



Grégory Beller

Janvier 2005

La musicalité de la voix parlée

IRCAM

Equipe Analyse Synthèse

Groupe de recherche : La voix

● Plan

Description sommaire des différentes parties

Introduction

I	État de l'art...	7
I.1	La prosodie:	7
I.1.a	Remarques globales :	7
I.1.b	Le cas du Français :	8
I.1.c	les modèles accentuels de la phrase française :	11
I.1.d	La prosodie dans la synthèse de la parole :	13
I.1.e	Automatic prosody generation using suprasegmental Unit Selection :	15
I.2	Synthèse de la parole :	17
I.2.a	Contexte du stage :	17
I.2.b	Système mis en place TALKAPILLAR :	20
II	Des oeuvres artistiques...	22
II.1	La voix comme matériau sonore :	23
II.1.a	Le rapport musique/verbe dans l'opéra :	23
II.1.b	La voix dans la musique contemporaine :	25
II.1.c	La voix comme processus compositionnel :	29
III	Pièces écrites:	33
III.1	Fo-low :	33
III.2	DisCutIon :	39
IV	Questions ouvertes :	41
IV.1	Qu'en dit notre cerveau...	41
IV.2	Le langage émotif musical :	47
IV.3	la musicalité ?	52
IV.3.a	Peut-on quantifier la musicalité ?	52
IV.3.b	Transformation de la musicalité de la voix	53
IV.3.c	Ecoute(s) :	56
IV.3.d	Entretien avec un conteur, Abbi Patrix :	59
V	Le rapport parole/musique	62
V.1	Vers une identité musicale ?	62
V.2	Similitudes structurelles entre la parole et la musique :	65
V.3	Prosodie musicale :	70

Conclusion

Bibliographie

Autres références

Arborescence du CD-ROM joint

Exemple d'une segmentation temporelle de la parole

• Description sommaire des différentes parties:

- I État de l'art... 7
- Dans cette partie introductive, nous dressons un panorama des différents moyens de synthétiser la prosodie du Français. Après l'énumération de nombreux modèles, nous terminons par celui que nous avons retenu pour le système TTS (Text-To-Speech) développé dans l'équipe Analyse Synthèse, appelé TALKAPILLAR. Cet état de l'art représente une base scientifique et bibliographique non exhaustive de l'étude de la prosodie du Français.
- II Des oeuvres artistiques... 22
- Afin de tisser le lien entre la parole et la musique, cette partie propose un survol de l'emploi de la voix parlée à des fins musicales. Son exploitation par de nombreux compositeurs durant le siècle dernier s'étant accrue, nous proposons deux axes de regroupement: L'utilisation de la voix comme matériau sonore et l'introduction de la voix dans les processus compositionnels. Nous ne tentons pas de dresser une liste complète d'œuvres mais plutôt d'en aborder quelques unes sous forme d'exemples.
- III Pièces écrites: 33
- Sont exposés ici les concepts et mises en œuvre de deux pièces, créées par Grégory Beller durant l'année 2004. Si l'une n'est pas proprement liée à la voix (Fo-Low), l'autre en revanche (DisCutIon) propose l'utilisation directe du système TALKAPILLAR et quelques pistes pour la création musicale à partir de la prosodie.
- IV Questions ouvertes : 41
- Avant de rapprocher l'écoute de la musique à celle de la voix parlée, nous dressons en premier lieu un état de l'art des neurosciences qui laisse entendre qu'un tel rapprochement est possible dans l'état actuel de nos connaissances. Puis nous traitons du langage émotif musical, pont entre la psychophonétique et la musicologie. Enfin nous tentons de définir ce que veut dire la musicalité par des entretiens avec des gens du théâtre et du conte.
- V Le rapport parole/musique 62_
- Cette dernière partie est riche en ouvertures et en parallèles entre la parole et la musique. Les similitudes structurelles et le fait que nous écoutions ces deux modes de communication de la même oreille sont deux exemples cités. En tout dernier lieu, nous tentons de donner des pistes pour de futures recherches tant scientifiques que musicologiques s'appuyant sur l'examen des comparaisons entre la prosodie verbale et la prosodie instrumentale (lié à l'interprétation lors de performances).

• Introduction:

Au début, il y a le son...

« La musique est notre plus ancienne forme d'expression, précédant le langage et l'art; Cela commence avec la voix, et par notre désir accablant de joindre les autres. En effet, la musique remonte à beaucoup plus loin que les mots. Ces derniers étant des symboles abstraits qui véhiculent des pensées factuelles. La musique touche nos sentiments bien plus que ne le font les mots, et elle nous fait réagir de tout notre être. »¹

La voix est depuis toujours, l'instrument de musique le plus utilisé et le plus répandu. En effet, chacun la possède et nul n'a jamais chanté durant sa vie. Certains diront que le chant ne leur a jamais plu et que par conséquent, ils peuvent affirmer n'avoir jamais émis une note de leur bouche. A cela, je leur dirais qu'il est trop tard... Car ils viennent de le dire, et donc, d'émettre un son qui de part sa nature possède une hauteur et une temporalité qui la font ressembler fortement à ce que l'on appelle communément une note de musique. Chanter lorsqu'on parle... Est-ce une hérésie ?

Une chose est sûre, notre organe vocal nous permet à tous de communiquer par l'intermédiaire du son qu'il émet, des informations mises en forme par notre cerveau. Que ce soit par le chant ou par la parole, nous sommes capables d'exprimer nos émotions et bien d'autres choses encore. De manière complémentaire, notre oreille assure la réception du message et notre cerveau le décode. Quelle est alors la véritable nature de cette information lors de la communication acoustique. Est-ce du chant, est-ce de la parole ? Dans tous les cas, le message est sonore. D'où vient alors cette distinction entre voix parlée et voix chantée ?

« La voix chantée est caractérisée par la hauteur, l'intensité et le timbre. La boucle audio-phonatoire a un rôle capital dans la voix chantée. La voix chantée représente en général 90 à 95 %, le silence et la phase de récupération 10 % pendant la performance vocale. La voix parlée peut se diviser habituellement en 2 facteurs temporels: la voix sonore elle-même 85 %, et le silence entre les mots 15 %. »²

Serait simplement une différence dans sa temporalité? Il est évident que non. Les distinctions généralement admises entre langage et musique s'établissent dans leurs fonctions,

¹ Y. MENUHIN, 1979, Menuhin and Davis

² J.M. ALBY, C. ALES, P. SANZOY, 1988, L'esprit des voix

largement considérées comme différentes. Nous y reviendrons plus tard. Mais comment figer un instrument aussi versatile que la voix dans l'une ou l'autre de ces fonctions? Voici par exemple deux de ses capacités:

« La voix est un instrument fondamental de la sociabilité humaine: Elle est le moyen, le canal le plus usité, le plus courant de toute relation entre plusieurs personnes. Plus facile dans la proximité, elle peut cependant également plus ou moins vaincre les distances en jouant sur son intensité. Dans un dialogue, dans une conversation, elle permet d'entamer, d'entretenir, de mettre fin au rapport que l'on a avec l'autre par l'intermédiaire de la parole. »³

« La voix est l'instrument le mieux partagé entre les êtres humains. Il n'existe pas de peuple qui ne chante pas. Chaque société a son chant dans l'espace et dans le temps. La musique vocale est le plus important et le plus ancien de tous les arts. L'histoire de la musique, tout au moins jusqu'au XV^{ème} siècle, se confond avec l'histoire de la musique chantée. »⁴

Instrument de musique pour tous et moteur de notre sociabilité, la voix est multifonctionnelle dans son usage. Mais ne le serait-elle pas aussi lorsque nous l'employons seulement pour la communication verbale. Qu'en est-il de la représentation acoustique du langage ? Reste-il langage lorsque il est exprimé physiquement ? Ou bien est-il son, son musical avant que notre cerveau ne le ramène à l'état de substance sémantique ?

La première partie de ce mémoire met en valeur le travail scientifique que j'ai effectué durant cette année au sein de l'IRCAM. Après avoir décrit scientifiquement ce qu'est la prosodie, nous expliquons comment nous procédons pour la synthétiser en vue de reproduire l'expression d'un locuteur spécifique.

Puis la seconde partie propose de dresser un état de l'art non exhaustif de l'emploi de la voix parlée dans la composition musicale. Nous verrons comment la musique contemporaine s'est appropriée cet instrument, tout en montrant que les liens entre la musique et la prosodie existent depuis plus longtemps.

La troisième partie reflète les pièces musicales que j'ai composées durant l'année et en explique les fondations conceptuelles ainsi que les moyens utilisés pour les exprimer.

Enfin, la quatrième et dernière partie de ce mémoire tente d'initier une discussion sur les liens entre prosodie et musique en ouvrant la problématique sur plusieurs domaines d'étude : Des neurosciences à la philosophie.

J. ABITBOL, 1983, Voyage au centre de la voix

³ A. CHARVET, 2004, La voix et ses métamorphoses dans « les métamorphoses » d'Ovide

⁴ Q.H. TRAN, 1980, Les musiques vocales

Ce mémoire a pour but d'explorer les liens entre la prosodie et la musique où plus généralement, entre le langage et la musique. Sans vouloir dresser des conclusions, il permet toutefois de nuancer des frontières communément acceptées et ceux, en traversant plusieurs domaines d'étude assez éloignés au premier abord. Nous verrons que de nombreux chercheurs et musiciens d'horizons variés se posent des questions qui se ressemblent et les rassemblent. Tout est lié...

I État de l'art...

I.1 La prosodie:

« La voix prend corps qui trahit un sujet et un être parlant. »⁵

Dans cette partie, nous allons tenter de dresser un portrait scientifique de la prosodie sans chercher à la définir. Nous nous sommes largement appuyés sur des papiers rédigés par des spécialistes⁶ en traitement de la parole afin de dresser un cadre le plus rigoureux et le plus précis possible. Ce faisant, nous pouvons tout de même apercevoir qu'il existe plusieurs modèles prosodiques, tous différents de part leurs interprétations et que par conséquent, il n'existe pas aujourd'hui de définition absolue de ce qu'est la prosodie. De plus, nous cantonnons cette description à l'étude des faits prosodiques du Français afin de restreindre notre domaine de recherche.

I.1.a Remarques globales :

« L'intonation joue des rôles multiples dans le langage de tous les jours. Elle reflète la structure hiérarchique de la phrase, et au-delà de la phrase, celle du discours. Elle distingue une question d'une réponse. Elle « désambiguïse » des séquences telle que « Je ne veux pas mourir idiot » (la prosodie doit préciser lequel des deux est l'idiot, le locuteur ou l'interlocuteur). Elle exprime des attitudes, des émotions. Elle n'arrive pas, cependant, à représenter des objets, des structures, des événements. Elle n'a pas de fonction représentative, comme les mots, et pas même une fonction figurative, non conceptuelle comme les gestes. »⁷

La structure prosodique résulte d'interactions complexes entre différents niveaux d'organisation sémantico-pragmatique, syntaxique et rythmique. Elle se manifeste par le jeu simultané de plusieurs paramètres acoustiques : la fréquence fondamentale F_0 , le timbre, l'intensité, la durée des phonèmes. Perceptivement, la hauteur et son évolution, le rythme et le tempo (débit), le registre et le timbre mais aussi les pauses et les silences nous permettent la

⁵ P.L. HASSOUN

⁶ Anne Lacheret-Dujour et Frédéric Beaugendre. "La prosodie du Français". CNRS Langage. CNRS edition.

⁷ I. FONAGY, 1983, la vive voix, essais de psycho-phonétique

compréhension d'informations au-delà des mots prononcés. C'est cette deuxième partie du **double codage de la parole** qui lui confère un caractère "naturel" et évite la monotonie. Elle permet entre autre de véhiculer des informations ectolinguistiques ou phonostylistiques (expressivité, sentiments), de lever des ambiguïtés de sens entre deux phrases phonétiquement similaires et de structurer l'énoncé.

La variation de hauteur est certainement l'indice acoustique le plus important dans la prosodie. Le registre couvert par la plupart des locuteurs est souvent divisible en 4 niveaux perceptivement distinguables : Nous les nommerons:

H+H+ : niveau le plus haut

HH

LL

L-L- : niveau le plus bas

La fréquence fondamentale F_0 évolue dans ce registre. Son évolution au cours du temps décrit des contours. Une phrase est généralement composée d'une suite de contours qui ne suivent pas nécessairement la même orientation de pente. On observe cependant une déclinaison générale qui correspond à un abaissement de F_0 du début à la fin de l'énoncé. La hauteur la plus basse correspond donc à la fin de cet énoncé et constitue ainsi un bon indice de segmentation. Ce phénomène à priori universel est de nature physiologique, mais il est géré par le locuteur à des fins linguistiques; il permet de délimiter la fin d'une phrase syntaxique. Il faut remarquer que l'on ne peut évaluer cette fréquence fondamentale que sur les segments voisés (voyelles et quelques consonnes...). Aussi, nous extrapolons celle-ci durant les segments non voisés afin d'avoir des contours continus. De plus, La hauteur de la voix étant fondamentalement différente selon le locuteur, on ne peut associer aux niveaux décrits précédemment des valeurs de fréquence fixes.

I.1.b Le cas du Français :

Le français est une langue à accent fixe ou accent de groupe (de mot). Elle se distingue ainsi des langues à accent libre comme l'anglais. L'anglais est une langue très musicale, caractérisée par de fortes variations de hauteurs et couvrant une large tessiture. Il utilise principalement les variations de hauteur et d'intensité. Les tons mélodiques sont très difficiles à acquérir pour les Français dont la tessiture est restreinte. D'autre part,

l'organisation rythmique de l'anglais est complètement différente de celle du français. L'anglais est une langue stress timed⁸ où l'accent n'est pas prédictible, mais l'espace entre deux pics accentuels est à peu près stable. A l'inverse, la place de l'accent tonique en français est totalement prédictible puisqu'elle affecte toujours la dernière syllabe du groupe rythmique.

On distingue deux types d'accents qui mettent en relief la phrase :

-L'**accent primaire** (ou tonique) se traduit par un allongement de durée et une variation significative de Fo. Il a une fonction structurante et peut se déduire de la syntaxe.

-L'**accent secondaire** se manifeste par des variations plus subtiles de Fo et de l'intensité. Il a une fonction focalisante, rhétorique ou expressive.

Cette distinction est fondamentale car elle met en valeur la différence fonctionnelle de ces deux accents. Ils sont les marqueurs temporels et acoustiques de deux types de groupes :

-Les **groupes intonatifs** (qui se terminent par un accent primaire). Ils expriment la modalité de la phrase. Ils ne sont pas congruents à la syntaxe mais la syntaxe est congruente à l'intonation.

-Les **groupes accentuels** (qui contiennent un accent secondaire). Ils mettent en relief des mots.

Les groupes intonatifs comprennent généralement un ou plusieurs groupes accentuels. Mais cet imbriquement et cette différence de durée n'impliquent en rien une hiérarchisation entre ces deux éléments car ils ne possèdent pas la même fonction :

-L'intonation :

Expression de la structure, modalité

=>invariance monotonie

Expression normalisée

=>cadre normatif

Adhésion à une structure sociale

Force de cohésion

Globale (phrase)

⁸ Pike, 1947

-L'accent :

Expression individualisée

=>variabilité, subjectivité

Excursion mélodique

=>expression originale

Rupture fondatrice d'une individualité

Force de dissociation

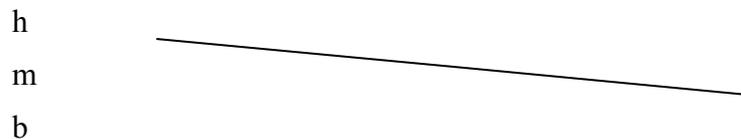
Locale (mot)

L'intonation permet de manifester la modalité de la phrase en français :

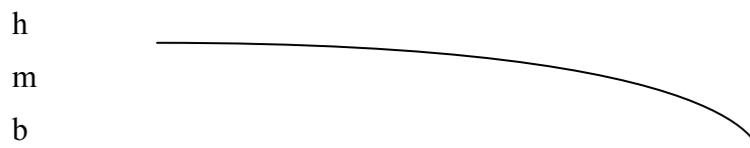
-phrase assertive : contour descendant du niveau haut au niveau moyen



-phrase impérative : contour descendant linéairement du niveau haut au niveau bas



-question partielle ou interrogation : contour courbe descendant du niveau haut au niveau bas



-question totale : contour courbe montant du niveau bas au niveau haut

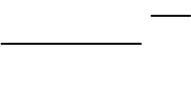


Le laboratoire de morphosyntaxe de Paris III (1991-1997) propose d'affiner cette description par contours en positionnant le locuteur sur ce qu'il dit : Chaque niveau de hauteur étant le reflet de ce positionnement :

- H+H+ : Mise en place de la co-énonciation
- HH : Consensualité acquise
- LL : Niveau neutre
- L-L- : Rupture de la co-énonciation/égocentrage

L'attitude monologique (contour descendant du LL au L-L-) et l'attitude dialogale (contour montant du HH au H+H+) deviennent des descripteurs de contour efficaces pour exprimer les modalités suivantes :

-l'incise (parenthèse) : accélération du débit ()

-la négation : * discordance de point de vue
 * Désengagement du locuteur 

-la question : * valeur neutre et consensuelle

 ? =>  .

* Changement de thème, demande de confirmation

 ? =>  .

* Suscite une réaction, énonciative

 ?

-l'exclamation : * appel à une convergence de point de vue

 !

* Ironie, égocentrage suivant un consensus

 !

* Surprise, discontinuité dans le fonctionnement de la pensée

 !

I.1.c les modèles accentuels de la phrase française :

Les recherches sur la prosodie ont abouti à de nombreux modèles accentuels de la phrase française :

- Ph. Martin (1980) construit des règles fondées sur les corrélations entre relations syntactico-sémantiques et le caractère accentuelle des unités. Des relations de dépendances syntaxiques (simple et à distance) construites “linéairement” (de gauche à droite) permettent de discerner l’unité accentuable.
- S.P. Verluyten (1982) établit une hiérachisation de la phrase en groupes intonatifs (déduits de la syntaxe) et en groupes accentuels (déduits de la longueur des mots) et ajoute à ces règles lexicales des règles de réajustement rythmique pour tenir compte du principe d’isochronie de la langue française (égalité de durée entre deux syllabes non marquées (NA)).
- F. Dell (1984) prédit le degré accentuel grâce à un réajustement rythmique par degré d’eurythmie (qui n’autorise pas que deux accents forts se suivent par exemple).
- Une extension du modèle de Ph. Martin (1986-1987) s’affranchit de la structure syntaxique. Elle propose plusieurs découpages en groupes accentuels et choisit celui ayant la meilleure eurythmie.
- A. Di Cristo et D. Hirst (1993-1996) construisent une grille métrique grâce à des règles eurythmiques permettant l’attribution des accents. Leur démarche diffère dans le sens où ils définissent le rythme comme l’évolution simultanée de la durée et du ton.
- Padeloup (1990) prétend que l’accent est un processus de groupement rythmique. Il instaure une hiérarchie en quatre niveaux allant de la phrase à la syllabe. Quatre règles génèrent en suite la prosodie.
- E. Delais-Roussarie (1995) utilise la théorie de l’optimalité : La génération accentuelle s’effectue en trois étapes :

- Génération de candidats (Groupes intonatifs et groupes accentuels)

- On fait passer ces candidats dans trois modules parallèles de contraintes hiérarchisées (modules syntaxe, rythme et sémantique)

- On évalue le candidat optimal selon une hiérarchie de contraintes.

Tous ces modèles sont issus d’observations et aboutissent pour la plupart à des jeux de règles. Qu’ils partent d’analyses syntaxique, phonologique ou rythmique (psycho-acoustique), ils permettent de mieux comprendre d’où proviennent les paramètres acoustiques de la prosodie.

Cependant, il convient de se demander s’ils sont adaptés à la prédiction

prosodique pour une génération automatique qui se veut naturelle et surtout personnalisée. Peut-t-on envisager une construction de la prosodie par règles dans le cadre de la restitution de la voix d'un locuteur spécifique ? L'élaboration de tous ces modèles vise à obtenir une vision globale et généraliste de la structuration prosodique. Dans tous les cas, ces modèles ont été élaborés dans l'optique de prédire l'évolution des paramètres acoustiques de n'importe quel locuteur. Cela revient à dire que, par conception, ces règles ne peuvent aboutir qu'au caractère normalisé de notre expression.

En effet, de nombreux modèles ne cherchent à prédire que l'apparition des accents primaires, qui sont les indices de la modalité (frontières des groupes intonatifs). Elles ne mènent que rarement aux marqueurs accentuels (accents secondaires) propres à l'expressivité et dont les apparitions révèlent la "personnalité prosodique" de chacun. Une approche par règles nous est donc prohibée si nous voulons restituer dans des phrases synthétisées, la personnalité d'un locuteur spécifique.

I.1.d La prosodie dans la synthèse de la parole :

La synthèse de la prosodie apparaît clairement indispensable pour tout système TTS (Text To Speech) qui désire véhiculer des informations que ne peut contenir les mots seulement. On distingue dans la littérature, trois méthodes pour la génération de la prosodie :

- L'approche par règles.
- L'approche basée sur l'apprentissage à partir de corpus :
 - * par réseaux de neurones
 - * par HMM (Hidden Markov Models)
 - * par d'autres méthodes statistiques...
- L'approche par sélection d'unités.

La connaissance de patrons intonatifs ou contours types permet aux domaines de la reconnaissance et de la synthèse de la parole d'élaborer des modèles de l'intonation française :

- Au CNET (1977-1989) : On étudie un corpus pour en extraire un jeu de règles qui attribue un patron intonatif en fonction de la syntaxe.

- Chez IBM (1971-1980) : On construit un jeu de règles statuant 9 contours types selon le nombre de syllabes, le nombre de mots... On distingue quatre niveaux dans une phrase : phrase, proposition, groupe et mot. L'auteur précise que les niveaux phrase et groupe suffisent pour la majorité des énoncés. Cela revient un peu à négliger les accents secondaires.

- G. Bailly (Grenoble) (1983) : Il segmente aussi la phrase en groupes de respiration, de phonation, de sens. Leurs tailles sont généralement comprises entre 8 et 12 syllabes.

Pour générer les contours, il utilise le modèle de H. Fujisaki.

Le modèle de H. Fujisaki :

La continuité des contours est de nature physiologique. Ils répondent à des commandes discrètes :

- commande de groupe : réponse d'un 2nd ordre à un Dirac.



- commande d'accent : réponse d'un 2nd ordre à un Echelon.



Ce second ordre modélise le muscle crico-thyroïdien (en translation et en rotation). L'avantage de cette modélisation est qu'elle présente des coefficients constants (fittables pour chaque locuteur). Seuls l'amplitude et le temps de déclenchement varient. Les trois commandes de groupe sont : Initialisation, réinitialisation, Finalisation. Ils correspondent à l'expression de la modalité.

- V. Aubergé (Grenoble) (1991-1997) : Création d'un lexique de contours. Il part de l'autonomie entre syntaxe et prosodie. Grâce à un réseau de neurones entraîné sur un corpus, il crée un lexique faisant le lien entre syntaxe et contours prototypiques. C'est aujourd'hui le modèle le plus abouti.

