

UNIVERSITÉ PARIS 8 – VINCENNES- SAINT-DENIS

UFR ARTS, PHILOSOPHIE, ESTHÉTIQUE

Département de Musique

Master Arts, mention Musique

Spécialité : Musicologie, création et société

**LE TRAITEMENT EN TEMPS RÉEL DE LA VOIX CHANTÉE :
PERSPECTIVES NOUVELLES POUR LA RELATION MUSIQUE ET TEXTE**

Mémoire de Master 1, présenté et soutenu par

Hadas PE'ERY

Le 8 octobre 2010

Réalisé sous la direction de Professeur Anne SEDES

Année universitaire 2009/2010

TABLE DES MATIÈRES

INTRODUCTION	2
I. LA VOIX CHANTEE ET LA MUSIQUE MIXTE	4
I.1 La perception de la voix	4
I.1.1 Éléments acoustiques et phonétiques	4
I.1.1.1 Sources	
I.1.1.2 Résonateurs et filtres	
I.1.1.3 Phonation	
I.1.2 Éléments psychologique	9
I.1.3 Conclusions	10
I.2 Le traitement du texte	11
I.2.1 Mise en contexte	11
I.2.2 <i>Strette</i> – Hector Parra, Paul Celan	13
I.2.3 <i>Poetry for // dark-/ dolls</i> – Marco Suarez, Christophe Tarkos	16
I.2.4 <i>Thema (Omaggio a Joyce)</i> – Luciano Berio, James Joyce	18
II. LE TRAITEMENT ELECTRONIQUE	23
II.1 La séparation des sons	23
II.1.1 <i>Yin~</i>	23
II.1.2 <i>Zerox~</i>	24
II.1.3 Cautions d'emploi	27
II.2 Le traitement des consonnes – la granulation	28
II.2.1 <i>Rogs~</i> et <i>sogs~</i>	28
II.2.2 <i>Munger~</i>	31
II.3 Le traitement des voyelles	33
II.3.1 <i>Psychoirtrist~</i>	33
II.4 Conclusions	35
III. REALISATION D' <i>AUTO-DETERMINATION</i> POUR VOIX DE FEMME ET DISPOSITIF ELECTRONIQUE EN TEMPS REEL	36
III.1 Le texte	36
III.1.1 L'interprétation du texte	37
III.1.2 Conception de la pièce <i>Auto-détermination</i>	39
III.2 L'écriture vocale	40
III.2.1 Général	40
III.2.2 La construction des gestes	41
III.3 Le traitement électronique dans <i>Auto-détermination</i>	44
III.3.1 Le traitement des consonnes	45
III.3.1.1 <i>Munger~</i>	45
III.3.1.2 La spatialisation	47
III.3.2 Le traitement des voyelles	47
III.4 Conclusions	48
CONCLUSIONS	50
BIBLIOGRAPHIE	52
ANNEXES	

INTRODUCTION

En écrivant une œuvre vocale, le compositeur est confronté aux problématiques qui ne se présentent généralement pas le travail instrumental. Parmi celles ci, la tradition vocale et l'emploi du texte méritent une attention particulière.

L'enseignement, l'exécution et, par conséquence, la composition vocale, se concentrent autour de certaines traditions : celle du style lyrique de l'opéra romantique et, dans une moindre mesure, celle du style expressionniste de la deuxième école de Vienne. L'omniprésence des paradigmes créés par ces styles délimite considérablement le travail d'exploration habituellement entrepris par un compositeur. Ainsi, en composant une pièce pour la voix, le compositeur doit se libérer non seulement de ses propres habitudes, mais aussi de ceux imposés par le genre.

L'emploi du texte est à la fois un grand avantage et un grand défaut du médium vocal. En utilisant un texte, le compositeur peut réunir deux moyens d'expressions, et ainsi atteindre un niveau de profondeur qui est rarement présent dans la musique instrumentale seule. Cependant, la présence des paroles entraîne souvent l'établissement d'une hiérarchie marquée, qui privilégie soit le texte, soit la musique. Historiquement, jusqu'au XX^{ème} siècle le texte dicta la structure de l'écriture vocale ; dans le XX^{ème} siècle, surtout à partir la deuxième guerre mondiale, les compositeurs commencent à ignorer ou à éviter l'emploi d'un texte afin de libérer l'écriture vocale. Ni l'un ni l'autre de ces extrêmes atteint la complémentarité de musiques et de textes proposés par le médium vocal.

