERE ETAPE**

**MATERIEL**

**ET**

**EXPERIMENTATION**

**DU STAGE**

# **Chapitre IV : Protocole et Base de Données**

## **IV.1) Objectifs du stage :**

Le stage s'inscrivant dans la continuité d'une étude débutée au sein du laboratoire par N. Henrich, nous sommes partis de la base de données d'exemples vocaux de 18 chanteurs de toutes tessitures enregistrée au cours de sa thèse. Un des objectifs du stage était d'étudier d'un point de vue acoustique certaines qualités vocales retenues pour leur pertinence à l'issue de l'étude linguistique réalisée en 2003 par M. Garnier à partir d'extraits vocaux de 3 chanteurs masculins de la base de données citée plus haut. Cette démarche donne toute l'originalité à notre étude sur la qualité de la voix car elle s'appuie sur un lexique utilisé par les professeurs de chant. Les termes ont été sélectionnés pour leur pertinence et le consensus qui règne entre eux.

Dans un second temps, il était prévu de réaliser des tests perceptifs de catégorisation libre à partir des exemples sonores.

Nous allons voir comment nous avons abordé le sujet dans la partie expérimentale de notre stage.

## **IV.2) Utilisation de l'Ancienne Base de Données :**

### **IV.2.1) Test exploratoire de catégorisation libre**

En début de stage, la première orientation était de voir si nous pouvions utiliser le corpus d'exemples de chanteurs du laboratoire pour continuer l'étude de la qualité vocale et réaliser ainsi un test de catégorisation. La catégorisation libre d'extraits vocaux devrait nous permettre d'accéder aux critères perceptifs pertinents pour l'auditeur pour discriminer ou regrouper des qualités vocales différentes. Cela permet également de mieux comprendre la façon dont il hiérarchise ces critères et se construit des représentations mentales de la qualité vocale. Par exemple, le vibrato et l'intensité sont ils des critères pertinents de catégorisation de la qualité vocale? Existe t'il un critère qui intervient prioritairement ou d'entrée de jeu dans la discrimination de qualités vocales ?

Afin d'examiner la validité de cette base de données pour effectuer un test de catégorisation , nous avons proposé à 3 sujets non experts du chant lyrique mais musiciens (bénévoles), d'effectuer cette tâche de catégorisation en pré-test avant de savoir s'il pouvait être mené à plus grande échelle.

Pour cela, nous avons réutilisé les exemples vocaux des 3 chanteurs masculins sur lesquels se basait l'étude antérieure (24 voix au total) :

NK-normale piano	BC-normale piano	AZ-normale piano
NK-normale mezzo	BC-normale mezzo	AZ-normale piano 2
NK-normale forte	BC-normale forte	AZ-normale mezzo
NK-très accrochée	BC-derrière et gros	AZ-normale mezzo 2
NK-chanson	BC-nasale	AZ-normale forte
NK-claire	BC-serrée	AZ-bouche en hauteur
NK-nasale	BC-soufflée, avec de l'air	AZ-voix de fausset
NK-à la russe	BC-variété	AZ-voix de variété

Toutes les voix ont été mises strictement au même diapason avec le logiciel AudioSculpt car il est impératif d'avoir des exemples sonores qui soient tous au même diapason. En effet, la hauteur affecte de façon non négligeable notre perception de la qualité : un exemple plus bas qu'un autre est jugé souvent plus sombre ou moins dynamique. A l'inverse, un exemple légèrement plus haut paraît consensuellement plus « brillant » ou plus « clair » (Etude des flûtes, M. Castellengo). Ce phénomène perceptif apparaît dès  $\frac{1}{4}$  de  $\frac{1}{2}$  ton (1/20 cents). Or les meilleurs algorithmes de

« *pitch-shift* » sont efficaces jusqu'au ½ ton grand maximum. La transposition de hauteur dans le cas de la parole ou du chant est encore plus problématique que dans le cas de sons instrumentaux du fait de la présence des formants vocaliques (également transposés, induisant une déformation perceptive de la voyelle produite) ainsi que de l'importance perceptive du vibrato sur la qualité vocale (un vibrato ralenti ou accéléré par l'algorithme changera beaucoup la qualité vocale).

Les sons étaient présentés sur une interface graphique (P. Gaillard). Les sujets ont été placés dans de bonnes conditions d'écoute, dans un studio isolé, avec restitution sur enceintes Tanoy via une carte son 24 bits externe reliée à un ordinateur portable silencieux.

La consigne pour le test était : « *Regroupez les sons que vous entendez en catégories, auxquelles vous donnerez un commentaire pour les désigner. Ecoutez les sons autant de fois que vous le désirez. Faire autant de groupes composés d'autant de sons que vous voulez.* »

S'en suivait un bref entretien pour préciser avec le sujet sa stratégie d'écoute.

#### **IV.2.2) Résultats et conclusions**

Ce test, d'envergure volontairement restreinte, a apporté certains résultats intéressants et des conclusions concernant la base de données.

Tout d'abord, les sujets n'avaient pas toujours regroupé les voix « normales » des 3 chanteurs entre elles.

Certaines qualités ont été reconnues par rapport à ce qu'avait fait le chanteur. C'est le cas de « voix claire ».

Le nombre moyen de groupes formés par sujet est de 8. De plus, sur les 24 groupes réalisés, nous notons une prédominance des groupes de 1 à 3 sons (plus de la moitié) : 4 de 1 sons, 7 de 2 sons et 6 de 3 sons. Seulement 7 groupes dépassent 3 sons : 2 de 4 sons, 2 de 5 sons et 3 de 6 sons. Ceci peut dénoter une trop grande variété d'exemples sonores non homogènes, rendant difficile la catégorisation.

Une autre remarque importante concerne les groupes composés exclusivement d'exemples d'un même chanteur. Ils représentent la moitié des groupes soit 12. Ils se répartissent de la manière suivante : 7 groupes sur 9 pour le premier sujet, 2 groupes sur 7 pour le deuxième et 3 sur 8 pour le troisième, impliquant un regroupement par chanteur.

Ensuite les quelques commentaires oraux viennent confirmer certaines remarques :

- D'une part la grande diversité des voix, souvent constituées d'exemples caricaturaux qui sont alors reconnaissables instantanément et qui sortent un peu du cadre de la voix lyrique. Les nombreuses occurrences de « caricatural » (4 fois), aussi bien pour catégoriser un groupe entier que comme commentaire général, celles de « trop » (11 fois) souvent associées à une qualité bien identifiée telle que « vibrato » ou « forcée », les jugements de valeur « agréable », « pas agréable », « pas naturel » ou enfin la référence à un genre bien caractéristique tel que « voix opéra » voire à d'autres types de chants tels que « un peu type variété », ont montré que ces exemples s'écartaient du cadre de l'étude. Il est important d'avoir une ligne directrice, fixée sur des qualités vocales bien déterminées et que l'exagération amène généralement l'exclusion d'un exemple par rapport aux autres créant un groupe de sons « évidents », ou amenant un jugement de valeur.
- D'autre part, le fait que les sujets catégorisent les voix en essayant de regrouper les chanteurs qu'ils pensent être identiques ensemble. On a alors affaire à une catégorisation du type reconnaissance plutôt que de nature qualitative, il est donc difficile de faire ressortir, avec les exemples proposés, une qualification de la qualité des voix qui soit commune ou propre à l'ensemble constitué. On voit apparaître alors une confusion dans la stratégie de classement : soit les sujets regroupent les voix par chanteurs, soit ils font abstraction des chanteurs et se focalisent sur la qualité de la voix mais cela est plus difficile.
- Les quelques exemples regroupés dans des catégories identiques chez tous les sujets ne sont

en général pas qualifiés de la même manière : 1 son « nasal » et 1 « bien accroché » ont été regroupés comme « nasillard, agressif » pour un sujet, « bien équilibré, expressif, vibrato bien adapté » pour le deuxième et « voix opéra, assez colorée dans les hauts médiums ». Cette remarque rejoint aussi ce qu'on avait dit dans la problématique Chapitre I : la question de référence pour juger d'une qualité vocale. Ici, comme les sujets n'ont pas toujours pu déceler de voix « normale » selon leur référence, ils ont eu du mal à identifier le processus des modifications des chanteurs. Ils ont donc remarqué les exemples les plus caricaturaux.

- Un dernier point concerne le « r roulé » qui a dérangé l'appréciation des voix, et qui n'était pas toujours prononcé de la même manière par les chanteurs (cf. choix de la mélodie).

### **Conclusion :**

Nous avons conclu que cette base de données n'était pas adaptée à un test de catégorisation de grande ampleur, d'où de la nécessité de refaire une base de données. Celle-ci a été motivée par :

- le besoin d'établir un protocole ciblé sur la qualité vocale et surtout une liste de qualités à imposer aux chanteurs pour avoir une homogénéité
- ne pas tomber dans le caricatural pour rester dans le cadre du chant lyrique
- la nécessité de contrôler la justesse des chanteurs sur l'attaque, car la perception se trouve changée pour quelques sons qui ont été remontés de plus de 100 cents
- nous avons compris qu'il fallait avoir assez d'exemples pour un chanteur afin de pouvoir réaliser un test par chanteur et non plus tout chanteur confondu (sinon l'écoute « causale » vient perturber la catégorisation)

De plus, nous avons aussi déduit de ce test que le vibrato était soit prédominant dans la perception, soit que l'auditeur essayait de juger la voix sans ce paramètre car celui-ci changeait tout le temps.

Ce test nous a donc inciter à concevoir un nouveau protocole destiné et pensé spécialement pour l'étude des qualités de la voix chantée en constituant le nouveau corpus d'exemples.

## **IV.3) Elaboration d'un Protocole pour notre Etude :**

### **IV.3.1) Protocole**

#### *a – objectif du protocole*

Le but de notre protocole n'est pas de repartir de zéro mais de respecter les observations tirées de l'étude antérieure sur l'ancienne base de données. Nous avons donc sélectionné les qualités vocales les plus pertinentes et les plus consensuelles de l'étude de M. Garnier. Nous demandons aux chanteurs de produire une liste (cf. liste p.24) de 16 qualités vocales réparties sur 31 séquences. Dans le but de chercher des corrélats acoustiques à ces qualités, nous avons trouvé intéressant de demander aux chanteurs de produire plusieurs graduations d'une même qualité vocale. De plus, nous ne cherchons pas à lui imposer notre propre conception des qualités demandées. ***Il est ainsi le seul juge*** de ce qu'il fait, c'est pourquoi nous trouvons pertinent de lui donner la possibilité de se réécouter et de recommencer l'exemple jusqu'à être en adéquation entre l'intention de production de la qualité vocale et la perception de cette qualité produite. L'objectif de l'enregistrement de cette base de données est l'analyse acoustique des diverses qualités vocales pour un ensemble de chanteurs masculins<sup>4</sup>, la constitution d'exemples sonores en vue d'un prochain test de catégorisation plus ciblé et par chanteur.

---

<sup>4</sup> Nous n'avons pris que des hommes, comme dans l'ancienne base de données, pour faciliter l'étude des formants. En effet, chez les femmes, le F1 est souvent en dessous de la fréquence fondamentale. De plus pour réaliser des tests de catégorisation, il est difficile de comparer à la fois des voix d'hommes et des voix de femmes.

### *b – choix des qualités vocales*

Parmi les qualités vocales dégagées en (A-III) et les études antérieures, nous remarquons que certaines jouent sur la prononciation du texte, notamment au niveau de l'articulation des voyelles. Notre choix va se porter ainsi sur « claire », « sombre », « ouvert » et « couvert » que l'on comparera aux émissions « antérieure » et « postérieure » pour voir les évolutions des voyelles. On cherche aussi à valider la pertinence de certaines observations sur les qualités évaluées d'après la précédente base de données : « brillante / sourde », « timbrée / détimbrée », « avec / sans souffle sur la voix ». La qualité vocale « nasale » a été reprise de l'étude antérieure bien que son évaluation n'ait pas été consensuelle pour toutes les personnes. Ensuite d'autres qualités ont été ajoutées car elles étaient ressorties de la verbalisation des experts lors de l'étude précédente : c'est le cas de « métallique », de « bâillée » et de « en poitrinant ».

### *c – déroulement du protocole*

Pour que les chanteurs ne cherchent pas consciemment à jouer sur leur prononciation des voyelles pour modifier les qualités vocales, il est préférable que ils ne soient pas au courant des recherches que l'on veut effectuer. Pour cela, on ne parlera pas de qualité vocale directement, ni des voyelles.

De plus, l'étude antérieure a exprimé certaines limites quant à la constitution de couples de qualités vocales opposées, ces antonymies n'étant pas toujours valides. Le protocole prévoit donc que les qualités vocales, qui jusque-là fonctionnaient par couples (notamment sur des échelles bipolaires lors des évaluations (A-III)) soient dissociées. On détermine ainsi un ordre dans lequel les qualités vont être demandées en prenant soin de séparer le plus possible les qualificatifs antagonistes comme « timbré/détimbré », « clair/sombre » ... (cf. liste p. suivante).

Nous cherchons au maximum à accéder aux conceptions spontanées des chanteurs sur les qualités vocales. C'est pourquoi, nous n'avons pas envoyé le protocole et la mélodie à l'avance au chanteur pour éviter qu'il ne demande des avis extérieurs ou ne perde en naturel à force de réflexion. Nous demandons en premier lieu une émission « normale » sur les 3 niveaux d'intensité (moyenne, ff et pp) avoir la référence du timbre non modifié du chanteur. Ensuite chaque qualité est demandée une par une en commençant par une modification légère (« un peu ») suivie d'une modification plus marquée (« très »). Dans tous les cas, les qualités vocales ont été réalisées sans trop d'exagération pour rester dans le cadre d'un emploi au sens lyrique et pour éviter de tomber dans l'écueil du caricatural, point faible de l'ancienne base de données.

Après chaque qualité vocale, on demande au chanteur de donner un commentaire concis sur la stratégie qu'il a adoptée pour réaliser l'émission modifiée. Tous les commentaires constituent ainsi une seconde base de données pour une prochaine étude linguistique. Elle permettra d'une part, d'analyser à nouveau une verbalisation cette fois ciblée sur la qualité vocale pour obtenir de nouvelles pistes quant aux relations sémantiques entre les qualités vocales (séparées et non plus couplées) et aux significations des termes pour les chanteurs ; d'autre part, d'accéder à ce que font, ou pensent faire les chanteurs pour rendre la qualité perceptible. Les chanteurs s'expriment librement sans être influencés dans leurs propos.

Enfin nous avons pris soin de ne placer qu'à la fin les séquences qui faisaient clairement référence à la prononciation des voyelles, à savoir l'antériorisation et la postériorisation des voyelles ainsi que la voix fry pour déterminer les formants.

**DEROULEMENT DE L'ENREGISTREMENT**  
**Protocole témoin non montré au sujet**

A. Emissions vocales normales :

- 1) Emission en voix **normale** et **intensité normale**
- 2) Emission en voix **normale** et **intensité ff**
- 3) Emission en voix **normale** et **intensité pp**

B. Emissions modifiant la qualité vocale :

- 4) Emission un peu **claire**
- 5) Emission très claire (brefs commentaires)
  
- 6) Emission un peu **nasale**
- 7) Emission très nasale (commentaires)
  
- 8) Emission un peu **sombre**
- 9) Emission très sombre (commentaires)
  
- 10) Emission un peu **métallique**
- 11) Emission très métallique (commentaires)
  
- 12) Emission avec un peu de **souffle**
- 13) Emission avec beaucoup de souffle (commentaires)
  
- 14) Emission un peu **brillante**
- 15) Emission très brillante (commentaires)
  
- 16) Emission plus **timbrée** que voix normale
- 17) Emission très timbrée (commentaires)
  
- 18) Emission plus **bâillée** que voix normale
- 19) Emission au maximum bâillée (commentaires)
  
- 20) Emission un peu **sourde**
- 21) Emission très sourde (commentaires)
  
- 22) Emission un peu **détimbrée**
- 23) Emission très détimbrée (commentaires)
  
- 24) Emission plus **ouverte**
- 25) Emission plus **couverte** (commentaires)
  
- 26) Emission en **poitrinant** un peu plus
- 27) Emission en poitrinant beaucoup (commentaires)
- 28) Emission en coupant les résonances de poitrine (commentaires)
  
- 29) Emission des **voyelles** de façon plus **antérieure**
- 30) Emission des **voyelles** de façon plus **postérieure** (commentaires)

C. Détermination des formants vocaliques :

- 31) Emission voyelles en voix **fry** : " i , o(le) , e , a , o , u , ou , é , â , è "

D. Conclusions – avis – commentaires



#### *a – Carte et logiciel d'acquisition*

Le modèle de carte utilisé est le *Kay CSL 4400* avec son logiciel relié à un PC *Pentium III*. Il nous a servi à faire une acquisition des exemples pour leur restitution instantanée destinée aux chanteurs. Pour perdre le moins de temps possible nous avons réalisé une *macro* pour automatiser l'affichage, l'enregistrement et les raccourcis clavier.

#### *b – Microphones*

Ont été utilisés :

Un microphone de pression Brüel et Kjaer de type 2669 avec pré-ampli Nexus. Il sert pour réaliser les mesures du niveau SPL à 30 cm de la bouche du chanteur.

Un couple de cardioïdes Neumann pour la stéréo avec alimentation fantôme. Il sert à effectuer une prise de son stéréo de qualité destinée à être écoutée ensuite par des auditeurs lors de tests perceptifs.

Un dB-mètre a été utilisé pour mesurer les intensités maximales au niveau des micros (soit à 30 cm de la bouche des chanteurs), de façon à calibrer les enregistrements.

#### *c – Electroglottographie*

Nous avons utilisé un electroglottographe de Glottal Enterprises Inc.

Un contrôle sur l'oscilloscope a été effectué pour vérifier l'absence de bruit dû à un mauvais positionnement des électrodes et voir si la forme d'onde était satisfaisante. Notamment lors de passages pianissimo, on peut contrôler par ce signal que le chanteur reste bien en mécanisme laryngé 1 ( et qu'il ne passe pas en mécanisme 2 (*falsetto*)).

#### *d – Support D.A.T.*

Pour enregistrer les signaux du micro SPL et de l'EKG : *PORTA DAT PDR 1000*

Pour la stéréo : *TASCAM DA 40*

Cassettes DAT SONY (95, 65, 35 minutes)

#### *e – Numérisation des DAT*

Un travail assez fastidieux a été ensuite réalisé pour découper les exemples et séparer les commentaires. L'acquisition des DAT a été faite sur Mac avec un carte son à entrée optique.

Puis nous nous sommes servis essentiellement des logiciels Pro Tools et Adobe Audition pour le montage des exemples.

### **IV.4.2) Chanteurs**

Les chanteurs qui se sont portés volontaires pour effectuer les enregistrements sont professionnels (ou semi) âgés de 22 à 51 ans : **quatre barytons** et **un ténor** représentés dans les exemples par **Bar1 à 3** (le 4<sup>ème</sup> n'étant pas dans nos analyses) et **T1**.

Un bon accueil est important pour placer le chanteur dans de bonnes conditions d'enregistrement. D'autant que les 31 séquences (sans compter les reprises, les réécoutes et les commentaires) demandaient beaucoup d'efforts de concentration, de technique et de réflexion dans un temps assez court sur des notions parfois délicates. De plus le chanteur devait rester debout pendant environ deux heures à une distance d'environ 30 centimètres des micros avec les électrodes de l'électroglottographe autour de son cou (ce qui est assez pénible).

**PARTIE C**

**ANALYSES**

**ACOUSTIQUES**

**ET**

**RESULTATS**

# **Chapitre V : Analyses acoustiques des voyelles**

## **déplacement des formants vocaliques suivant les qualités vocales**

N.B. : Règle d'écriture pour les voyelles : Comme nous n'avons pas tous les caractères phonétiques dans Matlab et sur notre environnement pour rédiger, nous utiliserons dans nos figures et dans le texte qui suivent, les voyelles écrites en français de la manière suivante : /i/ de « il », /u/ de « jusqu'à », /é/ de « oublier », /a/ de « jusqu'à », /o/ de « haut » (« o fermé »), /o(le)/ pour le « o ouvert » du mot « vole » et /ou/ de « oublier ».

Rappelons aussi que nous avons pensé le texte et la mélodie afin d'obtenir deux groupes de voyelles sur des hauteurs identiques pour pouvoir mieux les comparer. Ainsi, les voyelles antérieures /i/, /u/ sont sur un sol 2, et le /é/ sur un fa 2 ; les voyelles postérieures /ou/, /o/, /o(le)/ sont sur un la 2 et le /a/ sur un si bémol 2.

D'après les résultats de l'étude antérieure et les commentaires des chanteurs sur leur production, la prononciation des voyelles semble jouer un rôle important sur les qualités « claire », « sombre », « antérieure », « postérieure », « ouverte » et « couverte ». C'est pourquoi nous avons concentré les analyses acoustiques des voyelles sur ces qualités, en les comparant à la voix « normale » de chaque chanteur. Les qualités « ouvertes » et « couvertes » ayant été rajoutées au protocole en cours de stage, ces qualités n'ont pas été enregistrées pour les chanteurs Bar1 et T1.

Avant de débiter le dépouillement des analyses, nous avons une remarque préliminaire qui peut sembler anodine mais qui est à la base de l'étude. Elle concerne, la perception des différentes voyelles. Une tendance générale s'observe si l'on écoute les voyelles séparément. On perçoit une grande variation de prononciation et d'articulation suivant les qualités vocales. Pourtant, si l'on n'écoute la phrase que dans son ensemble, on ne perçoit pas autant la variabilité de certaines voyelles. Dans le cas des voyelles séparées, on a au contraire toute une gamme de voyelles qui apparaissent où parfois la voyelle d'origine est presque changée en une autre, et d'autres fois, elle devient inclassable par rapport à notre langue. C'est peut-être là une des richesses du chant lyrique par rapport à la voix parlée ?

### **V.1) Méthode de mesure :**

- Pour analyser le fry, nous avons utilisé un algorithme de détection de formants programmé sous Matlab par N. Henrich. Nous avons choisi d'estimer la position des deux premiers formants vocaliques à l'aide d'une méthode LPC, tout en étant conscient des limites de cette approche dans le cas de la voix chantée. En effet, comme nous l'avons rappelé en II.1.2), une modélisation de l'appareil vocal par une source et un filtre indépendants est assez valide dans le cas de la voix parlée mais ne tient pas compte d'interactions entre source et filtre importantes dans le cas de la voix chantée. Une étude est en cours actuellement au L.A.M. pour comparer la précision de la mesure des résonances acoustiques du conduit vocal avec d'autres méthodes (LPC, DAP, Coefficients cepstraux, ...) pour estimer les formants vocaliques dans le chant lyrique. Notre choix de la méthode LPC a été motivé par notre intention d'observer les déplacements de ces formants entre la qualité « normale » du chanteur et une qualité vocale modifiée, ou encore entre deux qualités considérées comme opposées. Nous ne considérons pas l'estimation des formants par LPC comme exacte mais cette méthode a le mérite de nous permettre d'estimer assez valablement le déplacement de ces formants en relatif.

- Représentation des formants vocaliques : nous avons adopté la norme utilisée dans de nombreuses publications internationales. Elle considère la représentation graphique des formants comme une interprétation des mouvements de la bouche : à savoir, une ouverture vers le bas (resp. une fermeture vers le haut) donc les premiers formants F1 croissants sont sur l'axe des ordonnées dirigés vers le bas ; et l'antériorité vers la gauche, caractérisant la cavité buccale la plus petite (resp. la postériorité vers la droite pour une cavité buccale plus grande) donc les deuxièmes formants F2 croissent vers la gauche sur l'axe des abscisses.