- F. Beaugendre (LIMSI) (1994) : Reconnaissance de contours perceptivement pertinents. 30 règles pour la génération de mouvements standards.

Dans le contexte d'une synthèse par concaténation d'unités (allant du semi-phone au mot ou plus), il semble "logique" de sélectionner aussi des unités prosodiques... Mais ce choix vient en fait de motivations plus profondes. En effet, cette approche permet tout d'abord une plus grande variété prosodique que les approches par règles. De plus elle permet de refléter le "caractère prosodique" de l'individu (chacun ayant ses modes d'intonation,

registre...), ce qui est essentiel compte tenu de notre but artistique : Restituer la voix d'un locuteur spécifique. Enfin, l'introduction de contours réels de Fo sur des blocs de parole permet de conserver la structure micro mélodique⁹.

I.1.e Automatic prosody generation using suprasegmental Unit Selection (Malfrère, Dutoit et Mertens) 1998 :

Le système de l'université polytechnique de Mons que nous allons décrire repose sur la sélection d'unités prosodiques. Il utilise le générateur LIPSS du projet EULER qui génère une description symbolique de la prosodie à partir d'un texte (fichier .txt.mlc) :

-une étude syntaxique donne les accents finals qui délimitent les unités :

-NA : syllabe non accentuée

-AF : syllabe accentuée (accent final=accent primaire)

-UNDEFINED : pause (silencieuse)

-une étude de la modalité donne la hauteur du ton final :

-déclaration : L-L-

-interrogation : HH

-exclamation : H+H+

-temps de pause : P1 ou P2

Ce générateur est appliqué aux phrases de la source comme à celles de la cible. Il permet de créer des unités descripteurs prosodiques de longueurs variables et dont les frontières sont les accents finals. Ainsi, chaque unité descripteur prosodique possède une clé propre représentant :

-L'index de l'unité dans la phrase

-les tons des accents finals de début (qui appartient à l'unité précédente) et de fin d'unité

-le nombre de syllabes neutres, inaccentuées dans l'unité

Cette clé peut ressembler par exemple à : "2 FA1NA1NA2NA3FA2 " ou FA1 et

⁹ A. DI CRISTO, 1982, Prolégomènes à l'étude de l'intonation – Micromélanges

FA2 prennent leur valeurs dans {HH, H/H, L-L-, H+H+, N} dans lequel N représente le début d'une phrase. On ajoute aux clés des unités de la source des marqueurs en liens avec le fichier audio aligné qui nous permettent de retrouver les paramètres acoustiques comme l'évolution réelle de Fo durant l'unité... Le choix de l'unité optimale s'effectue en minimisant une fonction de coût. Comme pour le choix d'unités segmentales, cette fonction de coût résulte de l'addition de deux coûts :

-coût de distance à la cible :

- les « tones » des premier et dernier accents doivent correspondre
- une pondération est ajustée en tenant compte du nombre de syllabes
- une autre est fonction de la position de l'unité dans la phrase

On obtient ainsi une présélection de plusieurs unités candidates.

-coût de concaténation :

Il est seulement basé sur la proximité des valeurs moyennes de Fo de deux unités consécutives.

On aboutit grâce à un algorithme de Viterbi à la sélection finale des unités en choisissant celles dont l'enchaînement présente le coût le plus faible. Puis on va extraire des unités suprasegmentales de la source choisies, les paramètres acoustiques (l'évolution de Fo). Ensuite, on les fournit à l'organe de synthèse (MBROLA) pour que celui-ci applique des transformations élémentaires à l'enchaînement des unités segmentales choisies en parallèle. Ainsi la phrase synthétisée présente une courbe intonative semblable à celle qu'aurait pu produire le locuteur lui-même.

Enfin le rythme est généré par règles grâce à CART, module du système FESTIVAL (système américain).

De manière à s'écarter de ces définitions scientifiques, voici une définition plus littéraire qui nous permet de rejoindre nos questions fondamentales concernant les liens entre la prosodie et la musique.

« Les éléments prosodiques du langage sont ses éléments musicaux, intonation et schéma rythmique. Ils sont conventionnels mais motivés, car leur caractère gestuel n'est jamais complètement aboli. »¹⁰

¹⁰ A. CHARVET, 2004, La voix et ses métamorphoses dans « les métamorphoses » d'Ovide

Le champ sémantique musical utilisé ici montre bien la dimension musicale que possède l'intonation. Lorsque nous parlons de durée, de hauteur, de rythme, d'intensité... Nous sommes en train de parler de notes de musiques.

On pourrait rétorquer que les différences de hauteur n'entrent pas dans le registre musical à cause de leurs faibles ambitus; Que les variations du signal de parole sont trop faibles pour être considérées comme un signal musical. On pourrait le faire, mais ce serait négliger toute l'évolution musicale du siècle dernier. En effet, la musique contemporaine dans son ensemble a vu naître des écritures à registres restreints comme la micro-tonalité (quart de tons, huitième de tons) et des exploitations diverses de la voix parlée pour ses qualités musicales. Nous verrons un peu plus loin de nombreux exemples de son utilisation dans des pièces contemporaines.

I.2 Synthèse de la parole :

Ce chapitre décrit le système réalisé lors d'un stage à l'IRCAM dans le laboratoire Analyse-Synthèse sous la direction de Xavier Rodet¹¹.

La synthèse de la parole par sélection d'unités est aujourd'hui la manière la plus efficace de synthétiser la parole. En effet, l'intelligibilité du résultat confère pour l'instant à cette méthode, un intérêt plus grand que les synthèses basées sur des modèles (modèles physiques paramétriques) ou celles à base de transformations. Aussi désire-t-on maintenant ajouter à la voix ainsi générée, un aspect naturel. Cet enjeu passe par la génération de la prosodie.

I.2.a Contexte du stage :

I.2.a.i de CATERPILLAR à TALKAPILLAR :

La synthèse musicale par sélection d'unités consiste à choisir dans une large base de données les unités sonores les plus appropriées pour construire, par concaténation et

¹¹ Le rapport de stage ainsi qu'une présentation « Power-point » figurent sur le CD-ROM joint.

modification, la phrase musicale à produire. La thèse de D. Schwarz sur ce sujet s'est terminée en 2004.

Elle présente :

- Constitution d'une large base de données par alignement de partitions.
- Création d'un système de gestion et de sélection : CATERPILLAR
- Applications musicales.

Conséquemment à ses travaux, une application en voix parlée a été aussi envisagée dans le cadre d'un projet de reconstitution de la voix d'un locuteur disparu. Ce stage s'inscrit dans ce projet de synthèse de la parole de haute qualité : TALKAPILLAR. Ce projet vise à synthétiser la voix de locuteurs spécifiques (Jean Cocteau et Xavier Rodet) pour rendre audibles des textes jamais prononcés par ces locuteurs.

Il a donc une double vocation :

- artistique, tout d'abord.
- scientifique, dans la mesure où il participe aux recherches effectuées dans ce domaine.

La génération de la prosodie par sélection d'unités est à mi-chemin entre la synthèse de la parole et la synthèse musicale. D'ailleurs, on parle également de prosodie instrumentale lorsqu'on veut décrire des nuances de pitch expressives (vibrato, pitch bend...) ou des variations de durée propres à l'interprétation d'un instrumentiste.

I.2.a.ii synthèse concaténative par sélection d'unités :

Dans un système de synthèse par sélection d'unités, des segments audio de tailles variables sont sélectionnés dans un grand corpus de parole puis concaténés pour synthétiser un signal de parole extrêmement naturel. La première étape indispensable est l'indexation et la segmentation de la source. La deuxième étape consiste en l'évaluation du meilleur candidat correspondant le mieux possible avec la cible.

On distingue actuellement deux tendances pour le système de sélection :

La première, issue des travaux de Black, Hunt, et Campbell (1996) et utilisée principalement par AT&T (US) et ATR (Japan), procède par minimisation dynamique d'une

fonction de coût, estimée à partir de la phrase à produire (et de ses caractéristiques linguistiques) et des phrases enregistrées dans une base de données (ces phrases étant elles-mêmes analysées en fonction des mêmes critères linguistiques que la phrase à produire). La base de données n'est pas organisée de façon particulière. Les unités disponibles ne sont pas regroupées en fonction de leurs similitudes spectrales. Cette approche, utilisée par Diemo Schwarz est à la base de CATERPILLAR.

La seconde, qui résulte d'une thèse de doctorat déposée par Robert Ed. Donovan à Cambridge en 1996, organise au contraire la base de données de façon à pouvoir choisir rapidement l'unité requise, à partir de ses critères linguistiques. Le plus souvent, il s'agit d'une classification en arbre, effectuée une fois pour toutes, lors de la conception du synthétiseur. La taille de l'arbre est représentative de la finesse de la modélisation et peut donc être adaptée à l'inventaire des segments disponibles. La sélection d'unités ne se fait qu'entre classes dont les contextes sont adéquats par opposition à une sélection globale.

Enfin, la dernière étape est la synthèse par concaténation des unités sélectionnées. Les algorithmes de concaténation sont conçus pour modifier les segments sélectionnés (par des transformations de base sur la hauteur et la durée des unités) et les concaténer de façon que les discontinuités (énergie, F0, formants, qualité de la source...) au point de concaténation soient réduites avec le moins d'artefacts possibles pour ne pas dégrader le naturel des segments de départ. Une des dimensions principales qui influe sur la dégradation audio est la distance entre la courbe de F0 originale et la courbe cible. L'algorithme TD-PSOLA (IRCAM) qui est utilisée dans CATERPILLAR possède cette propriété intéressante que si le mouvement est nul, la dégradation de qualité est nulle, ce qui n'est pas le cas de l'algorithme MBROLA (Dutoit 1996) issu du laboratoire de la faculté polytechnique de Mons, qui introduit une dégradation constante quel que soit la modification de F0. Nous utilisons donc TD-PSOLA dans notre système TALKAPILLAR.

I.2.a.iii Remarques individuelles :

Dans la mesure où nous déployons un système qui cherche à tout sélectionner, aussi bien au niveau segmental que supra-segmental, il me paraît dommage d'utiliser encore des règles pour construire le rythme. Aussi choisissons nous de ne pas rectifier les durées issues directement des unités segmentales.

Nous pourrions aller plus loin dans cette idée de proscription totale de règles, en bannissant l'étape fournie par EULER qui visent à déduire de la syntaxe une description symbolique de la prosodie. Il est certain que si cette étape donne une mauvaise description, alors on choisira de mauvaises unités. On pourrait envisager de construire une fonction de coût ne dépendant que des syntaxes des unités source et cible.

Cependant, dans la mesure où cette description ne dépend que du lexique et de la modalité, on peut soupçonner qu'elle traduira bien le caractère normalisé de la prosodie, c'est à dire l'aspect conventionné de notre expression. Comme nous l'avons vu, on est en droit de modéliser par des règles l'apparition de groupes intonatifs car la place des accents finals est systématique. Par contre, on peut difficilement modéliser l'expressivité issue des accents secondaires et de fluctuations plus fines et individuelles. Et c'est en cela que l'approche par sélection d'unités est intéressante. Elle utilise comme descripteurs, les traits communs de tous, pour donner accès aux variations intimes de chacun. Nous choisissons donc de garder EULER, ceci afin de faciliter aussi la description des unités suprasegmentales.

Dans le cadre de CATERPILLAR, de nombreux descripteurs bas niveaux ont été créés et peuvent se révéler très intéressants pour l'évaluation du coût de concaténation : Par exemple la concavité des courbes de F_0 ...

Le fait que l'on ait la liberté de forcer l'apparition d'une unité plutôt qu'une autre est très important dans notre optique artistique. Cela nous permet de choisir perceptivement les contours les plus vraisemblables.

Enfin, je pense qu'il serait intéressant d'effectuer la sélection des unités prosodiques avant de sélectionner les unités segmentales. En effet, une fois un contour choisi, on peut affecter aux unités segmentales appartenant à ce contour, des poids plus faibles que les autres de manière à favoriser leur apparition. Plus largement, il sera préférable de choisir des unités segmentales dont le F_0 sera proche de celui fourni par l'unité prosodique retenue, ceci afin de minimiser la transformation effectuée par l'algorithme TD-PSOLA.

I.2.b Système mis en place TALKAPILLAR :

Le cadre étant défini, rentrons dans les détails... TALKAPILLAR est l'adaptation de CATERPILLAR aux signaux de parole. Nous utilisons donc tout l'environnement créé par Diemo Schwarz comprenant la gestion d'une base de données

relationnelle PostgreSQL (Dbi), Les algorithmes de sélection (Viterbi...)... Pour en savoir plus : voir la thèse de Diemo Schwarz.

Il ne nous reste plus qu'à créer des unités prosodiques. Puis il faut adapter ou créer des descripteurs adaptés aux signaux de parole et à la prosodie. Et enfin, il faut régler leurs poids dans la fonction de coût pour la sélection.

La base de données relationnelle nous permet de créer des relations de congruence entre les unités. Nous allons donc créer les unités prosodiques à partir des unités déjà dans la base, grâce aux descripteurs accents et tons, comme dans le système de Dutoit, Malfrère et Mertens.

Ce système de synthèse nous permet aussi de composer de la musique. Rappelons qu'à la base, TALKAPILLAR est né de CATERPILLAR. Outil créé par Diemo Schwarz pour la synthèse de phrases musicales par concaténation. Nous verrons plus loin qu'il est possible d'utiliser un tel outil afin d'obtenir une « hybridation » entre signaux de parole et signaux musicaux. Ceci sera développé durant la troisième partie lorsque nous aborderons DisCutIon, pièce réalisée à l'IRCAM.

II Des oeuvres artistiques...

« L'harmonie et le rythme doivent s'accommoder aux paroles. »¹²

A la fin de la partie 1 du chapitre précédent, nous mentionnons des pièces utilisant la voix parlée à des fins musicales. Qu'elle en soit le matériau sonore principal, partiel ou bien même sous-jacent à la composition d'une oeuvre, de nombreux paradigmes la plaçant au coeur de la musique ont vu le jour durant le siècle dernier. Mais ce rapprochement n'est pas nouveau et il serait même plutôt originel.

Mondher Ayari travaille dans l'équipe PCM (Perception et Cognition Musicales) de l'IRCAM. Il analyse les musiques extra-européennes. Il m'explique que depuis tout temps, dans la musique classique orientale, le processus primaire de création repose souvent sur un texte. Ce texte lu, le compositeur extrait de la structure métrique de la langue, un paradigme rythmique qu'il exploite durant l'oeuvre.

L'histoire de la musique occidentale est très intéressante à cet égard. Au départ, les chants grégoriens utilisaient les accents du latin pour la mise en place des crescendo et autres « accents musicaux ». Ainsi, le chant occidental semble avoir parcouru durant sept ou huit siècles, un chemin initiatique entre la voix parlée et la voix chantée. En privilégiant le paramètre de la hauteur, l'Occident affirmait la prédominance de la mélodie et son aptitude à la vocalisation instrumentale comme à l'instrumentation vocalisante. Partant de l'intonation verbale, les lignes mélodiques ont quitté la voix parlée progressivement au détriment de l'intelligibilité du texte mais au service de l'expressivité musicale. Durant de nombreux siècles, le chant s'est éloigné de la parole. Mais l'histoire est un cycle... Et voilà que les compositeurs contemporains réutilisent la voix parlée pour sa qualité musicale propre. Explorons dans le temps cette évolution.

¹² PLATON, La république

II.1 La voix comme matériau sonore :

II.1.a Le rapport musique/verbe dans l'opéra :

[Extrait du livre de M. Poizat : Frénésie, entre parole et cri : le chant de la Diva]

« Toute l'histoire de l'Opéra peut être envisagée comme un long cheminement de la parole chantée au cri musical et finalement au cri pur. Ce cheminement s'inaugure avec Monteverdi dans le regard d'Orphée renvoyant au néant d'Eurydice, il trouve un aboutissement avec Alban Berg et un autre regard, armé celui-ci d'un couteau, celui de Jack l'éventreur poignardant Lulu. C'est dans le déroulement non linéaire, de ce parcours qui nous conduit de la triste mélodie d'Eurydice, encore proche du phrasé de la langue, jusqu'au terrible cri de mort de Lulu, que l'on peut cerner cette problématique clé de l'Opéra, celle du **rapport parole-musique**.

Les débuts de l'Opéra se caractérisent avec Monteverdi par une ligne mélodique qui épouse très étroitement le phrasé de la langue. Le souci d'intelligibilité contraint d'abandonner le système polyphonique antérieur et se donne pour projet de mettre au point tout ce qui permet de « parler en chantant ». Mais très vite, cet idéal de fusion laisse la place à une autonomisation radicale de l'aspect vocal musical. C'est, avec l'art des castrats, la visée d'une jouissance de l'objet-voix pur avec ipso facto la chute totale de toute préoccupation quant au texte et au drame. Ces développements, quelque peu paroxystiques par moments, appelèrent par réaction la première réforme de l'Opéra qui se posa comme telle, celle de Gluck. Ils avaient entre-temps alimenté l'étonnante « Querelle des Bouffons ». Ce débat idéologique d'une ampleur inouïe qui cliva complètement les cercles intellectuels et mêmes politiques de la société française du milieu du XVIII^{ème} siècle, justement sur cette question des rapports langue-musique (débat Rameau-Rousseau, opéra français contre opéra italien) dans son articulation avec la jouissance et la fonction morale de l'art lyrique. Gluck quant à lui réintroduit la prééminence du texte et du drame et sa réforme accompagne (historiquement, cela ne préjuge pas d'un lien de cause à effet) la chute des castrats. Ce qu'il est intéressant de repérer, c'est que dans chacune de ces oscillations de la forme, on retrouve à la fois le dispositif de production de la jouissance de la voix et le dispositif de contrôle et de limitation de cette jouissance.

Si, pendant la grande période des castrats, la jouissance vocale atteignait de tels sommets, la scansion de l'oeuvre par les récitatifs était là pour ramener sur terre les

spectateurs (lesquels le prenaient parfois fort mal, exigeant alors qu'on passe immédiatement à tel ou tel autre air, même si c'était l'air du troisième acte alors qu'on n'en était encore qu'au premier !). A l'inverse quand le compositeur se donne pour projet de brider le développement lyrique en l'assujettissant à la scansion signifiante de la parole, il produit une forme musicale dite par exemple « récitatif accompagné » chez Gluck, ce qui deviendra « mélodie continue » chez Wagner, tentant à abolir la distinction air-récitatif et donc la limitation de la jouissance que cette scansion opérait. Or ces ormes musicales ont pour effet de produire une tension, une captation continue source à son tour d'émotion et de jouissance profonde, repérée par tous les commentateurs, contemporains ou non, ceux de Wagner comme ceux de Gluck. Et cela d'autant plus que, même dans les oeuvres composées selon ce projet esthétique, une large place est quand même toujours gardée pour les développements lyriques rompant toute attache avec le verbe et amenant l'auditeur à cette jouissance vocale dont nous parlons. Ce mouvement pendulaire de la Flûte enchantée au Sprechgesang de Schoenberg et Berg, du bel canto italien au récitatif mélodique de Pelléas, jusqu'à devenir lui-même le thème d'un opéra, le capriccio de R. Strauss, créé en 1942, dont l'argument est précisément la discussion des rapports musique/verbe.

Cette problématique du rapport musique-parole ne s'organise pas de façon analogue dans le chant masculin et dans le chant féminin. L'attribution de plus en plus systématique à l'héroïne, c'est à dire à la femme qui sera sacrifiée au dernier acte, de la tessiture de soprano, ainsi qu'à l'intérieur de cette tessiture l'inscription du chant de la soprano dans des zones de plus en plus limites vers l'aigu et le suraigu, ont pour conséquence inéluctable la liquidation de l'intelligibilité de la parole de la femme. On sait en effet qu'il n'est acoustiquement pas possible, à partir d'une certaine hauteur de chant, d'articuler, tout en chantant juste, voire d'articuler tout court, certains phénomènes. Or plus on avance dans l'histoire de l'opéra plus les grands développements lyriques se situent dans cette zone d'inintelligibilité obligée. Or non seulement on ne constate pas la même évolution dans le chant masculin sauf dans certains grands airs de ténor où il subit alors un traitement qui s'en rapproche, mais au contraire il tend et c'est frappant chez Wagner et chez Berg à se réintroduire comme parole pure: Le duo de la scène finale de Lulu entre Jack et Lulu est à cet égard tout à fait significatif puisqu'il s'agit véritablement d'un duo entre un homme qui parle et une femme qui en retour module un cri culminant en son cri d'agonie brut, auquel succède la parole brut de Jack. »¹³

¹³ M. POIZAT, 1986, *Frénésie, entre parole et cri : le chant de la Diva*

Le final de Lulu est à cet égard révélateur du souci de la recherche d'un équilibre entre voix parlée et voix chantée. « Cet équilibre est difficile à obtenir entre raison et magie; J'entends par « raison » le mot, l'importance du mot, et j'entends par « magie » le besoin instinctif de l'être humain de se libérer par l'acte magique qu'est le chant, le chant vocalisé, que l'on retrouve aussi bien dans le culte qui garde certains mélismes, comme l'ambrosien, que dans la vocalise de l'opéra italien ou dans le flamenco »¹⁴. Ce sont les fameux « deux pouvoirs » du mot, qui ainsi que se plaisait à le répéter Marcel Beaufils¹⁵, est en même temps véhicule d'une idée, et son pur, donc magie.

Cette évolution du rapport musique/verbe possède des ramifications dans la musique contemporaine. Celle-ci s'approprie, le plus souvent sous la forme d'une dialectique, ce rapport pour le pousser à son paroxysme. De la même manière, l'instrument-voix lui-même, est repensé sans cesse afin de le renouveler, voire même de le réinventer.

II.1.b La voix dans la musique contemporaine :

« Depuis le début du XX^{ème} siècle, on assiste à une grande mutation, aussi bien dans les arts plastiques que dans le domaine sonore, comme il ne s'en était pas produite depuis la Renaissance. Cette remise en cause généralisée des « lois » qui ont régné pendant plusieurs siècles correspond non seulement à un désir de refonte du langage musical, mais également à celui de renouveler le matériau sonore en exploitant les possibilités des instruments d'une manière non traditionnelle. Il semblerait que les compositeurs aient découvert une mine de virtualités inexploitées dans l'instrument le plus malléable qui soit, puisque jaillissant directement du corps humain, la voix. »¹⁶

En 1911, A. SCHOENBERG, dans sa pièce intitulée Pierrot Lunaire, utilise le Sprechgesang, au sujet duquel il écrit dans sa préface : « La note chantée maintient une hauteur fixe sans variation, la note parlée attaque une hauteur fixe, mais la quitte aussitôt, soit en montant, soit en descendant ».