Le deuxième de ces enjeux paraît particulièrement intéressant, car l'emploi du texte et aussi porteur de la tradition. Ainsi, dans ce travail, nous proposons une approche de la composition vocale qui tente, à travers l'emploi de l'informatique, à établir des nouvelles perspectives sur ce rapport entre voix et paroles. Le panel indéfini de possibilités offert aux compositeurs par l'emploi de moyens électroniques en temps réel, peut permettre de composer avec le médium vocal en tant que tel, sans assumer les conventions de la tradition vocale.

Ce mémoire est organisé en trois parties distinctes, mais non indépendantes, culminantes à la réalisation d'une création musicale. A l'issue de chacune des premières

deux parties nous présentons les conclusions tirées des recherches menées jusqu'ici et leur contribution aux recherches qui suivent.

La première section expose des recherches menées sur l'appareil vocal et sur l'évolution et l'application de la corrélation entre texte et musique. En discutant les qualités acoustiques, phonétiques et psychologiques de l'appareil vocal, nous souhaitons déterminer les attributs de la voix à partir desquels nous pourrions établir des relations entre le texte, l'écriture vocale, et l'électronique. En analysant des œuvres pour voix et électronique, nous souhaitons comprendre l'influence que l'électronique sur le développement du rapport texte-musique.

La deuxième section expose une gamme d'outils informatiques, sélectionnés selon les conclusions des recherches menées dans la première section. Ici nous discutons les méthodes de séparation et les traitements possibles des sons harmoniques et inharmoniques. En expliquant le fonctionnement, les avantages et les désavantages de chaque outil, nous espérons donner une notion du caractère de l'écriture vocale qui peut être traité avec les dits outils.

La dernière section précise les étapes de notre création artistique. Cette partie du mémoire permet de comprendre comment les recherches exposées dans les premières deux sections ont été appliquées afin de composer la pièce *Auto-détermination* pour voix de femme et dispositif électronique en temps réel.

I. LA VOIX CHANTÉE ET LA MUSIQUE MIXTE

I.1. LA PERCEPTION DE LA VOIX

Nous croyons important de commencer ce travail par une discussion de divers attributs de la voix qui peuvent servir ou desservir à la composition d'une musique mixte. La prise en connaissance des dits attributs peut révéler de nouvelles possibilités ou faciliter le travail du compositeur, ainsi que l'aider à éviter des démarches potentiellement infructueuses.

En premier lieu nous aborderons quelques propriétés physiques et acoustiques de base. En suite nous discuterons des caractéristiques psychologiques jouant sur la perception de la voix chez l'auditeur.

I.1.1 ÉLÉMENTS ACOUSTIQUES ET PHONETIQUES

Dans les limites de notre travail, nous considérons la voix comme une ou plusieurs sources sonore soumis à un système de filtres et résonateurs.

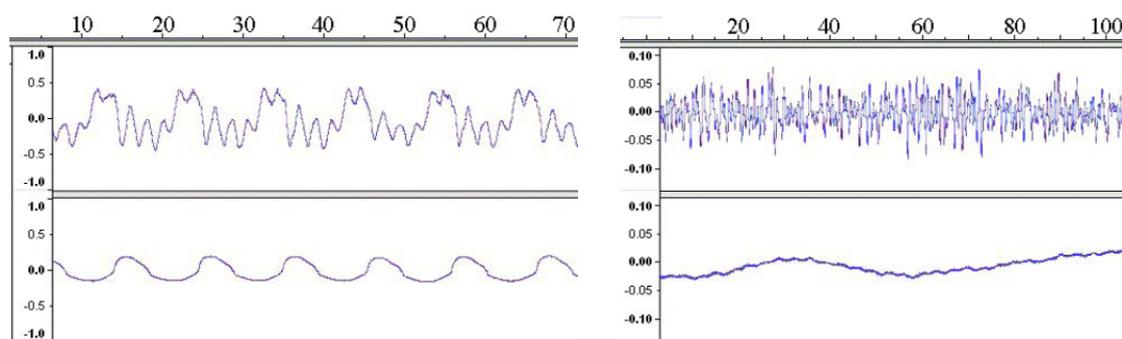
I.1.1.1 SOURCES

La source est le mode de production du son. Nous définissons deux types de sources : 1) la mise en vibration des cordes vocales par l'air expulsé de poumons, et 2) le passage irrégulier de l'air à travers les cordes vocales comprimées.