## **V.2) Comparaison de la voix fry avec la voix normale :** (cf. 4 figures des 2 pages suivantes)

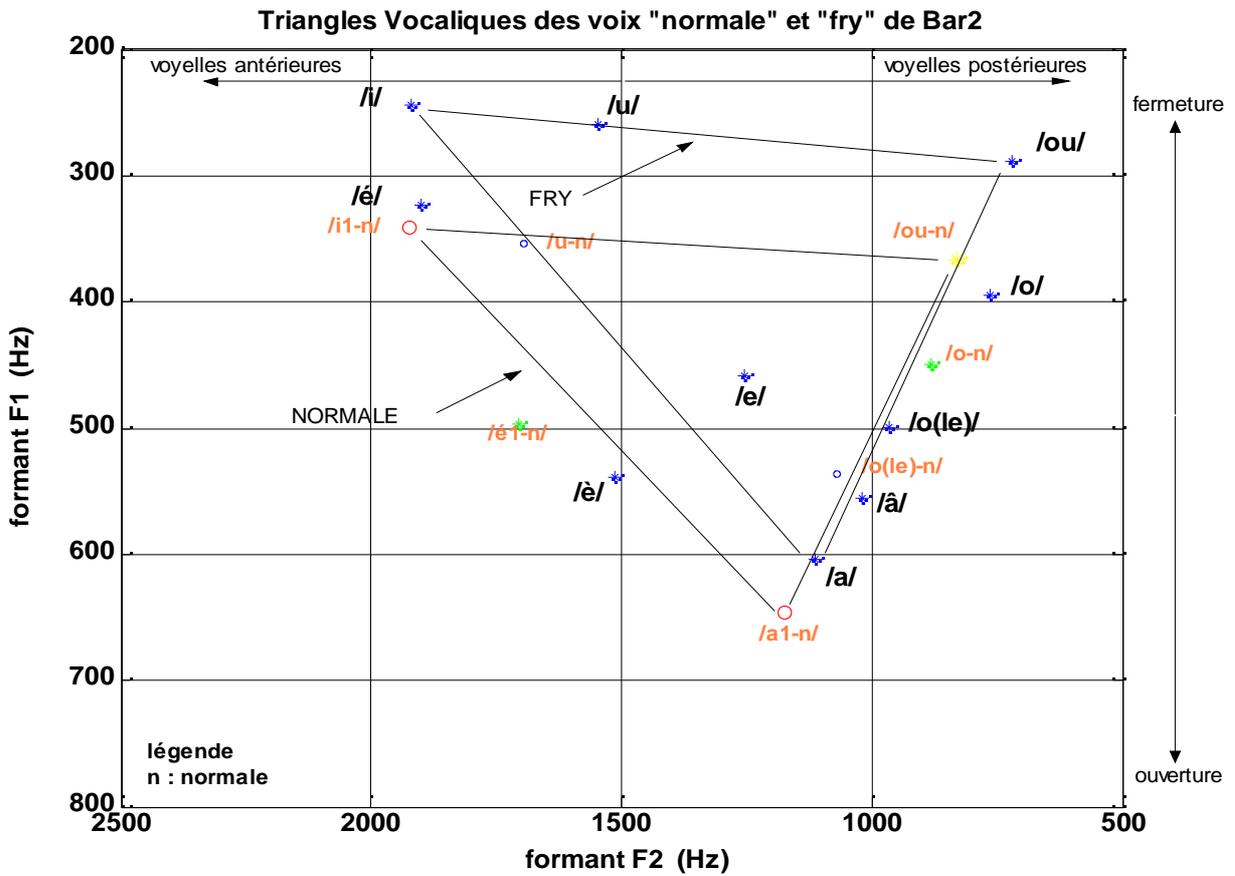
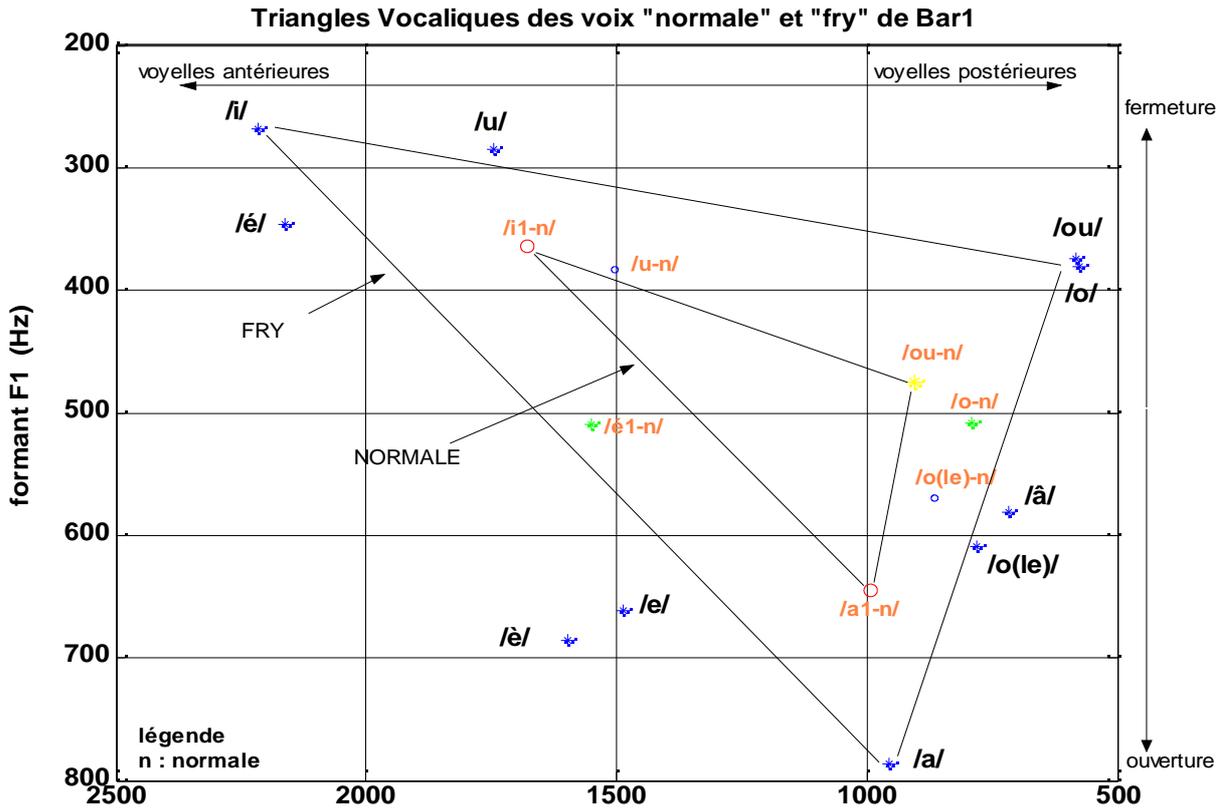
Dans un premier temps, il semblait utile de comparer la production des voyelles dans la situation de voix chantée « normale » avec celle réalisée en voix fry à la fin de l'enregistrement. L'analyse des formants du fry a donné des valeurs sensiblement différentes de l'analyse des formants de la voix « normale » comme le montrent les quatre figures précédentes. On peut proposer plusieurs explications à cela. Tout d'abord la réalisation du fry est difficile chez certains chanteurs peu habitués, donc ils vont chercher une configuration qui les aide à émettre dans ce mécanisme (d'autant que le fry était à la fin de l'enregistrement, donc les cordes vocales étaient sûrement un peu tendues et fatiguées. Ainsi la configuration des cavités devait être un peu différente lors de la production « normale » chantée au début par rapport à celle du fry plus proche de la voix parlée. En prenant donc toutes les précautions quant à l'interprétation des résultats donnés par la LPC, ces figures nous donnent des renseignements très intéressants sur les tactiques adoptées par les chanteurs dans le chant lyrique. Ces tactiques sont liées à l'apprentissage de la technique du chant lyrique et montrent les tendances et les « tricheries » de prononciations (ce n'est pas un terme péjoratif dans le chant lyrique et il est couramment employé pour désigner justement la tactique de compromis pour préserver la technique vocale sans dénaturer complètement le texte). Au moment où nous avons débuté les enregistrements, nous ne nous attendions pas à ces résultats, il est donc intéressant de les mentionner ici.

Nous avons dans un premier temps pensé à comparer les voix « normales » de différentes intensités ensemble pour voir si les formants bougeaient. Effectivement, les formants changent de place, et cela est assez cohérent avec les discours de professeurs de chant expliquant qu'un son pianissimo est très dur techniquement, car il fait appel à la « couverture » des sons. De plus les émissions que nous avons demandées étaient assez extrêmes en intensité. Ainsi nous prendrons comme référence la voix « normale » d'intensité moyenne.

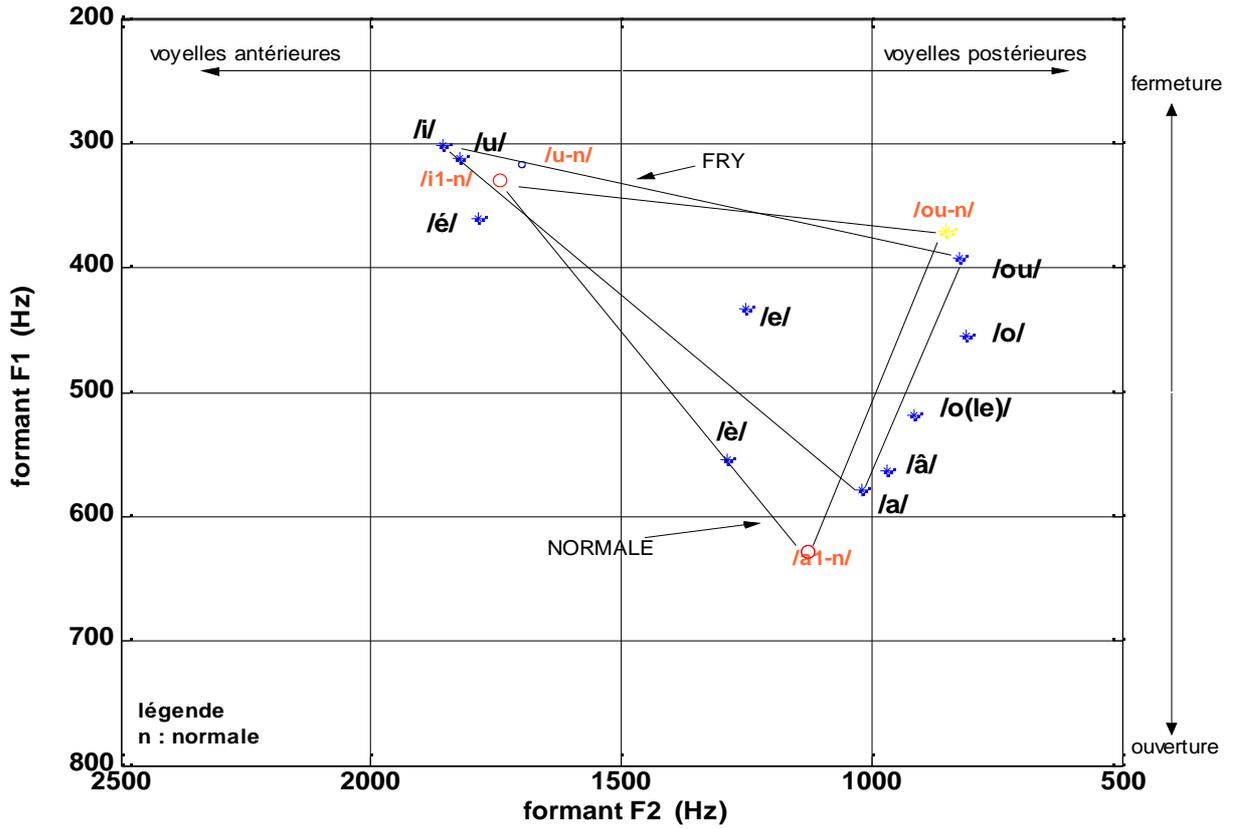
Tout d'abord, on voit qu'au niveau du fry, Bar1 et T1 couvrent des étendues à peu près similaires. Celles de Bar2 et Bar3 sont plus resserrées. On peut considérer que les émissions en fry vont se rapprocher plus de la voix parlée, il est donc normale de trouver des surfaces assez vastes par rapport au chant qui développe plus dans la verticalité de la bouche et dans le soulèvement du voile du palais. En langage parlé, les voyelles sont relativement bien distinctes et impliquent aussi bien une horizontalité de la bouche qu'une verticalité.

Mais c'est en observant les triangles vocaliques dus à la qualité « normale » des chanteurs que l'on peut faire des remarques intéressantes.

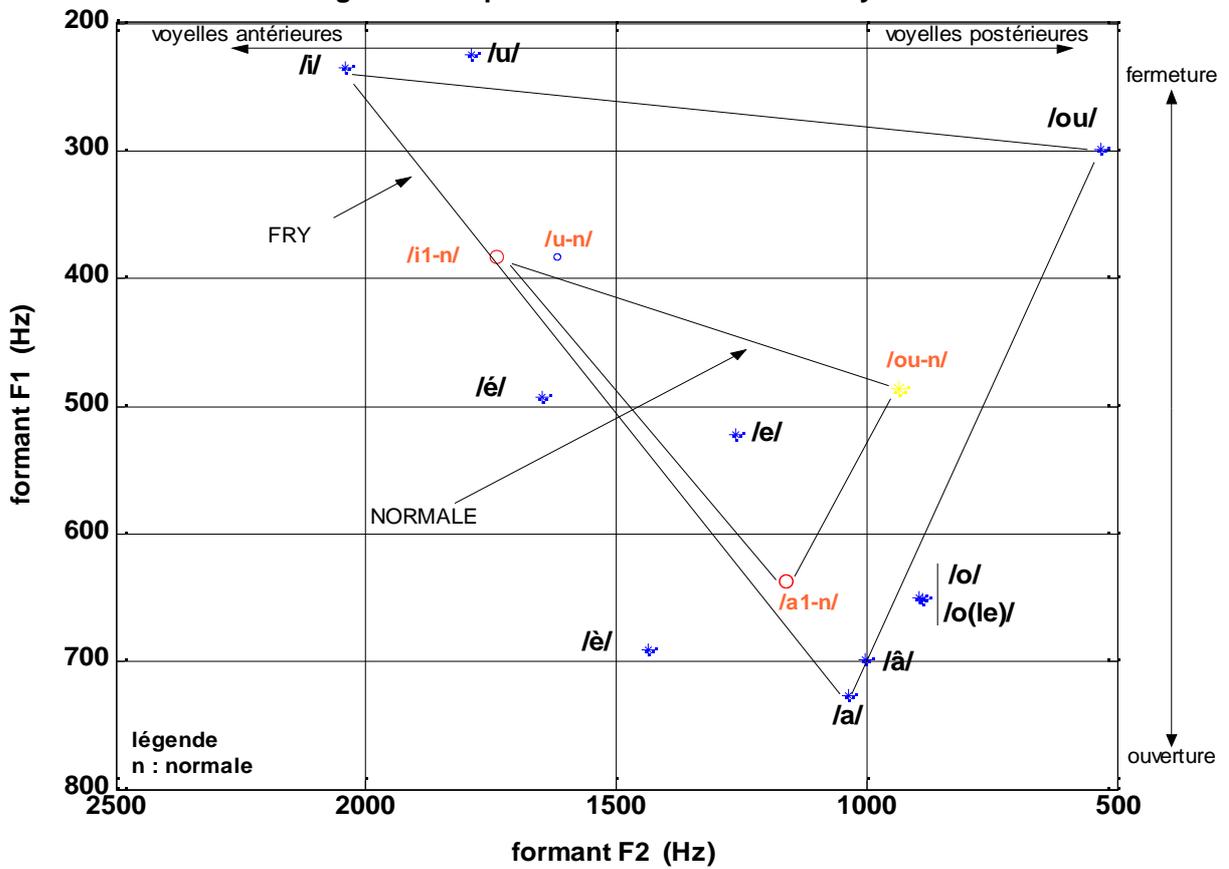
On voit que pour **Bar1**, le triangle « normal » s'inscrit dans celui du fry en suivant bien les mêmes directions et bien centré par rapport au fry. Cela veut dire qu'il resserre toutes les distances entre les différentes voyelles. Ainsi les /i/, /u/ /ou/, /é/, /o/ sont bien plus ouverts dans sa façon de chanter, et les /a/ et /o(le)/ sont un peu moins ouverts. De même les voyelles s'éloignent des extrémités antérieure et postérieure pour trouver un plus juste milieu.



Triangles Vocaliques des voix "normale" et "fry" de Bar3



Triangles Vocaliques des voix "normale" et "fry" de T1



Ces pratiques sont courantes dans l'apprentissage du chant, lorsqu'il ne faut pas trop serrer un /i/ ou bien le /ou/ qui se rapproche souvent du /o/ (déjà dans le fry le /ou/ est loin de l'horizontale du /i/, ceci s'observe plus ou moins pour tous les chanteurs).

Dans le cas de **T1** on retrouve un peu la même situation du triangle de la voix « normale » inscrit dans celui du fry. Cependant l'axe antérieur reste proche de celui du fry. Par contre il y a une plus grande ouverture des voyelles /i/, /u/ et /ou/ que pour Bar1. Le /ou/ va se retrouver plus proche d'un /o/.

**Bar2** adopte une tactique mixte par rapport aux deux autres. Il a une tendance à ouvrir plus les voyelles fermées que dans sa voix fry et à antérioriser plus les voyelles antérieures. Cependant il ne faut pas perdre de vue les échelles absolues. Bar2 aura quand même tendance à prononcer ses /i/ plus fermés que ceux de Bar1 et T1, pour prendre un exemple. Et plus antérieurs aussi. Par contre par rapport à T1 et Bar1, il va garder ses voyelles antérieures proches de celles du fry. On remarque notamment que le /ou/ est proche de l'horizontale du /i/.

Pour **Bar3**, la différence entre les voyelles fermées postérieures (/ou/) et ouvertes antérieures (/a/) n'est pas très grande comparée aux autres chanteurs. La voix normale a tendance à les écarter un peu plus mais la signification est peut-être dérisoire étant donné les écarts-types de la LPC. Néanmoins, pour ce chanteur, on note moins de différence entre les formants issus du fry et ceux de la voix « normale ».

### **Conclusion :**

Ces différences nous incitent à regarder les autres qualités vocales par comparaison avec la voix « normale » des chanteurs. De temps en temps, les points du fry peuvent apparaître en gras sur les figures (et étoiles bleues) quand ils nous indiquent un repère.

### **V.3) Comparaison des formants vocaliques pour les voyelles /i/, /u/, /é/ :** (cf. 12 figures sur 4 pages suivantes)

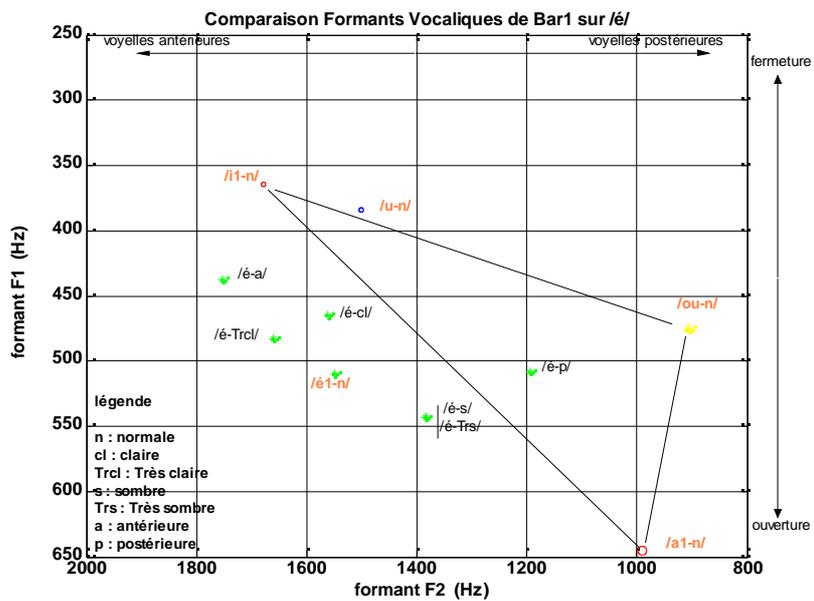
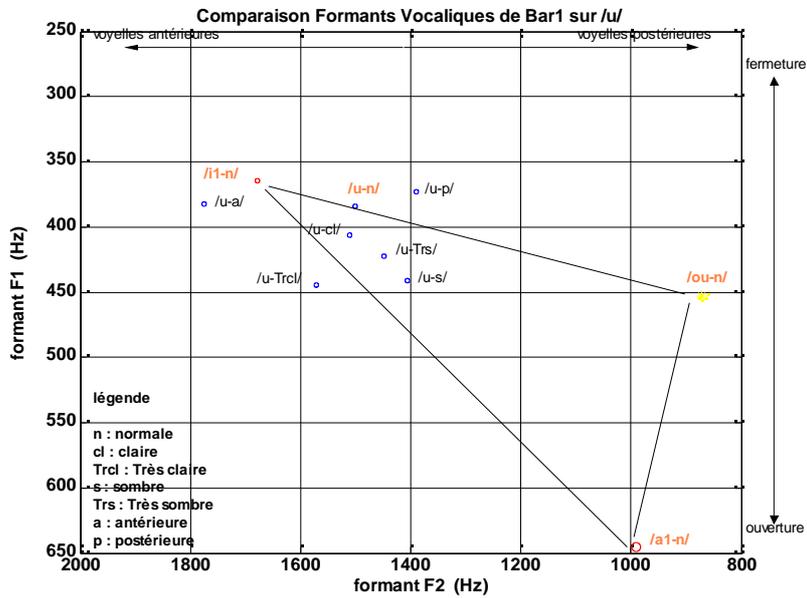
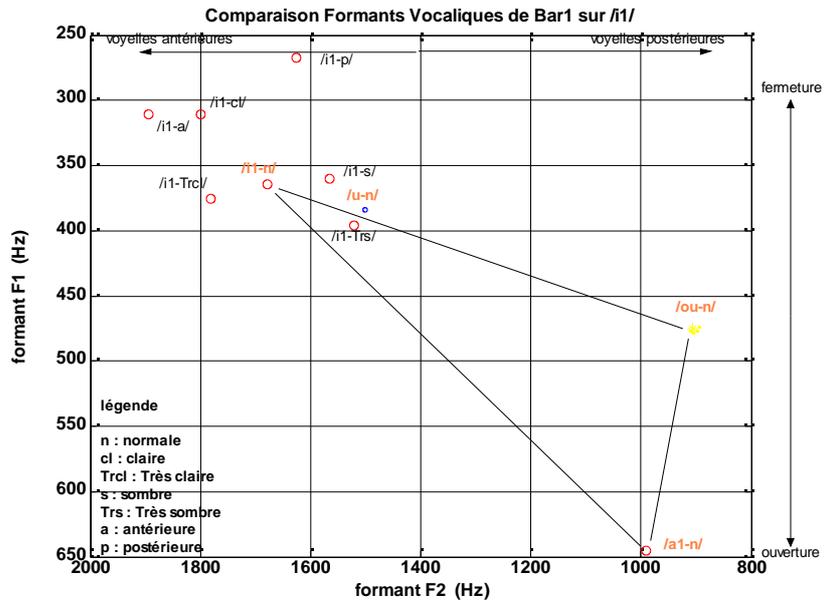
On cherche à voir si, pour ce groupe homogène de voyelles caractérisées par une descente de F2 du /i/ au /é/, on obtient des règles de variations entre les qualités « antérieure », « claire », « ouverte », « couverte » et « sombre », « postérieure » pour les 4 chanteurs.