¹⁴ R. MANCINI, 1991, la voix dans la culture occidentale

¹⁵ M. BEAUFILS, 1954, Musique du son, Musique du verbe

¹⁶ Madelaine GAGNARD, 1987, La voix dans la musique contemporaine et extra-européenne

II.1.b.i Sprechgesang :

« Avec le Pierrot Lunaire (1912), Schönberg inaugura le travail sur les inflexions de la voix chantée unies à celle de la voix parlée. Les tessitures de chacun de ces modes d'émission restant parfaitement distinctes, cela pose un problème d'homogénéité. La voix humaine, pareille aux instruments qu'elle précède, n'est pas une unité timbrale clairement délimitée. Dans l'accouplement du parler avec le chanter, il y a l'irrégularité du phénomène vocal. C'est à partir de Pierrot Lunaire que la voix parlée, en tant qu'elle «dérape» sur le chanter, est devenue partie intégrante de la musique. L'étrangeté du son que l'on entend est perçue comme un bruit totalement dégagé des interprétations futuristes et anecdotiques. Il est une nouvelle base harmonique, une nouvelle expression liée au concert. Schönberg redonne vie à la «consonne», alors que la musique vocale du siècle précédent (essentiellement italienne) privilégiait la «voyelle». A cet esthétisme du chant, il oppose une sensualité abrupte et directe. Composée à la demande de la comédienne « diseuse » Albertine Zehme sur des poèmes d'Albert Giraud, Pierrot Lunaire illustre l'union du parler et du chanter parce que Schönberg appelle le «Sprechgesang» - littéralement : Le chanter/parler. »¹⁷

La beauté de l'expression et de la forme est synonyme de pulsion. Elle cesse de représenter une symbolique du sens. Maintenant, « le son est sens immédiatement. »¹⁸
 « Au départ, le rythme devait être observé strictement comme s'il s'agissait de chant, mais alors que la voix chantée maintient la hauteur du son, la mélodie parlée (Sprechgesang) ne fait que l'indiquer pour la quitter aussitôt de façon ascendante ou descendante. »¹⁹
 Ce n'est que bien plus tard que pour « tuer dans l'œuf » toute tentation ou envie de chanter, Schönberg note le Sprechstimme (voix parlée sur une seule ligne). Il adopte un mode intermédiaire : la « mélodie parlée » qui, à vrai dire, ne se parle pas plus qu'elle ne se chante, mais procède par trémolo articulé (mis à part quelques sons isolés au-dessus desquels le compositeur a écrit « parler » ou « chanter »). La compréhension du texte est fonction de la déclamation. Mais pour Schönberg, il s'agit là d'un faux problème. Connaître le poème, en suivre la narration ne change en rien la texture homogène d'une oeuvre. Les correspondances « illusoires » entre musique et texte ne peuvent que masquer la profondeur de ce que le compositeur offre. Le contenu véritable est dans la musique.

¹⁷ D. COHEN-LEVINAS, 1987, La voix au-delà du chant

¹⁸ C. POTOCKI, 1968; L'inouï et l'indicible

¹⁹ A. SCHÖNBERG, 1912, préface de Pierrot Lunaire

II.1.b.ii Dans la continuité :

Son disciple et ami Alban Berg reprendra à son propre compte ce type très particulier et récitatif lorsqu'il écrira son opéra Wozzeck. La déclamation est coulée dans le moule d'une pensée musicale qui s'attache non seulement au mot-à-mot, mais aussi aux ressorts psychologiques, aux états intérieurs de chaque personnage, à la progression dramatique, que le compositeur a transposés vocalement. Par la musique et plus spécialement par la voix, Berg transcende ici le texte et va beaucoup plus loin en profondeur.

Sylvano Bussoti use lui aussi d'un mélange fréquent de parlé et de chanté, malgré une certaine discontinuité qui fait que de temps à autre surgit un son pointé dans l'aigu, ou une exclamation, alors qu'on ne s'y attendait pas. La diversité des techniques vocales employées dans Rara Requiem (trilles, tremolos, parlando, petites notes) et la manière de faire fusionner le tout confère à cet ouvrage une originalité, une personnalité.

Dans Madrigal, oeuvre vocale, Méfano déploie un remarquable esprit inventif quant à l'emploi des ressources de la voix. Mais il ne se contente pas de plaquer du Sprechgesang ici et là, il combine successivement et simultanément les diverses formes de déclamation en fonction du contenu expressif du texte, en dosant tous les degrés existant entre chanté et parlé, selon chaque mot.

K.H. Stockhausen ritualise la voix dans Am Himmel Vandre Ich. Le parlando clamé, lancé de la gorge possède un timbre qui le situe très loin du Sprechgesang. Le type de son qu'on rencontre le plus fréquemment dans ce cycle est celui qui correspond à une voix flûtée, très légère, un peu détimbrée et à peine vibrée.

Après avoir examiné comment Boulez traite la voix dans Le marteau sans maître, il est évident qu'il y a là un type de déclamation fort original. Cette écriture presque plus instrumentale que vocale aboutit à une voix tendue à l'extrême; Paradoxalement, on arrive à une émission flexible, d'une grande souplesse et Boulez se montre ici l'héritier à la fois de Debussy et de l'école de Vienne, mais il est complètement repensé, repris à son compte personnel le principe du Sprechgesang.

Dans Sequenza III, L. Berio coupe le texte par des rires, des cris, des halètements. C'est là la marque d'une fusion totale entre matière verbale et matière vocale. Sur la partition, figurent des mots dont le rôle est d'orienter la sensibilité de la cantatrice : Elle doit savoir se montrer, avec sa voix, « réservée, sereine, tendre, de plus en plus désespérée » ou prendre un air égaré. Le chant cohabite alors avec des bruits vocaux pour

faire ressortir l'émotion désirée.

M. Ohana compose Trois contes pour l'honorable fleur en 1978. Ce récitatif prend des couleurs changeantes selon le texte ou le point où en est l'action : Il se fait litanie sur un son, avec une voix localisée à l'avant du palais; il devient parfois confidence, toujours avec une émission palatale, ou encore déclamation haut-perché, proche du Sprechgesang; enfin il prend par moments la forme de lamentations, et même de pleurs sur le souffle, avec participation de la glotte. Avec ou sans mots, ce parlando, avec soudain des tremblements, comporte des inflexions qui semblent provenir d'une autre culture, de la patrie du Nô.

« Vox, Vocis, 1985, d'Ivo Malec, est une pièce dans laquelle l'unique voix féminine s'incarne dans un autre groupe, un autre genre de voix, celle des neuf instruments (cinq cordes, trompette, harpe et deux percussions). La rencontre de l'un avec les autres ne signifie pas pour autant que l'entité voix perd de son unicité. Elle reste une et unique. L'idée initiale du compositeur est d'extrapoler l'unité sémantique de la voix/langage vers un groupe dont le signifiant ne renvoie pas inéluctablement au signifié, mais exprime un imaginaire sonore indécodable par les voies de l'analyse structurelle. La voix lui permet d'élargir l'espace dynamique, ainsi que, le timbre à une écriture instrumentale polyphonique. Syntaxe musicale et syntaxe verbale s'unissent et se raréfient mutuellement sans jamais se confondre. L'identité de la voix enracine toutes références au langage et à l'écriture. Ivo Malec tente d'associer la voix-chant à la voix génératrice, la voix-source, car (dit-il) « c'est bien la seule qui chante vraiment. »²⁰ ».²¹

François Bernard Mâche innove aussi dans Danaé du point de vue matériau verbal, avec des éclats de mots, sans chercher à les rattacher à une langue quelconque, mais en les considérant, comme l'auteur les a pris, en tant qu'éléments préformateurs de la voix, en constatant également l'importance évidente des consonnes; On a des GAB-DE-GEB-DI-GIB-DU ou bien KIUX... DI GI DI KA, ou l'accent est mis sur les explosives.

²⁰ I. MALEC, 1985, Ivo Malec à propos de Vox, Vocis, F

²¹ D. COHEN-LEVINAS, 1987, La voix au-delà du chant

II.1.c La voix comme processus compositionnel :

« En 1964, François-Bernard Mache composa une pièce orchestrale : Le son d'une voix. À partir d'une analyse spectrographique d'une phrase parlée. Il essaya de reproduire le spectre de la voix grâce à l'orchestration instrumentale. En 1977, Michel Decoust réalisa : Interphone. Partant de l'enregistrement de la soprano Irène Jarsky déclamant un poème de Claude Minière, il utilisa les courbes dérivées des contours de Fo et de l'amplitude, (extraites de l'enregistrement par une analyse) afin de contrôler des sons synthétisés par ordinateur. Dans Different Trains, Steve Reich se servit d'un sampler digital pour répéter des phrases parlées : Un quatuor à cordes faisant écho en jouant des phrases musicales suivant la mélodie et le rythme de ces phrases. »²²

II.1.c.i Different trains de Steve Reich:

Different trains (1988) de Steve Reich est une oeuvre dont le processus créatif utilise largement la prosodie. Cette pièce majeure de l'oeuvre de S. Reich, pour quatuor à cordes et bandes, est née de l'idée de combiner des enregistrements de parole avec des instruments à cordes. Ainsi, ces enregistrements ont généré le matériau musical lors de la phase d'écriture de la pièce.

« J'utilise dans Different Trains, une nouvelle manière de composer qui a ses origines dans mes compositions antérieures pour bandes magnétiques : It's Gonna Rain (1965) et Come Out (1966). L'idée générale est d'utiliser des enregistrements de conversations comme matériau musical.

Pour combiner les conversations sur bande magnétique et les instruments à cordes, j'ai sélectionné des exemples brefs de discours, aux différences de ton plus ou moins marquées, et je les ai transcrits aussi précisément que possible en notation musicale.

Ensuite, les instruments à cordes imitent littéralement la mélodie du discours. Les exemples de conversation et les bruits de trains ont été transférés sur bande magnétique à l'aide d'un échantillonnage de claviers, les sampling keyboards, et d'un ordinateur. Trois quatuors à cordes séparés ont aussi été ajoutés à la bande magnétique pré-enregistrée et le quatuor final, joué par des musiciens, vient s'ajouter lors du concert.

22 J-C. RISSET, 1990, Speech and music combined: An overview

Different trains comprend trois mouvements - mouvement étant pris ici au sens large du terme car les tempi changent fréquemment dans chaque mouvement :

L'Amérique - Avant la guerre

L'Europe - Pendant la guerre

Après la guerre

Cette composition a donc une réalité à la fois sur le plan documentaire et sur le plan musical et ouvre une nouvelle direction. C'est une direction qui conduira sous peu, je l'espère, à une nouvelle sorte de théâtre multi-média combinant documentaire, musique et vidéo. »²³

II.1.c.ii Locus de Claudie Malherbe :

Locus pour voix réelles et virtuelles :

« La musicalité naturelle des langues est une chose connue depuis longtemps. Aujourd'hui, les outils d'analyse du son montrent qu'elles contiennent non seulement une mélodie mais aussi une harmonie, en même temps qu'ils révèlent avec précision leurs qualités rythmiques et dynamiques comme la composition de leurs timbres. Plus précisément, cette technologie, qui donne une représentation cohérente et manipulable d'une réalité sonore complexe, permet maintenant de lier ce qui autrefois était distinct. En effet avec ces nouvelles possibilités, la mise en oeuvre d'un matériau sonore n'est plus irrémédiablement coupé du travail symbolique et, à l'inverse, le maniement des signes musicaux n'exclut plus une approche concrète du son. Ainsi dans Locus, musique écrite pour des voix réelles et virtuelles grâce à l'informatique musicale, la parole énoncée est mise en musique au moyen de ses propres contenus une fois ceux-ci reconfigurés dans un espace abstrait.

Le déroulement est précisément balisé par cinq états remarquables du matériau. Les trois premiers qu'il est possible de qualifier de naturels : La parole, le chant et les bruits vocaux; Un autre, qui fait entendre des reconstitutions artificielles (plus ou moins décalées) de la voix; Enfin, une représentation synthétique de cette voix, au moyen de l'électronique, qui en révèle la part constitutive mais ordinairement cachée. Notons que dans ce contexte, le chant devient une catégorie parmi d'autre et n'est plus dominant comme habituellement dans

23 Traduit du texte figurant sur la pochette du CD.

la musique vocale occidentale.

La parole, forme la plus courante d'un flux vocal, sert de point de départ. A la parole s'ajoutent le chant et les bruits vocaux (toux, rires, pleurs, souffles, etc...) qui constituent les trois catégories principales de l'émission vocale humaine. Les outils informatiques (échantillonnage) permettent de disséquer puis de reconstituer un flux parlé, en une sorte de parole virtuelle, pendant de la parole réelle recomposée à partir d'éléments de celle-ci (analyses et échantillons). Les analyses temps/fréquence d'un matériau vocal permettent d'en révéler puis de faire entendre les éléments constitutifs ordinairement cachés; la nature électronique de cette restitution, désignée ici par sinus, la place dans un autre plan, construisant avec cette troisième dimension un tétraèdre.

Ainsi configurés, les éléments en jeu définissent cette figure géométrique qui va régler les différentes péripéties du matériau et va définir l'entier déroulement de la pièce: L'espace entre deux états remarquables (sommet du tétraèdre), définit un champ (arêtes du volume) et une organisation (qui détermine un paramétrage des matériaux syntaxiques et musicaux constitué d'états intermédiaires (trois par arêtes qui règlent les enchaînements et les transitions (ce nombre d'éléments qui est arbitraire, offrant un choix suffisant d'étapes caractéristiques et différenciées).

Par exemple, le champ du babil situé entre la parole et le chant induira une organisation syntaxo-mélodique (mixte de langage parlé et de mélodie chantée) caractérisée par trois états intermédiaires permettant de passer par transitions successives de la parole au chant : voix théâtrale (paroles aux intonations accentuées, sprechgesang (chanté-parlé) et babil (parole enfantine dénuée de sens précis et proche du chant). D'une semblable façon, entre la parole et le bruit s'établit le champ du grommelot (de grommeler, parole peu articulée et en partie indistincte) appuyé sur une organisation syntaxo-rythmique (précédemment le chant appelait le mélodique, ici le bruit convoque le rythme) réalisée à son tour au moyen de trois états intermédiaires : bégaiements, balbutiements et grommelot (déconstruction progressive d'une parole qui se transforme en bruit vocal). La transition du rire au pleur (du rythme au mélodique) est un autre exemple de ces transitions d'un état vocal à un autre : Les rires constitués de saccades rythmées sont progressivement transmués en plaintes qui allongent les sons, introduisant ainsi des hauteurs de plus en plus perceptibles qui s'agglomèrent en mélodies.

Enfin le parcours qui constitue la grande forme de la pièce est donné toujours par la même figure selon la règle suivante : Une fois un point d'entrée choisi sur le sommet, l'ensemble des arêtes doit toujours être parcouru une seule fois selon un trajet sans

discontinuité; Un sommet pouvant être énoncé plusieurs fois. Avec cette figure à quatre sommets distribuant chacun un nombre pair d'arêtes (quatre), il n'existe que deux points d'entrée à partir desquels il est possible de parcourir une seule fois l'ensemble des arêtes. »²⁴

Après l'explication de la constitution de la base de données, Claudy Malherbe nous renseigne sur les processus compositionnels qu'il a utilisé pour extraire d'un flux parlé, un matériau musical : « Il s'agit de prendre comme point de départ un énoncé parlé : When speech sounds are made, the larynx may or may not itself be vibrating to produce an oscillatory flow of air. A partir de l'analyse temps/fréquence de cette phrase, une segmentation syllabique et un lissage des partiels sont d'abord effectués. Ce résultat est ensuite resynthétisé puis transformé par réductions successives pour constituer un matériau musical original qui reste corrélé à la phrase parlée constituant son origine. »

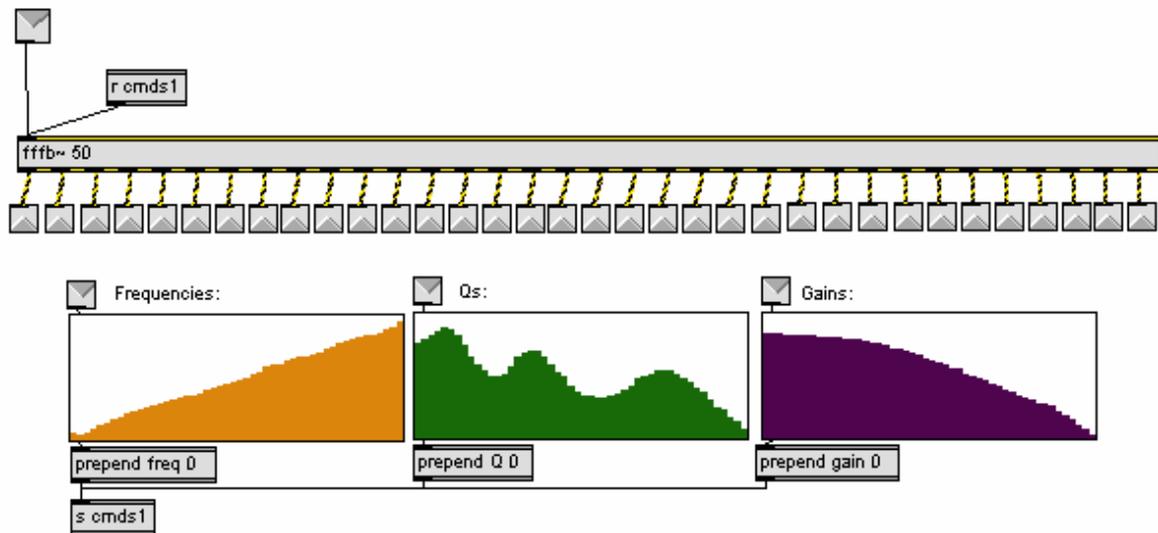
Quelle que soit leur provenance, tous ces compositeurs manifestent un besoin d'aller au-delà de tout ce qui avait été tenté auparavant, et ce désir de dépassement permet de constater que les ressources de l'organe vocal sont beaucoup plus vastes qu'on ne le supposait, les limites du réalisable ont reculé.

²⁴ C. MALHERBE, 1997, Locus

III Pièces écrites:

III.1 Fo-low :

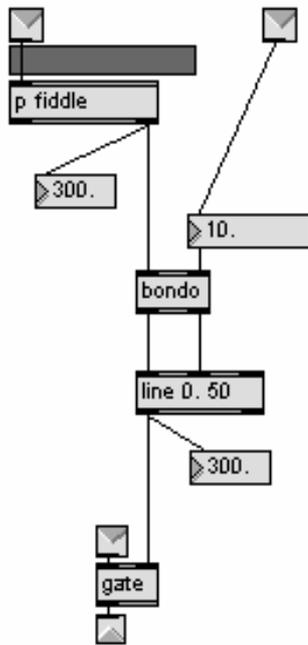
Fo-low est une pièce créée dans un contexte universitaire (Université PARIS VIII Saint-Denis). Elle s'inscrit dans le cadre des cours de « composition instrumentale » de José Manuel Lopez Lopez et ceux de « composition avec l'informatique » d'Anne Sédès. C'est une pièce mixte écrite pour flûte, accordéon, violoncelle et MAX-MSP.



Chacun des instruments dirige un méta-paramètre d'un instrument virtuel résultant. Le « moteur » de cet instrument repose sur le principe de la synthèse soustractive. C'est un peigne de 50 filtres passe-bande dont les fréquences sont en rapport harmonique entre elles durant les deux premiers tiers de la pièce (ces rapports s'effondrant à la fin pour laisser place à l'inharmonicité). Le résultat sonore est l'écoute d'une note à partir de n'importe quel spectre en entrée. De manière à accentuer la coupure des filtres du second ordre de l'objet « fffb~ », nous avons choisi d'en mettre deux en série.

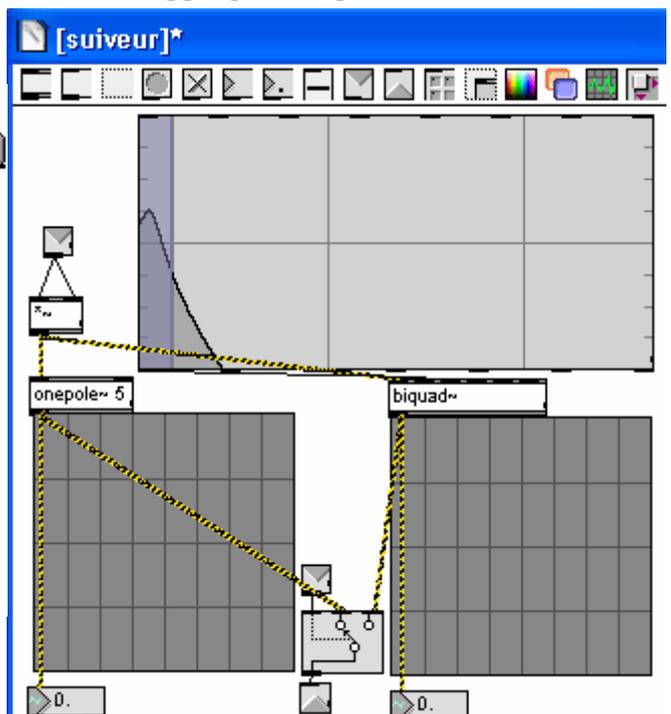
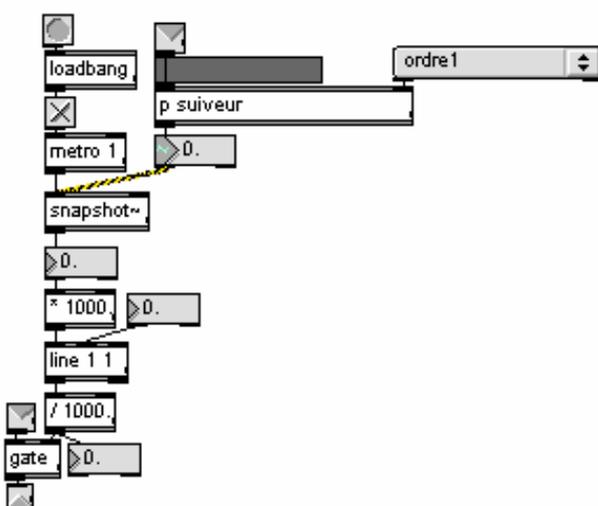
La hauteur de cette note, appelée F_0 , est donnée par la flûte basse au moteur par l'intermédiaire de l'objet externe « fiddle~ ».

Cet estimateur de pitch possède un registre limité et condamne donc F_0 à être bas. De plus, on cantonne les fréquences centrales des filtres au domaine audible, c'est à dire que la fréquence centrale du 50ème filtre ne doit pas excéder 22050 Hz. Or si elle atteint cette fréquence, cela veut dire que la fréquence centrale du premier filtre est 441 Hz



(22050/50). Par conséquent, le registre alloué à la flûte ne doit pas dépasser le LA3. D'ou le titre : Fo-low. J'ai choisi la flûte basse pour sa pureté spectrale dans le registre utilisé. Lors du concert, j'ai finalement utilisé une flûte en Ut de manière à ce qu'on entende mieux la ligne mélodique (elle se démarque mieux du violoncelle de part sa brillance et sa dynamique). La ligne mélodique, au départ écrite dans un registre inférieur au LA3, s'est retrouvée octaviée à cause du changement d'instrument. J'ai donc divisé l'estimation du pitch de la flûte en UT de manière à respecter la contrainte issue des filtres. La plus grande richesse harmonique de la flûte en UT par rapport à la flûte basse a permis à l'estimateur une plus grande stabilité. Ainsi entendions nous moins de sauts de fréquences (à l'octave ou mauvaise estimation). J'ai renforcé cette stabilité grâce à un « line » en sortie de l'estimateur.