La première de ces sources génère un son *harmonique*, dans lequel les fréquences des partiels sont des multiples entiers du son fondamental. La deuxième génère un son *inharmonique*, dans lequel les fréquences des partiels ne sont pas des multiples entiers du son fondamental.

Dans le domaine de traitement du signal, nous désignons ces sons : *stable* et *transitoire* respectivement. Leur classification est déterminée par le taux de changement

de signe du signal¹ ou *zéro crossing rate (ZCR)* - faible pour ces premiers, qui sont plutôt périodique, et ample pour les derniers où l'oscillation est irrégulière (Fig. 1). Dans le domaine vocal nous désignons les catégories : *voisé* et *non-voisé* (voir la section « phonation » ci-dessous). Nous percevons cette distinction facilement dans la voix chantée. Étant éphémères, peu résonants et de hauteur indéfinie, les sons inharmoniques ne peuvent qu'être employés dans une manière restreinte. En revanche les sons harmoniques - d'une souplesse de durée, niveau sonore et tessiture, et d'une richesse



spectrale – constituent 90% du répertoire vocale².

I.1.1.2 RESONATEURS ET FILTRES

Par la suite, le son est modifié par la cavité du pharynx qui fonctionne comme un résonateur. Ainsi, certaines bandes de fréquences sont amplifiées, d'autres atténuées. En sortant du pharynx, le son est sculpté par le palais, la langue et les dents, qui fonctionnent

Figure 1: Gauche : l'oscillograme et l'électroglottogramme (EGG) des voyelles du mot anglais « heard », voisés (stable). Droit : l'oscillograme et l'EGG du même son non-voisé (turbulente)³.

comme des filtres.

Ce phénomène d'amplification et de filtrage crée plusieurs pics de résonance dans le spectre vocal : ceux sont les *formants*. La capacité de comprendre la parole ou le chant est largement dûe aux formants, qui varient en fonction du phonème et de la manière dont il est prononcé (Fig. 2). Le placement et la transition⁴ des deux premiers

¹ BACHU, R.G. « Separation of Voiced and Unvoiced Using Zero Crossing Rate and Energy of the Speech Signal », p. 2-3.

² WOLFFE, Joe. Voice Acoustics, an Introduction.

³ SWERDLIN, Yoni. Glottal Signal, Vocal Tract Resonances and Output Sound.

⁴ L'évolution de changements dans les fréquences des formants de voyelles

formants, par exemple, déterminent le timbre d'une voyelle dans les langues non-toniques, et donc permettent de distinguer une voyelle d'une autre. Egalement, le filtrage effectué par la langue et les dents permet de reconnaître les voyelles et notamment les consonnes.

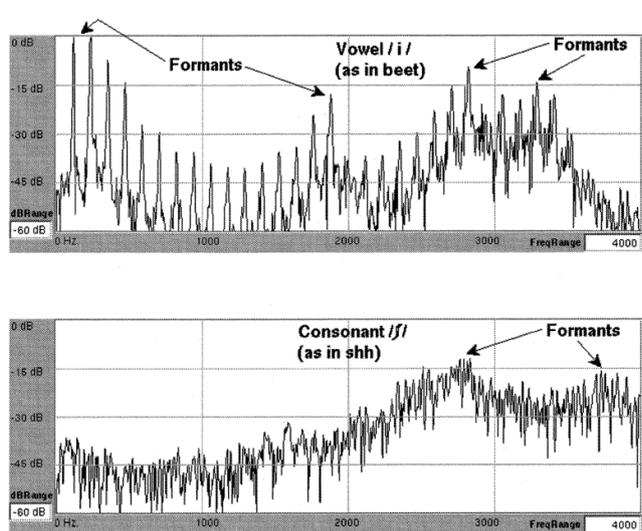


Figure 2 : Haut : Spectre de la voyelle [i], montrant les formants d'un signal harmonique. Bas : Spectre de la consonne [ʃ], montrant les formants d'un signal inharmonique⁵.