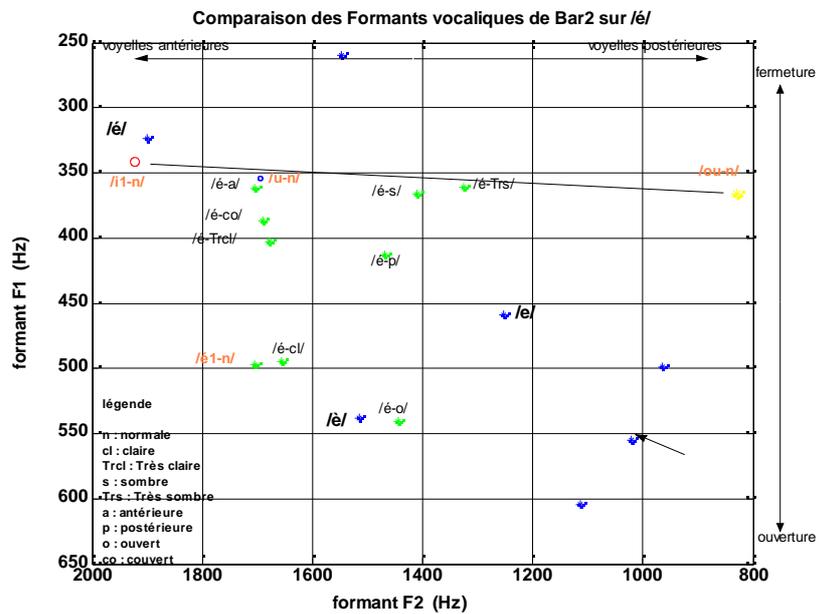
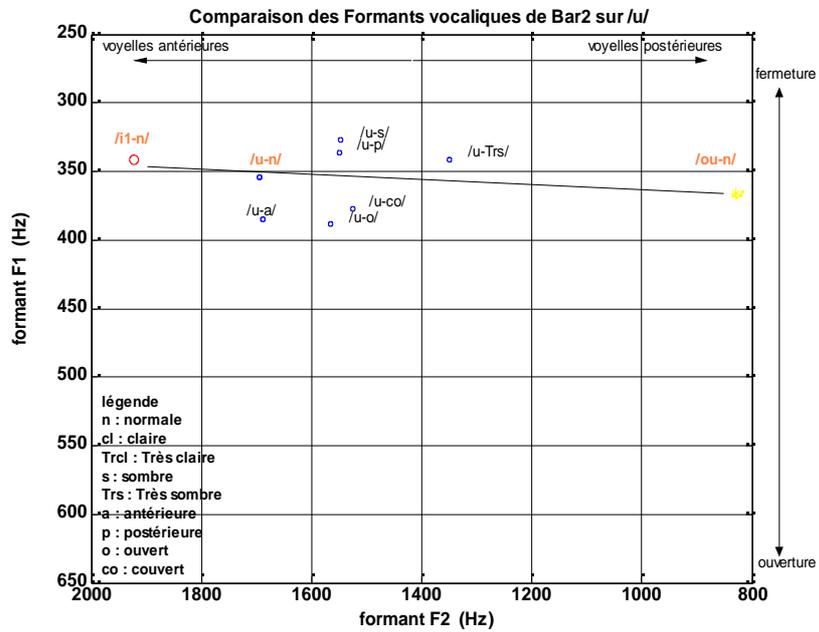
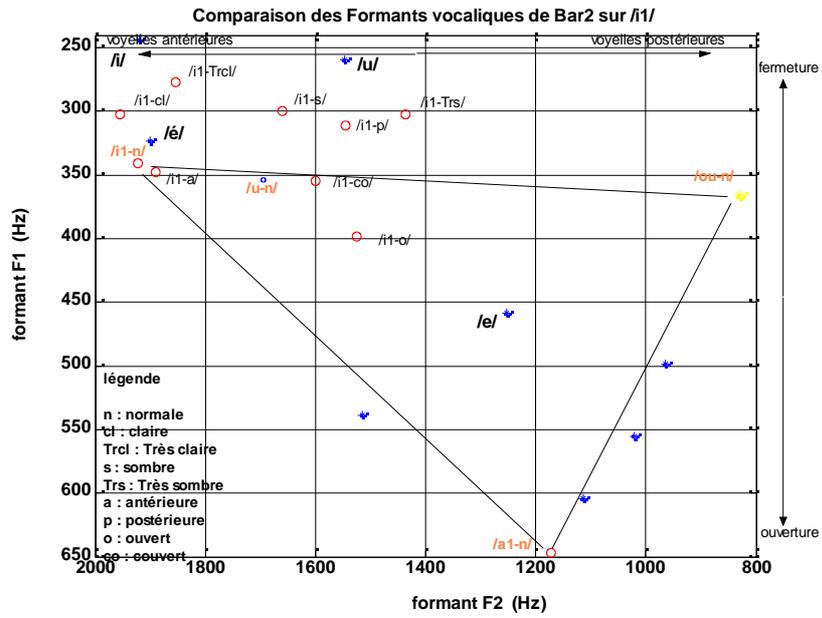
Les 3 voyelles du texte considérées sont les suivantes : « **il** vole là haut **jusqu'**à oublier ... ». Pour une plus grande clarté dans notre propos, nous considérons les chanteurs un par un puis nous concluons sur le groupe de voyelles.

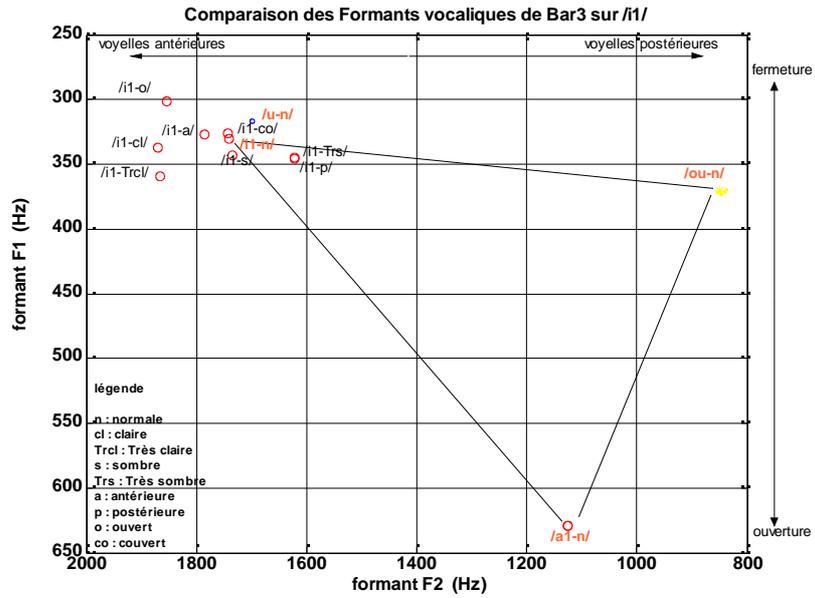
#### *a – Comportement pour Bar1*

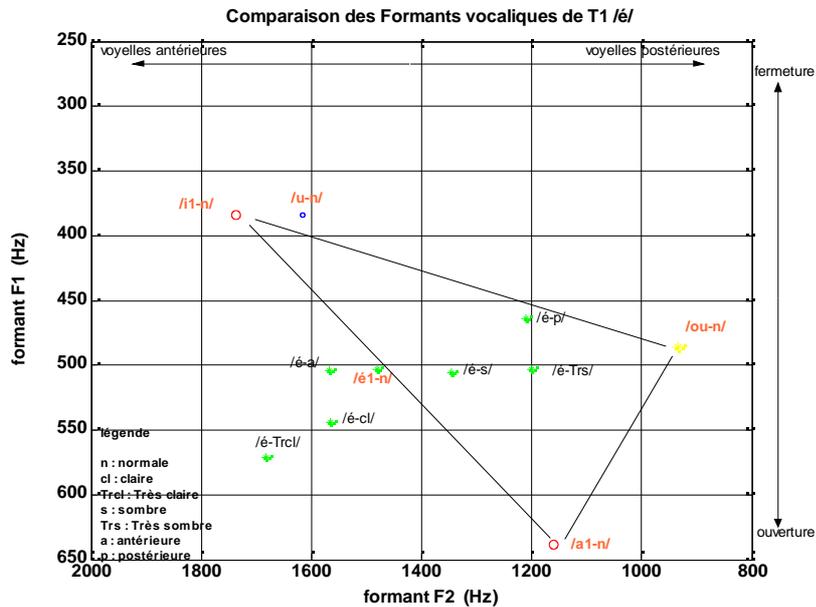
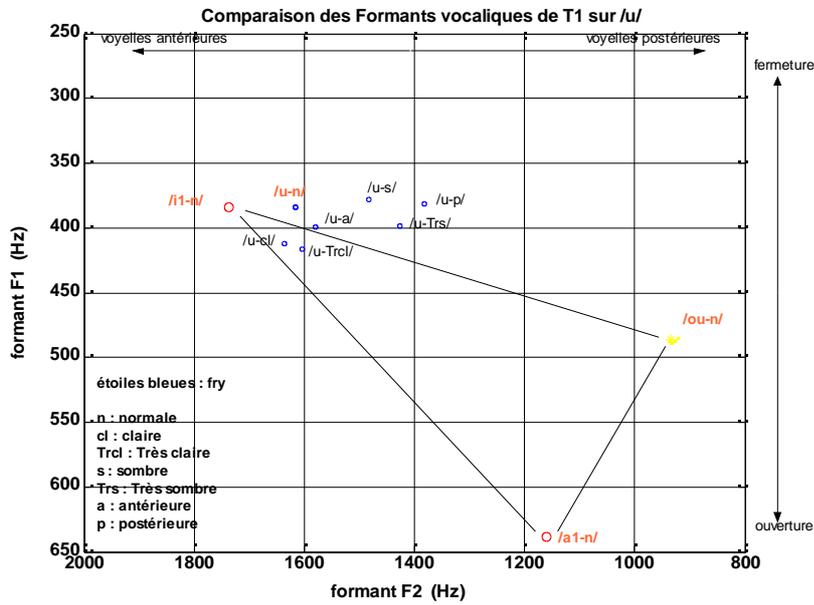
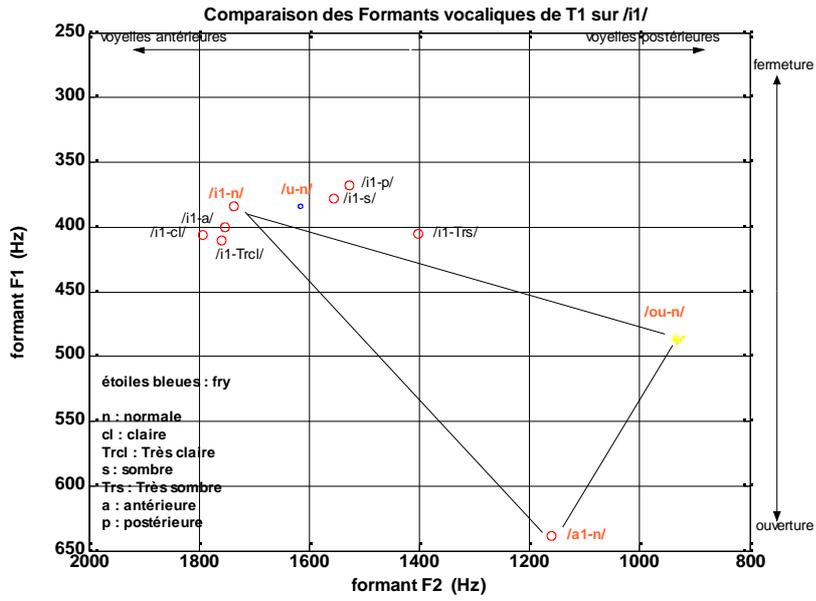
On observe sur les 3 figures de Bar1, une tendance générale à faire apparaître deux groupes de qualités dont les formants se déplacent dans des sens différents.

Les émissions « claires » semblent jouer pour les 3 voyelles plus sur F2 que sur F1 dans un premier temps, puis sur F1 pour augmenter encore la clarté (pour « très claire »). Cependant les voyelles /i/ et /é/ ont un F1 qui baisse dans la qualité « claire » par rapport à la voix « normale » (déplacement vers une articulation plus parlée), alors que pour le /u/ l'augmentation de F1 et de F2 se voient ensemble tout de suite (déplacement vers les /é/). De même, l'émission « antérieure » montre des variations communes avec les précédentes. Elle est bien caractérisée par une augmentation de F2 : pour /u/, on observe un déplacement vers le /i/ de la qualité « normale », et pour le /i/ et le /é/ on remarque en plus une diminution de F1 (plus fermées) allant dans le même sens que pour la voix « claire », c'est à dire vers une articulation plus marquée.









Un deuxième groupe se dégage avec les émissions « sombre » et « postérieure ». Alors que pour les voyelles /u/ et /é/ de l'émission « postérieure » on a bien un déplacement dans le sens contraire que pour « antérieure », le /i/ est plutôt fermé que postériorisé. Cependant, toutes voyelles des émissions « sombres » confirment un déplacement vers les voyelles postérieures avec une diminution de F2 conséquente : /i/ tend vers le /u/, /u/ plutôt vers le /e/ ou /ö/ (car il s'ouvre) et /é/ vers le /ö/.

La qualité « sombre » n'est pas toujours le contraire de « claire » comme pour le /u/ où il y a ouverture de la voyelle dans les deux cas ; pour le /i/ et le /é/ elles sont bien opposées.

#### *b – Comportement pour Bar2*

Ce chanteur semble emprunter une autre stratégie pour réaliser les émissions « claires ». Nous n'avons pas eu de valeurs avec les émissions « claire » et « très claire » pour la voyelle /u/ (le chanteur les a réalisées très piano et le /u/ est entouré du « j » et du « s » sur une valeur de croche). En effet, les /i/ et /é/ dans la qualité « claire » ont tendance à avoir un F1 qui baisse de plus en plus que le caractère « clair » augmente. La qualité « antérieure » semble jouer un rôle différent de celui attendu. Pour le /é/ « antérieur » on a plutôt une fermeture, pour le /i/ cela reste identique et pour le /u/ « antérieur », F1 est augmenté par rapport à celui de la voix « normale ».

Par contre on retrouve la même stratégie pour les qualités « sombre » et « postérieure ». Le caractère sombre décale tout d'abord vers les F1 et F2 plus bas. Mais ce qui est remarquable, c'est que pour rendre « très sombre » la variation ne se joue plus que sur F2 en le baissant (vers voyelles postérieures). /i/ tend vers le /u/, /u/ vers le /ou/ et le /é/ remonte aussi vers le /ou/.

On observe moins de règle pour les qualités « ouverte » et « couverte » qui ne semblent pas s'opposer. Ces qualités diminuent F2 et la qualité « couverte » implique un son de voyelle plus fermée que la qualité « ouverte » qui augmente aussi F1 (le /é/ « ouvert » tend vers le /è/ du fry).

La qualité « claire » semble jouer sur un écart plus important entre F1 et F2 et « sombre » sur un écart moins important, mais pas forcément dans les sens attendus.

#### *c – Comportement pour Bar3*

On observe ici aussi une tendance commune des qualités « claire » et « antérieures ». Elles augmentent beaucoup F2 par rapport à la voix « normale » avec un F1 à peu près semblable. Donc, pour ce chanteur, le /i/ s'antérriorise un peu plus, le /u/ tend vers le /i/ (le /u/ de la qualité « ouverte » est situé sur le /i/ « normal ») et le /é/ tend vers celui du fry avec la « clarté ».

Le caractère « sombre » s'affirme toujours dans le sens d'une postériorisation par rapport à la voix « normale », surtout pour le /é/ et le /i/.

Quant aux caractères « couvert » et « ouvert », on ne retrouve pas de comparaison avec Bar2. Pour le /i/ et le /é/, on a plutôt une antériorisation et une fermeture de la voyelle pour la qualité « ouverte » et seulement une fermeture pour la qualité « couverte ».

#### *d – Comportement pour T1*

Dans ce cas encore, on a une nette séparation, par rapport à la voix « normale », des groupes composés des qualités « claire » et « antérieure » d'une part, et de « sombre » et « postérieure » d'autre part.

Les /i/ et /u/ « clairs » et « antérieurs » sont assez proches de la voix « normale » (Peut-être parce

que sa voix de ténor est déjà « claire »). Néanmoins ils ont un F1 qui augmentent un peu. Les /é/ « clairs » augmentent bien F1 et F2 par rapport à la voix « normale » et le /é/ « antérieur » augmente seulement F2.

Pour les caractères « sombre » et « postérieur » on note encore des similitudes : l'accentuation de « sombre » déplace les voyelles suivant les axes du triangle « normal » vers les voyelles postérieures correspondantes.

### **Conclusions pour le groupe des voyelles /i/, /u/, /é/ :**

- Le /é/ n'est bien souvent pas celui du langage parlé dans le chant lyrique, comme le /i/ car ce sont des voyelles très fermées et antérieures donc elles vont un peu à l'encontre de la place de voix lyrique. Le /é/ est peut-être même plus modifié que le /i/, car cette voyelle est moins présente dans les autres pays occidentaux ayant développé la technique lyrique (Italie, Russie). De plus, en français elle est vraiment pronocée avec une bouche très horizontale. C'est pourquoi, on ne s'étonnera pas de voir une grande différence de place des formants du /é/ entre les voix « normales » des chanteurs. Elle va se retrouver quelque part à l'intérieur d'un triangle dont les sommets sont le /é/, le /e/ et le /è/ du fry. Ce point est très important car il peut avoir une grande influence sur la perception de la voix « normale ».

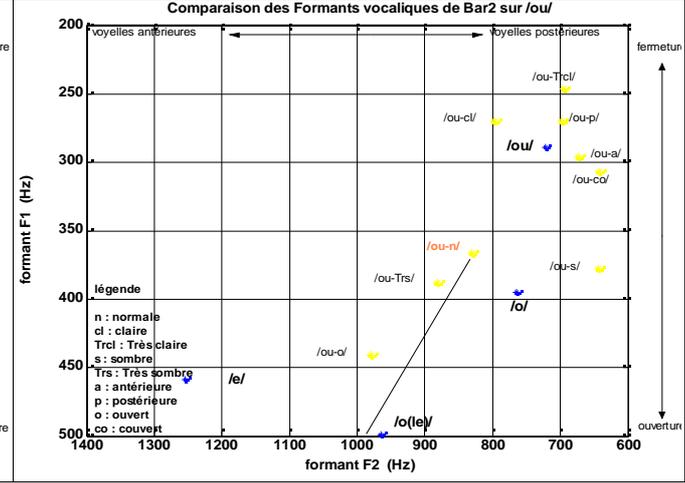
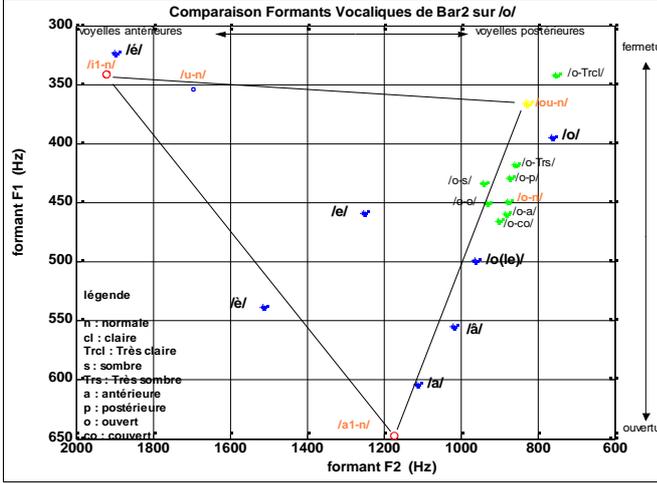
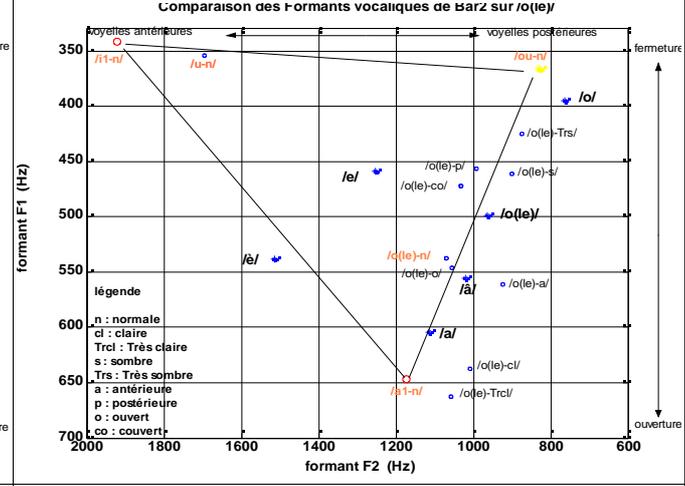
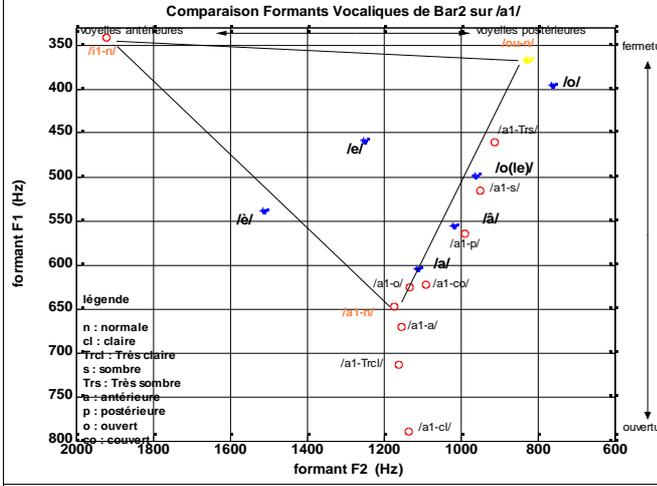
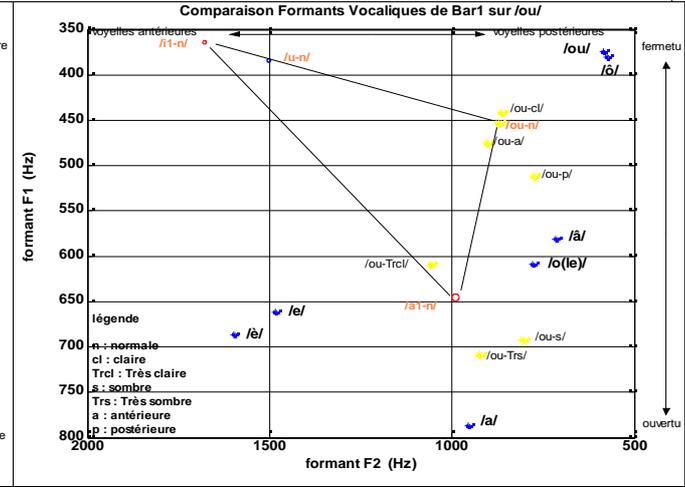
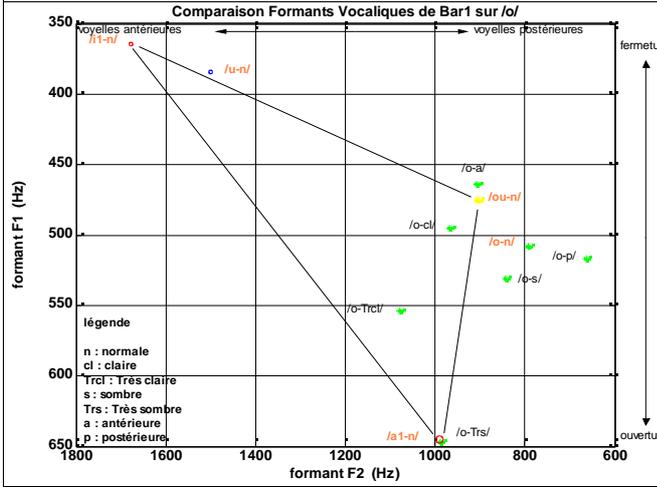
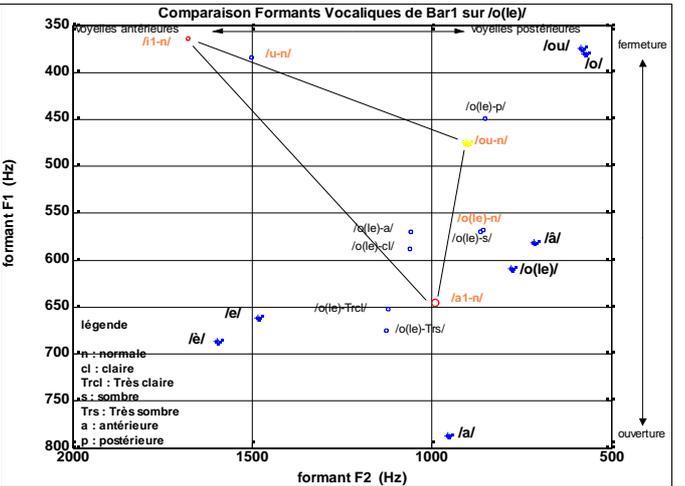
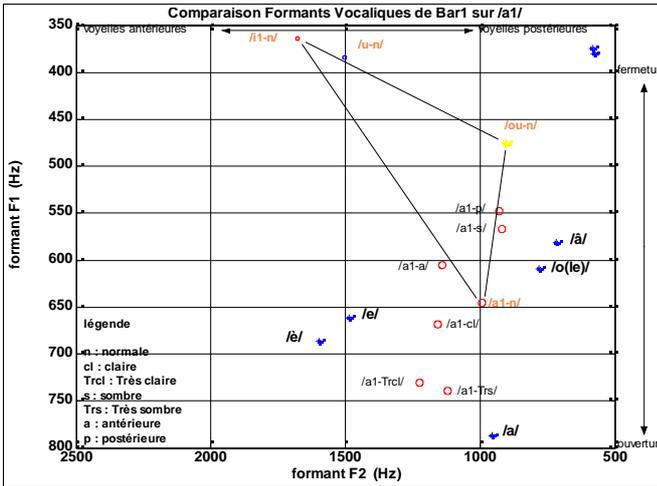
plusieurs stratégies se dégagent au sein des ressemblances :