Nous avons ainsi une structure de filtres en peigne dont nous connaissons, à présent, les fréquences respectives. Le gain de tous ces filtres et donc, l'amplitude de sortie résultante est modulé par le violoncelle. Il est donc le conducteur de l'enveloppe temporel du son du méta-instrument. Cette relation fait encore intervenir le concept du suivi de paramètre. De la même manière que la hauteur de la flûte donne la hauteur au méta-instrument, le suivi en temps réel de l'enveloppe du violoncelle gère l'amplitude instantanée du méta-instrument. Cela se fait par une détection d'enveloppe quadratique via MAX-MSP, comme le montre les deux figures suivantes.



Le violoncelle permet de nombreuses variations dans son évolution temporelle grâce aux nombreux modes de jeux qu'il offre.

Jusqu'ici, notre méta-instrument est structuré d'un filtre en peigne dont la « hauteur » est donnée par la flûte et dont le gain est modulé par le violoncelle. Mais jusque là, il ne sort aucun son de notre instrument virtuel. Il faut pour cela le « nourrir » d'un spectre. Le choix s'est d'emblée porté sur l'accordéon pour sa richesse timbrale, sa polyphonie, et aussi pour la stabilité de son enveloppe d'amplitude qui permet ainsi d'éviter de « trop marcher sur les plate-bandes » du violoncelle. Même si lors de la pièce, on peut se rendre compte que j'ai joué sur cette double modulation (à la source et lors du traitement).

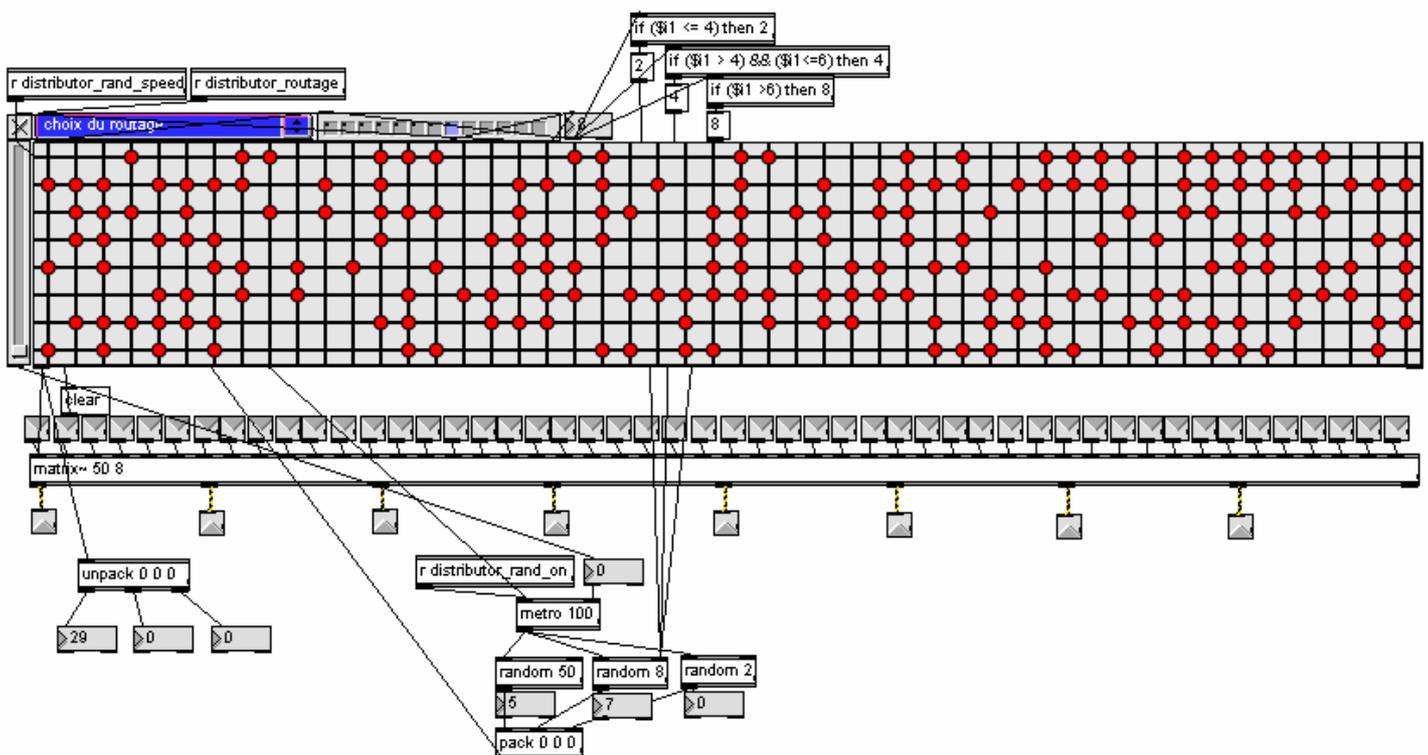
Voici donc présenté notre méta-instrument :

☞ dont la hauteur est donnée par la flûte basse.

☞ dont le volume instantané est donné par le violoncelle.

☞ dont le spectre de base est celui de l'accordéon.

Il est à remarquer qu'il existe en réalité un triple modulation d'amplitude dont la troisième modulante est due à la spatialisation de notre méta-instrument. En effet, la pièce est conçue pour une diffusion octophonique. Chacun des 50 filtres voit sa sortie continuellement et aléatoirement changée et est ainsi routée sur les huit haut-parleurs.



Ainsi, il arrive perpétuellement que quelques partiels disparaissent momentanément puis réapparaissent sur un autre enceinte. On a donc une légère modulation

d'amplitude résultante due aux $50 \times 8 = 400$ modulations binaires aléatoires réalisées à l'aide d'une matrice. Cependant, si on tient compte du nombre important de filtres (50) et de l'ergodicité ainsi engendrée, on peut négliger cette modulation qui en moyenne ne s'entend pas (aucune chute brutale du volume global).

L'écriture de la partition instrumentale s'effectue de concert avec l'élaboration du patch MAX. De nombreux paramètres changent lorsque le chef d'orchestre ou la flûtiste appuie sur une pédale midi. L'écriture de la pièce a donc plusieurs supports : La partition, tout d'abord pour les instrumentistes et le patch lui-même comprenant une collection gérant l'évolution temporelle du comportement du méta-instrument. Il n'a pas été choisi de système compositionnel particulier (sérialisme...) pour l'écriture instrumentale. Ceci car la contrainte de la manipulation du méta-instrument par les instruments réels est suffisamment forte pour définir un espace des possibles. En effet, chaque partition instrumentale est pensée de manière à ce que de la combinaison des trois, s'érige une partition virtuelle : Celle du méta-instrument.

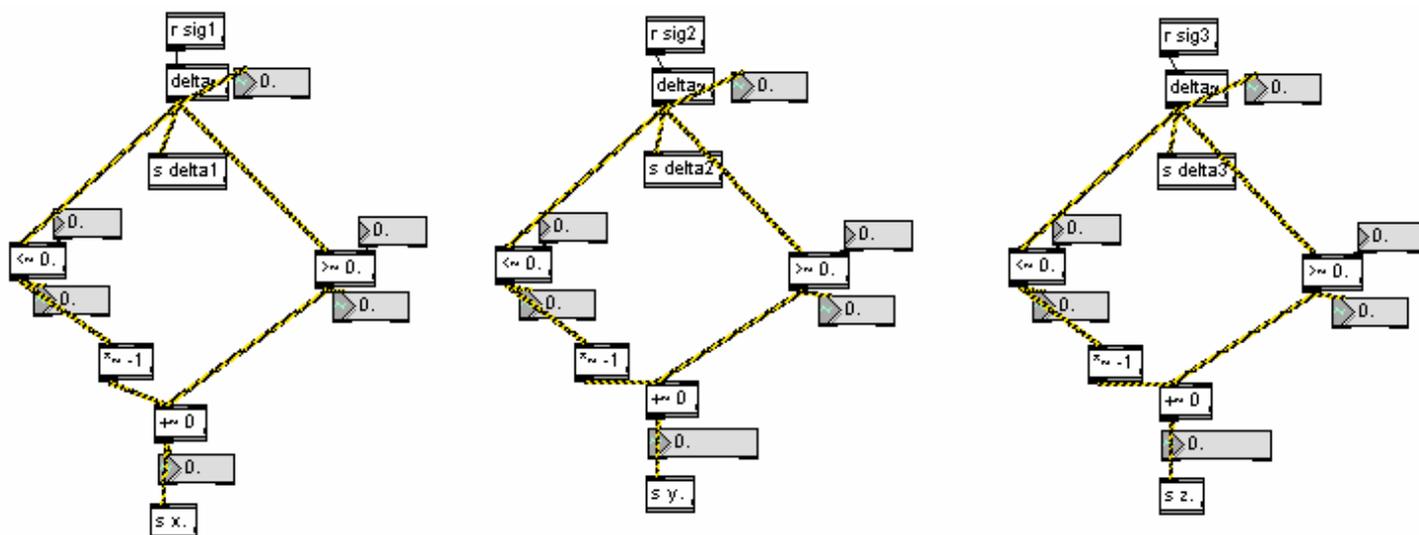
Ainsi la ligne mélodique de la flûte n'est pas seulement propre à cet instrument mais aussi au méta-instrument. Ne reste alors que d'autres paramètres comme le timbre, le rythme... comme degrés de liberté pour la flûte. De même pour le violoncelle qui doit céder son enveloppe au méta-instrument, ce qui fera apparaître de nombreux modes de jeux dans sa partition. De nombreux clusters sont présents dans la partition pour accordéon de manière à livrer au méta-instrument un spectre assez riche et inharmonique.

Mais la véritable contrainte ne se situe pas véritablement dans le seul fait que chaque instrument se voit « déposséder » d'une partie de son espace compositionnel, mais bien dans la difficulté d'obtenir, au final, une cohérence musicale entre tous les instruments réels mais aussi entre les instruments réels et l'instrument virtuel, tout en gardant une certaine indépendance de chacun des instruments face au méta-instrument (indépendance rythmique pour la flûte et mélodique pour le violoncelle principalement). De part la structure du méta-instrument (synthèse soustractive), la pièce a tout de même été élaborée dans une pensée spectraliste. D'ailleurs, son déroulement en est révélateur.

Quatre moments sont bien discernables dans Fo-Low. Du bruit naît une note accompagnée de l'harmonicité. Puis cette harmonicité se « dégrade » jusqu'à ce que lâche la relation harmonique fondatrice de la sensation de hauteur du méta-instrument. La relation rigide entre les fréquences centrales du filtre en peigne se rompt laissant libres les filtres d'évoluer aléatoirement dans le domaine fréquentiel. S'en suit une phase d'inharmonicité qui va brutalement replonger dans le domaine de l'aléatoire et du bruit. Fin de la pièce.

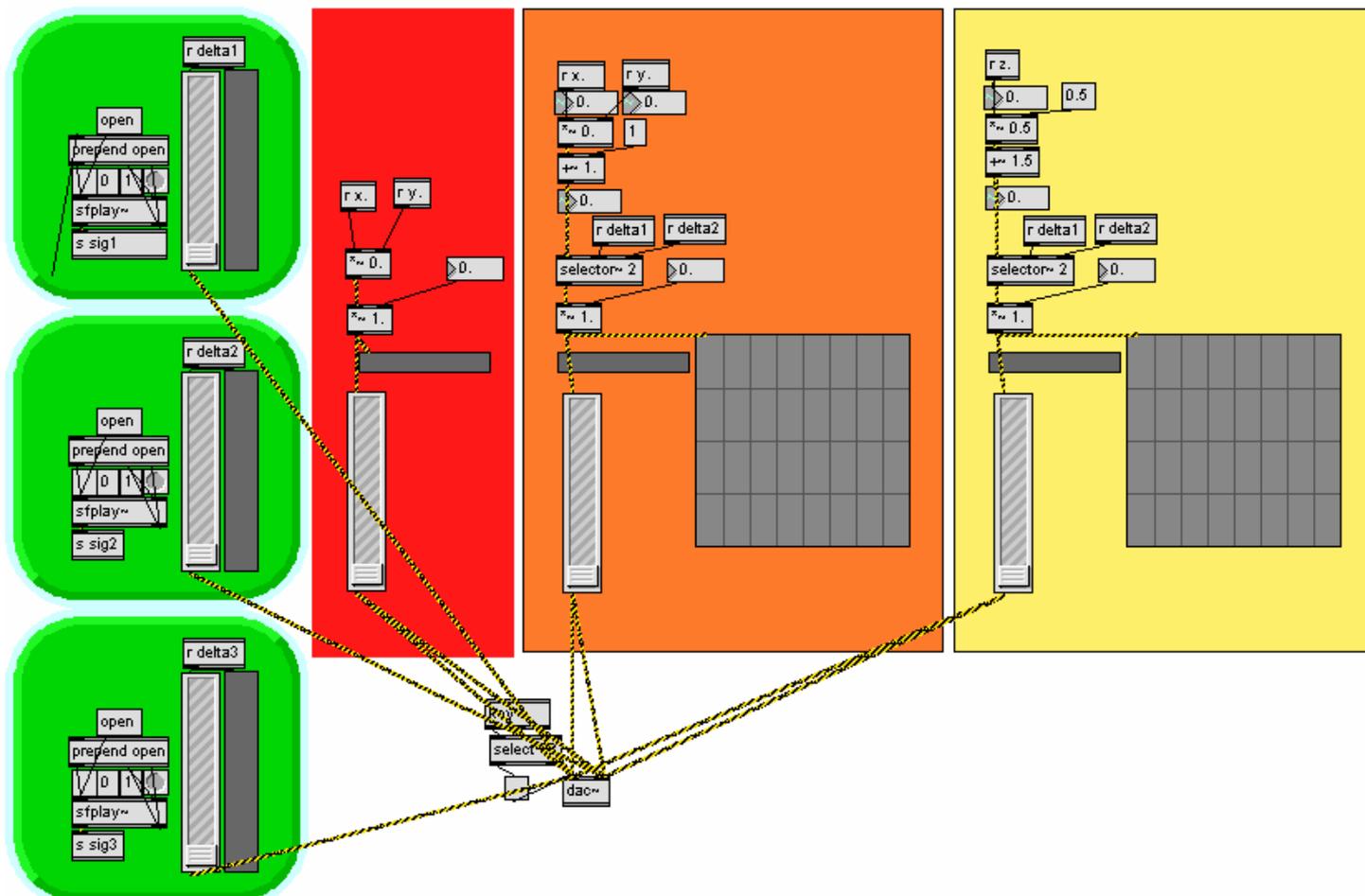
Le déroulement temporel est donc principalement structuré sur une dialectique entre l'ordre et le désordre, le déterminisme et l'aléatoire, le suivi et la divergence. Le titre Fo-Low (suivre) décrit ce qu'essaye de faire le méta-instrument durant toute la pièce : Sorte d'alien instrumental, glouton des paramètres des instruments réels, il va tenter de les suivre sans jamais y arriver jusqu'à l'implosion finale.

Les phases bruitées du début et de la fin de la pièce sont enregistrées au préalable car elles sont réalisées en temps différé. Là encore, intervient cette dialectique entre « sons purs » et bruits. Le patch Derivate.mxb utilisé pour générer des séquences de sons bruités a été réalisé avec MAX. C'est une sorte de synthèse croisée non linéaire. Je désirais obtenir des sons bruités à partir de sons harmoniques. Il m'est naturellement venu alors à l'esprit d'utiliser la dérivée d'un signal qui est très sensible aux bruits. Mais je ne voulais pas simplement rehausser les composantes hautes fréquences du bruit présent dans de mauvais enregistrements. J'ai donc effectué une transformation non linéaire entre deux sons pour en obtenir un seul en sortie. Le principe est simple : grâce à l'objet `delta~`, on observe les variations instantanées des deux sons en entrée, d'un échantillon à un autre.



Selon les signes des dérivées de ces deux sons (signes de « x » ou « y »), le son résultant va tantôt évoluer comme le son 1 ou comme le son 2. Par exemple, si le son 1 « monte » ($\text{delta1} > 0$) et que le son 2 « descend » ($\text{delta2} < 0$). Alors le delta du son résultant (delta3) sera égal à delta1 :

- * Si les deux sons évoluent dans le même sens ($\text{delta1} * \text{delta2} > 0$), $\text{delta3} = \text{delta2}$.
- * Si les deux sons évoluent dans les sens contraire ($\text{delta1} * \text{delta2} < 0$), $\text{delta3} = \text{delta1}$.



C'est donc une transformation non linéaire de bas niveau, mais qui a pour avantage de donner un résultat sonore assez intéressant puisqu'elle réalise une sorte de mix instantané des deux sons en entrée, chargé de bruit. Résultat assez frappant lorsqu'on choisit une percussion et un son harmonique. On obtient alors une percussion teintée du timbre du son harmonique.

Le module jaune à droite fait intervenir un troisième son. Selon le signe de sa dérivée « z », on « switch » entre delta1 et delta2.

Cet instrument m'a permis de réaliser les parties introductrice et conclusive de la pièce.

Fo-low a été jouée par l'ensemble Futur-Musiques, lors de deux concerts :

œ Le 17 Juin 2004 à l'amphi X de l'université

œ Le 18 Juin 2004 au théâtre Gérard Philipe²⁵

Cette pièce ne s'inscrit pas vraiment dans la problématique de ce mémoire. D'ailleurs, elle ne fait pas intervenir la voix, car aucun interprète n'était disponible dans la formation de départ. Cependant le procédé de suivi de hauteur utilisé s'applique très bien à la voix. D'ailleurs la flûte est un instrument vocal.

²⁵ Enregistrement et partitions « Finale » sont sur le CD-ROM joint.

III.2 DisCutlon :

Je souhaite plutôt désigner cette pièce par le terme : expérience sonore. En effet, elle propose d'expérimenter par l'écoute ce que peut comporter la prosodie comme processus compositionnels. Ses buts, car elle a des buts (et c'est en cela qu'elle ne peut être une pièce musicale, une oeuvre artistique), sont :

œ de nous faire entendre l'agencement structurel d'un matériau sonore connu par l'intonation pour voir si oui ou non, on peut assimiler la prosodie à un processus compositionnel.

œ d'essayer de reconnaître les intentions et autres informations délivrées par les locuteurs lors de la conversation réelle, or contexte sémantique. Peut on tirer des informations de la prosodie sans les mots ?

Ainsi, cette aventure nous permet d'apprécier pragmatiquement si oui ou non, la prosodie peut être assimilée à la composition musicale. De plus, elle met en évidence le type d'informations que peut véhiculer l'intonation par l'écoute.

Tout commence à la maison... Une situation quotidienne mettant en scène une conversation entre plusieurs personnes est enregistrée. Le choix d'une telle situation provient de la spontanéité qu'elle implique. Afin d'éviter tout cliché mélodique ou autre prosodie exacerbée, on effectue cet enregistrement secrètement (avec l'accord des participants bien entendu).

Une fois cette conversation qui se veut la plus naturelle possible enregistrée, nous en extrayons le contenu prosodique grâce au système présenté précédemment. La segmentation nous donne le rythme, l'analyse de la fréquence fondamentale nous donne les variations de la hauteur et différents descripteurs nous donnent accès à l'intensité et au timbre des conversants.

Cette structure prosodique est ensuite utilisée pour agencer des matériaux musicaux issus de bases de données. Chacun des interlocuteurs est représenté par un « instrument » ou plus précisément par une base de données constituée de l'échantillonnage d'un instrument. Le père est le violoncelle, la mère la flûte et l'enfant, Le piccolo ? Ce choix étant bien sûr arbitraire, mais tente tout de même de garder une certaine cohérence au niveau des registres de chacun des personnages.

A ce jour, cette pièce n'a pu être créée car le système mis en place ne l'a pas encore permis (Problème d'optimisation qui ralentit beaucoup trop la création d'une base de donnée). Cependant, nous avons travaillé sur des enregistrements d'un locuteur. Nous avons mis en rapport sa prosodie avec une base de données composée d'échantillons de violons.

Ces essais sonores sont écoutables sur le Cd-rom joint au mémoire. Ils permettent de donner un avant-goût de ce que sera DisCutIon. Une composition originale nommée Xavier-Cocteau-Violon²⁶ constitue un agencement primaire des séquences les plus pertinentes.

²⁶ Voir -dossier DisCutIon sur le CD-ROM joint.

IV Questions ouvertes :

IV.1 Qu'en dit notre cerveau...

Nous voyons que traiter des rapports entre le langage, l'intonation et la musique recouvre bien des domaines d'études: Phonétique, linguistique, psycho-phonétique, musicologique, physique et bien d'autres... Nous nous trouvons ici à la frontière des sciences humaines et des sciences exactes. De nombreux chercheurs approuvent ce point-de-vue pluraliste.

En montrant des analogies frappantes entre grilles métriques utilisées par les linguistes pour définir la prosodie et grilles rythmiques musicales, F. Lerdhal et J. Halle²⁷ arrivent à la conclusion que l'étude des relations entre musique et parole conduit directement à des questions de psychologie fondamentale.

E. Narmour²⁸ est conduit à des conclusions similaires à F. Lerdhal et J. Halle. La comparaison, cette fois-ci, des structures mélodiques du langage à celles de la musique l'amène à la modélisation du comportement neurologique. C'est pour cela que nous introduisons dans ce chapitre une vision supplémentaire : Celle des neuro-sciences.

En effet, avant même de se poser la question : Y'a-t'il des liens entre les perceptions de la parole et de la musique ? Il nous faut savoir comment d'un point de vue « pratique », notre cerveau traite ces différentes informations. D'ailleurs, sont-elles si différentes ?

Dans La voix dévoilée²⁹, R. Cross interroge Mr. Lhermitte, spécialiste du cerveau, sur le rôle que joue celui-ci lors de la réception de signaux sonores.

Richard Cross demande au professeur Lhermitte :

« Quelles sont donc les caractéristiques de l'hémisphère droit dans le cadre des processus du chant et de la parole ? »

Le professeur Lhermitte répond : « Je crois qu'on peut recourir à deux exemples. L'un est un exemple pathologique, je vais commencer par celui-là parce qu'il est

²⁷ F. LERDHAL, J. HALLE, 1990, Somme lines of poetry viewed as music

²⁸ E. NARMOUR, 1990, The melodic structures of music and speech: Applications and dimensions of the implication-realization model

²⁹ R. CROSS, 1991, La voix dévoilée

grossier. Reprenons un patient qui a une lésion très importante de l'hémisphère gauche, il est donc aphasique, il est incapable de parler, il dit : « ah da ya ya ... » et quand vous lui demandez ou que vous lui faites comprendre qu'il faut qu'il chante, il se met à vous chanter tous les airs qu'il connaît. Bien sur, il aura du mal mais très rapidement vous allez obtenir des chansons qui vont stupéfier tout l'entourage puisque ce malade, absolument incapable d'émettre dans le langage parlé le moindre son articulé, devient capable d'émettre une suite mélodique avec toutes les composantes musicales. Bref, il ne parle pas mais il chante ! C'est vraiment extraordinaire.