Nous avons ainsi une notion simplifiée du modèle source-filtre⁶ (Fig. 3). Bien que ce modèle ne soit qu'une approximation du vrai fonctionnement de la production vocale, nous pouvons, en s'appuyant sur lui, supposer les qualités acoustiques et phonétiques d'un son. En observant la source, par exemple, nous pouvons déterminer si un son est voisé ou non-voisé, et donc si son signal est régulier ou turbulent. En observant les filtres et les résonateurs, nous pouvons raffiner notre prédiction afin de déterminer le phonème qui sera produit.

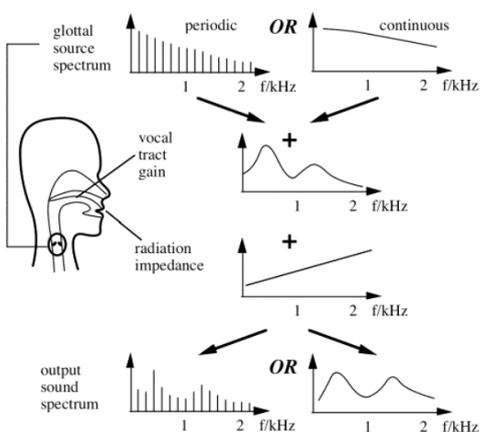


Figure 3 : le fonctionnement théorique du modèle source-filtre.

: : History Current Work, and Future Directions ». p. 39.
 unction.

I.1.1.3 PHONATION

Jusqu'à présent, nous avons traité la voix comme une abstraction acoustique. Or, nous discuterons le rapport entre la dite conception acoustiques et la langue parlée.

Nous commençons avec un simple tri des voyelles et des consonnes. Les voyelles, générés par la vibration régulière des cordes vocales, sont harmoniques, et donc voisées (Fig. 4). Les consonnes, issues de mélanges différents de vibrations régulières et turbulentes des cordes vocales, nécessitent une classification plus précise. Nous diviserons donc les consonnes en fonction de leur mode d'articulation : les *occlusives*, produites par le blocage complet et puis la relâche de l'écoulement d'air ; les *fricatives*, produites par une constriction qui provoque des turbulences dans l'écoulement d'air sans le bloquer complètement ; les *nasales*, produites par l'abaissement du voile de palais, qui bloque l'air dans la bouche sans empêcher son écoulement par la cavité nasale. Bien que ces trois types de consonnes ne constituent qu'une partie délimitée des sons employés dans les langues occidentales, ils nous suffiraient pour l'étude de textes abordés dans notre travail. ⁷

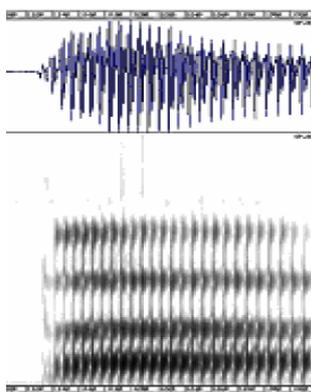


Figure 4 : Spectre de la voyelle [a]⁷

Les consonnes occlusives ([b]/[p], [d]/[t], [g]/[k] en français), caractérisées par le silence crée lors de l'occlusion, contiennent forcément une partie harmonique : une voyelle porteuse de l'éclat lors de la désocclusion (Fig 5). Nous distinguons entre les occlusives voisées ([b], [d], [g]) où les cordes vocales vibrent au début de l'éclat, et les

⁷ HUCKVALE, Mark. « Lecture 2-5 : Formant Transitions ». p 6.

non-voisées ([p], [t], [k]) où les cordes vocales sont comprimées et un signal turbulent apparaît au début de l'éclat. (Fig 5).