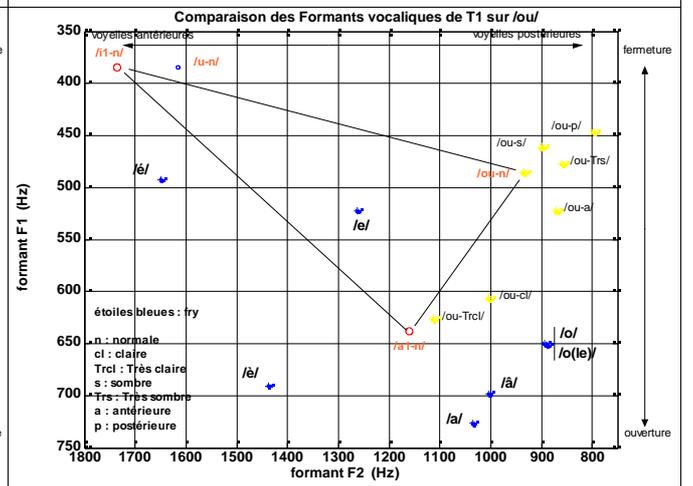
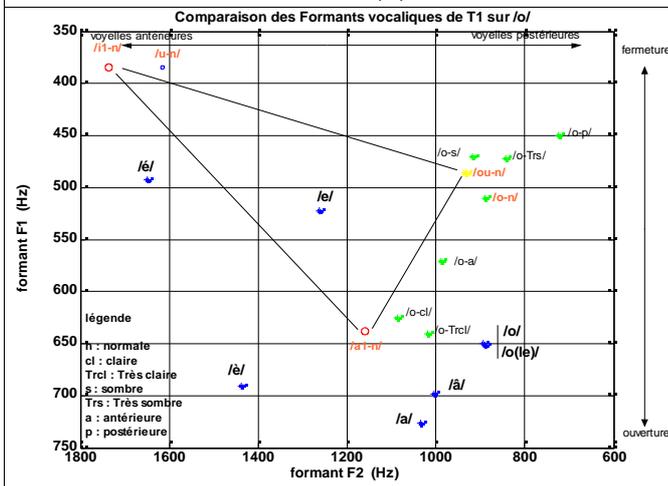
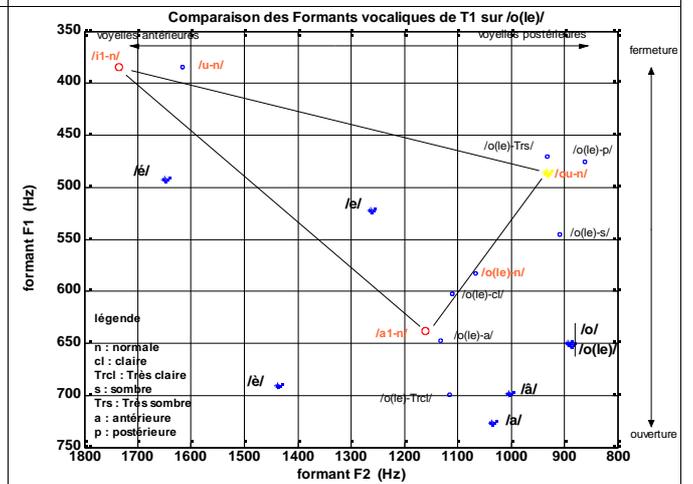
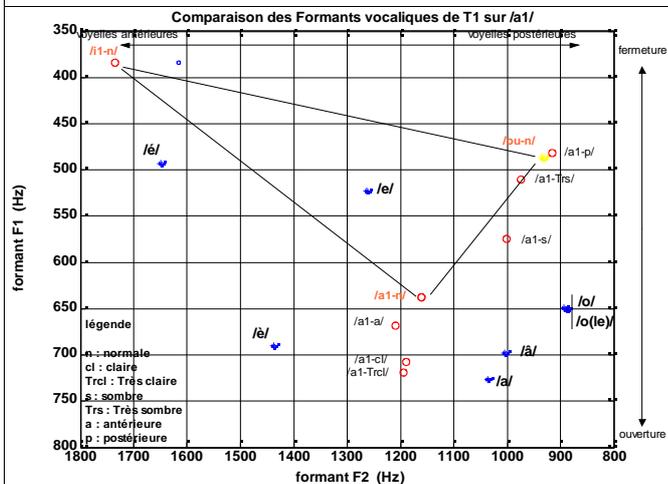
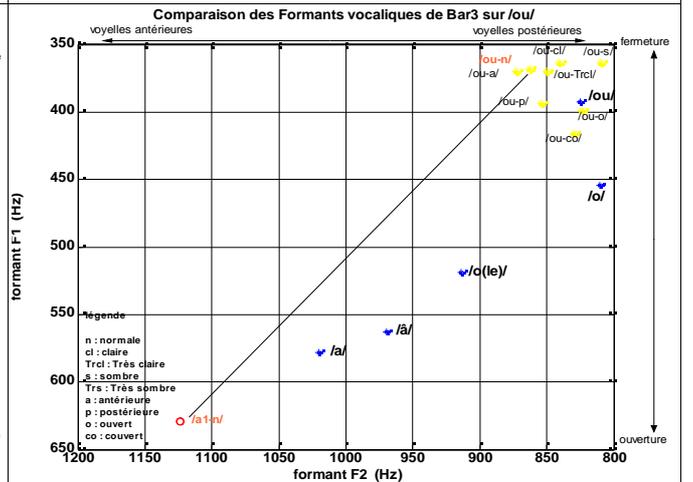
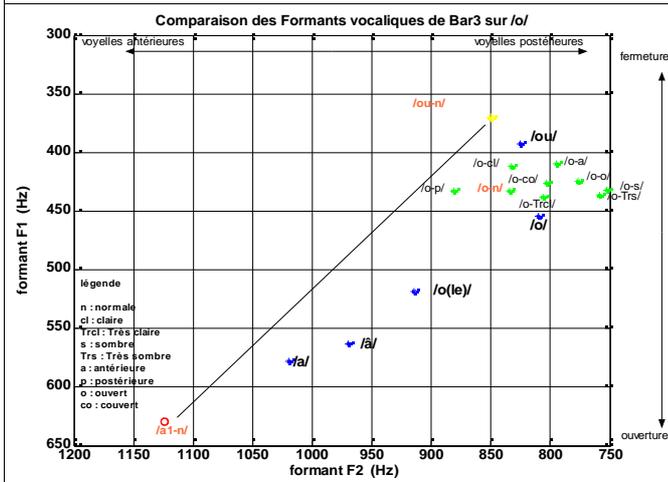
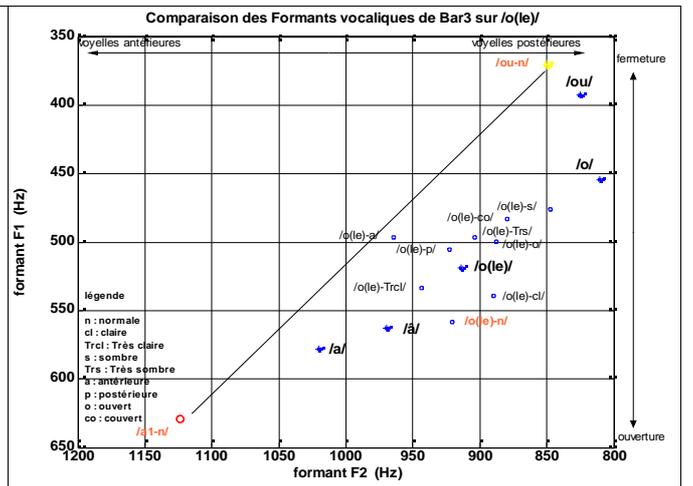
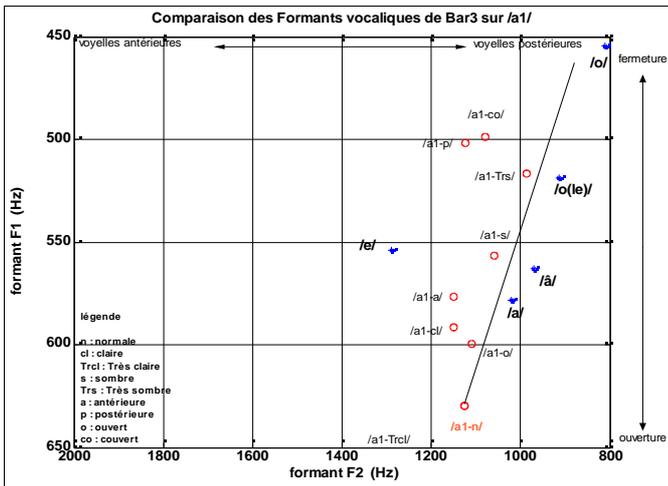
- le /i/ de caractère « sombre » comparé à la voix « normale » joue plutôt sur F1 (même tendance que « postérieure ») et se rapproche dans un cas plutôt du /u/ chanté « normal » et dans d'autre plutôt du /u/ fry
- le /i/ de caractère « clair » comparé à la voix « normale » joue à la fois sur F1 et sur F2 (même tendance que « antérieure »), mais il y a deux approches pour faire encore plus clair, soit plus fermé vers le /i/ fry par exemple, soit plus ouvert vers un /é/ chanté
- Enfin, suivant le type de voix, il est naturel d'avoir des rapprochements de qualités vocales : nous pensons au T1 (/i/ « normal » proche du /i/ « clair ») qui serait sûrement jugé plus « clair » que les barytons. Cela ne va pas à l'encontre de certains commentaires apportés par les chanteurs lors de l'enregistrement de la qualité « claire » : 2 des 5 chanteurs ont employé immédiatement le mot « ténorisant » pour synonyme de « claire ». Cette remarque montre bien la problématique de la qualification des qualités vocales comme nous le montrions à la fin du chapitre I.

### **V.4) Comparaison des formants vocaliques pour les voyelles /a/, /o(le)/, /o/, /ou/ :** (cf. 16 figures sur 4 pages suivantes)

*a – voyelle /a/ (« l<sup>h</sup> haut »)*

L'évolution de **Bar2** est particulière : on peut séparer une première droite suivant à peu près le même F2 (reste à la même place dans la bouche), qui commence par le /a/ de la voix « normale » (F1 le plus haut donc voyelle la plus ouverte) et qui se poursuit avec ceux des voix « Très claire », « antérieur », « claire » et « ouvert » (les classant par ordre décroissant de F1 donc du plus ouvert au plus fermé). Puis, on observe une bifurcation suivant les F1 et F2 décroissants (prenant la branche des /o/) à partir de « couvert », suivi de « postérieure » « sombre » et « très sombre ». Tout se passe comme si la première droite représentait des /a/ plus ou moins proches d'un /a/ parlé ou du fry représenté par « ouvert » (ce qui paraît logique) et qu'à partir de la « couverture », le caractère « sombre » s'affirme en transformant progressivement le /a/ en /o/ plus fermé.





Pour les 3 autres chanteurs on observe des groupements similaires entre eux, à savoir : le /a/ qui tend vers la branche des /o/ pour les qualités vocales « sombre », « postérieur » ou « couvert » (F1 et F2 plus petits que les « normaux ») et le /a/ qui se décale vers les F2 plus grands pour les émissions « antérieure », « claire » et « très claire ». On voit que pour une plus grande clarté, l'émission « très claire » se différencie souvent de « claire » par une augmentation de F1 donc de l'ouverture de la voyelle.

Le /a/ de la qualité « ouverte » de Bar3 se rapproche aussi de celui du fry.

### **Conclusion :**

- on voit que la clarté joue une fois de plus sur F2 en les rendant plus aigus (renforce antériorisation) et aussi sur F1 quand on demande encore plus clair. Donc le caractère « clair » a tendance à tout remonter. Pour « sombre » c'est plutôt l'inverse les formants diminuent en fréquence.

*b – voyelle notée /o(le)/ (« vole »)*

On remarque à première vue qu'il n'y a pas une grande homogénéité entre les chanteurs pour cette voyelle. Peut-être est-ce dû au fait que ce n'est pas une voyelle « cardinale » comme /i/, /ou/, /a/ et qu'elle va donc varier suivant les origines de chacun ?

Pour **Bar3**, toutes les qualités vocales se retrouvent dans une zone assez confinée, où on ne peut pas tirer de grande conclusion mis à part que tous les /o(le)/ des différentes qualités ont un F1 inférieur à celui de la voix « normale ». On voit quand même que pour les 2 voix « sombres », « couverte » et « postérieure » ils sont voisins et se rapprochent du /o/, et aussi que le /o(le)/ de l'émission « ouverte » est le plus proche de celui du fry. La qualité « claire » et « très claire » n'est pas très éloignée de la voix « normale » pour le /o(le)/. On observe que celui de « couvert » a un même F2 que celui de la voix « normale » mais avec un F1 plus faible comme nous l'avions déjà remarqué pour les autres voyelles (simple fermeture).

Pour **Bar2**, il y a une séparation en deux zones autour de la voix « normale ». D'une part, les qualités « couverte » et « postérieure » affectent F1 en le diminuant par rapport à « normale », et les deux « sombres » diminuent aussi F2 tendant vers le /o/. D'autre part, les deux émissions « claires » augmentent plutôt F1 se rapprochant ainsi d'un /a/. Une tendance similaire est remarquée chez T1 où les caractères « sombre » et « postérieurs » diminuent à la fois F1 et F2 vers /o/ et /ou/. Au contraire « claire » et « antérieure » remontent un peu F2 et surtout F1 dans la direction du /a/. Seul **Bar1** n'adopte visiblement pas la même stratégie pour le caractère « sombre » où il ne cherche pas à postérioriser les voyelles (il serait intéressant de voir l'influence que cela peut avoir sur la perception lors de tests) puisqu'au contraire il augmente F1 ce qui rapproche le /o(le)/ de la voix « très sombre » de celui de « très claire ». A l'écoute on perçoit le /o/ de « vole » comme presque identique au /e/. Peut-être est-ce parce que sa prononciation en voix « normale » est déjà proche de celle du « sombre » que le chanteur change de stratégie ? Par contre on observe une progression similaire aux autres chanteurs pour les émissions « claires » et « antérieure » qui augmentent F2 de la même manière, et augmentent progressivement F1 pour aller vers plus « clair ».

### **Conclusion :**

- les résultats semblent indiquer que le /o(le)/ se trouve essentiellement entre deux pôles représentés par les voyelles /a/ et /o/ et qu'en règle générale, malgré des stratégies différentes, les voix « sombre » et « postérieure » referment la voyelle /o(le)/ alors que « claire » et « antérieure » l'ouvrent un peu plus.

*c – voyelle /o/ (« haut »)*

Nous observons une similarité avec la voyelle précédente. Nous pouvons faire les mêmes remarques pour **Bar1** sur sa différence de stratégie pour assombrir : le caractère « sombre »

entraîne le /o/ vers le /a/ de la voix « normale » et ne va pas dans le sens de la postériorisation de la voyelle.

**Bar2** affiche aussi un changement de stratégie sur l'exemple « très clair » en diminuant les deux formants (peut-être) par un manque d'attention sur cette voyelle au moment de l'enregistrement. De même, le caractère « sombre » augmente F2 par rapport à la voix « normale ». C'est effectivement ce qui ressort lors de l'écoute des deux extraits : sur le terme « haut », les qualités semblent « inversées » par rapport à ce que l'on attendrait (il y a sûrement une explication).

Pour **Bar3** on retrouve les /o/ assez resserrés autour d'une position centrée sur la production « normale ». On ne peut pas faire les recouvrements à de si faibles distances : nous avons « postérieure » à l'opposé de « sombre », et « antérieure » à l'opposé de « claire ». Cela dénote une faible variabilité de prononciation pour cette voyelle.

**T1** a une évolution à peu près similaire pour le /o/ que pour le /o(le)/. Les deux émissions « sombres » jouent sur F1 puis F2 par rapport à « normale » rendant le /o/ moins sonore (plus fermé), de même que le /o/ de « postérieure ». Pour les deux « claires » et « antérieure » la voyelle descend plutôt vers le /a/.

Nous pouvons faire une suggestion pour expliquer le manque d'unité dans la stratégie : « haut » représente un point d'appui dans la phrase, tout comme le /ou/, et sont placés sur un temps fort. Les chanteurs cherchent peut-être, de manière consciente ou inconsciente à éviter des variations trop exagérées sur les temps marqués, et changent ainsi leur stratégie sur cette voyelle?

*d – voyelle /ou/ (« oublier »)*

On observe trois tendances pour le /ou/ suivant les chanteurs :

La première concerne Bar3 qui ne varie pas trop son /ou/ : les /ou/ de « (très) claire », « antérieure », « normale » et « sombre » ont le même F1 et seule « sombre » diminue F2 ; « postérieure » et « ouverte » se rapprochent de la voix fry, tandis que « couverte » augmente un peu F1 (ouvrant un peu plus la voyelle).

La deuxième concerne T1

La troisième concerne Bar1 et Bar2 qui réalisent l'inverse de T1 pour « sombre » (augmente F1 donc la voyelle est plus ouverte) et pour « claire » on a une diminution des deux formants pour Bar2, mais Bar1 rejoint la stratégie de T1 pour « très claire » augmentant F1 et F2.

### **Conclusions sur les quatre voyelles /a/, /o(le)/, /o/ et /ou/ :**

- On dénote plusieurs tendances suivant les chanteurs : pour Bar3, toutes les voyelles postérieures et un peu fermées subissent peu de variations et restent autour de la voix « normale » ; pour T1, on observe deux groupes fonctionnant de telle manière que les voix « claires » augmentent les 2 formants et antériorisent les voyelles, et les voix « sombres » diminuent les 2 formants, tendant vers la voix « postérieure » ; Bar1 et Bar2 montrent des stratégies un peu différentes notamment pour « sombre » où l'on observe l'inverse que pour T1

## **V.5) Quelques conclusions générales :**

- Nous avons remarqué l'homogénéité de l'influence de la « **couverture** » sur les voyelles de **Bar3**. En effet, sur presque toutes les voyelles étudiées ici, l'analyse a montré qu'elles subissaient un abaissement de F1, donc une fermeture, avec la qualité « couverte » de Bar3. Cela affecte dans une proportion plus grande les voyelles ouvertes telles que /a/ et /o(le)/, dans une proportion moyenne les voyelles « semi-ouvertes » telles que /é/ et /o/, et dans une faible proportion les voyelles fermées /i/ et /u/. Seule /ou/ subit un effet inverse dans l'émission « couverte », marquant donc une limite dans l'effet d'abaissement. Cela rejoint, en quelque sorte, le discours de certains professeurs de chant, expliquant la « couverture » comme s'il fallait penser toutes les voyelles un peu comme le /ou/ chanté, mais en faisant percevoir quand

même la bonne voyelle. D'où la limite sur le /ou/ lui-même.

- Ce qui différencie bien souvent l'émission « très sombre » de « sombre », c'est la diminution supplémentaire de F2, à F1 constant : donc une simple postériorisation supplémentaire qui a pour effet de resserrer l'écart entre F1 et F2. Donc F2 semble bien avoir un rôle dans le caractère « sombre ».
- Le caractère « très clair » semble lui s'affirmer plus avec l'augmentation de F1 à F2 constant par rapport à la voix « claire ».
- Tout se passe donc comme si le chanteur effectuait une diminution de F1 pour « sombre » et des 2 formants pour « claire » définissant ainsi la nouvelle place de la voyelle pour ces qualités, et qu'ensuite, pour accentuer l'un des deux caractères ce n'est qu'un formant qui varie le plus : dans un cas pour diminuer F2 – F1, dans l'autre pour l'augmenter.
- Tout ces résultats ne sont à ce stade de l'étude que des tendances qu'il faudrait valider peu à peu avec un grand nombre de chanteurs.

## **V.6) Remarques :**

Nous ne nous sommes pas attachés dans ce rapport à l'étude des consonnes. Néanmoins, le texte de la mélodie a aussi été prévu pour une éventuelle étude de consonnes voisées. Quelques remarques peuvent être énoncées ici :

- l'articulation, cette fois celle des consonnes, varie extrêmement suivant les qualités vocales, notamment pour la qualité « sourde »
- dans certains cas, la perception de consonnes voisées est clairement changée en perception d'une autre consonne, d'une façon parfois surprenante : c'est le cas de [v] dans les émissions « claires » et « très claires » qui se transforme à l'oreille en [f] (avec du souffle), on voit sur un sonagramme des harmoniques qui disparaissent et du bruit aigu qui apparaît plus ; de même pour les émissions « soufflées » où cela est plus compréhensible.
- d'autres fois du bruit apparaît plus à certains endroits qu'à d'autres, surtout au niveau du [ʒ] (« **j**usqu'à ») ou du [z] (liaison de « **n**os **â**mes ») qui devient parfois [s]

Il est possible d'étudier plus profondément cet aspect pour en tirer des conclusions significatives au point de vue acoustique.

# Chapitre VI : Analyses de la répartition spectrale de l'énergie pour quelques qualités

Nous avons utilisé des fonctions développées sous Matlab au LAM. Elles permettent de calculer le pourcentage de l'énergie totale répartie suivant des bandes de fréquences. Nous nous sommes servis aussi des spectres moyennés sur la phrase entière. Nous présentons ici quelques résultats concernant les qualités « détimbrée », « timbrée », « brillante » et « sourde » que nous confrontons avec l'émission « normale » et les émissions « très timbrée », « très détimbrée », « très brillante » et « très sourde ».

## **VI.1) Qualités vocales « détimbrée » et « timbrée » :** **(cf. spectres moyens et répartitions spectrales de l'énergie pour ces qualités pages suivantes)**

Commençons par observer la qualité « **timbrée** ». On observe deux tendances : la première (Bar1 et T1) joue énormément sur le formant du chanteur pour timbrer la voix, confirmant ce que l'on connaît déjà. Pour Bar1, il se situe entre 2000 et 3000 Hz et pour T1 il est un peu plus proche de 3000 Hz. On remarque que le maximum du formant du chanteur est à - 12 dB de H1 pour Bar1 « normale » alors qu'il n'est plus qu'à - 8 dB pour « timbrée » et - 7 dB pour « très timbrée » ; de même pour T1 on a une évolution de - 11 dB à - 5 dB.

Pour timbrer encore plus, on voit que le formant du chanteur n'augmente pas plus mais c'est plutôt la bande de fréquences 5000-7000 Hz qui semble se renforcer par rapport à l'émission un peu « timbrée » (on le voit sur les spectres moyens). De plus, on observe aussi une évolution entre les bandes médiums 400-800, 800-1200 Hz qui diminuent un peu par rapport à la qualité un peu « timbrée » pour augmenter la zone précédent le formant du chanteur 1200-1800 Hz (voire 2000 Hz) pour « très timbrée ».

La seconde tendance (Bar2 et Bar3) joue moins sur le formant du chanteur (peu de différence par rapport à la voix « normale ») mais plutôt sur la bande du médium 400-800 Hz correspondant à peu près aux H2, H3 et H4. On voit sur le spectre moyen de la voix « timbrée » de Bar2 par exemple que les harmoniques (de H2 à H6) sont bien remontés par rapport à la voix « normale ». H2 dépasse même H1 en moyenne.

Pour Bar3, le formant du chanteur augmente néanmoins sensiblement avec les émissions plus « timbrées ». Les bandes aiguës semblent quand même jouer un rôle (répartition spectrale).

**Conclusion** : Donc nous avons là deux stratégies qui se recoupent un peu, l'une jouant plus sur l'aigu (formant chanteur) et l'autre plus sur le médium (avec les premiers harmoniques).