Le deuxième exemple est beaucoup plus subtil, il fait référence à des expériences chez des sujets normaux. Je vous avais dit toute à l'heure que pour apprécier les capacités de l'oreille droite et gauche, de l'hémisphère droit et gauche, il fallait adresser des messages concurrentiels dans chaque oreille. A ce moment-là, effectivement, vous mettiez en compétition les deux oreilles. Imaginez que dans un casque, avec des paramètres physiques et acoustiques, équivalents, vous adressiez dans l'oreille droite un signal parlé (vous racontez pendant 10, 15 ou 20 secondes une histoire) pendant que, dans un même laps de temps, vous envoyez dans l'oreille gauche un signal musical (1, 2 ou 3 airs de musique). C'est ce que l'on appelle l'écoute dichotique. Ensuite vous pouvez demander au sujet : « ... finalement qu'est que vous avez entendu ? ». En fonction de sa réponse, vous allez voir s'il a entendu préférentiellement le langage ou la musique dans la mesure où le message sonore linguistique était adressé à l'oreille gauche et le message musical à l'oreille droite. Vous pouvez contrôler en faisant la stimulation inverse : Vous adressez à l'oreille gauche de la musique et à l'oreille droite du langage. Il est sur que, dans la mesure où vous adressez à l'oreille droite du langage, ce dernier va être immédiatement capté par l'hémisphère gauche et la musique sera scotomisée, effacée.

[...] Toutes les informations qui nécessitent un traitement par le langage, donc raconter une histoire, expliquer quelque chose, donner un raisonnement, et bien sur, penser de manière philosophique, tout ceci ne peut se faire ailleurs que dans l'hémisphère gauche qui dispose non seulement du langage, mais du couple langage/pensée. »

A ce stade de l'interview, on est en face de déclarations manifestes d'une profonde différence entre les traitements de la parole et de la musique. Il semble que les décodages de ces deux signaux se situent dans des zones bien distinctes. Ce qui laisse à penser qu'ils ne peuvent exprimer la même chose; Et que toutes recherches concernant un espace sous-jacent à ces deux modes de communication n'est pas justifiée. Mais le fonctionnement de notre cerveau n'est pas si simple. Tout n'est pas blanc, ou noir.

« [...] Je crois que l'ensemble de la population sait que chez un droitier, le système de la motricité, c'est à dire tout ce qui nous permet d'agir et de mouvoir le membre supérieur droit et le membre inférieur droit, est en rapport avec l'hémisphère gauche et seulement l'hémisphère gauche. C'est vrai aussi pour le champ visuel. Tout le champ visuel droit se projette dans l'hémisphère gauche et vice-versa. Mais, dans ce qui touche l'oreille, il existe deux représentations, à savoir un faisceau relativement mince, plus petit que le second, qui va directement de l'oreille gauche à l'hémisphère gauche (donc sans croiser) et un gros faisceau qui se projette de l'oreille droite à l'hémisphère gauche croisé cette fois-ci. Si bien que là intervient une subtilité : Lorsqu'on adresse un message sonore seulement dans une oreille, il ne va pas seulement dans l'hémisphère du côté opposé, comme c'est le cas lorsqu'on touche une partie du corps, il va aussi directement dans l'hémisphère du même côté. »

A la lumière de cette indication, il nous est encore permis d'imaginer des liens physiques entre intonation et musique. En effet, on sait que seul l'hémisphère gauche du cerveau est capable de gérer le couple langage/pensée. C'est donc l'oreille droite qui est la mieux prédisposée à recevoir un signal de parole. On admet, par déduction, que c'est alors le cerveau droit qui s'occupe de la musique et donc l'oreille gauche. Mais comme il existe des liaisons directes et non croisées entre les oreilles et les hémisphères, il paraît possible que les écoutes de la voix parlée et de musique activent les mêmes zones de notre cerveau. Notre espoir de liens physiques entre ces deux mondes n'est donc pas vain. Et cela se confirme lorsque R. Cross demande :

« Si j'introduis la notion de hauteur du son, de timbre, et d'intensité, quelle compréhension peut nous apporter sur ces phénomènes la neuropsychologie ?

Plus nous nous enfonçons dans l'infiniment petit, plus le mystère s'opacifie. Je ne crois pas qu'on puisse répondre précisément à votre question car vous touchez à des domaines du cerveau qui sont totalement inconnus. Il est clair que c'est le cortex cérébral qui développe tous les montages permettant à la fois le décodage de tous les paramètres physiques du son. Prenons le cas des phonèmes, nous allons élaborer dans certaines régions du cortex tous les systèmes autorisant cette jonglerie motrice extraordinaire que de savoir prononcer les 52 phonèmes de la langue française, dans d'autres régions corticales, les 52 phonèmes de la langue anglaise, et encore dans d'autres régions, les différents paramètres physiques du chant. Le problème du rythme est très intéressant parce qu'il y a plusieurs régions impliquées. Il y a aussi d'autres systèmes sensoriels que l'oreille qui vont intervenir, par exemple, la vue (un rythme peut être visualisé sur un tableau avec des points, des traits...). »

Finalement, on s'aperçoit que nos connaissances sur le cerveau ne sont pas assez solides pour affirmer s'il existe ou non des processus communs au décodage de la parole et de la musique. Mais ce qui est important ici, c'est de constater que l'hypothèse d'un traitement commun à ces deux modes de communication n'est pas caduque et reste pour le moment viable d'un point de vue des neuro-sciences.

« Vous avez dans l'hypothalamus, une énergie de faim, une énergie de soif, une énergie de défense et une énergie sexuelle. C'est lui qui gère le registre émotionnel. »³⁰

La « clef » est certainement là, dans l'hypothalamus ! Si l'on peut être ému par une belle musique, l'on peut aussi l'être en entendant un enfant nous parler (par exemple) au cours d'une discussion. Ces deux « genres » de signaux sonores activent alors probablement les mêmes zones de l'hypothalamus, mettant plus ou moins en résonance telle ou telle partie de notre être émotionnel. Cette partie du cerveau se révèle alors être ce lien physique que nous cherchions. Il existe bel et bien.

Ne rentrons pas tout de suite dans le registre des émotions, bien que nous nous apercevions de son poids dans notre question. Il semble que des liens entre la parole et la musique s'établissent à un autre niveau et dans d'autres zones du cerveau. C'est ce que nous explique Amélie Racette, étudiante qui effectue un doctorat en neuropsychologie clinique.

La musique peut-elle aider à récupérer la parole ?

« Il nous est tous arrivé d'utiliser la musique pour retenir un texte. Par exemple, nous apprenons plus facilement l'alphabet en le chantant. Ce rôle de la musique ne date pas d'hier. Dans la tradition orale, les chansons ont longtemps été utilisées pour transmettre des histoires d'une génération à l'autre. Ceci soulève cependant certaines questions : comment la musique peut-elle aider à produire la parole ? Pourrait-elle aider les personnes aphasiques à récupérer leur langage ?

Depuis plus de deux siècles, la littérature scientifique fait effectivement mention de personnes aphasiques qui arrivent à chanter correctement les paroles de chansons connues alors que leur production spontanée du langage parlé est déficiente. Dès 1736, Olof Dalin a fait la description d'une personne aphasique qui ne répétait qu'un seul mot, « oui », mais qui, avec de l'aide, arrivait à chanter son hymne nationale, de même qu'à réciter certaines prières. La question est donc de savoir comment le support de type musical peut permettre la récupération du langage. »

³⁰ F. LHERMITTE, 1991, Le cerveau

Existe t'il une « aphasie musicale » ?

« Du point de vue neuropsychologique, le traitement du langage est connu pour être fortement latéralisé dans l'hémisphère cérébral dominant, plus souvent gauche. Le traitement de la musique, quant à lui, serait associé à l'hémisphère droit. L'interdépendance entre le langage et la musique est confirmée par le fait que l'un et l'autre peuvent être sélectivement touchés ou préservés chez des patients. Si l'aphasie est l'atteinte spécifique du langage, l'amusie est l'atteinte spécifique de la musique. Les personnes amusiques ne peuvent plus reconnaître des mélodies familières (alors qu'elles peuvent en reconnaître les paroles), discriminer des mélodies ou les chanter alors que leurs facultés verbales sont bien préservées. »

Donc, si la musique est dissociée du langage dans le cerveau, comment peut-elle lui venir en aide ?

« Au départ, un phénomène de plasticité est invoqué : la musique activant plus fortement l'hémisphère droit, celui-ci pourrait prendre en charge le traitement du langage normalement laissé à l'hémisphère cérébral gauche. Récemment, à l'aide de techniques d'imagerie cérébrale, des chercheurs ont effectivement observé que c'est d'abord l'hémisphère droit qui est le plus actif lors de la réadaptation de l'aphasie. Cependant, une récupération optimale ne se fait que par la réactivation de l'hémisphère gauche. Des composantes non verbales, voire musicales, qui relèvent de l'hémisphère gauche pourraient donc être à l'origine de l'effet facilitant. »

Qu'en est-il des chansons, un alliage habituel de langage et de musique, deux domaines que l'on sait distincts du point de vue neuropsychologique ?

« Dans le cas des chansons, certains ont avancé que la musique et le langage pourraient être intégrés sous un même code en mémoire. La musique pourrait alors donner accès à ce code intact qui contient aussi du langage. Ceci expliquerait l'aide que la musique peut apporter à la production de paroles. Par contre, d'autres raisons peuvent venir expliquer cet effet. Par exemple, en ralentissant le débit de parole, la musique pourrait aider certaines personnes présentant une aphasie d'expression. De plus, le caractère simple et répétitif de la plupart des mélodies de chansons représente une structure qui aide à emmagasiner en mémoire. Ou encore, on pourrait assister à une dissociation automatico-volontaire, chez des personnes aphasiques, entre la production de paroles de chansons connues, qui sont bien ancrées en mémoire, et la production spontanée, généralement plus problématique.

Il existe donc de bonnes raisons de croire que la musique puisse améliorer la production de paroles chez certaines personnes aphasiques. Des thérapies orthophoniques se sont d'ailleurs inspirées de ce principe pour la réadaptation de cette clientèle. »

Qu'est-ce que la thérapie d'intonation mélodique et rythmée (Melodic Intonation Therapy) ?

« La thérapie d'intonation mélodique et rythmée, qui en est la représentante principale, met l'emphase sur un patron exagéré et simplifié, progressivement diminué, du rythme, de la prosodie et des accents pour les phrases travaillées avec le thérapeute. Il en résulte une sorte de mélodie rythmée à deux tons. Cependant, cette thérapie ne semble obtenir de résultats positifs que chez certaines personnes aphasiques répondant à des critères stricts comme avoir une capacité d'autocritique, une bonne motivation ainsi qu'un discours sévèrement atteint (afin qu'une amélioration soit possible) et stéréotypé avec difficultés de répétition, caractéristiques des aphasiques de Broca. »

Où en sont vos recherches ?

« Même si la parole chantée a acquis une bonne réputation, l'effet facilitant de la musique sur le langage est encore peu connu parce que peu étudié. Pour l'instant, la recherche n'a pu démontrer que les personnes aphasiques produisent plus de mots lorsqu'elles chantent que lorsqu'elles parlent. Des études empiriques sont donc nécessaires pour valider et expliquer ce phénomène. En plus de faire avancer les connaissances sur la place de la musique et du langage dans le cerveau, ces études possèdent un intérêt clinique : elles permettront de vérifier s'il y a de véritables raisons de penser que des personnes aphasiques vont améliorer leur production du langage par la musique. L'idée m'est donc venue de faire apprendre des chansons à des personnes aphasiques afin d'étudier le rôle de la musique dans la récupération des paroles de chansons. »³¹

Les liens entre la musique et la parole intéressent aussi le monde médical. Cette méthode neuro-psychologique permettrait de soigner des patients atteints d'aphasie. Elle est donc très prometteuse et nécessaire. A cette thérapie, on peut ajouter celles des orthophonistes qui visent à réguler des problèmes de prononciation par le chant. A ce stade, on voit que de nombreux domaines assez éloignés, provenant des sciences exactes ou des sciences humaines, se rejoignent autour des mêmes questions. De cette symbiose naît d'autres domaines d'investigation. Voici maintenant l'avis d'un, sinon le, spécialiste de la psycho-phonétique : Ivan Fonagy.

31 http://www.aphasie.qc.ca/divers/la_musique.htm

IV.2 Le langage émotif musical :

Le mot compte, la manière dont il est prononcé aussi. Car le sens seul ne porte pas l'émotion, mais la mélodie du verbe, l'intonation prise par le locuteur, la prosodie autrement dit, est un facteur essentiel dans l'intensité de la charge émotive portée par le mot. Ivan Fonagy, dans la vive voix³², montre que l'existence et l'importance de la mimique vocale deviennent apparentes au cours d'une régression momentanée de la communication gestuelle, prélinguistique. Car le signal expressif appartient au système de communication le plus évolué, la langue, mais en même temps relève du stade archaïque, autistique de ce que l'on appelle « l'acting », ou les mouvements corporels servent à réduire directement, immédiatement la tension (ex-expression = élimination de tout ce qui crée une tension). Il semble en effet qu'autrefois, la musique et la parole n'étaient pas séparées : Ainsi fonctionnait le prélangage des désirs et besoins inassouvis, qui consistait en une suite de mouvements plus ou moins violents et coordonnés, accompagnés d'émissions vocales réduisant momentanément la tension psychique.

« Ce ne sont pas tant les mots qui importent, qu'une prosodie particulière : Car dans la régression passagère des crises émotives, ce sont toujours les éléments prosodiques qui dominent, tandis que la structure grammaticale est perturbée, et que les mots sont relégués au second plan, servant avant tout de support à l'intonation et au schéma rythmique. »³³

« L'intonation émotive doit être interprétée comme une règle de transformation, une distorsion que l'on fait subir à la phrase neutre, déterminée par des règles de grammaire. Ces règles de transformation sont probablement indépendantes de telle ou telle langue. La transformation est assez souvent assez importante pour que le message émotif soit accessible même à ceux qui ne comprennent pas la langue en question. Ces tendances paralinguistiques apparaissent même en dehors de la communication verbale, dans la musique européenne vocale et instrumentale à partir du XVIème jusqu'au XXème siècles »³⁴

« L'angoisse, par exemple, s'exprime en français, en anglais, en allemand, en hongrois et dans la musique vocale européenne, par une forte réduction de la gamme mélodique. Dans la Flûte enchantée de Mozart, quand la statue du Commandeur vient

32 I. FONAGY, 1983, la vive voix, essais de psycho-phonétique

33 FONAGY et MAGDICS, 1963, Emotional patterns

34 Meyer, 1956, Emotion

frapper à la porte de Don Giovanni, l'intervalle de la voix chantée de Leporello se réduit soudainement d'une septième à une seconde. (« Je ne veux plus voir l'ami, doucement, doucement, allons nous cacher », deuxième acte, quinzième scène) et Papageno réagit de la même façon à l'approche de Sraastro (« Oh, si j'étais une souris, je me cacherais ! », quatorzième scène). On rencontre la même réduction des intervalles dans Pelléas et Mélisande (« j'entends parler derrière cette porte », quatrième acte, première scène, ou « parle plus bas : que font-ils ? », troisième acte, quatrième scène), ou dans Boris Godounov quand Varlam et Mishail s'échappent furtivement dans le troisième acte. L'intonation et la mélodie suggèrent une attitude corporelle : Celle de l'homme traqué qui se blottit, s'immobilise pour passer inaperçu. Le débit accéléré, dans la parole comme dans la musique, reflète l'excitation qui à son tour pourrait être considérée comme une récapitulation de la fuite ancestrale exigeant, tout autant que le combat, une très forte consommation d'oxygène, donc une accélération du rythme pulmonaire et cardiaque. Sartre interprète la peur comme « un évanouissement joué, une conduite magique qui consiste à nier l'objet dangereux avec tout son corps. »³⁵

L'expression de la tendresse est plus riche, plus variée dans le langage musical que dans la parole. La tendance fondamentale (vouloir épargner toute peine, tout effort à la personne aimée pour la ramener à l'état de parfaite harmonie) y est encore plus apparente. Dans le chant comme dans la musique instrumentale elle inspire une ligne mélodie ondulante, des changements lents, graduels. Ainsi dans l'air d'Orfeo (« E piu felice », premier acte de l'opéra de Monteverdi), dans l'air d'Ottavio (Mozart, Don Giovanni, premier acte, neuvième scène) ou celui de Pelléas (au quatrième acte de Pelléas et Mélisande de Claude Debussy).

Une analyse statistique des chants populaires hongrois montre clairement la préférence pour les transitions lentes dans l'expression de la tendresse comparée à celle de la colère³⁶. La musique va cependant plus loin, mettant au service de l'expression, l'harmonistique et la composition : elle caresse et dorlote en évitant toute dissonance, même légère, en multipliant les parallèles de tierce et de sixte, les reprises, les variations de hauteur hautement prévisibles.

L'expression musicale de la colère contraste avec le style doux, mélodieux de la tendresse. Le schéma mélodique et généralement très pauvre: souvent réduit à une ligne droite interrompue par des sauts d'une quarte ou d'une quinte, ainsi dans l'air d'Osmin n°3

35 SARTRE, 1960, Esquisse

dans le premier acte de l'Enlèvement au sérail de Mozart, ou dans le finale de l'opéra où Osmin intervient comme trouble-fête pour proposer, une dernière fois, qu'on pendre et écartèle le héros et l'héroïne, en saccageant la ligne mélodique pour un rythme « staccato » et une série de petits accents aigus qui se dressent comme des épines enfilées. Malgré les divergences dues aux styles de composition, la colère se reflète d'une façon semblable dans la voix de Beckmesser berné par Hans Sachs (dans les maîtres chanteurs de Nuremberg de Wagner).

Ce qui distingue donc l'expression musicale de la colère ou de la peur de celle des sentiments affectueux, c'est surtout la tendance à réduire la mélodicité du chant au profit du rythme, trait déjà remarqué dans la parole agressive; Et la prédominance de l'élément consonantique, du bruit, au détriment de l'élément vocalique.

[...] L'expression du désir languissant se distingue de la coquetterie, par la lenteur (ritardanto). L'anacrouse fait partie intégrante de la formule mélodique du désir, non seulement dans les langues à rythme oxyton comme le français, mais aussi dans celles à rythme baryton, comme les langues germaniques ou le hongrois. Le témoignage de la musique nous a montré que cette tendance n'était pas liée à telle ou telle langue, mais correspondait à une expression naturelle du plaisir. Le freinage du débit, la diminution de l'intensité, le rétrécissement des intervalles de la voix qui ensuite, s'amplifient graduellement, caractérisent l'expression verbale et musicale. C'est de cette façon que le désir s'exprime dans le prélude de Tristan et Yseult de Wagner, dans le deuxième mouvement de la sonate Les Adieux de Beethoven, dans la chanson de Brahms O wusst ich doch den Weg zuruck, au début de l'Après-midi d'un faune de Debussy, dans la scène I de l'acte II de Pelléas et Mélisande (« elle est si loin »).

Le déplacement du sommet mélodique vers le début ou vers la fin de la phrase correspond visiblement à une certaine attitude, à une certaine tactique pulsionnelle. A quelle tactique pulsionnelle correspond cependant l'intonation, le mouvement mélodique des phrases en général, indépendamment de ses réalisations concrètes ? Mais pourquoi vouloir ramener à tout prix l'intonation à une base pulsionnelle au lieu de se contenter de déterminer ses fonctions linguistiques ? Par la simple raison que le rôle démarcatif (ou commence et finit la phrase), modal (s'agit-il d'un énoncé ou d'une interrogation?) ou expressif (la phrase exprime-t-elle une certitude ou un doute?) explique l'importance que l'on attribue à l'intonation, mais n'explique point le plaisir causé par le mouvement tonal, l'effet esthétique

de l'intonation dans la poésie et la prose artistique, et de la mélodie - ce jeu savant de tension et de détente - dans la musique vocale et instrumentale. Serait-ce par un pur hasard que les éléments prosodiques du langage humain, le rythme et l'intonation, ont pu fournir les éléments d'un « langage » sans mots, sans phrases, du langage musical qui ne communique que pour faire plaisir ?

Si nous substituons au centre auditif un laboratoire acoustique extrêmement modeste, avec un ingénieur ne possédant que des diapasons, on peut démontrer aisément³⁷ que, pour déterminer la structure acoustique d'un bruit, l'ingénieur sera obligé de faire 40 fois plus de mesures que pour déterminer le spectre sonore d'une voyelle parlée ayant la même durée. Il devra faire 250 fois plus de mesures pour déterminer le spectre d'une voyelle parlée que pour déterminer celui d'une voyelle chantée. Même sans recourir à des expériences imaginaires, on conçoit aisément qu'une structure aussi régulière que qu'une voyelle, chantée sur le même ton et n'ayant que des tons partiels harmoniques (multiples entiers du ton fondamental), exige ou permet moins de choix par centi-seconde qu'un bruit ayant une structure quasi aléatoire. La perception d'un ton musical est donc plus plaisante que celle du bruit ou de la voix parlée puisque son décodage exige beaucoup moins d'effort. Sachant quelle part importante l'économie intellectuelle joue dans le plaisir esthétique³⁸, on sera à peine surpris de retrouver ce même facteur à la source de la jouissance musicale. »³⁹

Les analogies évidentes entre les formes mélodiques de la parole émotive et celles qu'on retrouve dans la musique européenne montrent la profondeur de la régression. L'intonation nous conduit semble-t'il, vers l'époque précédant la séparation de la musique et de la parole. Le mot grec « mousiké » qui désigne à la fois la danse, la musique vocale et instrumentale, les structures métriques des poèmes et les éléments prosodiques de la parole, semble avoir gardé le souvenir de ce « langage » ancestral, hypothétique, dont l'unique fonction était la résolution des tensions biologiques et mentales causées par des appétences inassouplies (faim, désirs sexuels) ou par la douleur, la peur. Ce prélangage consistait peut-être (comme le suggère l'ontogenèse de la parole) en une suite de mouvements plus ou moins violents, plus ou moins coordonnés, et accompagnés d'émissions vocales réduisant momentanément la tension psychique.

« Il semble que l'intonation affective est à la périphérie du propre du langage et

³⁷ I. FONAGY, 1960, le ton et la musique

³⁸ FREUD, der Witz

³⁹ I. FONAGY, 1983, la vive voix, essais de psycho-phonétique

du propre de la musique. »⁴⁰

Dans l'interprétation artistique, l'intonation obéit à des lois assimilables à celles qui régissent les formes musicales. Par là, elle s'éloigne de sa fonction initiale qui est de transmettre, de façon intelligible à tous, des expériences et, avant tout, des sentiments. Cependant, tout en s'éloignant de la parole de tous les jours et en se rapprochant de la musique, cette intonation, loin d'appauvrir, enrichit le message; ajoutant les moyens d'expression musicaux aux possibilités de l'expression linguistique. Ainsi le message émotionnel est souvent mis en emphase par des distorsions plus importantes de la courbe neutre : en élargissant le registre. Les variations rythmiques plus importantes permettent aussi des mises en valeur d'accent, bien plus fortes que dans le langage parlé usuel. Un bon exemple est le Sprechgesang qui fait tendre la courbure prosodique naturelle vers le chant, combinant donc, ces deux modes d'expression.