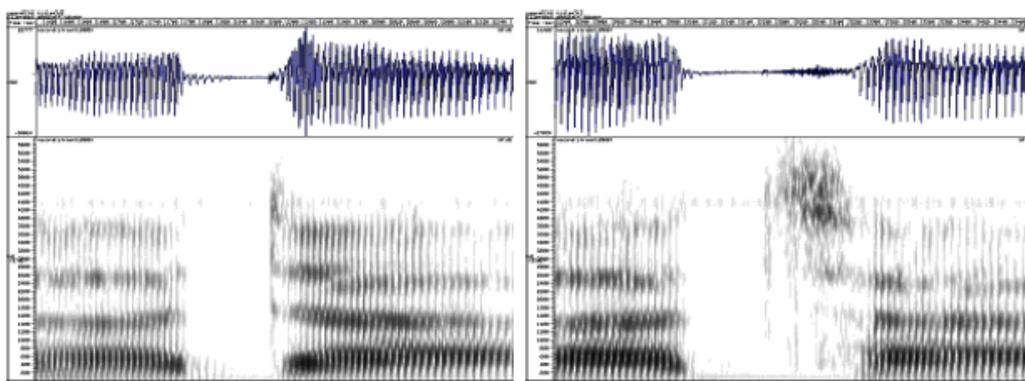


Figure 5 : Spectres de la consonne occlusive voisée [d] (gauche) et de la consonne occlusive non-voisée [t] (droit)⁸.

La distinction voisé / non-voisé est aussi valable pour les consonnes fricatives ([v]/[f], [z]/[s], [ʒ]/[ʃ]), qui génèrent un bruit turbulent continu. Les cordes vocales vibrent lors de la production d'une fricative voisée ([v], [z], [ʒ]) car la constriction principale est créée par les lèvres, les dents ou la langue. Ainsi ces consonnes sont légèrement harmoniques et plus résonnantes que leurs contreparties non-voisées ([f], [s], [ʃ]). Celles-ci, produites lors du resserrement des cordes vocales, génèrent un signal très irrégulier et peu résonnant. Elles sont ainsi presque entièrement inharmonique - plus que tout autre type de consonne discuté dans cette section (Fig. 6).

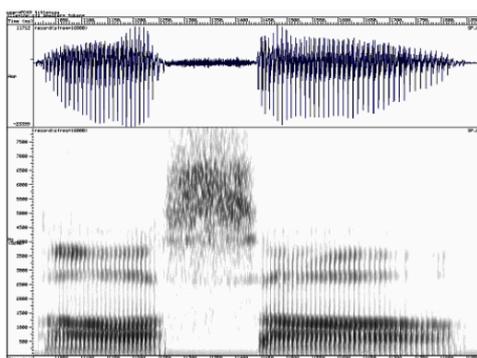


Figure 6 : Spectre du son [asa]⁹

8 HUCKVALE, Mark. « Lecture 2-6 : Plosives and Nasals ». p 3.

9 HUCKVALE, Mark. « Lecture 2-4 : Acoustics of Fricatives ». p 4.

Les nasales ([m], [n]) sont presque toujours voisées car l'occlusion faite par le voile du palais n'empêche ni la vibration libre des cordes vocales, ni le flux continue d'air. Nous constatons que leur spectre ressemble au celui de voyelles (Fig. 7), les différences principales étant le manque d'harmonicit  dans l'aigu et le placement des formants. 10

A partir de classifications des phon mes ci-dessus, nous pouvons  tablir des corr lations entre le son de la voix et sa manifestation en tant que signal. La compr hension de ces corr lations est propice au d veloppement d'une interaction voix- lectronique.

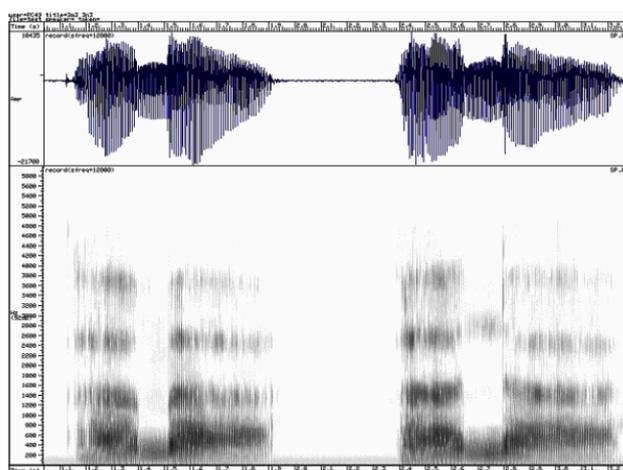


Figure 7 : Spectres des consonnes nasales [m] et [n]¹⁰.