Considérons maintenant ce qui se passe pour la qualité vocale « **détimbrée** ». A première vue, nous ne pouvons pas affirmer que les qualités « timbrée » et « détimbrée » sont opposées au niveau acoustique. Si l'on regarde la répartition spectrale des chanteurs, on remarque qu'il y a toujours une bande de fréquences qui ne présente pas de règle antagoniste entre ces deux qualités. Il s'agit souvent de la bande de fréquences 400-800 Hz ou 800-1200 Hz suivant les chanteurs. Cependant pour tout le reste du spectre, cela semble assez antagoniste. En effet, l'exemple le plus marquant est celui de T1, qui représente bien le fait que la seule bande de fréquences qui ne varie pas entre « timbrée » et « détimbrée » est celle de 400-800 Hz. Alors que pour « timbrée » c'est la partie grave – médium du spectre (50-800 Hz) qui semble être la zone où l'énergie est puisée pour alimenter toute la partie aiguë, ici bien que l'émission « détimbrée » soit plus faible en intensité (voir spectre moyen), c'est justement à partir de 400-800 Hz que l'énergie est prise pour augmenter les premiers harmoniques. Ils renforcent la sensation de hauteur fondamentale mais cela appauvrit le reste du spectre. Cette tendance se confirme avec les émissions graduées de « Très timbrée » jusqu'à « Très détimbrée ».

Nous évoquons le manque d'unité dans les stratégies de « timbrée ». Pour « détimbrée », l'atténuation du formant du chanteur est très importante chez tous les chanteurs (cf. spectre moyen et répartition de l'énergie). Pour Bar3, les bandes de fréquences supérieures à 1200 Hz représentent chacune seulement quelques pourcents de l'énergie totale : 6 à 7 % pour toute la partie 1200-15000 Hz.

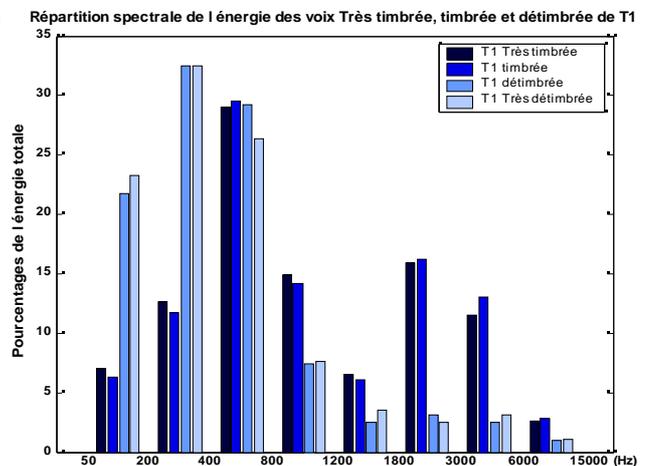
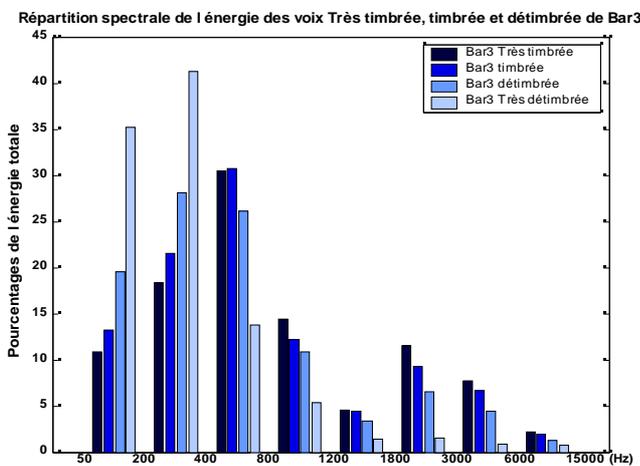
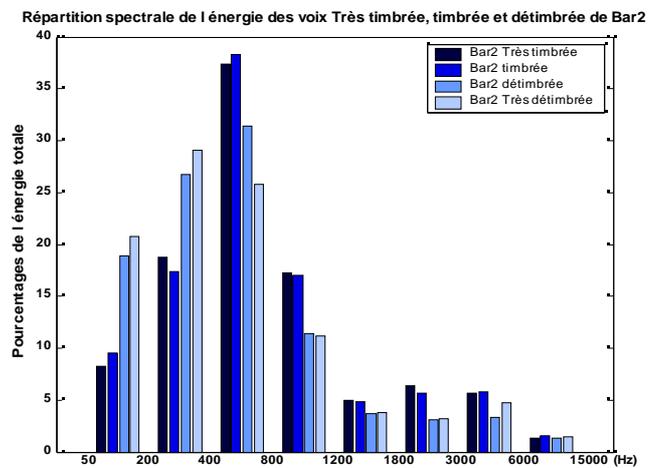
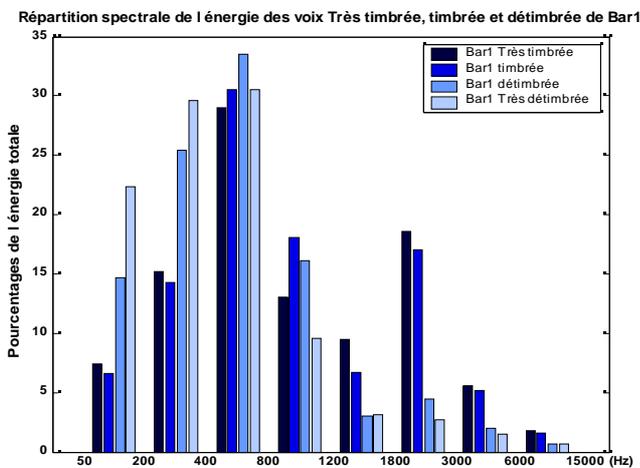
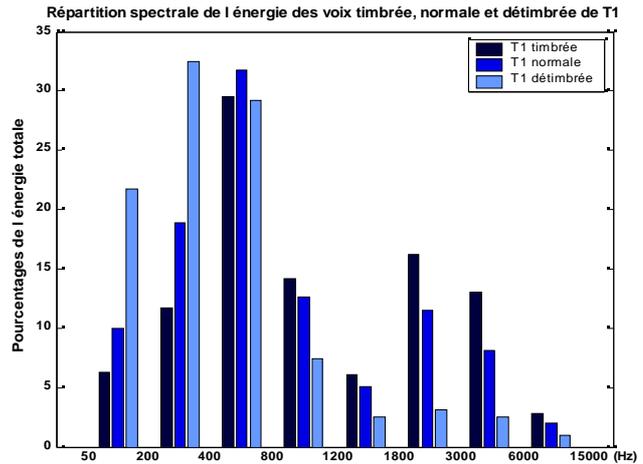
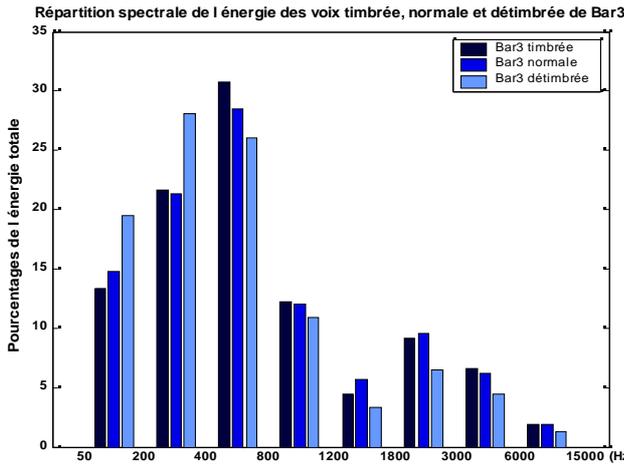
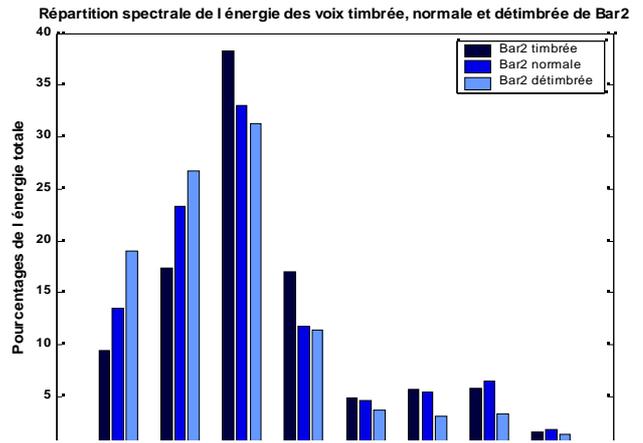
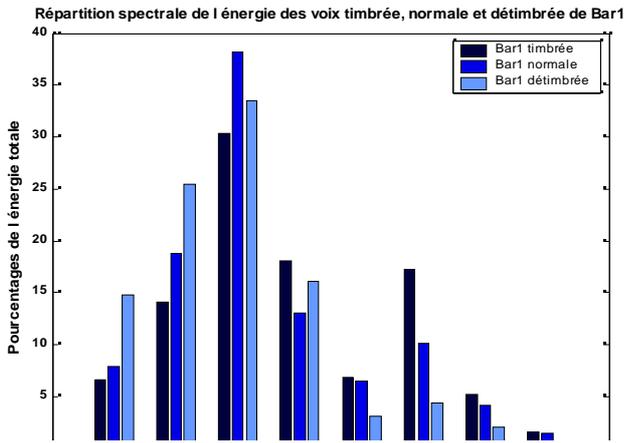
La décroissance des premiers harmoniques est très nette chez tous les chanteurs, surtout pour les voix « très détimbrées » (voir spectre moyen). Les atténuations des harmoniques sont de l'ordre de 6 à 7 dB/octave pour les 6 ou 7 premiers harmoniques en général. Pour Bar3, la voix « très détimbrée » indique une pente de – **18 dB/octave**.

Le formant du chanteur ne ressort presque plus surtout pour Bar3 mais pour les autres chanteurs également (de

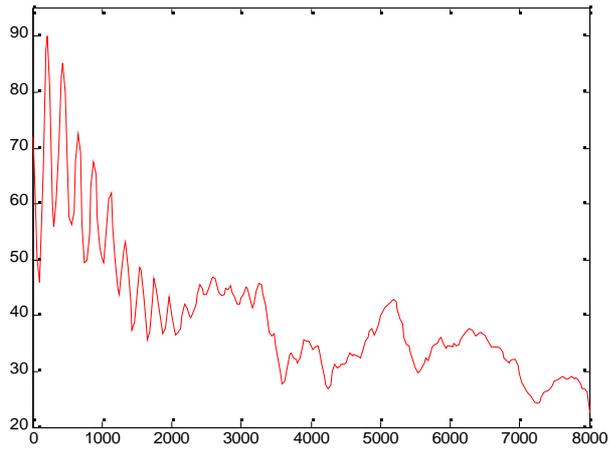
On peut noter une petite différence entre les chanteurs, déjà présente dans les voix « normale » et « timbrée » : par exemple les spectres moyens de Bar1, de Bar3 et de T1 montrent que différentes zones de l'aigu sont bien marquées et évoluent séparément (entre ~2000-3000 Hz, 3000-4000 Hz, 4000-5000 Hz, 5000-6000 Hz et 6000-15000 Hz) alors que pour Bar2 il y a un trou entre 4000 et 6000 Hz.

*Tableau de l'atténuation en dB/octave des x premiers harmoniques concernés :*

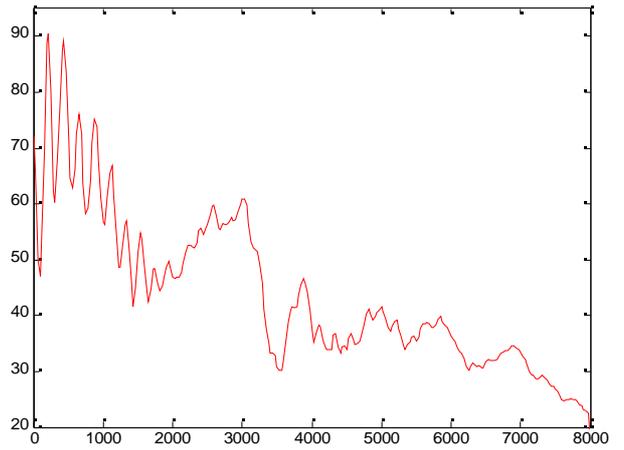
Bar1 détimbrée	Bar1 Très détimbrée	Bar2 détimbrée	Bar2 Très détimbrée	Bar3 détimbrée	Bar3 Très détimbrée	T1 détimbrée	T1 Très détimbrée
H2 – H3 = - 13 dB	- 7 dB/oct. (7 H)	- 6 dB/oct. (7 H)	- 7 dB/oct. (6 H)	H2 – H3 = - 19 dB	- 18 dB/oct. (3 H)	- 6 dB/oct.	- 7 dB/oct.