S'il existe comme le prétend I. Fonagy, un concept sous-jacent à la parole et à la musique qui implique la communication des émotions, alors il semble fort intéressant d'étudier ces deux voix d'expression de front et de manière complémentaire, afin d'extraire d'une telle étude comparative, des paradigmes concernant la communication sonore émotionnelle. Jusqu'à aujourd'hui, l'étude des émotions est envisagée comme appartenant aux domaines parapsychologiques. Peu de chercheurs en sciences pures s'y intéressent comme le souligne R. Carlson et B. Granstorm : « Les aspects émotionnels de la langue parlée n'ont pas été étudiés en détail. Dans le futur, cette aire de recherche permettra de mieux comprendre la relation entre la parole et l'expression musicale. »⁴¹

Le renoncement à s'engager dans ce domaine, pour l'instant parascientifique provient certainement de la versatilité des émotions, de la difficulté à les définir ou bien même, de les connaître. De plus, il existe dans cette communication des émotions, une indétermination intrinsèque qui la rend difficile à analyser : Le récepteur a la possibilité d'inflencer la nature du message par son interprétation. Nous ne sommes pas tous « touchés » de la même manière par une pièce musicale ou par la voix d'un acteur. Nous sommes sans cesse en train de filtrer des messages émotionnels faisant de chacun de nous, des antennes sentimentales dont les caractéristiques sont différentes et qui plus est, variables dans le temps. Pourquoi le discours d'un mendiant fera-t-il tirer des larmes aux uns tout en attisant la haine des autres, alors qu'il est le même pour tout le monde. Cette éventuelle multi-représentativité d'un message émotionnel par un récepteur entraîne une difficulté

40 J. RISCHÉL, 1990, Invariance in the linguistic expression, with digression into music

certaine pour l'analyse. Je pense d'ailleurs que nous éprouvons tous cette difficulté au jour le jour dans nos vies quotidiennes lorsque nous avons un "problème de communication".

IV.3 la musicalité ?

IV.3.a Peut-on quantifier la musicalité ?

Nous avons déjà introduit l'une des techniques évocatrices qui permettent à l'artiste vocal d'assimiler la parole à la musique à l'aide d'une mélodicité élevée qui prête à ses mots. La phrase semble se déplacer, monter et descendre en progressant dans le temps. On pourrait concevoir la musicalité comme une dimension de profondeur de la mélodie phrastique qui lui permet tantôt de s'approcher, tantôt de s'éloigner du plein-chant. La musicalité de la voix dépend de la régularité de la distribution des fréquences fondamentales à l'intérieur d'une syllabe. La mesure la plus appropriée pour mesurer la régularité d'une succession de fréquences est probablement celle de la redondance. En considérant les fréquences des vibrations successives comme une chaîne markovienne, il est possible de déterminer l'information ou la redondance de la distribution des fréquences. La musicalité paraît être directement proportionnelle à la redondance.

La voix humaine est perçue comme du chant s'il elle dépasse un certain degré de musicalité (c'est-à-dire de régularité), sinon, elle est considérée comme étant de la parole. Mais cette dichotomie n'est pas universelle. Les Maoris, par exemple, distinguent quatre catégories qu'ils désignent par quatre termes différents : Le koorero correspond à la parole, le haka à une parole « stylisée », mélodieuse, le karakia est de la mélodie rituelle et le waiata est du chant. On peut admettre, même dans le cadre des langues indo-européennes, des catégories intermédiaires entre chant et paroles quotidiennes. Klara Magdics⁴², par exemple, relève neuf degrés de musicalité, de régularité musicale dans la voix. La solennité, la tendresse, la prononciation enjouée rapprochent la voix du pôle positif, de la musicalité maximum. La mélodicité de la parole croît avec la tendresse et décroît brusquement dans l'expression des émotions agressives. Les clichés mélodiques du langage infantin (na-na-nère) ou du français parisien sont toujours chantonnés. Cette troisième dimension des mouvements mélodiques joue dans toutes les langues un certain rôle, sans avoir de fonction

41 R. CARLSON, B.GRANSTROM, 1990, Performance rules in a text-to-speech system

distinctive. Toutefois, selon List⁴³, le Nyanguamata, une langue de l'Australie occidentale, signale l'intensif sémantique par une plus grande musicalité. Dans cette langue, la musicalité équivaut au suffixe du superlatif. La musicalité de la voix confère une nouvelle dimension à l'interprétation, et que, dans cette dimension, le poème récité, sonore, évoluant entre un pôle positif et un pôle négatif, s'enrichit d'un contenu défini, différencié, mais difficilement définissable. Ce qui est évident, c'est que la musicalité, dans un sens plus large du terme, est indépendante de la régularité des vibrations, de la musicalité dans le sens technique du terme. Le chuchotement ou l'absence de musicalité physique est aussi expression musicale, non verbale.

On peut se demander, d'après cette dernière remarque, à quoi peut bien servir une mesure technique de la musicalité, si celle-ci n'est pas représentative du sens musical ? Citons Abbi Patrix : « Lorsque je suis sur scène, que je raconte, l'ensemble des paramètres est conduit de manière à converger pour exprimer une intention. »⁴⁴ La mesure citée précédemment serait alors peut-être un bon indice pour un conteur lorsqu'il exerce sa voix. Imaginons qu'un compositeur désire utiliser cette mesure de la musicalité à des fins musicales comme contrôleur temps réel...

IV.3.b Transformation de la musicalité de la voix

Lors des réunions sur la voix qui se sont tenues à l'IRCAM durant cette année, Emmanuel Nunes a émis le souhait de pouvoir passer d'une voix parlée à une voix chantée de façon continue et ce, par le biais d'un programme informatique. Cette idée d'un Sprechgesang contrôlé manuellement à posteriori, voudrait que d'une phrase parlée puissent émerger des moments chantés de manière très réaliste. Ce contrôle ressemble fort à un contrôle de la musicalité, dans le sens technique que nous avons décrit jusqu'ici. Dans Max/MSP, j'ai tenté de réaliser un tel programme. Grâce aux objets Yin~ (estimateur de la fréquence fondamentale) et psych~ (Synthèse PSOLA), ce patch nous permet de modifier en temps réel la fréquence fondamentale d'un signal de parole. J'utilise une table pour décrire la distorsion. La courbe dessinée dans cette table ressemble d'ailleurs fortement à une courbe de distorsion puisqu'elle est en réalité une courbe de transfert entrée-sortie (Fo réel en entrée et Fo de la synthèse en sortie). On peut faire ainsi saturer Fo de manière à ne percevoir plus

⁴² K. MAGDICS, 1963, Melody of speech

⁴³ LIST, 1963, Speech and Song

qu'un ton monocorde. On peut au contraire accentuer les accents ou même les inverser. Tout ceci en temps réel. Une courbe en escalier fait de la phrase parlée, une vocalise. Si l'on place en parallèle une table linéaire, on peut faire un mix entre les deux et ainsi passer de manière continue et fluide d'une voix "normale" à une voix "chantée" ou harmonisée. Ce programme est très simple et possède l'avantage d'influer sur la fréquence fondamentale en temps réel. En revanche, il ne peut agir sur le rythme et la durée. Il est en effet très difficile de pouvoir jouer sur ces paramètres temporels en temps réel tout en gardant un aspect naturel à la transformation. Ce désir est d'ailleurs paradoxal.

Benoit Meudic, assistant musical à l'IRCAM, travaille sur une pièce dont un passage est constitué d'un morphisme progressif du français parlé à l'anglais parlé, en temps réel. Outre la difficulté de la transition phonétique, se dresse un obstacle bien plus grand : Le changement de prosodie. Comme nous l'avons rappelé lors de la première partie, le Français et l'Anglais possèdent des structures prosodiques fondamentalement différentes. Le Français possède des accents fixes en fins de mots. L'Anglais est une langue à temps accentués. Cette différence dans les distributions des accents ne permet pas d'action en temps réel pour faire ressembler une langue à une autre sans passer par un apprentissage préalable. Prenons comme exemple le Hongrois et le Français. Toutes deux langues à accents fixes. Le Hongrois situe ses accents au début des mots. Si l'on essaye d'effectuer une telle transition progressive du Français vers le Hongrois, en déplaçant les accents de la fin du mot au début, cela suppose déjà de connaître les frontières des mots (reconnaissance de la parole), mais surtout de les prédire. On ne peut attendre l'information d'accent d'un mot Français pour déclencher une transformation puisqu'on aurait dû l'effectuer dès le début de ce mot. Cette causalité intrinsèque est due à la construction même des groupes accentuels qui s'élaborent certainement en amont de la phrase réalisée. Voilà pourquoi un tel exercice de traduction instantanée n'est pas réalisable en temps réel sans l'appui d'un modèle complexe et adaptatif. On fera alors appel aux précédents modèles cités dans la première partie. Une manière simple serait peut-être d'utiliser l'accent final du Français comme pré accent marqueur pour l'accent initial du Hongrois.

En ce qui concerne la transformation du rythme, elle nécessite aussi un découpage segmental et une prédiction d'événements qui s'accordent mal avec le temps réel. Il semble que l'agencement des paramètres temporels de la voix parlée ne puisse être aujourd'hui altéré que par des transformations globales comme des variations de débit par

⁴⁴ entretien avec Abbi Patrix : Figure sur le Cd-rom joint

exemple. Or ce genre de transformations sur la voix ne génère souvent qu'une qualité déplorable et souvent choquante, dépourvue de naturel. Cela vient du fait que "dans la parole emphatique, il y a généralement une corrélation étroite entre la suraccentuation (ou déplacement d'accent) et la fréquence des pauses emphatiques et des « attaques dures » qui prêtent au discours un caractère Staccato⁴⁵. Une transformation acceptable sur un signal de parole doit prendre en compte tous ses paramètres simultanément, sans quoi elle détruit ces relations de corrélation, marque de son aspect naturel.

Malheureusement, de telles combinaisons de transformations locales sont difficilement réalisables en temps réel sans l'établissement d'un modèle prosodique a priori. Ceci car les instanciations des structures des phrases parlées sont des mélanges de hors temps et de temps réel. Cela rappelle d'ailleurs la manière dont la musique est gérée lors d'une improvisation. Les articulations se callent sur une structure connue à l'avance. Alors le temps réel s'allie au temps syntaxique musical dans un rapport qui rappelle l'expression orale. Difficile alors de gérer les deux en même temps par des transformations globales. Romain Kronenberg, assistant musical à l'IRCAM, travaille sur une pièce de théâtre contemporaine inspirée de Fernando Pessoa et intitulée Le Privilège des chemins⁴⁶. Le metteur en scène désirait que deux personnes échangent de voix en temps réel dans le spectacle. Malheureusement, nous ne possédons pas encore les moyens techniques de le faire. Mais ce vœu soulève une question qui rejoint celle que se posait Benoît Meudic. Supposons que l'on arrive à transformer les paramètres acoustiques d'une voix pour la faire coïncider avec une autre: Le timbre, la hauteur... (Chose que l'on va savoir faire très bientôt grâce à Fernando Villavicencio Marquez, auteur d'une thèse sur la transformation de voix⁴⁷). Que resterait-il de l'individu dans cette voix. De façon certaine, sa personnalité transparaîtrait à travers son intonation et sa prosodie. Mais si à présent, on souhaite changer cette dernière en temps réel pour arriver à l'ultime transformation de l'identité vocale ; On se heurte à une dépersonnalisation de la voix ainsi synthétisée car nous ne possédons pas tous le même temps syntaxique. Aussi, ce genre de transformations prosodiques n'est réalisable qu'hors temps. De manière à tenir compte aussi du débit de parole... Toutefois ces transformations peuvent se révéler très intéressantes lors de leurs exploitations à des fins artistiques. Lorsque l'on désire justement faire briller le naturel par des distorsions audibles. Nous y reviendrons dans l'interview d'Abbi Patrix qui suit la prochaine partie.

⁴⁵ I. FONAGY, 1983, la vive voix, essais de psycho-phonétique

⁴⁶ Pièce jouée par la comédie française au Studio-Théâtre

⁴⁷ Thèse se déroulant à l'IRCAM

IV.3.c Ecoute(s) :

Durée, hauteur, mais aussi timbre et pause sont des paramètres de la musicalité. Leurs infinies combinaisons situées dans leurs contextes d'utilisation font de la musicalité, une notion complexe à définir et de surcroît mouvante. Sans oublier que comme pour la communication des émotions, la musicalité dépend fortement des oreilles du récepteur...

Lors d'un entretien avec Roland Cahen⁴⁸, nous nous sommes confrontés à la difficulté de définir deux des termes que nous employons ici : **spontanéité** et **musicalité**. Il avertit, « il n'y a pas de naturalisme. ». Lorsqu'on enregistre une discussion et que l'on la reproduit ensuite, le contexte s'en trouve bien sûr changé. On se retrouve dans une situation acousmatique telle que Jacques Perrioux la désigne : Lors de l'écoute de parole enregistrée, on ne peut être dans la situation de l'objet parlant. Il se crée un décalage dont le manque d'informations nous pousse à reconstituer à partir d'indices plus ou moins fiables ces informations manquantes. Comme il nous manque la causalité instrumentale, on ne voit pas la cause première, l'action qui produit⁴⁹, il ne reste qu'un son désincarné. On va tricher à l'écoute, de manière créative ou malhonnête. Comment la chose pourrait être entendue autrement ?

Lors de travaux avec des acteurs, R. Cahen leur propose d'enregistrer des scènes de tous les jours durant leurs vies quotidiennes. Puis il les invite à reproduire ces scènes dans deux contextes différents : Avec ou sans avoir écouté l'enregistrement avant l'action. Le résultat de cette expérience est que l'on constate de grosses différences dans le cas où les acteurs n'ont pas entendu l'audio préalablement. Il en conclut que les scènes réelles sont beaucoup plus expressives que les scènes « artificielles ». Il semble que la spontanéité traduise l'instantanéité dans ce cas. Jean Gilbert, dans son travail d'acteur, cherche à se défaire des clichés mélodiques de manière à faire resurgir une spontanéité. Stanislavsky fait de même.

L'autre avertissement de R. Cahen concerne la musicalité. La musique est l'organisation abstraite des sons, des relations qui articulent des objets hiérarchiquement organisés qui ne possèdent pas de sens et dont la forme se construit dans le temps. La parole, quant à elle, voit son articulation contrainte par des relations linguistiques et n'est pas faite

⁴⁸ L'entretien avec Roland Cahen figure sur le Cd-rom joint.

⁴⁹ acousmatique selon Schaeffer

d'abstractions. Mais si ces définitions semblent être plutôt relatives à la naissance et la création de la musique et de la parole, Il ajoute qu'elles dépendent aussi fortement de l'oreille qu'on leur prêtent. Schaeffer définit quatre types d'écoutes :

<p>4. COMPRENDRE</p> <ul style="list-style-type: none"> - pour moi : signes - devant moi : valeurs (sens-langage) <p>Emergence d'un contenu du son et référence, confrontation à des notions extra-sonores</p>	<p>1. ÉCOUTER</p> <ul style="list-style-type: none"> - pour moi : indices - devant moi : événements extérieurs (agent-instrument) <p>Emission du son</p>	1 et 4 : objectif
<p>3. ENTENDRE</p> <ul style="list-style-type: none"> - pour moi : perceptions qualifiées - devant moi : objet sonore qualifié <p>Sélection de certains aspects particuliers du son</p>	<p>2. OÛIR</p> <ul style="list-style-type: none"> - pour moi : perceptions brutes, esquisses de l'objet - devant moi : objet sonore brut <p>Réception du son</p>	2 et 3 : subjectif
3 et 4 : abstrait	1 et 2 : concret	

Pierre Schaeffer, Traité des objets musicaux, p. 116

Ecouter c'est rechercher dans les indices du son sa provenance possible (la causalité) ; *ouïr* c'est écouter le son sans se poser de question ; *entendre* c'est pratiquer l'écoute réduite ; *comprendre* c'est percevoir la signification des sons (langage). De ces quatre écoutes, nous ne retiendrons pas « l'ouïr » et renommons les trois suivantes :

-l'écoute causale : identification de bruits informatifs (klaxon, porte qui claque...)

-l'écoute linguistique : qui permet la compréhension du sens des mots...

-l'écoute musicale : des relations entre les sons...

De nombreuses pièces tournent autour et jouent avec l'écoute causale : Ainsi, Yann Parentoen nous fait explorer le voyage sonore d'une lettre. Ce travail narratif grâce aux ambiances sonores, aux bruits et aux sources informatives nous fait voyager de la gare Montparnasse jusqu'en Bretagne. De même, la lettre de Alain Savouret, tirée de sa sonate baroque, imite les bruits d'une machine à écrire. La même écoute est sollicitée, mais dans ce cas, les bruits sont reproduits par synthèse instrumentale, ce qui vient approfondir la recherche sonore sur le lien de causalité. Les bourdons de l'Odyssée réalisés par le GRM de

Bourges sont des ambiances de grillons harmonisées. Dès lors que l'illustratif devient fonction dans le paysage musical, il est difficile de savoir comment s'exerce le passage de l'écoute causale à l'écoute musicale.

Lorsque nous parlons, les mots suscitent sans cesse l'écoute linguistique de nos interlocuteurs, mais qu'en est-il de la prosodie ? A quelles autres écoutes nous référons nous lorsque nous entendons quelqu'un parler ?

Pour R. Cahen, on peut définir la musique par ce qu'elle comporte : des répétitions, des analogies, des différences, une conduite énergétique... La parole n'est pas vraiment de la musique mais elle possède une dimension musicale. D'ailleurs la musique vient originellement de la prosodie. Les chants grégoriens sont rythmés par les accents du latin. Le blues vient des palabres africains. Le tambour africain parle (« tais toi, tu parles trop »).

Peut-on alors dériver des structures musicales inédites de langues prosodiquement différentes. On peut imaginer un outil logiciel d'extraction automatique de lignes mélodiques à partir de contenus prosodiques pour composer musicalement dès lors que l'on aurait approfondi les liens entre prosodie et musique... Il me semble que ce serait très intéressant pour l'analyse des phénomènes paralinguistiques universaux. L'on se rendrait certainement compte de structures communes à tous et l'on pourrait alors comprendre pourquoi certaine musique nous « parle » à tous. Nous y reviendrons en dernière partie. On voit bien, à la lumière de cet entretien, qu'un son ne devient musical que selon l'écoute qu'on lui prête. C'est pour cela qu'on ne peut définir une musicalité au sens large et universel, car chacun possède sa propre musicalité, son propre sens esthétique qui fera tantôt intervenir l'écoute musicale, tantôt intervenir l'écoute causale. Nous avons vu à quel point le traitement de l'information sonore dans notre cerveau est complexe. Il ne semble alors pas dangereux de faire l'hypothèse que notre cerveau est capable de combiner différentes écoutes. L'écoute d'une pièce instrumentale permet de reconnaître l'arrivée soudaine d'un nouvel instrument dans un ensemble. Est-ce alors l'écoute causale qui joue ou bien seulement l'écoute musicale ? Il paraît plus prudent de penser que les deux participent à la détection d'un tel événement. Ce parallèle nous permet maintenant de revenir à nos considérations sur la voix et sur sa musicalité. Comment entendons nous une voix ? Est-ce que seul notre écoute linguistique participe à ce fait ? D'accord, cette écoute nous permet de reconnaître des mots et de les comprendre. Mais qu'en est-il de la participation de l'écoute musicale ? Ne serait-elle pas associée durant une conversation pour reconnaître des formes universellement connues et contenues dans l'expression intonative ?

On est à même de comprendre un étranger nous demandant l'heure dans sa langue. L'écoute causale et la vision nous font reconnaître la voix de celui-ci ainsi que sa montre autour de laquelle il fait un geste circulaire. Puis l'écoute linguistique est déroutée à cause de mots qu'elle ne comprend pas. Enfin l'écoute "prosodique" reconnaît la question et nous engage à y répondre. Cette écoute "prosodique" me paraît difficile à définir dans le sens de Schaeffer sans faire l'hypothèse que notre cerveau combine toutes les écoutes causale, linguistique et musicale en même temps pour en extraire les informations d'intention, d'émotion et d'intonation.

Cette musicalité identifiable dans les faits prosodiques est peut-être même à la naissance de ceux-ci. « Et l'on pourrait se demander si, en réalité, il ne serait pas plus juste de dire, du moins du point de vue physiologique, que dans l'élaboration du message parlé, ce sont les « timbres » et les bruits qui « accompagnent » les éléments musicaux. Car il est bien certain qu'en fonction d'un contenu psychologique donné, ce sont d'abord le rythme et la mélodie qui sont commandés et réalisés, en utilisant comme matière brute le son fondamental encore informe et polyvalent qui sera ensuite façonné par la bouche et flanqué de « bruits » qui aideront à faire de lui le noyau d'une syllabe. »^{50 51}

IV.3.d Entretien avec un conteur, Abbi Patrix :

Il suffit d'écouter Abbi Patrix parler⁵² pour se rendre immédiatement compte que la voix est musicale. Musicien de formation, musicien dans l'âme, il est devenu conteur car on ne voulait pas, à l'époque, de théâtre musical. Le travail du conteur sur sa voix consiste à améliorer sa qualité phonatoire. Chacun de nous possède une note autour de laquelle il oscille lorsqu'il parle. Cette note définit un registre dans lequel notre voix est timbrée. Quand je parle, quand ma voix est timbrée, je suis autour du RE. Lorsque je suis sur scène, que je raconte, l'ensemble des paramètres est conduit de manière à converger pour exprimer une intention. La voix fait partie de la composition de l'espace acoustique au même titre que les instruments de musique... C'est pourquoi, je « tonalise ». Sur scène, l'intonation devient musicale : Elle devient tonalisation. Et ainsi, elle s'accorde avec une percussion, car toute manifestation sonore a une tonalité.

⁵⁰ G. FAURE, 1962, Aspects et fonctions linguistiques des variations mélodiques dans la chaîne parlée

⁵¹ G. FAURE, 1962, Recherches sur les caractères et le rôle des éléments musicaux dans la prononciation anglaise

⁵² L'entretien avec Abbi Patrix figure sur le CD-ROM joint.

Le conteur est un musicien qui accorde sa voix, un chanteur qui s'ignore. Il raconte en musique, tantôt en mineur, tantôt en majeur et ainsi, véhicule un climat, une couleur. Mais Abbi nous prévient : Un ton, c'est énorme quand on raconte (écouter l'enregistrement pour s'en persuader). Un conteur va jusqu'à des sauts d'une octave. On est plus très loin du chant, même si chanter, c'est vibrer par la note, c'est exprimer une note. La tonalisation, tout le monde l'a de même que le sens du rythme. Il est intéressant de remarquer que lorsqu'on donne une autre note ou un autre rythme à quelqu'un, il n'arrive généralement plus à le suivre. (Vers une identité musicale ? Voir la remarque d'Edouard Sievers sur la mélodie du texte un peu plus loin). D'ailleurs, les occidentaux ont du mal à se concentrer sur une histoire à cause de la musique. Pour moi, on ne peut soustraire celle-ci au conte. C'est la musique de l'histoire. En studio, c'est notre principal activité : Amplifier la dimension musicale. J'écris des contes comme d'autres écrivent la musique. Le montage final ne pose jamais de problème et ce, grâce à la tonalisation. Le mixage est toujours fait par les musiciens, car ils savent bien gérer les plans.