I.1.2. ELEMENTS PSYCHOLOGIQUES

Dans cette sous-section nous discuterons des caract ristiques psychologiques jouant sur la perception de la voix chez l'auditeur.

Nous constatons que l'auditeur  coulant une pi ce de musique vocale se concentre sur certaines informations qui n'existent point dans l' coute de la musique instrumentale : le genre de chanteur, son  ge, son taille, pour ne citer que celles l . Nous voyons ainsi que l'auditeur cr e imm diatement des associations entre la sonore et le visible, en donnant une image, et m me une personnalit    la voix. En traitant la voix avec des moyens  lectroniques, il faut  tre conscient de cette association et de ses

10 HUCKVALE, Mark. « Lecture 2-6 : Plosives and Nasals ». p 6.

implications psychologiques. Si nous dénaturons le son d'un instrument, il est probable que cela ne sera perçu que comme un effet ; mais si nous dénaturons le son de la voix, il est possible que cela sera perçu comme une déformation de vocaliste même. Il faut se souvenir que – à cause de la palpabilité de ses qualités humaines – la transformation de la voix est aussi une transformation de la personne qui la porte.

Par ailleurs, à force de récurrence du traitement électronique de la voix dans notre société – soit dans des contextes commerciaux, populaire ou artistique – nous sommes habituellement exposés aux résultats sonores de certains traitements de la voix, tels que les vocoders et harmoniseurs dans la musique populaire ou de publicité. Par conséquent, un emploi similaire de ces mêmes traitements peut créer chez l'auditeur des associations qui nuisent à l'intérêt musical de l'œuvre. Nous mènerons donc notre travail avec les outils électroniques d'une façon qui cherchera à esquiver, dans une certaine mesure, les dites connotations. Ce choix n'exclut pas la mise en œuvre ni d'outils ni de traitements usuels, mais privilège une implémentation plus originelle.

I.1.3. CONCLUSIONS

Ayant approfondi nos connaissances de la voix parlée et chantée, nous arrivons à quelques conclusions propices à la suite de notre travail de création.

Sachant que la production des sons harmoniques et inharmoniques entraîne des phénomènes acoustiques distincts, nous pouvons déterminer quels traitements sont plus appropriés à chaque type de son. Le traitement des sons inharmoniques avec des outils qui modifient la hauteur, le contenu harmonique ou le contenu spectral du son, sera largement infructueux par exemple, car le signal de ces sons est trop chaotique. Par contre, nous pouvons envisager l'application des outils traitant la texture ou la temporalité – comme des lignes en retard ou des granulateurs – qui accentueront les qualités percussives et éphémères des consonnes ou des sons non-voisés. Également, nous augurons que l'application des outils temporels aux sons harmoniques et continus sera inféconde, mais que les traitements harmoniques et spectraux des dits sons seront réussis.

Par ailleurs, les observations ci-dessus peuvent nous aider à inventer de nouveaux systèmes d'interaction ou gestes vocaux. Nous pouvons envisager, par exemple, une

interaction dans laquelle le repère d'un son harmonique ou inharmonique par le patch déclenche un événement. Comme le passage entre des sons harmoniques et inharmoniques dans la langue parlée ou chantée n'est pas régulier, l'emploi proposé peut donner des résultats intéressants qui seront impossible à récréer autrement, et qui seront unique à chaque interprétation. Nous pouvons aussi envisager un patch contenant plusieurs outils, dont des outils pour le traitement de chaque type de son. Un tel système peut créer des juxtapositions étonnantes d'effets sonores en fonction du texte, ainsi qu'inspirer le compositeur à inventer des gestes vocaux passants des consonnes aux voyelles d'une manière favorable au système. Nous jugeons ainsi que la réflexion sur l'électronique en fonction de l'acoustique de la voix, et sur l'écriture vocale en fonction de la dite pensée sur l'électronique, peut engendrer des nombreuses possibilités compositionnelles.