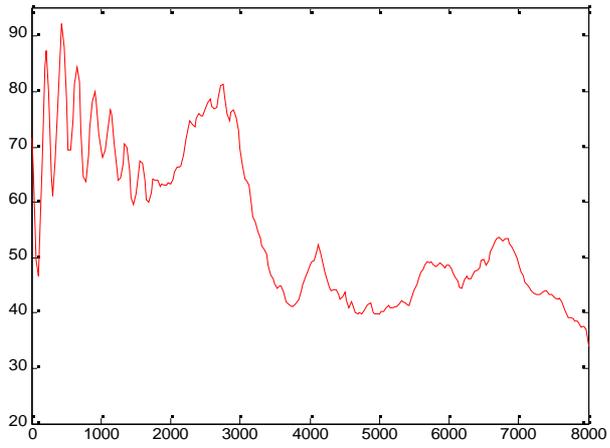
**Bar1 Tres detimbree**



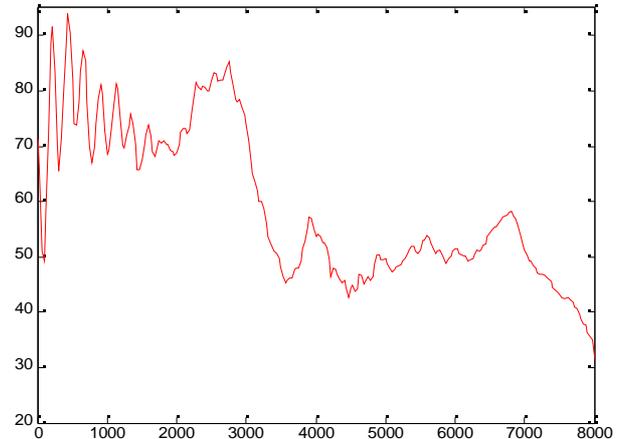
**Bar1 detimbree**



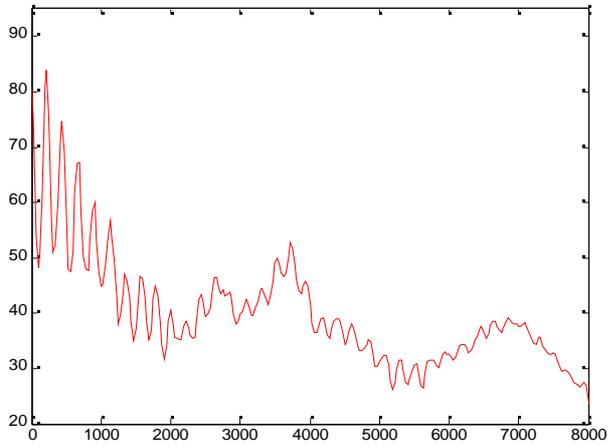
**Bar1 timbrée**



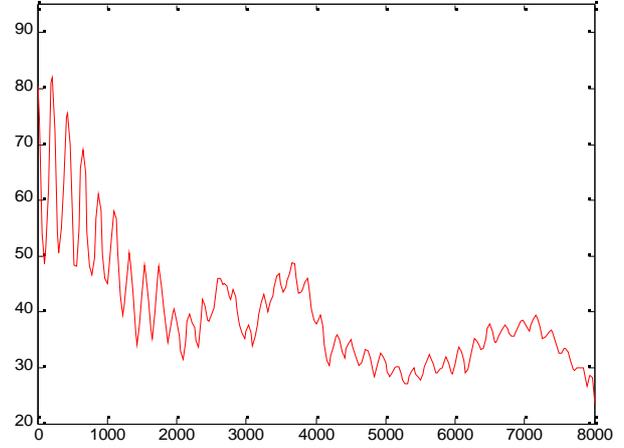
**Bar1 Tres timbrée**



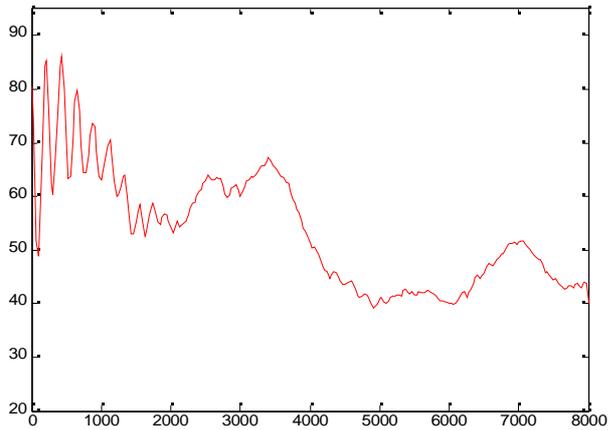
**Bar2 Tres detimbree**



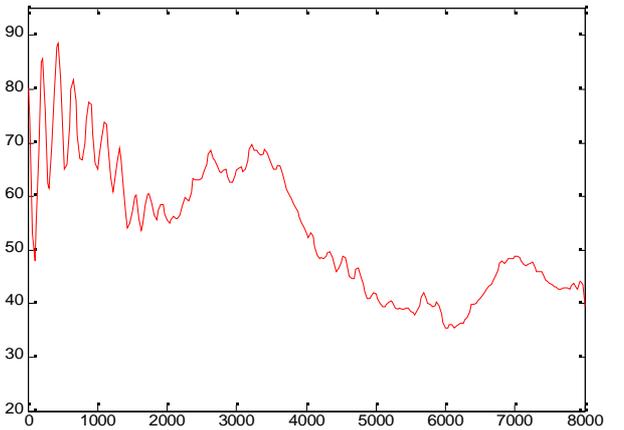
**Bar2 detimbree**



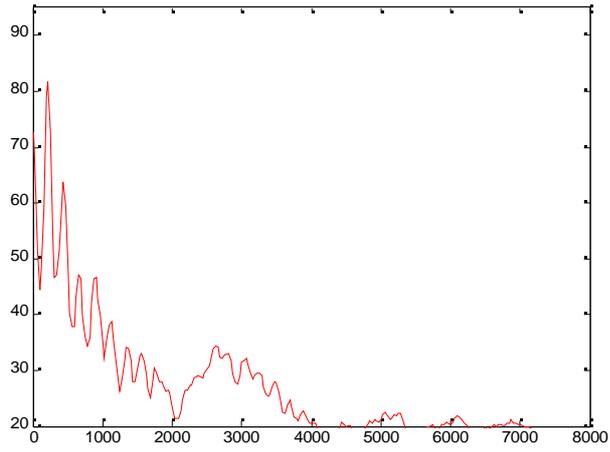
**Bar2 timbrée**



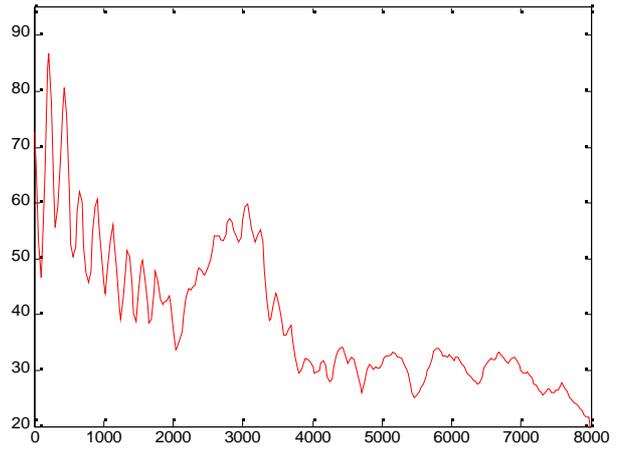
**Bar2 Tres timbrée**



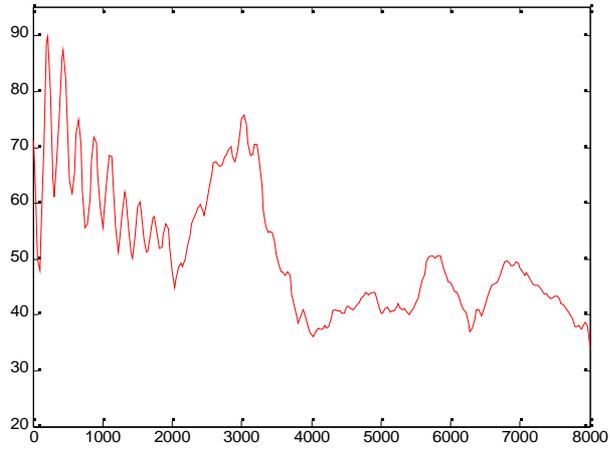
**Bar3 Tres detimbree**



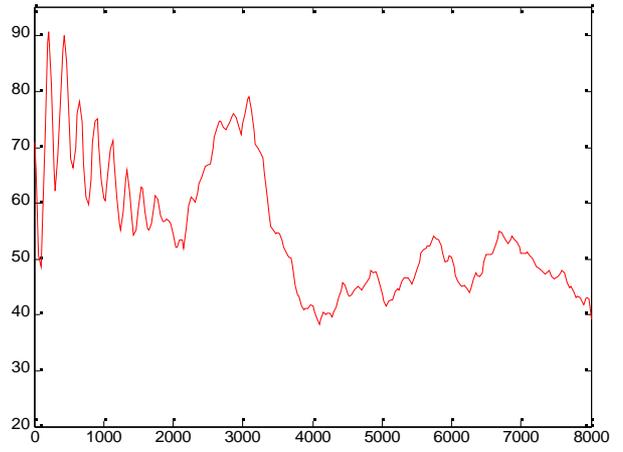
**Bar3 detimbree**



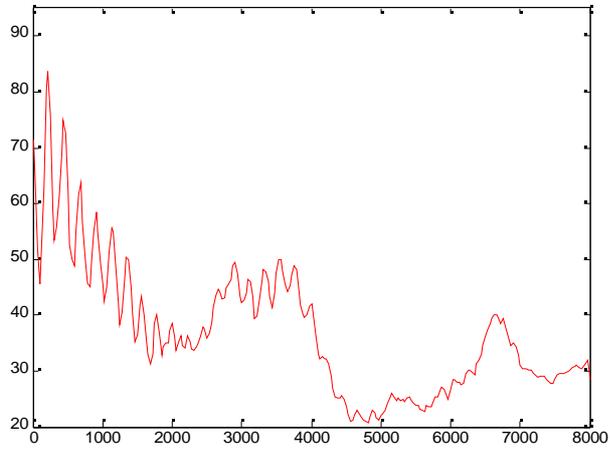
**Bar3 timbree**



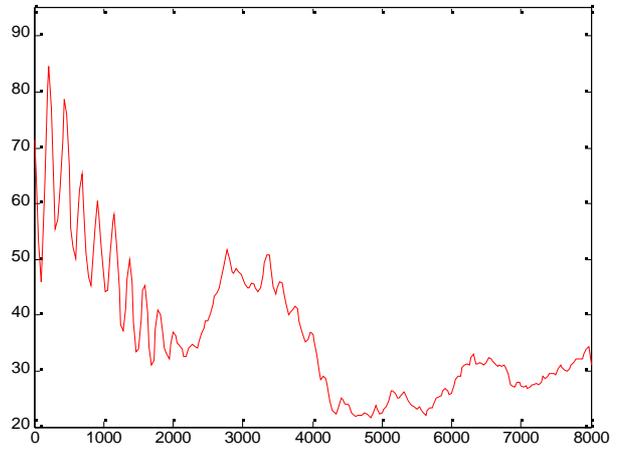
**Bar3 Ires timbree**



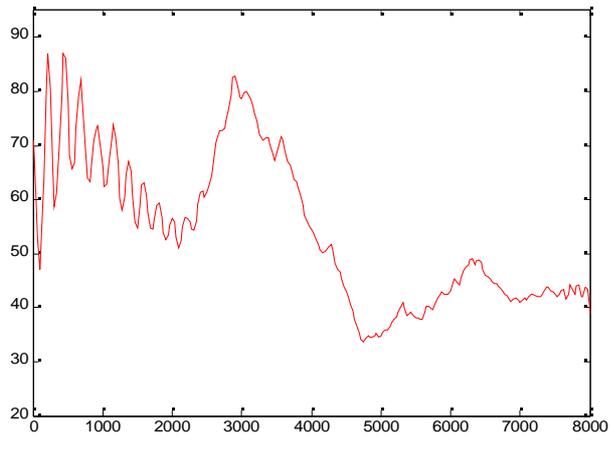
**T1 Tres detimbree**



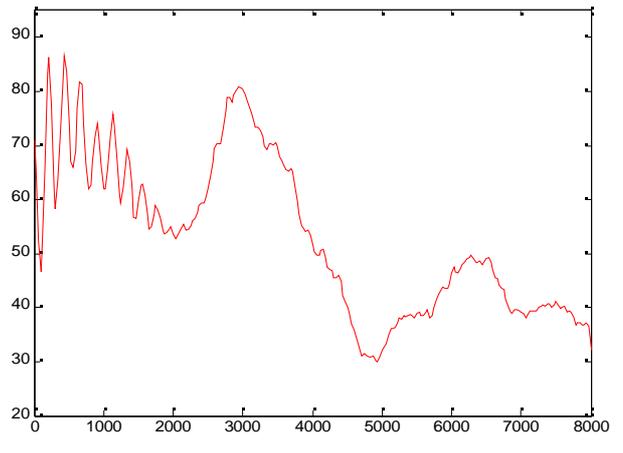
**T1 detimbree**



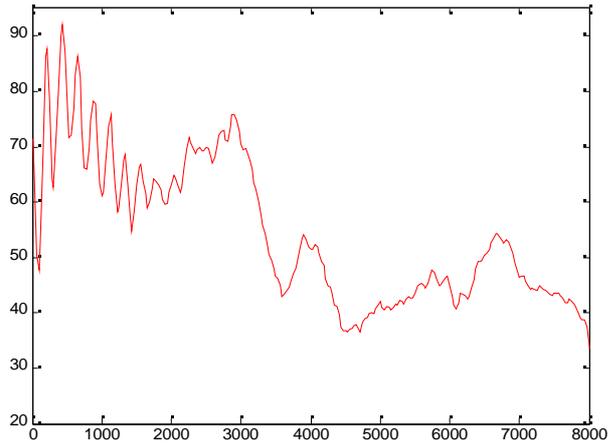
**I1 timbree**



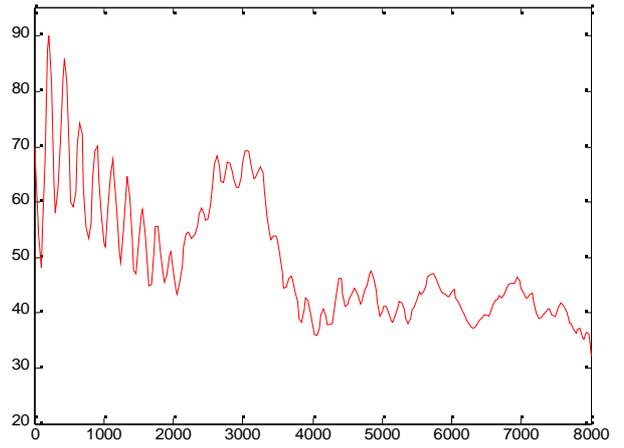
**I1 Ires timbree**



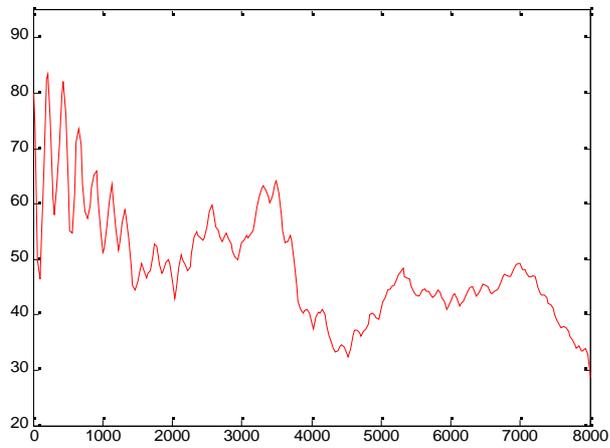
**Bar1 normale**



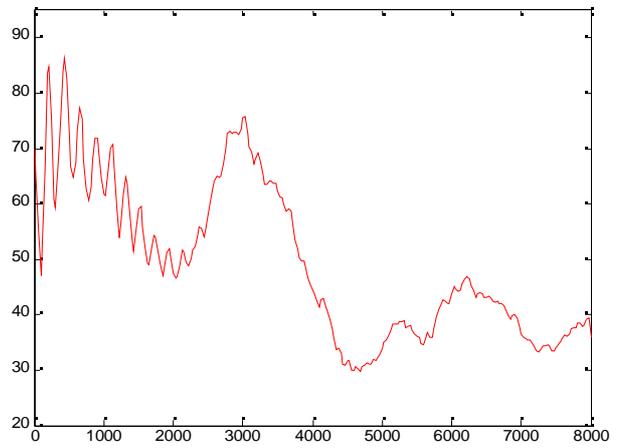
**Bar3 normale**



**Bar2 normale**



**T1 normale**



## **VI.2) Qualités vocales « brillante » et « sourde » :** (cf. spectres moyens en annexe et répartitions spectrales de l'énergie pages suivantes)

Ces deux qualités présentent des caractéristiques similaires à celles de « timbrée / détimbrée » mais ne sont pas tout à fait équivalentes.

Tout d'abord considérons la qualité « brillante » et comparons la répartition de l'énergie spectrale des voix « **brillantes** » avec celle des voix « timbrées ». Pour tous les chanteurs, on remarque que le rôle du formant du chanteur (zone 1800-3000 Hz) est moins important pour la qualité « brillante » que pour la qualité « timbrée », bien qu'il soit quand même une caractéristique de la voix « brillante » par rapport à la voix « normale ». L'examen des émissions « très brillantes » va aussi dans ce sens en révélant que le formant du chanteur n'augmente pas significativement entre « très » et « un peu ». Néanmoins on peut noter l'importance toute relative de la bande 1800-3000 Hz pour Bar1 (~3% pour « sourde », ~10% pour « normale » et ~14 % pour « brillante ») et T1 (~4% pour « sourde », ~12% pour « normale » et ~ 15% pour « brillante »).

La voix « brillante » se distingue bien de la voix « timbrée » par l'importance de la zone de 4000 à 7000 Hz que l'on peut observer sur les spectres moyens. En effet, si « timbrée » amène plus d'énergie dans ces zones que « brillante » (cf. par exemple T1), on voit suivant les chanteurs certaines zones précises qui semblent avoir une influence sur la « brillance ». En fait des pics ressortent plus que d'autres. Pour Bar2 on voit un réhaussement significatif juste au niveau de 4000 Hz qui sera complètement anéanti pour la voix « sourde », de même autour de la fréquence 7000 Hz on a un pic plus important que pour la voix « normale » et la voix « timbrée ».

On comprend que ceci peut avoir des conséquences sur la perception, en permettant de faire ressortir plus certaines parties du spectre localisées dans l'aigu, d'où une sensation de « brillance ». La voix « timbrée » permettrait au contraire d'avoir une sensation plus répartie dans l'aigu donc un spectre plus riche.

Enfin une petite différence entre les voix « timbrées » et « brillantes » se remarque au niveau des tous premiers harmoniques (cf. spectres moyens). Pour le timbrage, nous observons une tendance à bien remonter H2 voire au dessus de H1. Cette tendance s'observe moins dans le cas de la « brillance », notamment pour Bar2.

Considérons maintenant les voix « **sourdes** ». Nous avons un net changement dans la répartition spectrale de l'énergie par rapport à « brillante ». On remarque tout d'abord le lien avec « détimbrée » dans la décroissance harmonique des 6 ou 7 premiers harmoniques (cf. spectres moyens). De plus comparées aux voix « normales » et à ce que l'on a dit sur la voix « brillante », nous observons que le formant du chanteur joue ici un grand rôle dans la sensation de la voix « sourde ». Il est autant atténué que pour la voix « détimbrée » si ce n'est plus.

On peut remarquer que la voix « sourde » concentre son énergie dans les deux premières bandes de fréquences du spectre où elle dépasse largement les voix « normales » et « brillantes ». Ceci expliquerait pourquoi lorsqu'on demande aux chanteurs leurs impressions sur « sourd », ils parlent d'une impression de grave et d'autre part pourquoi ils ont tendance à baisser (en justesse), dès l'attaque, dans la réalisation des exemples de voix « sourdes » (nous l'avons remarqué à plusieurs reprises lors des enregistrements).

Une des principales caractéristiques de la voix « sourde » semble être pour tous les chanteurs, la diminution des bandes de fréquences aiguës, dans une plus grande proportion que pour la voix « détimbrée ». Celles-ci incluent sur le formant du chanteur qui d'une manière plus importante que pour « détimbrée ». Une fois de plus, on peut noter une petite différence de stratégie entre les chanteurs Bar1 / Bar2 et T1 / Bar3.

Pour les deux premiers, la bande 800-1200 Hz est affectée d'une diminution d'importance supérieure à celle observée pour leurs voix « détimbrées », au profit du grave du spectre : ex :

pour Bar1 les 4 premières bandes de fréquences représentaient ~ 15%, 26%, 33% et 17% pour « détimbrée », elles représentent maintenant pour « sourde » ~ 20%, 30%, 34% et 8% (contre ~ 8%, 19%, 38% et 13% pour la voix « normale » ; ~ 6%, 12%, 38% et 14% pour « brillante »).

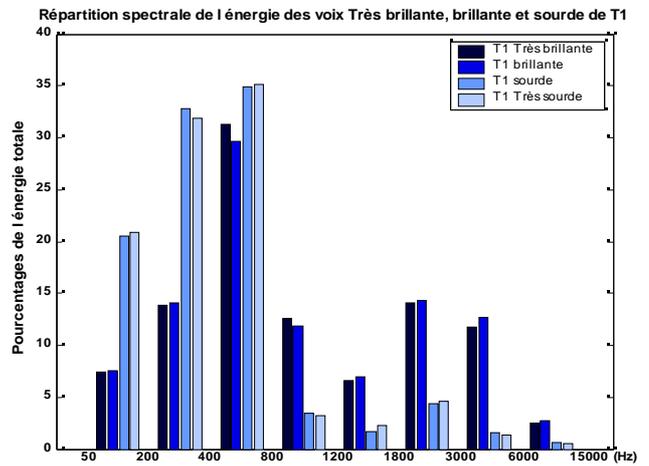
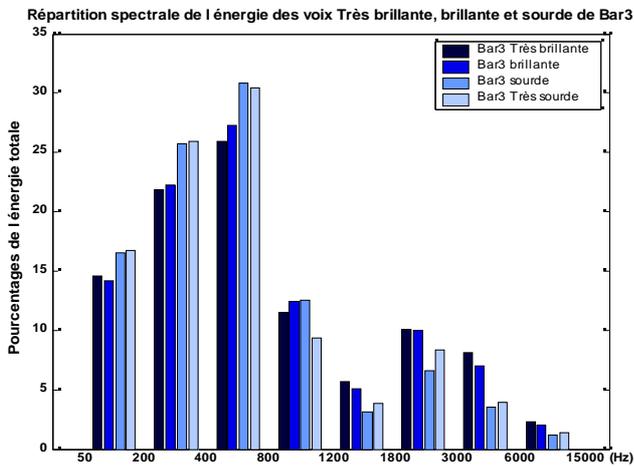
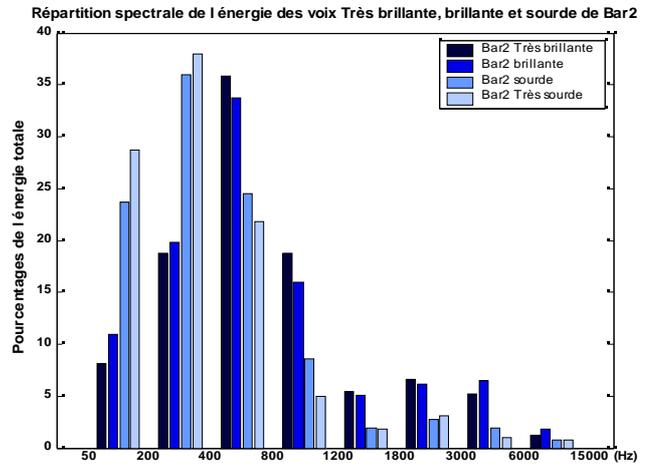
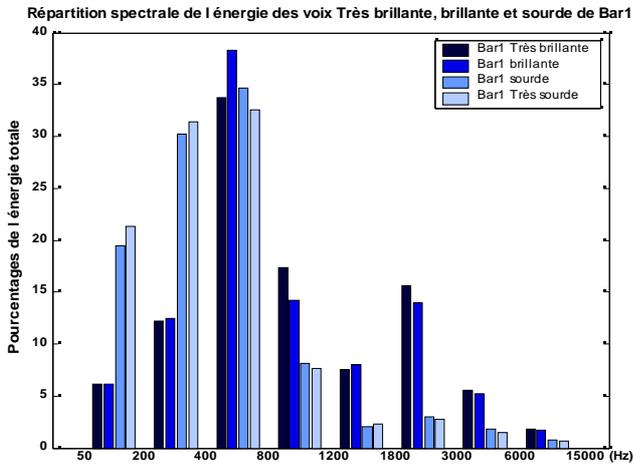
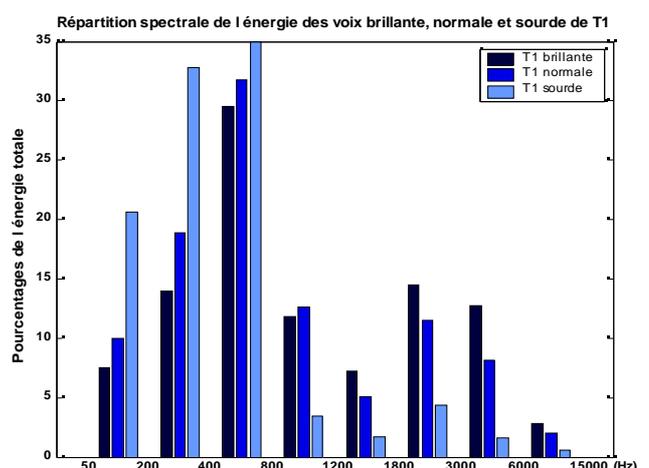
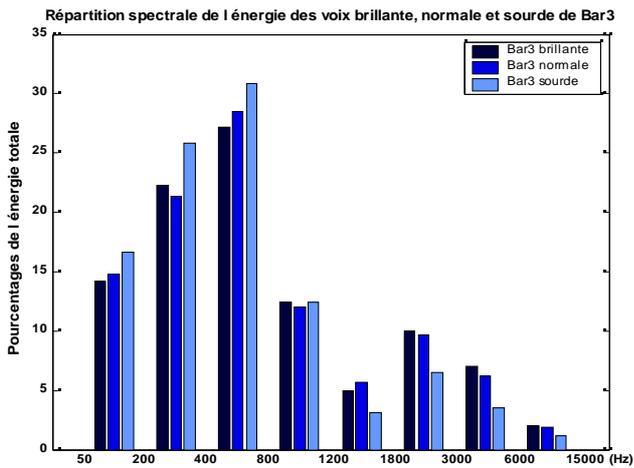
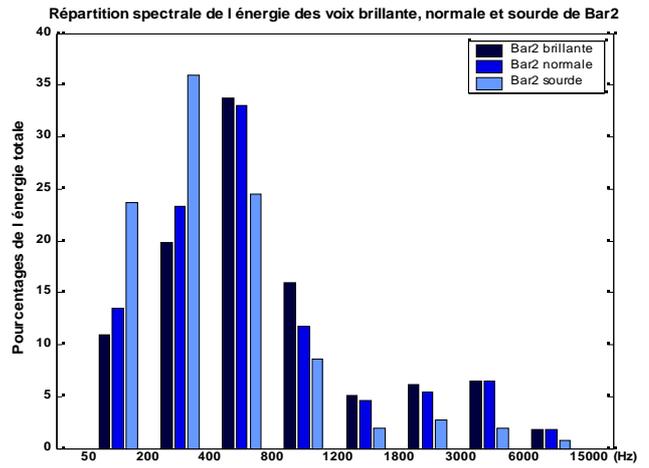
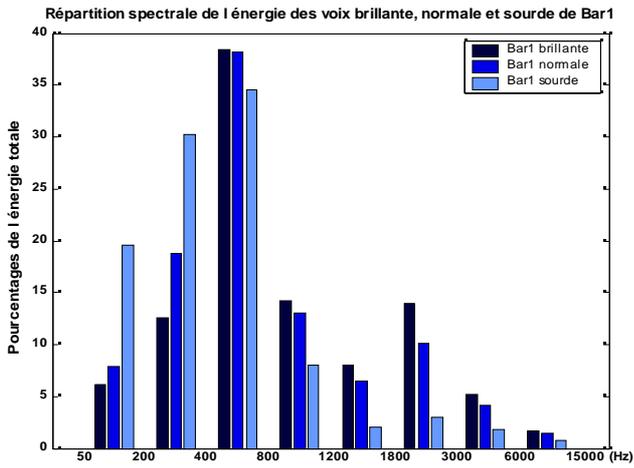
Les deux autres chanteurs renforcent énormément la bande 400-800 Hz par rapport à la voix « normale ». Ils ne réalisent cependant pas le même décalage des fréquences : pour T1 la diminution est très importante concernant la bande 800-1200 Hz comme pour Bar1 et 2 ; pour Bar3, c'est la répartition entre les 3 premières bandes qui est plus étalée (on le voit en comparant avec sa voi « détimbrée » qui privilégie plutôt les deux premières). Ainsi pour ce chanteur la bande des fréquences comprises entre 800 et 1200 Hz ne semble pas très affectée par la « voix sourde » ni par la voix « brillante », en terme de pourcentage énergétique.

Etrangement, la voix « très sourde » ne diminue pas davantage que la voix « sourde » les fréquences aiguës à partir de 1200-1800 Hz jusqu'à 15000 Hz. Pour 3 des chanteurs les zones 1200-1800 et 1800-3000 Hz (formant chanteur) remontent même un peu. C'est plutôt entre les zones 400 Hz et 1200 Hz que la répartition d'énergie diminue encore pour le caractère « très sombre », au profit des premières bandes de fréquences. On voit ainsi que ces deux bandes (400-800 et 800-1200 Hz suivant les chanteurs) jouent un rôle plus grand que lors des émissions « timbrées » et « détimbrées ».

Tout se passe comme si le caractère « sourd » causait en premier lieu la diminution des fréquences aiguës incluant le formant du chanteur, dans une proportion supérieure à « détimbrée », puis pour augmenter le caractère « sourd » encore un peu, ce sont les fréquences avant le formant du chanteur qui diminuent en proportion.

La voix « sourde » diminue aussi comme « détimbrée » le rapport H2/H1 (cf. Bar1).

Si l'on prend l'exemple de **Bar2** : « sourde » diminue la zone 2000-4000 Hz d'environ 40 dB par rapport au premier harmonique, mais il reste quelques zones privilégiées dans l'aigu, 4000 Hz, 5000 Hz, 6000 Hz et 7000 Hz. La voix « très sourde » va masquer encore plus la zone 4000 Hz et 5000 Hz. Pour **Bar1**, l'apparition de pics entre 4000-5000 Hz puis autour de 6000 Hz et de 7000 Hz est bien marquée. Bien que toute cette zone bénéficie de moins d'énergie pour les voix « sourdes » que pour les voix « brillantes » et « normale », cette observation marque une légère différence avec « détimbrée », et diffère aussi de la voix « normale ». C'est peut-être une raison pour laquelle la voix « sourde » se distingue en général assez bien de la voix « détimbrée » même dans les tests de perception.



# CONCLUSION

Nous avons tout d'abord abordé le sujet d'un point de vue perceptif et musical, en cherchant à comprendre les critères importants intervenant dans la perception de qualités vocales dans le chant lyrique. Cette étape a abouti à un test de catégorisation libre sur l'écoute d'une ancienne base de données. Elle a mis en évidence certaines stratégies d'écoute de la qualité vocale selon les auditeurs, des critères de regroupement ou de discrimination de qualités vocales, mais aussi les limites de cette ancienne base de données qui n'était à l'origine pas destinée principalement à l'étude de la qualité vocale.

Ce stage a permis d'enregistrer une nouvelle base de données de chanteurs lyriques masculins dédiée à l'étude acoustique et perceptive de la qualité vocale dans le chant lyrique. Les qualités vocales enregistrées ayant été sélectionnées suite à une étude lexicale des termes usités par les professeurs et chanteurs, cette étude et cette base de données s'intègrent donc dans une problématique plus large à laquelle sciences humaine et acoustique musicale tentent en association d'apporter des éléments de réponse.

Nous nous inscrivons dans un champ d'investigation novateur. Le Laboratoire d'Acoustique Musicale développe des méthodologies originales en ce qui concerne l'étude de la qualité vocale et celle d'instruments de musique.

Du point de vue acoustique, nous nous sommes dans un premier temps concentrés sur l'étude de l'influence de la prononciation des voyelles sur les qualités « claire », « sombre », « antérieure », « postérieure », « ouverte » et « couverte », que l'on a comparées avec les émissions « normales » des quatre chanteurs masculins de la base de données. Il y avait plusieurs possibilités pour analyser les résultats, nous avons opté pour l'observation de deux groupes homogènes de voyelles: /i/, /u/ et /é/ ; /a/, /o « ouvert »/, /o « fermé »/, /ou/. L'analyse acoustique a mis en évidence des corrélation entre les qualités « claire » / « antérieure » et « sombre » / « postérieure », au niveau du déplacement des deux premiers formants vocaliques par rapport à la production « normale » du chanteur. Nous avons également observé des invariants entre les différentes voyelles. Enfin, nous avons remarqué des différences de stratégies entre les chanteurs (déjà au niveau de leur production vocale par rapport à leur voix fry), mais aussi sur le déplacement des deux premiers formants vocaliques pour des qualités. Nous n'avons pas observé de corrélation entre les qualités « ouverte » et « couverte » et d'autres qualités.

Dans un deuxième temps, nous avons analysé la répartition spectrale énergétique et les spectres moyennés sur la totalité du signal des qualités « timbrée », « brillante », « détimbrée » et « sourde » pour déterminer les zones de fréquences renforcées ou atténuées en énergie jouant un rôle dans la perception de celles-ci. Nous avons retrouvé certains résultats d'études antérieures concernant l'importance de l'émergence du formant du chanteur dans les voix « timbrées » et inversement pour les voix « détimbrées » mais sans montrer une réelle opposition entre les deux termes. De même pour la « brillance » et le caractère « sourd » d'une voix, on observe des similarités avec la notion de timbrage auxquelles se rajoutent différentes stratégies entre les chanteurs pour rendre perceptible la « brillance ». Au contraire le caractère « sourd » atténue les aigus de manière assez forte. Nous espérons ainsi avoir participé à l'avancée des recherches sur l'influence de la prononciation dans la qualité vocale.

La principale difficulté rencontrée dans ce stage a été l'intégration, dans un temps assez court, de nombreuses notions, faisant appel à plusieurs domaines de compétences tels que la psycholinguistique, l'acoustique, le traitement de signal .... Indépendamment de la problématique de recherche, ce stage a été l'occasion de se former à de nombreux outils et d'acquérir de nombreuses connaissances très variées (montage sonore, transformation de sons sous

audiosculpt, techniques de prises de son, enregistrements egg, programmation de macros CSL, programmation Matlab, familiarisation aux notions de catégorisation...). Nous sommes loin d'avoir acquis toutes les connaissances nécessaires à la compréhension de la qualité vocale. De plus, de part définition, la recherche est un domaine exploratoire qui conduit parfois à des impasses. Durant ce stage nous avons exploré plusieurs pistes qui n'ont pas toutes abouties à des résultats exploitables, notamment lors des deux tests de catégorisation. Loin de considérer ces expériences comme décevantes, elles nous ont fait bénéficier d'une vision vaste et réaliste du métier de chercheur dans le domaine de l'acoustique musicale.