Le travail d'un ingénieur du son peut se dérouler de deux manières :

- œ Passive : Il enregistre simplement du mieux qu'il peut (ce qui peut être très complexe)
- œ Active : Il participe à l'écriture et devient musicien.

C'est dans cet esprit de collaboration qu'est pensé le dernier spectacle d'Abbi Patrix : Au bout du monde. Tout le spectacle est basé sur le travail de la voix. Se balançant toujours entre le parler et le chanter, Il utilise aussi des nouvelles technologies : Sampling et jeu en temps réel mais surtout, spatialisation. Alexis Blanchet, l'ingénieur du son a mis au point un système de sonorisation multi-diffusion totalement transparent, sans égalisation et sans retour sur scène. La cohérence acoustique qui en résulte rapproche tout le monde du conteur, qui se permet durant le spectacle d'alterner entre voix amplifiée captée par un micro HF frontal et son acoustique direct. Le but est d'établir un contact transparent, naturel avec le public pour jouer avec, après. Le spectacle se termine sur le souffle du conteur sonorisé de telle manière que le lien de proximité n'est plus choquant, devenu invisible. C'est ce genre de transformations dont nous parlions précédemment qui peuvent paraître gênantes si elle ne sont pas dirigées, fonctionnalisées musicalement car elles détruisent le naturel de la voix.

Abbi Patrix a fait de nombreuses autres expériences dans la relation qui lie la musique à la parole. Dans Le Compagnon, spectacle réalisé avec le violoniste Bernard Chaize, Le conteur parle et s'arrête progressivement, et c'est le violon qui continue le discours et qui raconte tout. Le musicien devient alors conteur. Dans l'histoire du galop de cheval, Le soubassement rythmique joué par une percussion représente l'éternité. Les

accélération et décélération du débit du conteur entraînent des décalages locaux avec ce rythme. Le conteur saute sur le cheval, en descend... Un deuxième percussionniste vient illustrer ces actions par des frappes isolées. L'auditeur est dans les espaces libres. Écouter : C'est résonner avec...

On ne peut qu'apprécier lorsque cette phrase sort de la bouche du conteur. Un bon conteur, c'est une vibration. Son meilleur partenaire est le son. Abbi Patrix fait du théâtre musical : Il joue du piano et de la tamboura durant l'interview. De nombreuses fois, j'ai remarqué qu'il me parle dans la peau du conteur et qu'il vit ainsi perpétuellement ce changement d'état qu'il décrit lorsqu'il monte sur scène. De la tête qui parle, on se déplace à l'intérieur, et c'est le corps qui parle. « Le corps est une forge, tous les éléments rentrent en action : L'air, le feu, le sang... A travers le larynx, on forge du son. La parole est un matériau forgé. La langue et les cordes sont un métier à tisser : La parole est un fil... La parole incarnée. »⁵³

⁵³ La parole du monde, mythe de création Dogons, Mercure de France

V Le rapport parole/musique

Poème de Verlaine :

« Et quand vous parliez, à dessein distrait,

Je prêtais l'oreille à votre secret :

Car la voix, ainsi que les yeux de Celle

Qui vous fait joyeux et triste, décèle,

Malgré tout effort morose et rieur,

Et met au plein jour l'être intérieur... »

V.1 Vers une identité musicale ?

A travers ce mémoire, on peut se rendre compte que la musicalité et l'émotion possèdent des canaux de transfert dont les fonctionnements sont similaires. Dans les deux cas, le récepteur peut influencer sur la nature même du message selon son interprétation. Certains diront qu'ils entendent du bruit là où d'autres entendront de la musique. Certains reconnaîtront de la perfidie dans la voix d'un innocent. Car ces deux canaux de transmission sont très sensibles au contexte de la situation comme les télécommunications le sont au bruit. Il ne faut pas négliger ce qu'il y a de culturel dans une voix, d'identitaire.

Un domaine intéressant est dans l'étude de l'influence de la langue maternelle d'un musicien sur sa manière de jouer lors d'une performance ou bien sur sa manière de composer. Il a été dit que les Français font attention à la couleur des sons comme Debussy, Berlioz, Ravel, Messiaen, Boulez, Bayle. Cela serait dû à l'importante place que prend la voyelle dans la langue française. D'ailleurs, ce fait rend difficile le chant des mélodies en Français. Il a aussi été dit que le phrasé de nombreuses musiques hongroises proviendrait du fait que l'accent tend à se trouver sur la première syllabe en hongrois.

Il y a donc déjà une identité nationale qui transparaît aussi fortement dans la musique que dans la voix. « Les différences entre les réalisations du chanté anglais et du chanté Bengali semblent être corrélées aux différences entre les structures prosodiques de

ces deux langues. »⁵⁴

On peut préciser géographiquement et surtout en France, que les régions possèdent aussi leurs identités musicales et accentuelles. De nombreux phonétistes, ethnologues et ethnomusicologues l'étudient, prêtant, pour le mieux, attention à la coordination de ces deux évolutions.

Mais je pense que l'on peut même aller plus loin avec notre télescope culturel en pointant l'individu. Chacun possède son identité musicale, sa connaissance, sa culture, sa manière de parler. Sans doute, sommes nous moins attentifs aux nuances prosodiques qu'aux différences entre les musiques. « Bien qu'il existe de nombreuses manières de parler, le domaine de la parole reste bien plus homogène que le domaine musical qui présente une énorme variété de types et de formes différentes. »⁵⁵

Cependant nous possédons tous une manière propre de s'exprimer vocalement. Même si celle-ci ne nous paraît pas perceptible à la première écoute, nous reconnaissons toujours nos interlocuteurs à leurs voix lors d'une conversation téléphonique. Notre expression personnelle se transcende en son à travers notre voix et peut devenir ainsi le phare d'autres voies d'expression artistique. Nous en parlions lorsque nous citions les moyens de composition des musiques extra européennes avec Monder Ayahri. Dans la letera Amorosa de Monteverdi écrite dans le style « rappresentativo », la voix, matériau phonique, conserve les inflexions expressives du texte. Elles sont stylisées, affinées par une extrême attention aux contours mélodiques et rythmiques de la phrase. L'inverse peut aussi se produire dans la littérature lorsque la voix influence l'écriture :

Edouard Sievers, un des précurseurs de la phonétique moderne, fut sans doute le premier à prendre au mot la métaphore de « mélodie du texte », bravant ainsi le sourire pas toujours indulgent des confrères, et à attribuer au texte, poème ou prose, une mélodie qui se réalisait au cours de la lecture. Il est allé encore plus loin, en postulant l'existence de constantes prosodiques individuelles, des structures rythmiques, des schémas mélodiques propres à tel ou tel poète ou écrivain à une certaine période de sa vie, constantes qui se reflétaient à travers ses textes. Selon Sievers, la lecture simple et fidèle d'un texte reproduisait involontairement ces caractéristiques vocales.

D'après cet écrivain, nous possédons tous une identité musicale et prosodique qui peut transparaître à travers l'écriture. Ce point de vue semble s'accorder avec celui

⁵⁴ B. HAYES, A. LAHIRI, 1990, Durationally specified intonation in English and Bengali

⁵⁵ K. SHERER, 1990, Emotion expression in speech music

d'Abbi Patrix pour qui, nous avons tous une note autour de laquelle nous évoluons de manière personnelle. Cette évolution est le reflet d'une identité prosodique qui s'accompagne d'une identité culturelle et musicale.

« Seule l'intériorité sans objet, la subjectivité abstraite se laisse exprimer par les sons. Subjectivité abstraite qui est un moi entièrement vide, sans autre contenu. La tâche principale de la musique consiste donc, non à reproduire les objets réels, mais à faire résonner le moi le plus intime, sa subjectivité la plus profonde.»⁵⁶

La musique pourrait être, en quelques sortes, une communication visant à mettre en liaison deux personnes dont les identités musicales sont semblables. Gardons à l'esprit qu'une des formes de rassemblement des êtres réside dans l'écoute d'une même musique. C'est sur ce principe que se créent des « ethnies » musicales d'adolescents qui se reconnaissent dans une même identité grâce à un groupe ou un courant musical. Remarquons que ces « ethnies » possèdent bien souvent une manière de parler, voire un vocabulaire propre. Cette manière exclusive de s'exprimer naît de déviations, de distorsions d'un patron standard toujours renouvelé.

« Le chant passe lui aussi pour une modulation de la voix; Don qui révèle une parfaite maîtrise du son, il apporte à ceux qui le possèdent l'amour de tous et un certain pouvoir sur le monde. Les composantes acoustiques (mélodie, intensité, timbre) et rythmiques (articulation, tempo ou vitesse d'élocution, durée des phones et des pauses) peuvent varier avec le ton de la voix, qui est fonction de l'usage que l'on fait de celle-ci : Car toute parole est orientée. »⁵⁷

En effet, toute voix porte la marque unique de l'intériorité de chacun. « Mallarmé entend dans la musique [...] la transcription sonore du verbe, la « voix intérieure de notre esprit ». Toute âme est une mélodie, qu'il s'agit de renouer; et pour cela, sont la flûte ou la viole de chacun. »⁵⁸

Rousseau identifiait la musique à la langue naturelle. C'est la raison pour laquelle il considère la mélodie comme inséparable du chant et de la parole. Le sens de la musique est dans la mélodie, le prolongement de la voix humaine. Les sons n'ont de réalité que parce qu'ils épousent et se fondent dans les « intonations » naturelles de la voix et dans les inflexions de la langue. Il identifie aussi la composition musicale à un unique principe vital et organique : Le langage parlé, sans s'interroger sur la déperdition sonore que cela

⁵⁶ HEGEL, Esthétique

⁵⁷ A. CHARVET, 2004, La voix et ses métamorphoses dans « les métamorphoses » d'Ovide
⁵⁸ D. DUCARD, 1980, Le chant perdu de la langue : voix et écriture poétique

entraîne inévitablement.

Au manifeste mallarméen joignons la confession de foi de l'écrivain contemporain Louis René des Forêts qui expose l'attraction de l'écriture pour la musique, souveraine. « L'emprise qu'a toujours exercée la musique sur moi tient peut-être à ce qu'elle procède plus que tout autre art des lois de la mort. Mais pour m'en tenir à la musique considérée comme moyen d'expression, j'y vois surtout le milieu conducteur ou le double courant de la pensée et de l'émotion a le plus de chance de s'établir, cet échange s'opérant par la médiation d'un idiome particulier, heureusement délivré de la scorie des mots (soumis, il est vrai, à une syntaxe non moins stricte et même plus savante que celle à laquelle est astreint l'écrivain), mais où le rythme a toujours valeur souveraine. »

Cette déclaration reflète une autre similarité entre musique et parole liée à la construction même de ces deux langages. Certes, il semble que les moyens de compréhension de ces deux formes de communication soient semblables de par leurs natures. Cela provient certainement du fait que ce sont les deux seuls moyens pour l'Homme de faire passer ses émotions dans un message sonore. Mais il est des ressemblances encore plus frappantes dans leurs élaborations, en ce qui concerne leurs structures.

V.2 Similitudes structurelles entre la parole et la musique :

Que la voix soit parlée ou bien chantée, elle relève de la même intention, de la même motivation d'exprimer des émotions. Mais la parole et la musique possèdent des fonctions qui les différencient.

« Les buts d'une théorie sémantique sont au nombre de trois : La théorie doit tenir compte du nombre et du contenu de chacune des manières de lire une phrase; Elle doit détecter les anomalies sémantiques; Et elle doit reconnaître les relations paraphrasées entre les phrases. »⁵⁹

W. Slawson⁶⁰ y voit une analogie forte avec une théorie d'association musicale s'il on remplace simplement le mot « phrase » par « passage ». Il est vrai que cette opinion reflète aussi la définition de la musique donnée précédemment par R. Cahen.

Cependant, si des similitudes structurelles existent entre la musique et la parole,

⁵⁹ KATZ, FELDOR, 1963, The structure of semantic theory

⁶⁰ W. SLAWSON, 1990, Structure and association in descriptions of music, with a coda on linguistic analogies

elles apparaissent tout de même attenantes à deux sortes de messages: Les messages sémantiques et les messages esthétiques. « Le point de vue sémantique est une question posée au monde extérieur, l'information sémantique doit préparer une décision des actes; Il y a un caractère nettement utilitaire, mais surtout logique; Elle constitue un code normalisé; Elle est traduisible et commutable d'un canal à l'autre. Par contre, l'information esthétique ne se réfère pas à un répertoire universel, est de caractère personnel; le point de vue esthétique n'a pas de caractère d'intentionnalité, il détermine en fait des états intérieurs, l'information esthétique est spécifique au canal qui la transmet, elle n'est donc pas traduisible. Les prescriptions réglementaires en cas d'incendie, les programmes communiqués aux ordinateurs, les commandements militaires véhiculent de l'information sémantique : Les oeuvres d'art apportent surtout de l'information esthétique. »⁶¹

D'autre part, Moles considère les structures abstraites, constantes, qui sous-tendent la communication comme messages sémantiques, les réalisations concrètes, individuelles, qui constituent l'acte de communication comme des messages esthétiques. Ainsi les règles de la syntaxe, les phonèmes, la partition musicale, le sujet que représente le tableau relèvent de la sphère sémantique, d'autre part, les phrases individuelles, les sons de la parole, une réalisation individuelle de l'oeuvre musicale, le maniement du pinceau qui caractérise un peintre véhiculent de l'information esthétique.

Les deux définitions sont à la fois complémentaires et contradictoires. L'oeuvre musicale s'oppose en tant que source d'information esthétique à l'oeuvre verbale, elle appartient en même temps, à la sphère sémantique, conformément à la deuxième définition. Pour éviter de tels chevauchements, on aurait intérêt à distinguer les messages stylistiques que constitue toute réalisation individuelle d'une structure virtuelle, verbale, musicale, picturale ou autre, d'une part, et les messages esthétiques qui relèvent d'un code non verbal, d'autre part. La confusion inhérente à la définition double proposée par Abraham Moles est, toutefois, stimulante et expressive. Elle met en évidence l'analogie profonde entre message artistique non verbal (musical, pictural, chorégraphique) et le message contenu dans les faits de style verbal. Les messages stylistiques communiqués à l'aide de la manière de parler, de la façon de s'exprimer ont un caractère préverbal, malgré le fait qu'ils sont parfaitement intégrés à l'acte verbal proprement dit. Les messages communiqués à l'aide du style musical sont, par contre, des messages préverbaux intégrés à un acte de communication non verbale, on pourrait dire préverbale.

⁶¹ A. MOLES, 1958, *Théorie de l'information*

Mais si ces messages préverbaux sont communs à ces deux formes de communication que sont la parole et la musique, il réside néanmoins des différences fondamentales dans les constitutions de ces deux langages.

« La différence la plus frappante entre musique et parole est le rôle sémantique. Une autre différence importante réside dans le rôle de la structure. Dans le langage, la structure a pour but de véhiculer le message linguistique. Dans la musique, le but de la structure est elle-même. Les compositeurs construisent de magnifiques structures de groupes de sons. Si la structure est un point capital de la musique, elle est secondaire dans la parole. A l'inverse, la sémantique inhérente au langage est en définitive très fragmentaire et optionnelle dans la musique. De nombreuses similarités entre musique et parole subsistent tout de même. Lors de performances musicales et durant une conversation, l'expression est véhiculée par des déviations de durée et de hauteur par rapport à une « ligne neutre » d'événements attendus. Le groupement et la catégorisation sont deux principes utilisés par notre perception durant l'écoute de musique et de parole. Ces similarités ne sont pas inattendues. La parole et la musique sont les deux formes principales de la communication organisée entre humain par le biais de signaux acoustiques. Le mécanisme qui les perçoit et notre manière de pensée sont les mêmes dans les deux cas. Il a été démontré que le code acoustique, utilisé dans la musique pour décrire les buts expressifs est similaire ou presque identique à ceux qu'utilise la prosodie. »⁶²

D'ailleurs, on peut faire référence à de nombreuses tentatives de création d'un support écrit permettant de transcrire aussi bien de la musique que de la parole.

« Nous avons établi qu'il était possible d'utiliser une partition pour transcrire une description détaillée du rythme de la parole parlée ou lue. L'interprétation d'une séquence de notes n'est pas une tâche triviale, et il n'est pas du tout évident que des sujets musicalement entraînés suivent les mêmes règles lors de la transformation d'un flux rythmique en valeurs discrètes de notes. Cependant, des conclusions concernant l'interprétation des « up-beats », des sauts de lignes et des syncopes peuvent être faites. On peut d'ailleurs faire une analogie avec la si proche transcription prosodique des phonéticiens.

L'utilisation de règles pour transformer la transcription en valeurs de durée est intéressante si l'on regarde les valeurs de durée des segments de parole. Néanmoins, ces règles ont été élaborées dans l'intention de transcrire un flux musical et n'ont pas été

⁶² J. SUNDBERG: « Music is a language of emotions. Speech is also a language of emotions. ».

destinées à un usage si particulièrement non musical. »⁶³

Une autre difficulté dans l'acheminement d'un système commun à l'écriture de la musique et à la transcription de la prosodie réside dans la disproportion entre l'ambitus restreint de la voix parlée et l'immense registre musical. De même, la précision temporelle qu'exige la parole peut être un obstacle.

Toutefois les outils logiciels sont capables de cette précision. Grâce à TALKAPILLAR (voir la partie sur la synthèse de la parole), il nous est désormais possible d'aboutir à de telles définitions. C'est pour cela que va être créé un nouveau genre de fichiers contenant des informations relatives à la prosodie : Fréquence fondamentale, segmentation temporelle, groupes prosodiques, syntaxiques... A ces données symboliques pourront être ajoutées n'importe quelles informations permettant de mieux décrire l'intonation. Des informations de nature à qualifier les émotions par exemple. Ceci est possible grâce au format SDIF qui permet de déclarer de nouvelles matrices dans un chunk initial. Les compositeurs désireux d'exploiter de tels fichiers « prosodiques », les chargeront par le biais de OPENMUSIC, AUDIOSCULPT ou DIPHONE⁶⁴ afin de les visualiser, de les analyser et de les transformer. Cependant, même si ce type de fichier permet aux musiciens d'utiliser ces données relatives à l'intonation et à la manière de s'exprimer et ainsi de mettre un pied dans le monde de la linguistique, il semble mal adapté pour la traduction d'événements musicaux.

« Les données expérimentales sur la production de la parole et la performance musicale montrent des similarités. Dans les deux cas, le système de production utilise un accès local à la structure prosodique qui est hiérarchique, mais pas récursive. Une performance musicale comme la parole est tout de même restreinte par l'aspect phonologique. En s'affranchissant des limites formelles de la phonologie, il peut être possible d'étudier la conjecture entre la parole et la musique. »⁶⁵

Devant ces disparités d'échelles temporelles et fréquentielles entre la parole et la musique, il semble que l'établissement d'un système d'écriture commun est vain. Andrezej Rakowski ajoute qu'il peut exister des systèmes phonologiques dans la musique. Mais ils s'appliquent respectivement aux domaines de la hauteur, du rythme et du timbre. C'est seulement dans cet esprit séparatiste qu'il peut s'établir des analogies entre musique et

⁶³ G. FANT, A. KRUCKENBERG, L. NORD, 1990, Stress patterns and rhythm in the reading of prose and poetry with analogies to music performance

⁶⁴ Logiciels de l'IRCAM

⁶⁵ J.B. PIERREHUMBERT, 1990, Music and the phonological principle: Remarks from the phoneticians's bench

parole. L'écriture doit décrire fidèlement les données acoustiques de ces deux signaux. Or la notation musicale n'est pas restreinte aux attaques des notes et à leurs hauteurs. Dans la Sequenza III, Luciano Berio décrit les états émotionnels dans lesquels l'interprète doit s'incarner. Il s'agit alors pour ce dernier, du même travail qu'un acteur déclamant des phrases avec tel ou tel sentiment dans la voix. Comme nous l'avons vu, le registre émotionnel peut s'exprimer autant dans la parole que dans la musique. C'est donc certainement dans l'écriture des émotions que pourrait se rejoindre les auteurs de tous horizons (compositeurs, écrivains...).

Cela dit, d'autres similitudes structurelles existent outre l'aspect directement émotionnel. Slawson⁶⁶ décrit « une analogie entre parole et musique provenant de remarques sur les grammaires des compositions tonales des XVIII et XIXème siècle. Mais il va plus loin en constatant que de nombreuses compositions du XXème siècle, aussi complexes soient-elles, possèdent aussi des caractéristiques systématiques comme les hiérarchies de règles... Un compositeur choisit généralement un système (dodécaphonisme...) dans lequel il peut s'exprimer ». Il pose au départ une contrainte lui permettant d'évoluer et de décrire un espace des possibles. Cette contrainte est systématique et garante de l'intégrité d'une œuvre. Le compositeur la crée le plus souvent avec des règles d'écriture ou bien avec un patch MAX... Puis il va s'exprimer en respectant ou non ce système initial. Il est important de remarquer que ces déviations sont très proches des distorsions prosodiques. En effet, on peut comparer ce système délimitant au cadre normatif de la grammaire d'une langue. Un compositeur crée ainsi son propre langage musical et l'expose aux autres soit par une description explicite et complémentaire, soit par une première phase énonciative. Puis il exprime son rapport personnel à ce système compris de tous par une utilisation plus ou moins détournée. Les déviations qu'il opère sont alors révélatrices de sa personnalité.

Si les liens entre ces deux structurations de langage ne sont pas directes. Cela est dû à l'incroyable diversité des langages musicaux qu'ont créés les compositeurs. Difficile d'y voir clair lorsqu'on est habitué à un système rigide et commun à tous comme peut l'être une langue... Il est d'ailleurs difficile la plupart du temps d'apprécier les déviations d'un compositeur par rapport au système qu'il a érigé tout en comprenant celui-ci simultanément. Cependant, pour exprimer des émotions, il possède la liberté d'élaborer cette structure librement. Elle peut aussi se révéler le véhicule d'une expressivité. Ce qui décuple le pouvoir expressif du langage musical par rapport à la parole et explique l'hétérogénéité du

monde musical dans son ensemble.

Si le compositeur possède la liberté de créer un langage. Il n'en est pas de même pour l'interprète qui doit se tenir à l'exécution du message. Cependant, il a la possibilité, lui aussi, d'exprimer sa personnalité émotionnelle au travers de sa réalisation. C'est pour cela que nous préférons des versions d'une même pièce à d'autres ; Ou bien même simplement des interprètes à d'autres. On parle alors de leurs sensibilités. Ce mot reflète peut-être leurs capacités à mettre en emphase l'émotion voulue par un compositeur. Mais il peut aussi cacher une autre forme de communication sonore relative au jeu instrumental...