Finalement, en réfléchissant sur les éléments psychologiques mentionnés ci-dessus, nous pouvons éviter les effets que nous considérons non souhaitables. Comme la plupart des dits effets sont issus de traitements spectraux – tel que les harmoniseurs, les flangers, les effets de chœur – nous devons faire attention de ne pas exagérer l'emploi de ces outils. Nous pouvons envisager, par exemple, de ne modifier que très peu certains paramètres, afin que notre musique soit écoutée pour son contenu et non pas pour ses effets électroniques connotés.

I.2. LE TRAITEMENT DU TEXTE

I.2.1 MISE EN CONTEXTE

Du moyen âge jusqu'au milieu du 20ème siècle, le répertoire vocale manifeste une hiérarchie évidente entre musique et texte. A travers les époques, les régions et les esthétiques, la musique a souvent été subordonnée aux paroles, son rôle étant cantonné au fait de clarifier, d'illustrer ou d'accompagner le discours. L'écriture de la musique a ainsi, souvent, suivi le même processus de corrélation avec le texte : la forme est dictée par la structure de vers, l'harmonie suit la dramaturgie du texte et les gestes sont déterminés par le figuralisme.

Dans le 20^{ème} siècle s'observe un changement de paradigme qui traverse les disciplines artistiques : l'artiste n'est plus regardé comme un prophète, chargé de livrer aux peuples la Vérité absolue à laquelle lui seul peut accéder. Une évolution similaire se constate dans la philosophie, qui s'éloigne des interrogations sur la vérité, la moralité et l'unité, pour s'orienter sur des réflexions sur la logique, la sémiologie et la multiplicité.

Pour ce qui concerne la poésie, ce développement se manifeste dans la disparition progressive du sens métaphysique qui transcende les paroles du poème ; et de plus en plus le sens se cherche dans les composantes concrètes du poème, comme l'enchaînement des mots ou des phonèmes, la prosodie, la syntaxe et le son, parmi d'autre paramètres. De la même manière, les allusions et les métaphores deviennent plus rares à partir du 20^{ème} siècle, alors que les répétitions et les jeux de mots sont davantage présents.

La poésie n'étant plus porteuse de la Vérité, la musique a pu se libérer de sa contrepartie littéraire, et à partir de la deuxième moitié du 20^{ème} siècle, les pièces ignorant ou évitant l'emploi d'un texte deviennent de plus en plus abondantes. Une des œuvres les plus importantes de ce genre, *Sequenza III* (1966) de Berio, n'utilise ni un texte ni la notation standard, mais est construite par des séries de bruits, de déclamations et de gestes vocaux (Fig. 8). *Il Canto Sospeso* (1956) de Luigi Nono, une de premières œuvre sériales pour la voix, emploie des extraits de lettres écrites lors de la deuxième guerre mondiale par des membres de la Résistance Européenne, avant leur exécution. Mais l'écriture vocale est telle que les paroles sont souvent incompréhensibles et semblent détachées du matériau vocal. Ces pièces, parmi de nombreuses autres, ont énormément contribué à l'évolution et au développement de l'écriture vocale : elles présentent un pôle opposé à celui de l'écriture vocale de jadis, et ouvrent ainsi tout le champ des possibilités entre ces deux extrêmes. Toutefois, l'approche de la composition vocale révélée au début de la deuxième moitié du 20^{ème} siècle n'aborde pas la problématique de l'intégration du texte et de la musique.