## PERSPECTIVES

La base de données enregistrées ne va pas être utilisée uniquement pour cette étude et offre de nombreuses perspectives de recherche.

- Tout d'abord, la base de données que nous avons créée est constituée du signal de pression, du signal egg et d'un enregistrement stéréo de qualité. Nous nous sommes focalisés sur le signal acoustique lors de cette étude. Avec la somme considérable d'informations qu'elle fournit, il convient déjà d'approfondir les analyses sur des critères plus précis, ou bien par d'autres outils qui doivent être développés spécialement. De nombreuses qualités restent encore à étudier telles que « métallique », « poitrinée » ... et l'on peut envisager d'étudier les voyelles sur toutes les qualités pour voir lesquelles les affectent le plus (notamment une voix « bâillée » ou suivant l'intensité).
- Une deuxième étude pourra s'intéresser aux mécanismes laryngés pour voir leur influence dans la qualité vocale, utilisant ainsi les signaux électroglottographiques (EGG) enregistrés simultanément avec les signaux acoustiques.
- De même une troisième tâche reste à faire au niveau des commentaires des chanteurs recueillis lors de l'enregistrement. Elle permettra de poursuivre l'analyse psycholinguistique initiée par M. Garnier. Elle apportera sans nul doute davantage de connaissances sur les stratégies adoptées par les chanteurs lyriques.
- Une autre orientation, amorcée dans notre stage, concerne la réalisation des tests de catégorisation libre, à grande échelle, et le tracé des arbres de catégorisation. Ils renseigneront sur les corrélations entre qualités vocales au niveau catégoriel et permettront ainsi d'identifier les corrélats acoustiques responsables de ces associations en dégagant des axes de pertinence à la fois verbale et acoustique. C'est dans ce but que nous avons enregistré les sons en stéréo pour les destiner à être diffusés à des auditeurs lors de tests perceptifs.
- Il reste aussi à aborder d'autres aspects de la qualité vocale comme la posture, le geste musical, et de les intégrer à un contexte musical, notamment dans les relations qu'ils peuvent entretenir avec d'autres instruments, avec l'orchestre, avec d'autres voix ou en s'intéressant à la production vocale d'un chœur.
- Il serait également souhaitable de poursuivre l'enregistrement du protocole avec d'autres chanteurs, de manière à pouvoir tirer des conclusions statistiques de nos analyses acoustiques. La difficulté étant que l'étude de la qualité vocale nécessite une grande maîtrise de la part des chanteurs. Il n'est donc pas aisé de convaincre des chanteurs de très bon niveau de participer bénévolement à ce test. On peut en outre imaginer de réaliser une étude similaire sur des voix de femmes, d'enfants ou de contre-ténors.
- Enfin, toutes ces recherches pourront être confrontées en réalisant une base de données mixte

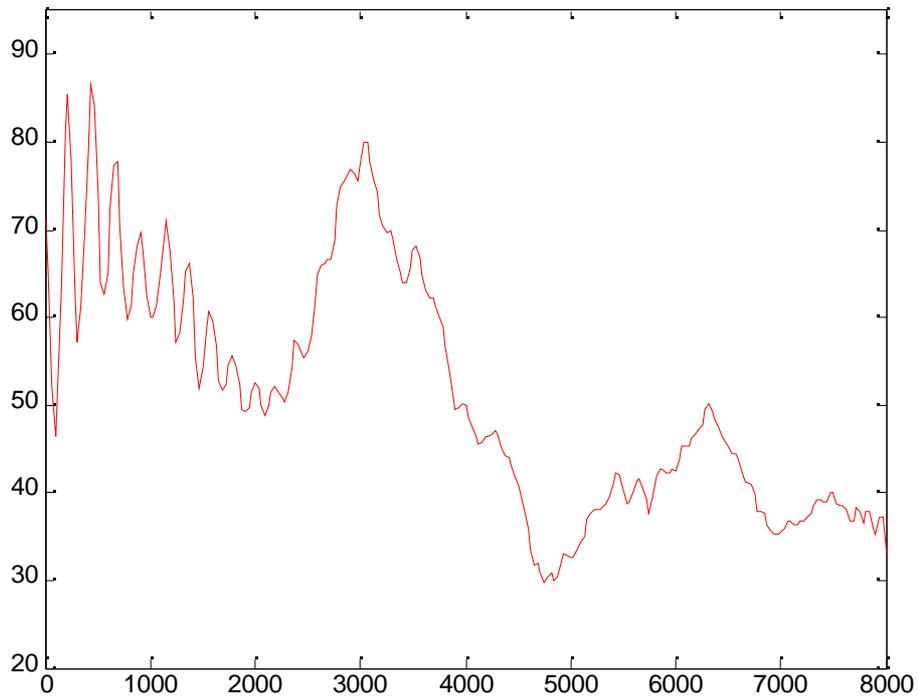
composée de signaux resynthétisés, contrôlés par les paramètres identifiés lors des études précédentes, et de signaux réels. Cette validation des descripteurs acoustiques pourra se faire au niveau perceptif par des tests de catégorisation ou sur des échelles d'évaluation. On pourrait imaginer d'ailleurs un test évaluant les voyelles séparées des exemples des chanteurs associées à des voyelles synthétiques.

# BIBLIOGRAPHIE

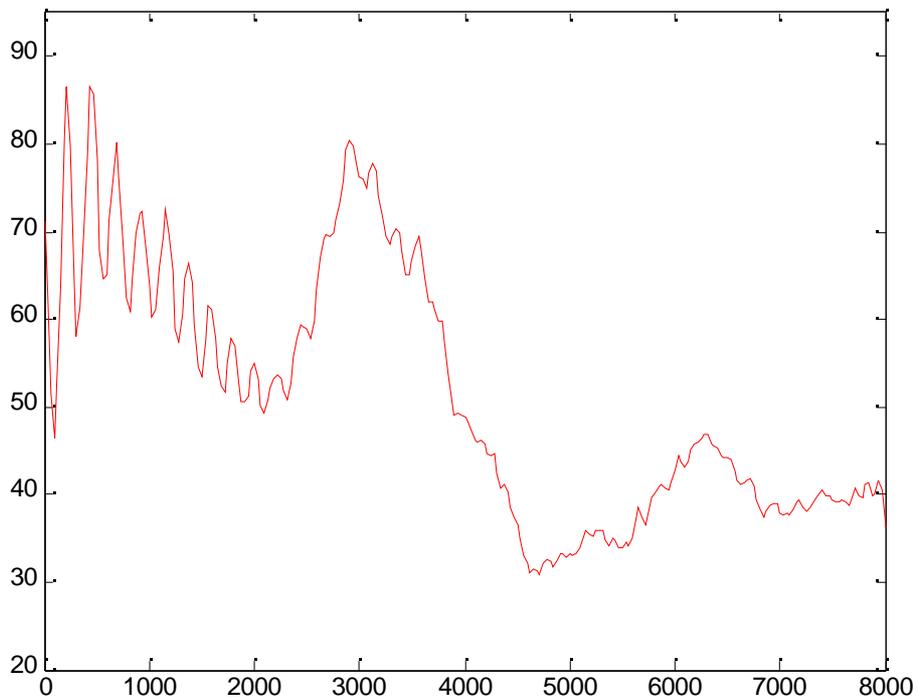
1. Bloothoof G., Plomp R. (1983), « Spectral analysis of sung vowels », *Journal of Acoustical Society of America*
2. Castellengo M. (1973), « Particularités acoustiques de la voix des chanteurs professionnels », *Bulletin du G.A.M.* 67
3. Castellengo M. (2004), cours de psychoacoustique du DEA ATIAM
4. Chapman Aua J. (2003), « Voice quality and the singing voice », ISCA, Voqual'03
5. Childers D.G., Lee C.K. (1991), « Vocal quality factors ; analysis, synthesis and perception », *Journal of Acoustical Society of America*
6. Chuberre B. (2000), « Les registres et passages dans la voix chantée », Thèse
7. Epps J. et al. (1997), « A novel instrument to measure acoustic resonances of the vocal tract during phonation », *M.S.T.* 8
8. Fant G. (1960), « Acoustic theory of speech production »
9. Garnier M. (2003), « Approche de la qualité vocale dans le chant lyrique : perception, verbalisation et corrélats acoustiques », Stage de DEA
10. Garnier M., Henrich N., Castellengo M., Dubois D. et Poitevineau J. (2004), « Perception et description acoustique de la qualité vocale dans le chant lyrique », *JEP* 2004
11. Garnier M., Dubois D., Poitevineau J., Henrich N. et Castellengo M. (2004), « Perception et description verbale de la qualité vocale dans le chant lyrique », *JEP* 2004
12. Guastavino C. (2003), « Etude sémantique et acoustique de la perception des basses fréquences dans l'environnement sonore urbain », Thèse
13. *Handbook of phonetic sciences*, 1997, 1999
14. Henrich N. (2001), « Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électrographiques, perception », Thèse
15. Klatt D.H., Klatt L.C. (1990), « Analysis, synthesis and perception of voice quality variations among female and male talkers », *Journal of Acoustical Society of America*
16. Kreiman J., Vanlancker-Sidtis D., Gerratt B.R. (2003), « Defining and Measuring voice quality », ISCA, Voqual'03
17. Léothaud G. (1999), « Théorie de la phonation », cours Sorbonne Musicologie
18. Maffiolo V. (1998), « Etude sémantique et acoustique de la qualité sonore de l'environnement urbain », Thèse
19. Miller R. (1986-1990), « La structure du chant », ed. Ipmc Paris
20. Molinier N. (2000), « Analyse, synthèse de la voix chantée en vue d'une étude perceptive du quotient d'ouverture », Stage DEA
21. Sakakibara K-I. (2003), « Production Mechanism of Voice Quality in Singing », *Journal of the Phonetic Society*
22. Sundberg J. (1987), « The science of the singing voice »
23. Winckel F. (1973), « Comment mesurer l'efficacité de la voix chantée professionnelle », *Bulletin du G.A.M.* 67

**ANNEXE : spectres moyennés des qualités « brillante » et « sourde » des 4 chanteurs (axe y = I (dB) ; axe x = F (Hz))**