V.3 Prosodie musicale :

« La musique n'est pas relative aux sons tels qu'ils sont réalisés
mais tels qu'ils sont intentionnés »⁶⁷

« Dans le passé, les recherches en analyse musicale se sont principalement préoccupées des aspects structurels et compositionnels. Une question également fascinante reste cependant à explorer : Comment un musicien peut créer une ambiance émotionnelle appropriée. A présent, nous travaillons sur un système de règle qui introduit automatiquement des effets d'articulation. Les premiers résultats semblent pénétrer les principes de groupement bas-niveau de la performance musicale. L'analyse des performances musicales a déjà contribué au lexique utilisé par les musiciens pour véhiculer des informations non écrites sur la partition (extrascorales). Il semble que cette voix de recherche met en valeur des similarités frappantes entre la musique et la parole. Cela n'est pas surprenant puisque ces deux systèmes de communication humaine par les sons doivent refléter les mêmes processus cognitifs et perceptuels. »⁶⁸

La première remarque concernant cette déclaration d'un éminent chercheur suédois est relative aux informations dites extrascorales. Celles-ci sont à mettre en parallèle avec les informations prosodiques et émotionnelles que nous citons auparavant. Il semble qu'il se dégage de cette similarité, un sujet de recherche passionnant visant à étudier le rapprochement entre ces deux types de distorsions (de la prosodie et du jeu instrumental).

⁶⁶ W. SLAWSON, 1990, Structure and association in descriptions of music, with a coda on linguistic analogies

⁶⁷ R. JAKOBSON, Musicologie et linguistique, musique en jeu N°5

⁶⁸ J. SUNDBERG, 1990, Music performance research: An overview

Une telle recherche conduirait très vraisemblablement à l'établissement d'un vocabulaire émotionnel ou du moins expressif. Cependant, rappelons nous que dès lors que ce genre de lexique sera établi, il deviendra probablement obsolète compte tenu de ce qu'il arrive aux clichés mélodiques dès lors qu'ils sont reconnus. Toutefois, selon Ivan Fonagy, nos émotions et leurs manières d'être exprimées remontent à la préhistoire. Il n'est pas de raison alors qu'elles changent de si tôt.

La seconde remarque et sans doute la plus intéressante car moins prospective, est que cette déclaration signifie qu'il existe un espace de communication lié à la performance musicale que l'on pourrait appeler : **Prosodie musicale**. Le parallèle avec la parole n'étant pas fortuit, puisqu'il s'agit là aussi de communiquer des émotions de manière sonore. Cette distinction permet aussi de rapprocher les aspects formels de la musique et du langage, le texte et la partition.

« Je ne crois pas que la frontière se trouve véritablement entre structure et expression musicale dans le cas de la musique. En écrivant la structure, on peut exprimer des émotions. Un rapide changement d'amplitude dans une séquence de chant synthétisée donne une couleur émotionnelle agressive. Alors qu'une progression lente donne un caractère plus paisible et amical. Il y a un mix entre les aspects structurels et émotionnels. Ils se croisent partiellement, mais pas entièrement. »⁶⁹

J'ose à dire que c'est la même chose lorsqu'on lit un texte triste par exemple. Le champ sémantique de la tristesse crée chez l'auditeur un rappel émotionnel qui le met en situation. Ceci est « appuyé » par la façon dont est lu le texte. Là aussi, le texte et l'expression sonore s'allient pour communiquer un sentiment, une émotion. De la même manière, l'interprète exprimera par son jeu, la pensée écrite du compositeur, de manière à amplifier le « transfert d'émotion ». A. Gabrielsson⁷⁰ ajoute que les « déviations expressives » de la performance musicale sont effectuées dans le but de clarifier la « structure émotionnelle ».

Ce qui importe dans toutes ces affirmations est l'émergence d'un nouveau champ de recherche dans l'analyse musicale. Il est important de considérer qu'il est un bon moyen de faire le lien entre la parole et la musique. D'ailleurs, lorsqu'on interroge J. Sundberg⁷¹ à ce sujet : « Quelles sont les fonctions de la musique ? » ; Il répond : « L'une d'entre elles est de nous entraîner à interpréter les aspects expressifs de la parole. »

⁶⁹ J. SUNDBERG, 1990, Music performance research: An overview

⁷⁰ WENNER-GREN INTERNATIONAL SYMPOSIUM SERIES, 1990, Music, langage, speech and brain

⁷¹ J. SUNDBERG « Music is a language of emotions. Speech is also a language of emotions. »

Pour le moment, l'Université KTH (Suède) s'est investie dans cette voie et à déblayer le terrain. L'IRCAM s'intéresse aussi à ce sujet, toujours dans le cadre musical. Des résultats sont déjà accessibles concernant la modélisation du jeu d'un interprète particulier par l'analyse. Cette analyse est issue de la comparaison entre la partition MIDI et des données de l'alignement avec l'audio. Autrement dit, le jeu d'un interprète est modélisé grâce à ses déviations par rapport à une partition qui se veut neutre et référentiel. Lors de ces analyses⁷², Le laboratoire TMH de l'Université KTH s'est rendu compte de l'importance que pouvait jouer les fins de notes et les variations locales de tempo dans l'expressivité du jeu d'un interprète. Les articulations interviennent entre deux phrasés selon la définition du musicologue allemand Hermann Keller et : « elles sont pour la mélodie, le meilleur moyen d'exprimer des émotions. »⁷³

Bien sûr, d'autres informations extracorales participent à la prosodie musicale comme les variations de timbre par exemple.

Dans le cas de la parole, rappelons que la difficulté est plus grande car nous ne possédons pas de courbe neutre pouvant servir à la comparaison. Si l'aspect structurel de la musique semble être plus riche que celui de la parole, il n'en est pas de même pour la prosodie musicale qui possède moins de degrés de liberté que la prosodie vocale.

Cependant, il faut mettre un bémol à ce genre de déclaration puisque les recherches visant à comparer ces deux types de prosodie sont encore à leurs phases initiales. Tout est à faire dans un domaine qu'il est difficile d'aborder car il requiert aussi bien un savoir musical que linguistique. De manière à pouvoir aboutir à des résultats fiables, il faut établir une réelle collaboration entre ces deux sciences. Cela pourrait être un sujet de recherche conduit par deux étudiants en thèse par exemple appartenant respectivement à ces deux chairs. Des travaux complémentaires et simultanés ne seraient alors possible que par le biais d'un support technique commun comme par exemple, le format de fichiers prosodiques que nous mentionnions dans la partie précédente.

⁷² J. JEKERT, 2003, Measurements and models of musical articulation

⁷³ H. Keller, 1965, Phrasing and articulation

● Conclusion

Ce mémoire a plusieurs vocations. La première est celle que j'ai introduite dès le début : Je n'ai en aucune façon tenté de tirer des conclusions, mais seulement de nuancer des frontières qui semblent encore aujourd'hui bien solides et bien définies. La musique est-elle un langage ? La parole est-elle musique ? Qu'importe les réponses à des questions qui n'ont finalement pas beaucoup de sens. Ce qui importe, c'est d'éviter toute catégorisation hâtive qui piégerait l'auditeur dans une écoute réductrice de ces deux phénomènes. La parole permet d'échanger des messages sémantiques dont on oublie trop la part musicale et esthétique, pourtant fondatrice de nos personnalités.

De manière à mettre en emphase ce propos, les compositeurs du XX^{ème} siècle ont créé de nombreuses pièces poussant au paroxysme la musicalité de la voix parlée en libérant souvent celle-ci de son rôle sémantique. Ils l'ont fait revenir dans la musique classique occidentale. Cette dernière possède d'ailleurs de nombreux « gênes » de la voix parlée puisque c'est elle qui l'a certainement engendrée.

On pourrait même aller plus loin dans les conséquences d'un tel « accouchement » en citant Askenfelt⁷⁴. Il mentionne un parallèle intéressant entre les dimensions timbrales de la voix chantée et celles des instruments à cordes frottées. Ce genre de parallèle pose une question fondamentale : Peut-être que les propriétés spectrales de la voix servent de références à notre perception du timbre ? Et par extension, est-ce que le rythme proviendrait du débit de parole, que la mélodie naîtrait de l'évolution de notre fréquence fondamentale ? A l'évidence, il ne faut pas tenter de répondre à une question qui ressemble fort à : « Qui de l'œuf ou la poule est naît le premier ? » Il est préférable, pour rendre compte des similarités entre la parole et la musique de constater que leurs évolutions sont très différentes, mais quelles sont toutes relatives à un domaine commun, dit de communication préverbale. « Une grande partie du langage musical dialogue avec le langage naturel tel quel: celui de l'homme. »⁷⁵ Je pense qu'il est très important dans les recherches futures d'analyser ce « canal de transmission » qui pour certains est déjà associé au « transfert d'émotions ». Que cette déclaration soit juste ou erronée, il n'en reste pas moins que la question subsiste et qu'elle est fondatrice de nouveaux champs d'investigations.

La deuxième vocation de ce mémoire est de vous exposer le travail que j'ai

⁷⁴ J. SUNDBERG: « Music is a language of emotions. Speech is also a language of emotions. »

effectué cette année. La partie sur la synthèse de la parole vous permet d'apprécier un travail dont le but est de synthétiser la prosodie d'un locuteur spécifique. Ce travail nous a amené à d'autres exploitations des données prosodiques que nous avons analysées. J'ai tenté de les utiliser à des fins purement musicales dans DisCutIon : Expérience sonore visant non seulement à associer la parole à un matériau compositionnel, mais à dénuder la parole de son rôle sémantique pour apprécier le rôle narratif et informationnel de la prosodie. J'ai aussi essayé durant cette année de transformer cette dernière en temps réel. Difficulté qui a m'a offert pour principale récompense, la promesse d'un défi passionnant dans les années à venir. Enfin, l'attrait du matériau prosodique pour les compositeurs m'a poussé à définir un format de fichiers, leurs permettant d'user des données relatives à l'intonation pour composer.

La troisième vocation de ce mémoire est d'ouvrir des pistes pour d'éventuels travaux futurs. Nous avons défini à présent ce format de fichiers prosodiques. La formalisation et la réalisation restent à être développées, ceci dans les prochains mois à venir grâce à l'aide de l'équipe analyse/synthèse de l'IRCAM. Non seulement, les compositeurs pourront les utiliser (Je pense notamment à Jonathan Harvey qui désire faire parler un orchestre l'été 2007), mais d'autres chercheurs pourront s'en servir comme des linguistes par exemple : En effet, ce format est un bon outil de description de corpus en vue d'études statistiques. Associé à un outil d'extraction automatique de ses données, il peut se révéler précieux pour quiconque désire étudier la prosodie.

Les résultats de ces diverses analyses me semblent de bons moyens de pénétrer le rôle de la communication préverbale. Mais seulement si on prend aussi en compte son existence au sein de la musique. C'est pourquoi je préconise une étude complémentaire visant à mettre en évidence des paradigmes communs à ces deux langages. Je pense que si l'on étudie, de front, le rôle des articulations dans la musique et dans la parole ; La manière dont la prosodie verbale et la prosodie musicale se construisent par distorsions d'un patron neutre fixé au départ (par la comparaison automatique de phrases mélodiques à des phrases intonatives par exemple); Et enfin, les similarités sonores (pauses, variations de débit, de timbres, de registres) qui existent entre ces deux modes d'expression ; Alors on pourra mieux connaître et tracer les frontières de ce domaine commun à la voix parlée et à l'interprétation musicale que certains comparent déjà au berceau de nos émotions.

⁷⁵ D. COHEN-LEVINAS, 1987, La voix au-delà du chant

● Bibliographie

- Dusterhoff, K. and Black, A., 1997, Generating F0 contours for speech synthesis using the Tilt intonation theory
 - Black, A., 1997, Predicting the intonation of discourse segments from examples in dialogue speech
 - Black, A. and Hunt, A., 1996, Generating FO contours from ToBI labels using linear regression
 - Black, A. and Campbell, N., 1995, Predicting the intonation of discourse segments from examples in dialogue speech, (Short version)
 - Black, A., 1995, Comparison of algorithms for predicting accent placement in English speech synthesis
 - Hunt, A. and Black, A., 1996, Unit selection in a concatenative speech synthesis system using a large speech database
 - **Thèse de Diemo Schwarz, 2004, Data-driven concatenative sound synthesis. IRCAM**
 - Thèse de Romain prudon, 2003, Synthèse de la parole multilocuteur par sélection d'unités acoustiques. LIMSI
 - Prudon R., d'Alessandro C., 2001, A selection/concatenation TTS synthesis system : Databases development, system design, comparative evaluation.
 - Prudon R., d'Alessandro C., et Boula de Mareüil P, 2002, Prosody synthesis by unit selection and transplantation on diphones
 - Bozkurt B., Dutoit T., Prudon R., d'Alessandro C., et Pagel V, 2002, Improving quality of MBROLA synthesis for non-uniform units synthesis
 - **Malfrère, Dutoit et Mertens, 1998, Automatic prosody generation using suprasegmental Unit Selection**
 - Thèse de Christophe Blouin
 - **Anne Lacheret-Dujour et Frédéric Beaugendre.”la prosodie du Français”. CNRS**
- Langage**
- Jacqueline Vaissière, 1997, Langues, prosodie, syntaxe (www.cavi.univ-paris3.fr/ilpga/ed/dr/jvdr2/articlesJV/vaissiereatala1997.pdf)
 - Y. MENUHIN, 1979, Menuhin and Davis
 - J.M. ALBY, C. ALES, P. SANSOY, 1988, L'esprit des voix

- J. ABITBOL, 1983, Voyage au centre de la voix
- CHARVET, 2004, La voix et ses métamorphoses dans « les métamorphoses » d'Ovide
- Q.H. TRAN, 1980, Les musiques vocales
- **FONAGY, 1983, la vive voix, essais de psycho-phonétique**
- DI CRISTO, 1982, Prolégomènes à l'étude de l'intonation – Micromélogie
- PLATON, La république
- M. POIZAT, 1986, Frénésie, entre parole et cri : le chant de la Diva
- R. MANCINI, 1991, la voix dans la culture occidentale
- M.BEAUFILS, 1954, Musique du son, Musique du verbe
- **Madelaine GAGNARD, 1987, La voix dans la musique contemporaine et extra-européenne**
- **D. COHEN-LEVINAS, 1987, La voix au-delà du chant**
- POTOCKI, 1968; L'inouï et l'indicible
- SCHÖNBERG, 1912, préface de Pierrot Lunaire¹
- MALEC, 1985, Ivo Malec à propos de Vox, Vocis, F
- COHEN-LEVINAS, 1987, La voix au-delà du chant
- J-C. RISSET, 1990, Speech and music combined: An overview
- F. LERDHAL , J. HALLE, 1990, Somme lines of poetry viewed as music
- NARMOUR, 1990, The melodic structures of music and speech: Applications and dimensions of the implication-realization model
- **R. CROSS, 1991, La voix dévoilée**
- LHERMITTE, 1991, Le cerveau
- http://www.aphasie.qc.ca/divers/la_musique.htm
- FONAGY et MAGDICS, 1963, Emotional patterns
- MEYER, 1956, Emotion
- SARTRE, 1960, Esquisse
- FONAGY, 1960, le ton et la musique
- C. MALHERBE, 1997, Locus
- FREUD, der Witz
- R. CARLSON, B.GRANSTROM, 1990, Performance rules in a texte-to-speech system
- K. MAGDICS, 1963, Melody of speech
- LIST, 1963, Speech and Song
- **P. SCHAEFFER, Traité des objets musicaux**
- FAURE, 1962, Aspects et fonctions linguistiques des variations mélodiques dans la

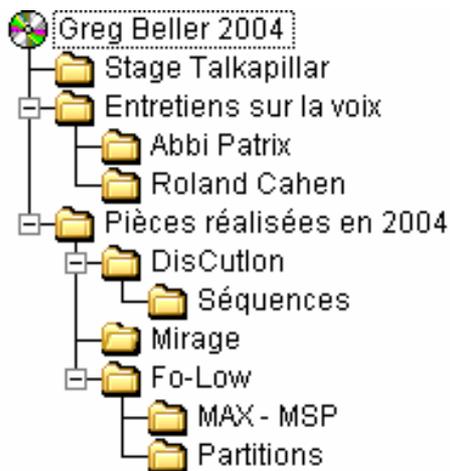
chaîne parlée

- FAURE, 1962, Recherches sur les caractères et le rôle des éléments musicaux dans la prononciation anglaise
- La parole du monde, mythe de création Dogons, Mercure de France
- HAYES, A. LAHIRI, 1990, Durationally specified intonation in English and Bengali
- K. SHERER, 1990, Emotion expression in speech music
- HEGEL, Esthétique
- DUCARD, 1980, Le chant perdu de la langue : voix et écriture poétique
- DUCARD, 1980, Le chant perdu de la langue : voix et écriture poétique
- KATZ, FELDOR, 1963, The structure of semantic theory
- W. SLAWSON, 1990, Structure and association in descriptions of music, with a coda on linguistic analogies
- A. MOLES, 1958, Théorie de l'information
- J. SUNDBERG: « Music is a language of emotions. Speech is also a language of emotions. ».
- FANT, A. KRUCKENBERG, L. NORD, 1990, Stress patterns and rhythm in the reading of prose and poetry with analogies to music performance
- J.B. PIERREHUMBERT, 1990, Music and the phonological principle: Remarks from the phoneticians's bench
- R. JAKOBSON, Musicologie et linguistique, musique en jeu N°5
- SUNDBERG, 1990, Music performance research: An overview
- **WENNER-GREN INTERNATIONAL SYMPOSIUM SERIES, 1990, Music, language, speech and brain**
- JEKERT, 2003, Measurements and models of musical articulation
- H. Keller, 1965, Phrasing and articulation

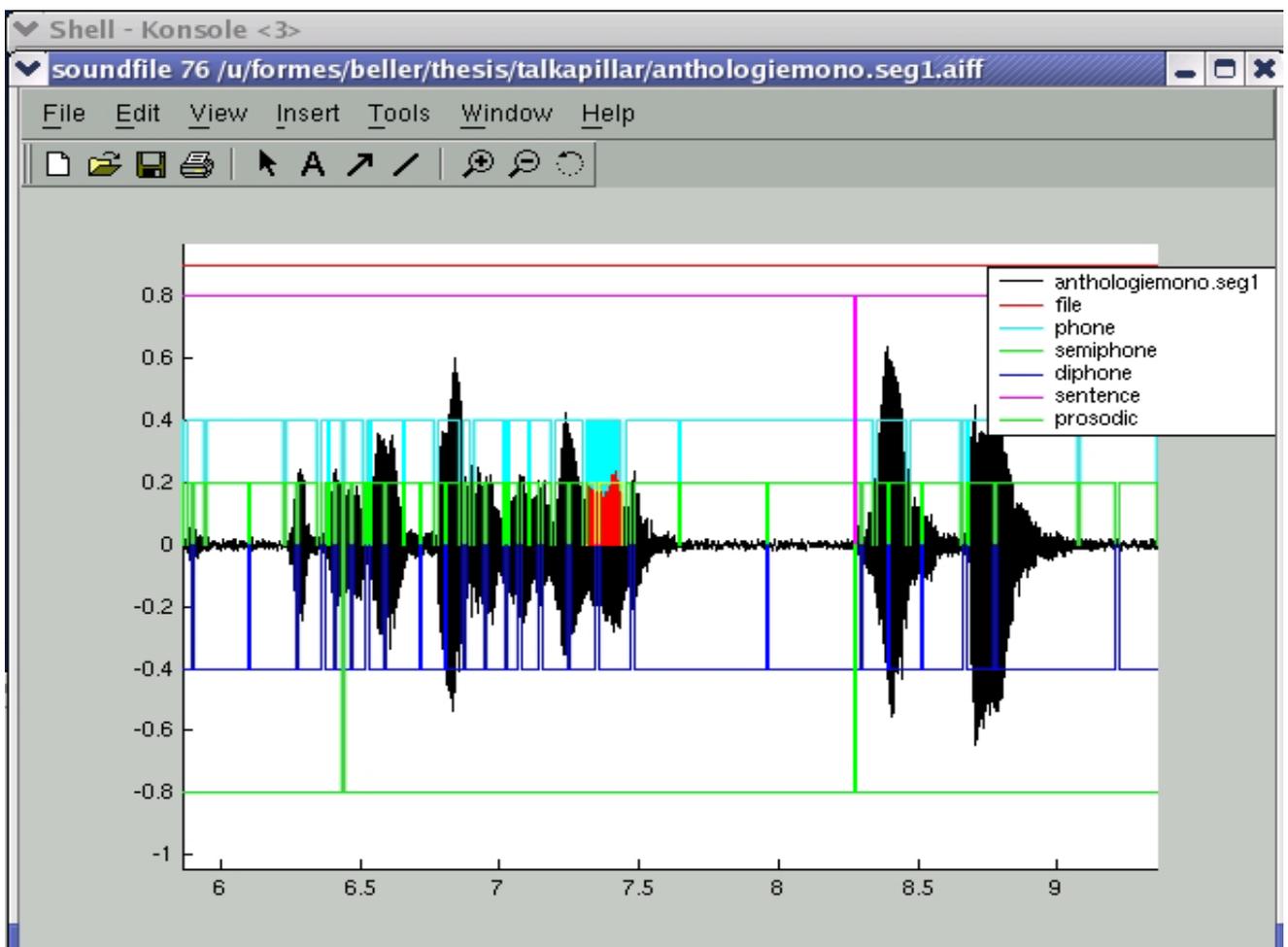
● Autre références :

- Entretien avec Roland Cahen (IRCAM), homme de théâtre et pédagogue.
- Entretien avec Abbi Patrix (Maison Du Conte), conteur.
- Entretien avec Claudy Malherbe, compositeur.
- Entretien avec Giordano Ferrari, spécialiste de l'Opéra et de la dramaturgie.
- seuils de Marc André Dalbavie
- Sara Dolce Tacere de Luigi Nono
- Sequenza III de Bérió
- Femme de Florence Baschet
- Filippo Filidei
- Come natura di foglia de Marco Stroppa
- Aventures et nouvelles aventures de G. Ligeti
- De la nature de l'eau de Aperghis
- Études sur l'E muet, timbre, durée, Intensité, hauteur (Fréquence Fondamentale) de J. Varnet Pleasants
- Prescurien N°2 de Ferrari : expérience musicale dans laquelle les sons et la parole font musique. Quelqu'un commente un univers en décrivant ces mouvements, ces visions... S'en suit la création d'une atmosphère agrémentée de sons.
- Pierre Boeswillwald (théâtre) met en scène une théatralisation du sonore ou des injonctions verbales se marient à des sons. (Tati).
- Bernard Heidsiek, Cisif : Voix enregistrée en décalage... Avant la source...
- Les Tocattanes du Groupe expérimental de Bourges.
- John Cage, Finegens Wake, Lis : Il réorganise l'ordre des phonèmes selon le Yi-kin. Rythme de la récitation.
- Artaud, Pour en finir avec le jugement de Dieu : Il fait sortir l'énergie profonde par le cri primal.
- François Bernard Mâche imite la prosodie des oiseaux fidèlement.
- La Dictée, Sonate Baroque : imite la machine à écrire.

- **Arborescence du CD-ROM joint :**



- **Exemple d'une segmentation temporelle de la parole :**



- **Partitions de Fo-Low :**