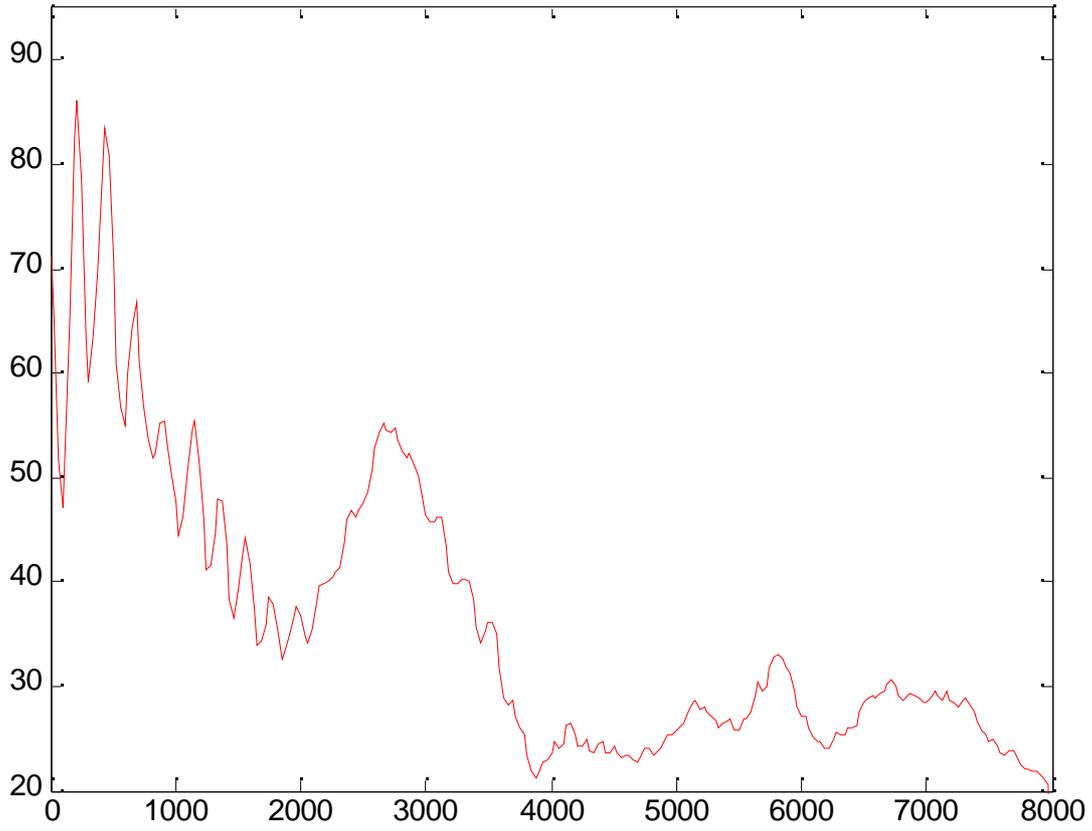
**T1 brillante**



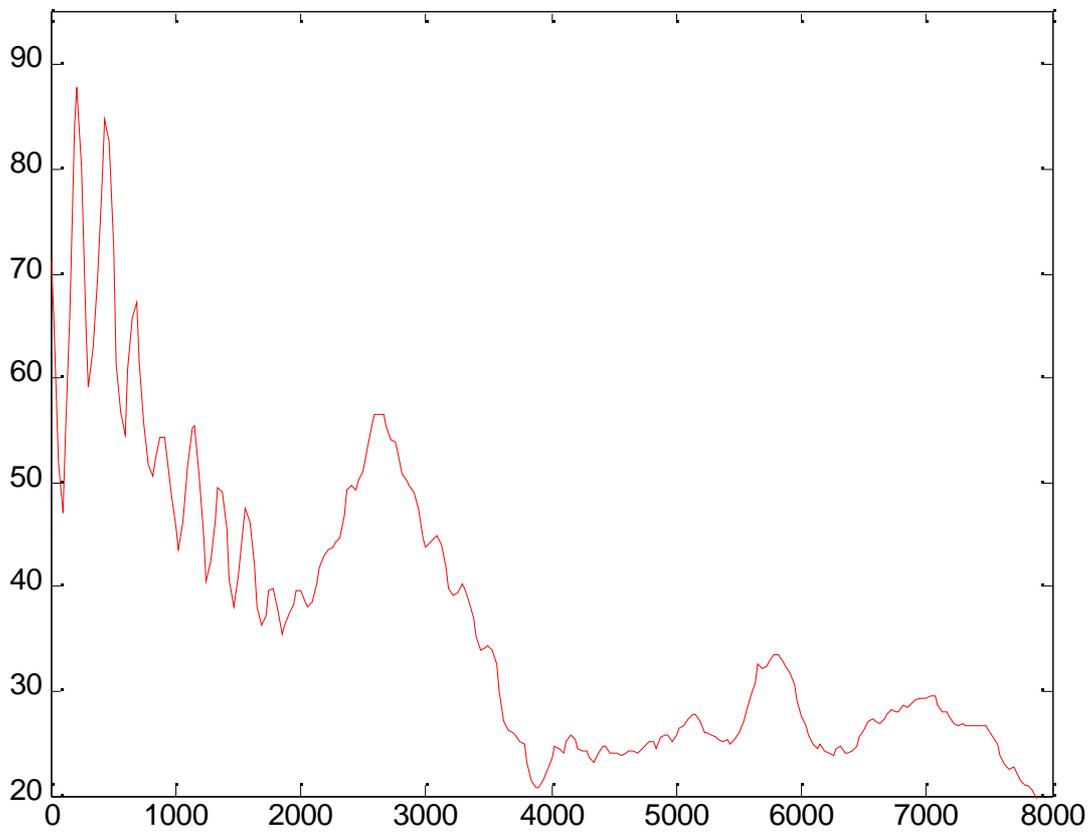
**T1 Très brillante**



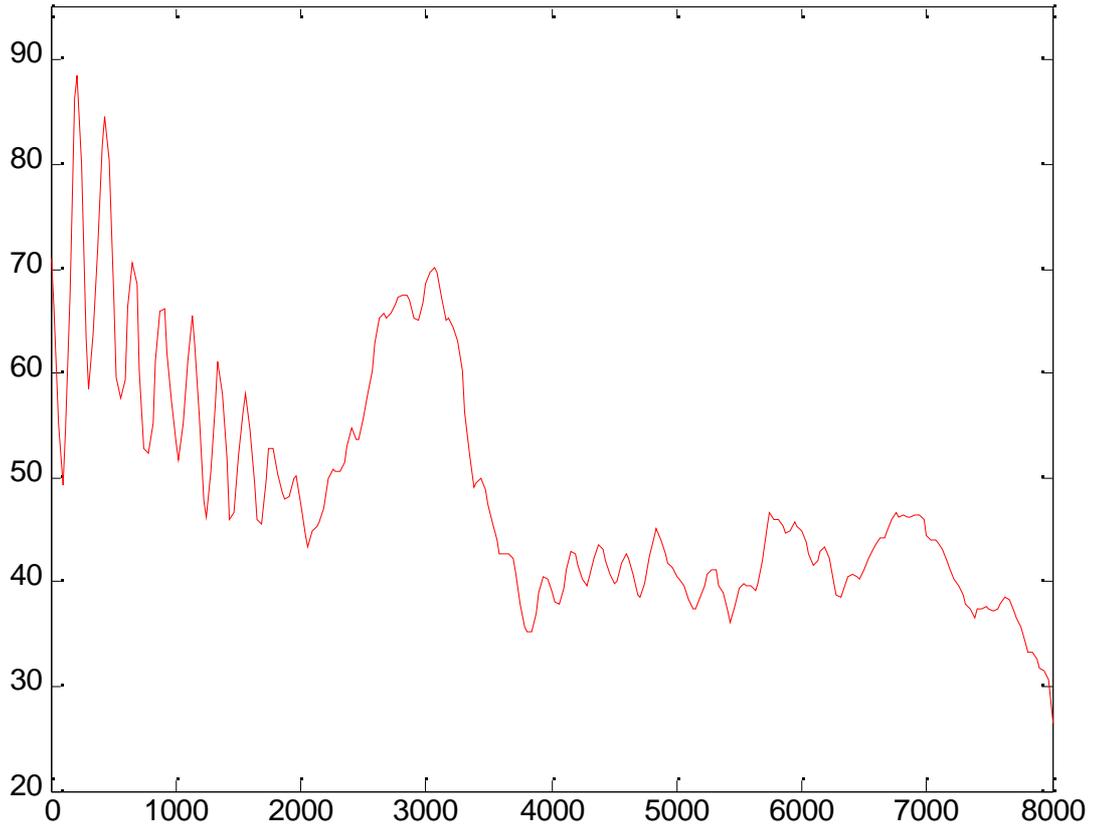
## T1 sourde



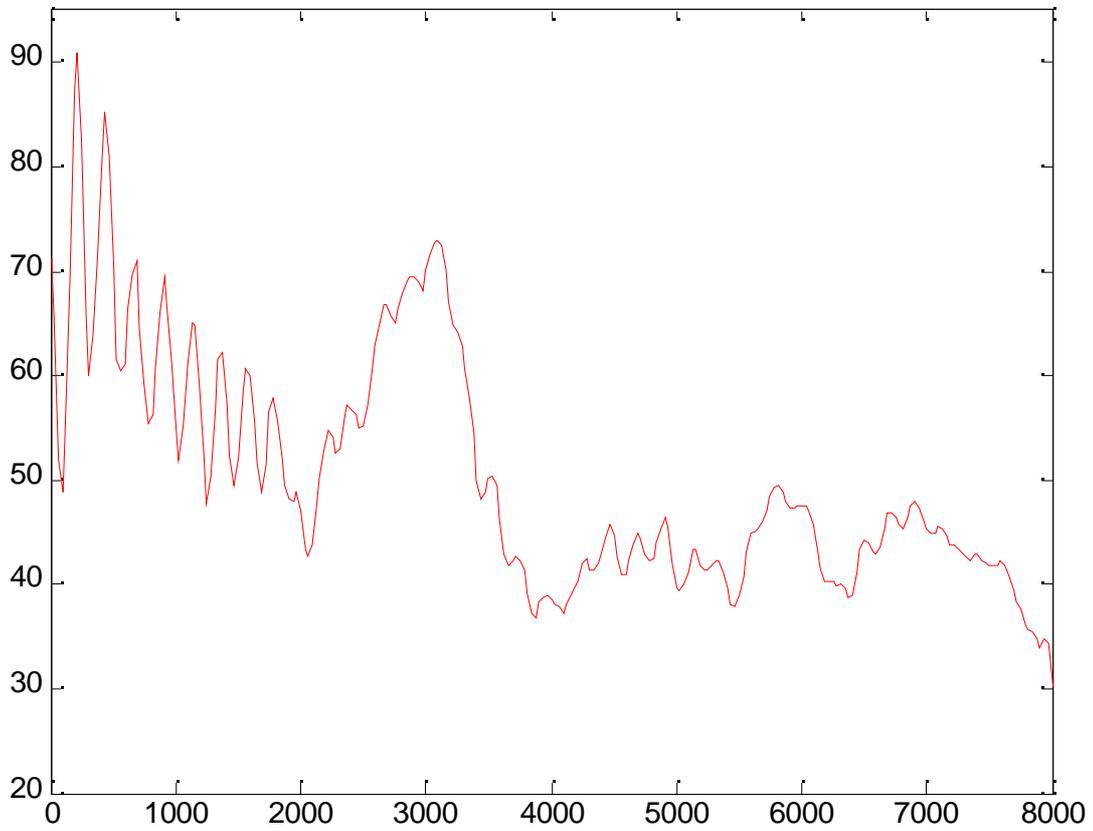
## T1 Très sourde



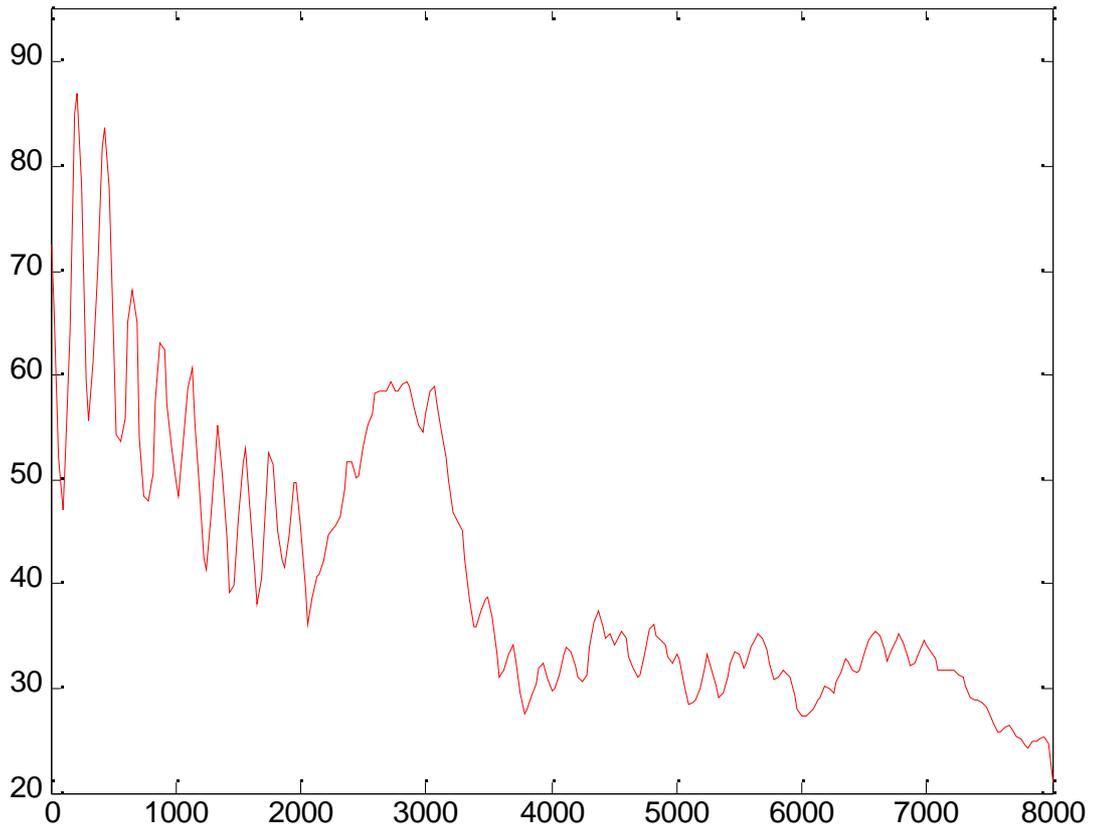
### Bar3 brillante



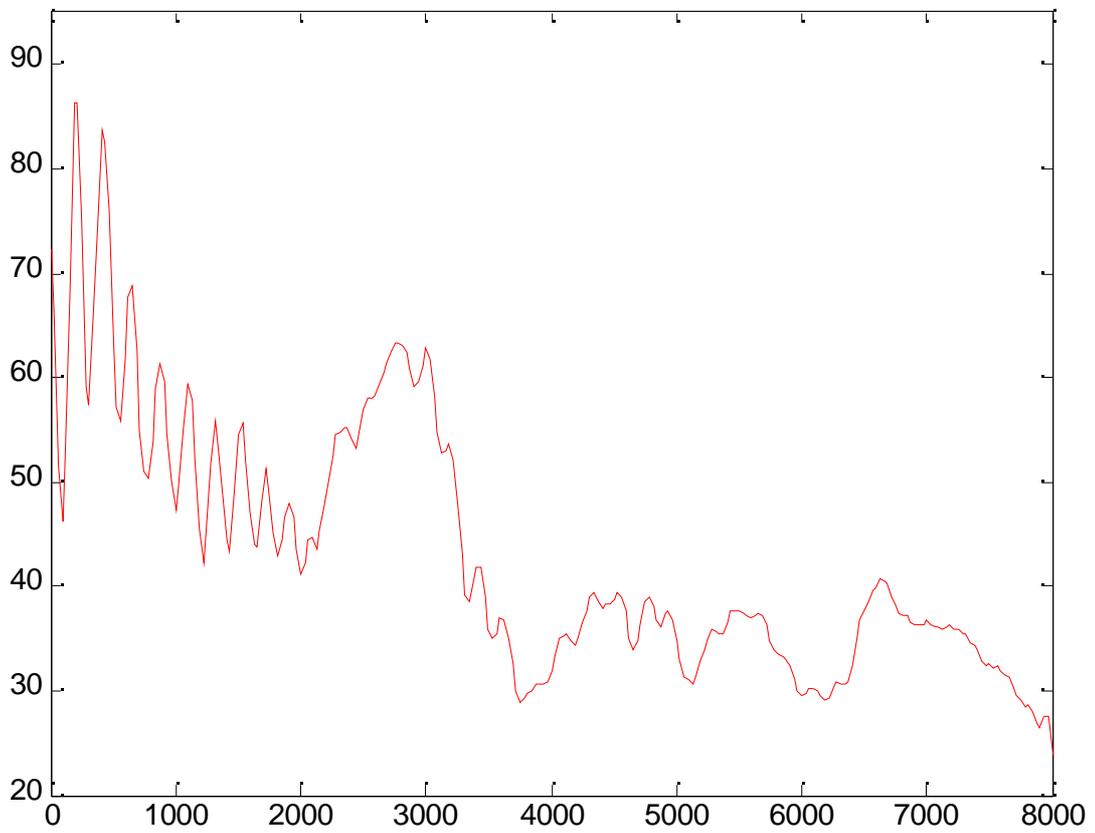
### Bar3 Très brillante



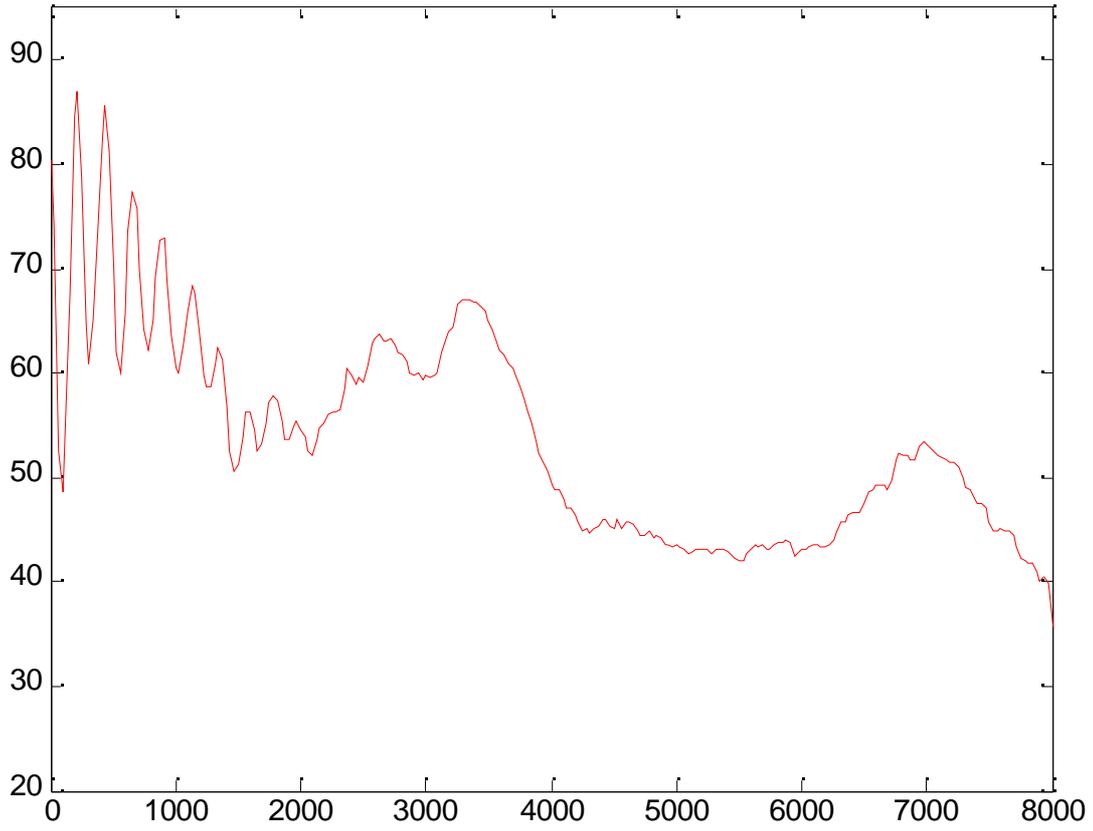
### Bar3 source



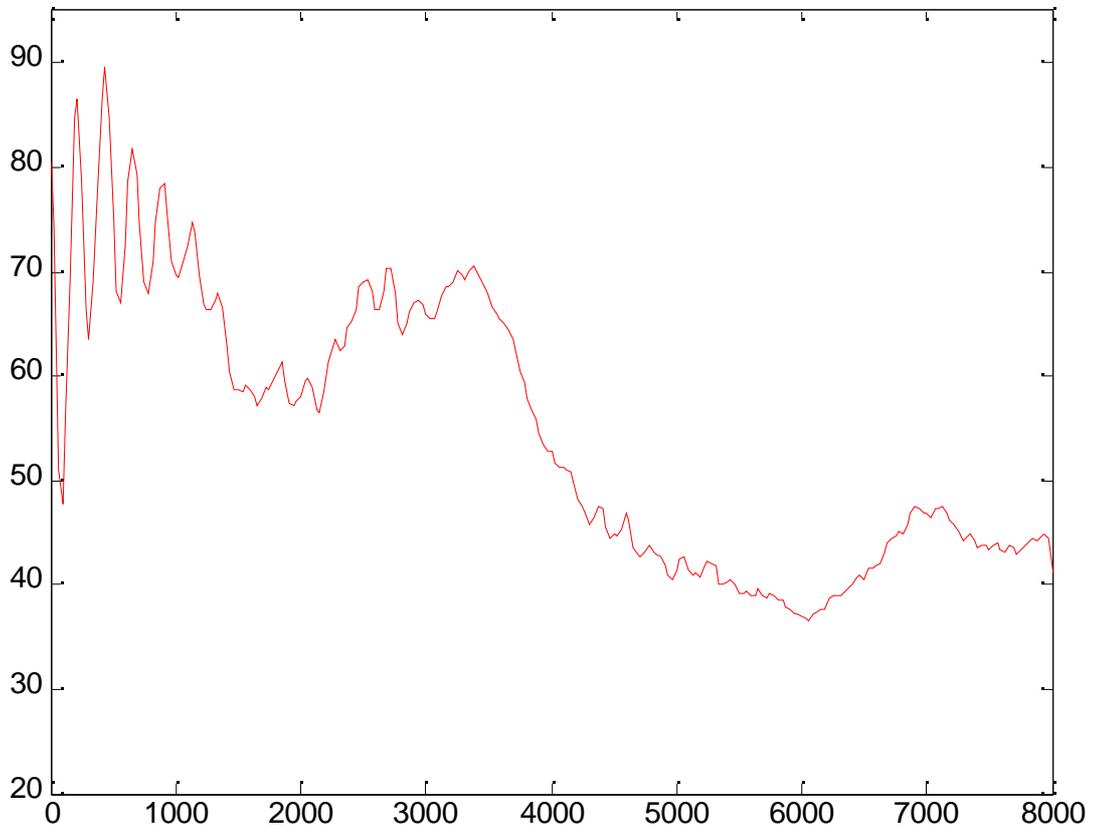
### Bar3 Très source



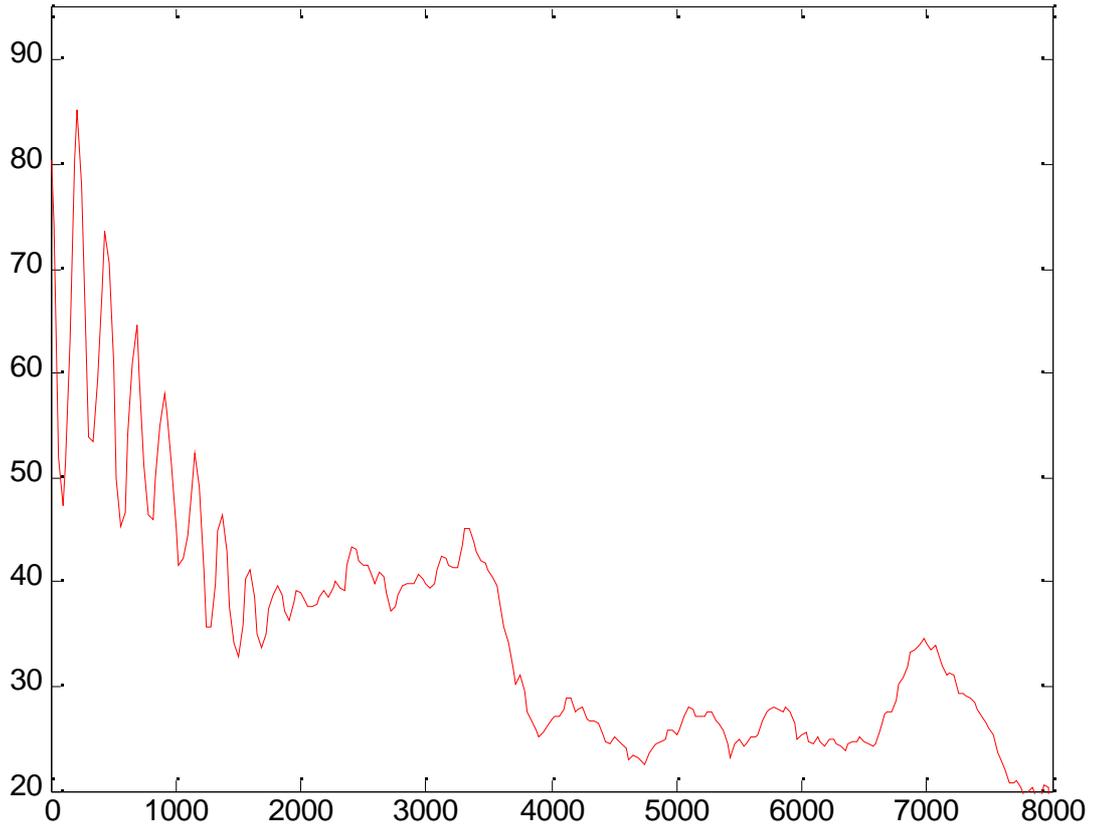
## Bar2 brillante



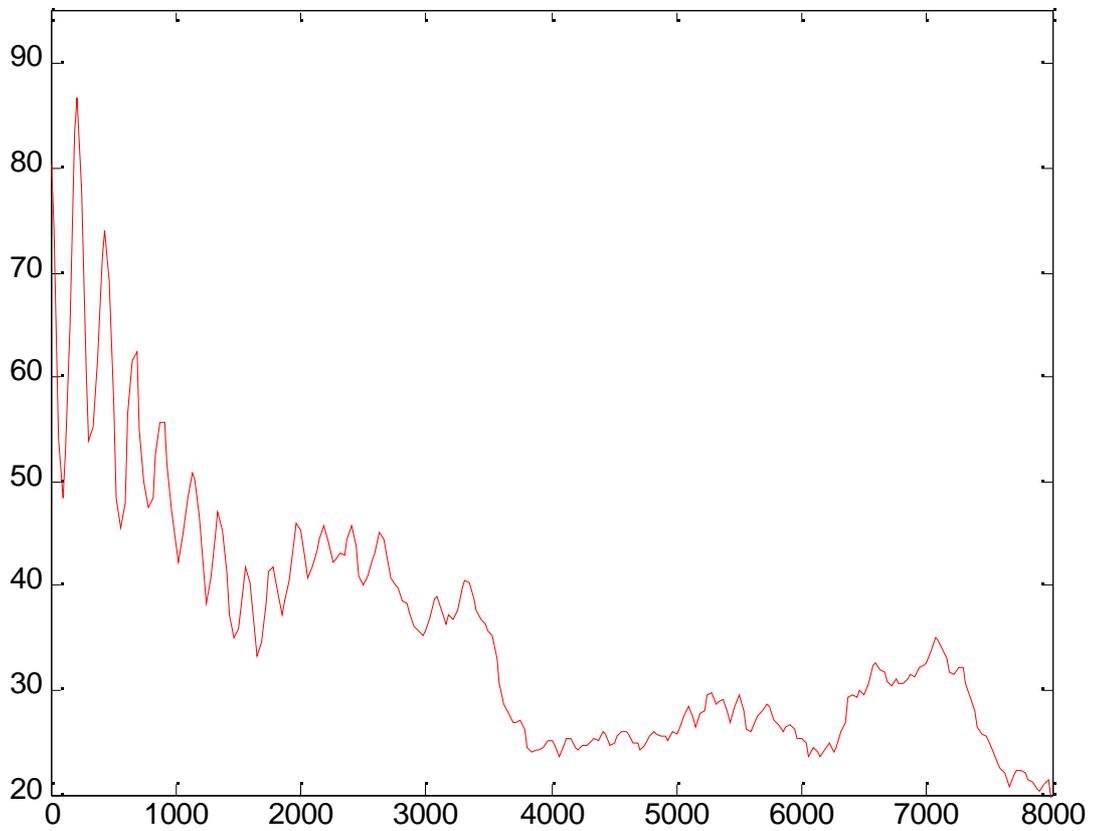
## Bar2 Très brillante



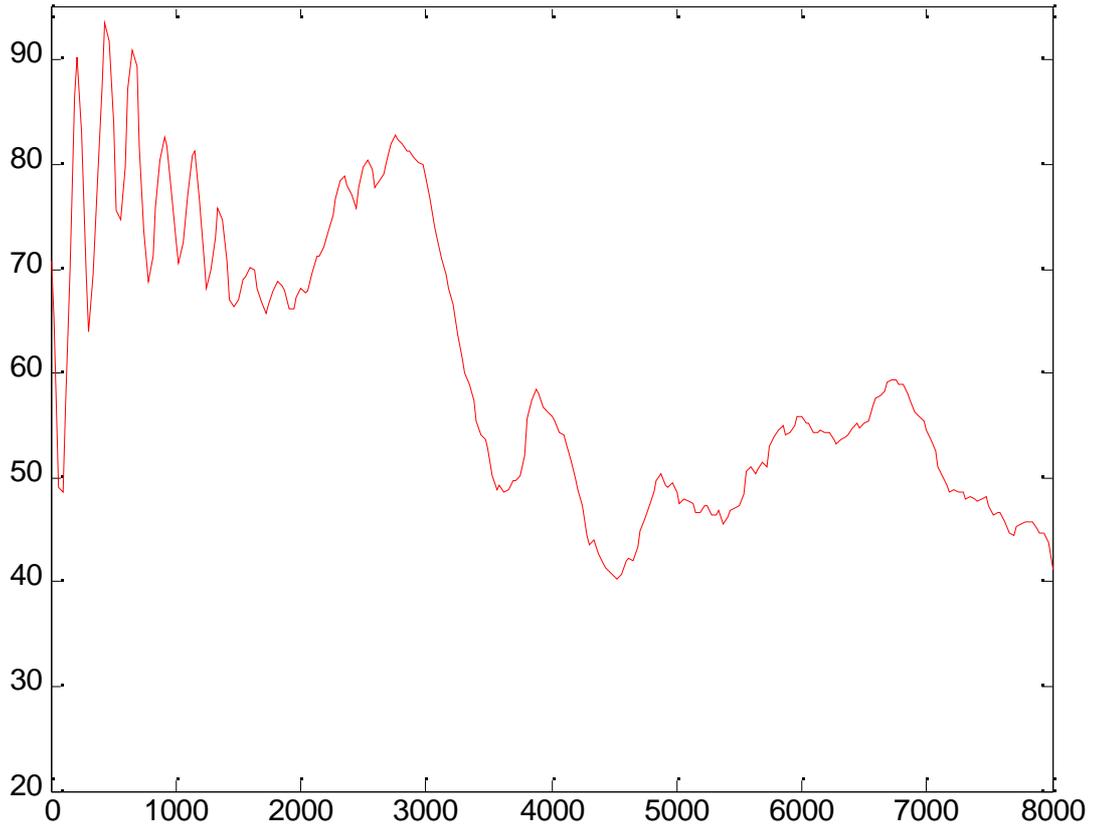
## Bar2 source



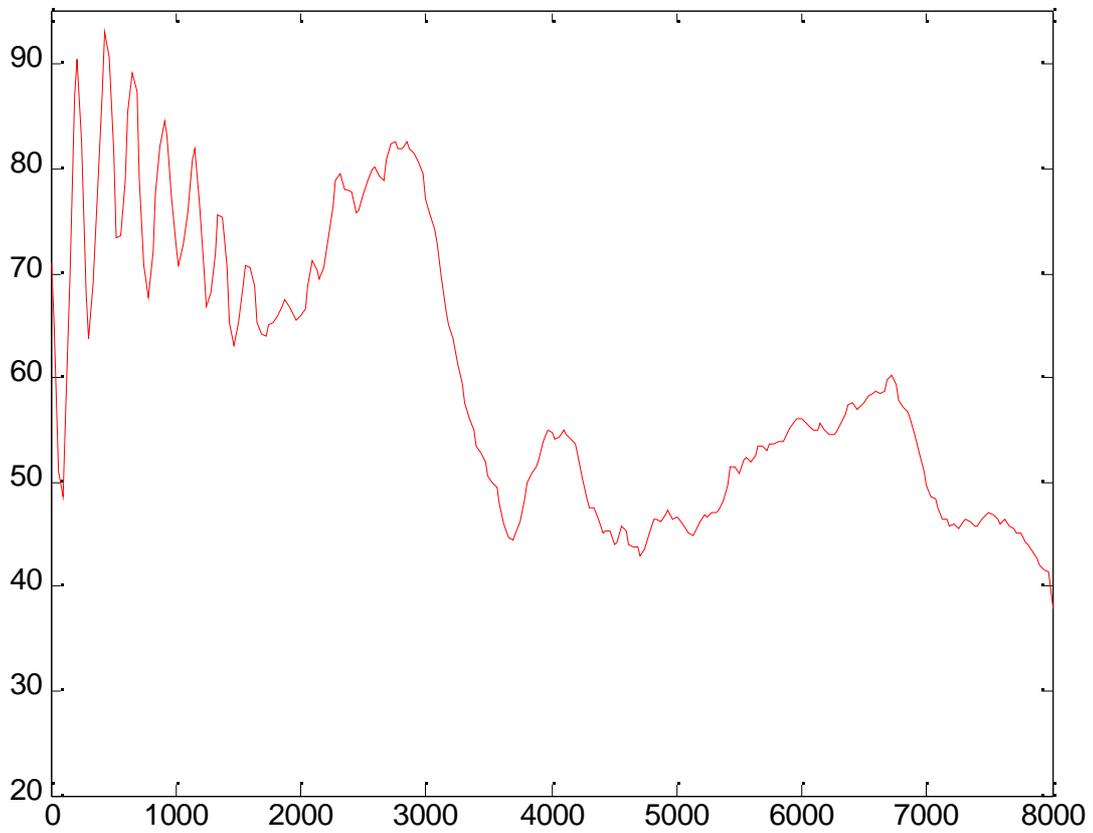
## Bar2 Très source



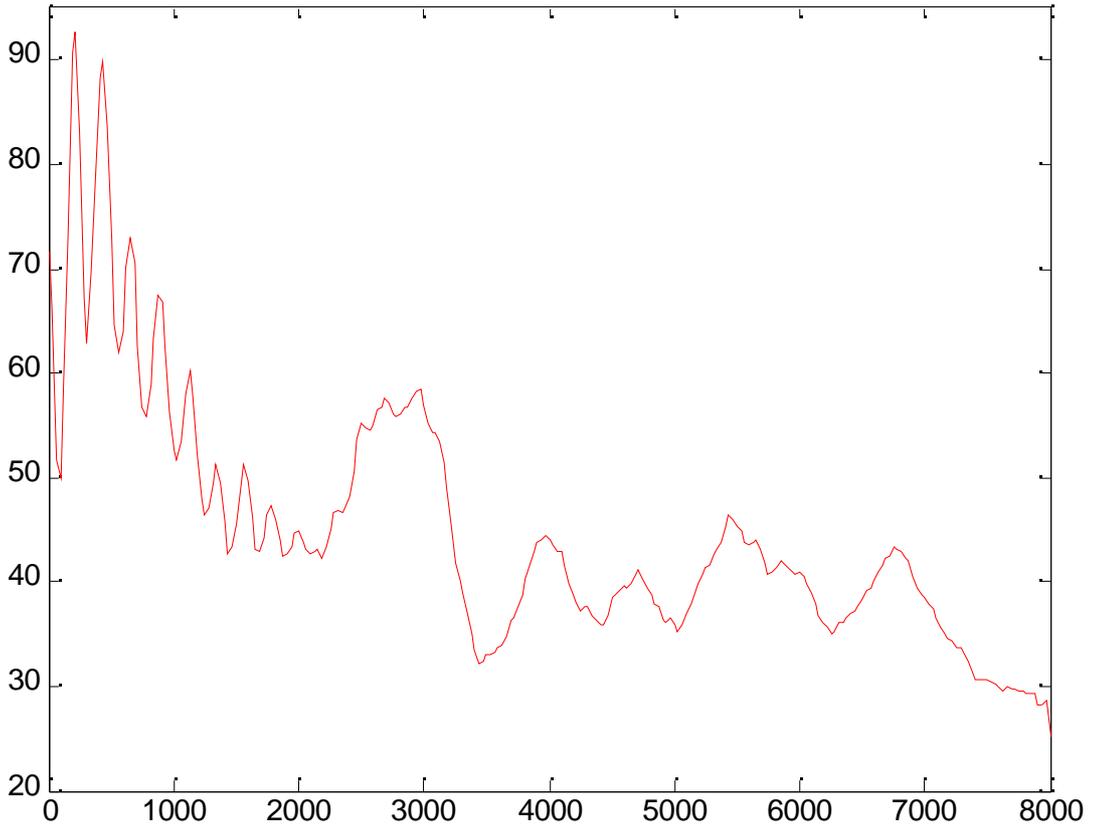
## Bar1 brillante



## Bar1 Très brillante



### Bar1 sourde



### Bar1 Très sourde