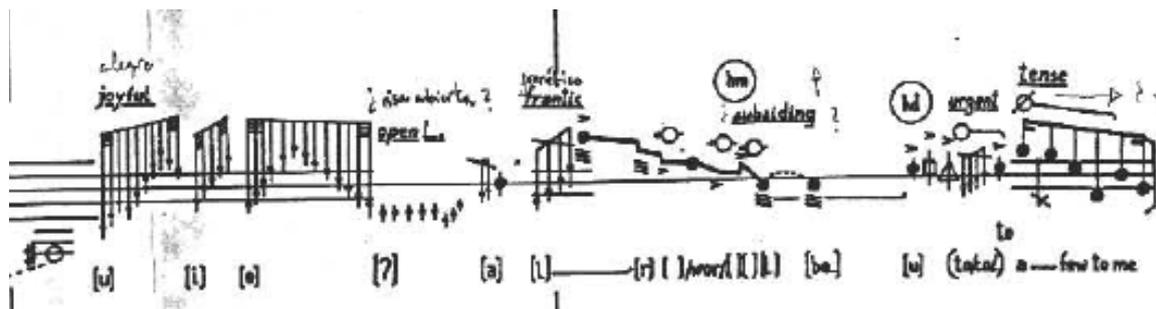


Figure 8 : extrait de *Sequenza III* de Berio.

Cette problématique est toujours présente aujourd’hui et, peut-être, encore plus difficile à affronter. Composer une pièce dans laquelle la musique suit servilement le texte présente désormais peu d’intérêt ; pourtant, employer un texte insignifiant ou écrire une musique osée qui se libère complètement des paroles est décidément moins passionnant aujourd’hui que dans les décennies d’après guerre. Il est donc inévitable que maints compositeurs s’engagent à chercher de nouvelles manières de lier texte et musique afin que les deux se complémentent sans se nier. Dans cette section de notre travail, nous exposerons comment certains compositeurs se sont servis des outils informatiques afin de trouver des nouvelles perspectives sur la question de la dynamique texte-musique.

Nous analyserons ici trois œuvres pour voix et électronique qui tentent d’établir une relation importante entre paroles et musique. Ces trois pièces représentent aussi trois manières différentes de hiérarchiser le texte par rapport à la musique : dans la première, le texte est pris comme point de départ pour la composition musicale et le traitement électronique ; dans la deuxième, la pièce est composée sans paroles, le texte n’ayant été ajouté qu’après, pour clarifier la volonté de compositeur ; dans la dernière, le traitement de texte, l’électronique et les éléments musicaux sont élaborés simultanément à partir des enregistrements et des transformations de la lecture du texte.

1.2.2 STRETTE – HECTOR PARRA, PAUL CELAN

Strette de Hector Parra est un monodrame pour soprano, électronique, vidéo en temps réel et lumières, basé sur le poème *Engführung* de Paul Celan. Le traitement électronique, qui a été réalisé avec Max/MSP, est largement en temps réel.

Le travail compositionnel de Parra est né du poème de Celan. Bien que le poème parle de l'Holocauste, on aurait tort de croire que son sens est simplement d'exprimer, à travers la poésie, l'atrocité des camps de concentration nazis. Au contraire, elle nous indique l'inexistence d'une langue qui permet de discuter un tel sujet, l'incapacité de l'aborder - surtout en employant l'allemand de jadis, désormais corrompue par le nazisme.

En composant *Strette*, Parra a cherché à mettre en relief les dits éléments linguistiques de l'écriture celanienne. A la sensibilité du compositeur, le langage de Celan prend des significations différentes au fil de l'œuvre : au début, la langue déchirée et incompréhensible marque l'imperméabilité de ce qu'a vécu le poète ; par la suite, elle « invite à pénétrer »¹¹ dans le monde du poète ; encore plus loin, elle devient un être en soi, qui permet de faire abstraction du sujet ; finalement, elle permet de faiblement franchir la réalité qui était infranchissable tout au long du poème. Cette interprétation du texte guide le déroulement dramaturgique de la musique ainsi que le choix de traitement électronique.

Dans les premières minutes de *Strette*, Parra voile et découpe la voix chantée afin de retrouver l'opacité du début de poème. L'écriture vocale - inspirée par la synthèse granulaire - et l'instruction du compositeur d' « exagérer les consonnes » limitent la continuité et la capacité de projection de la voix (Fig. 9). Ainsi, le lyrisme typique qui rend au chant sa clarté est évité ; la voix reste introvertie, inaccessible. Ce sentiment est amplifié par le traitement électronique en Max/MSP. L'objet *yin~* (voir section II.2.1), qui mesure le niveau d'harmonicité du signal, permet de détacher les voyelles et consonnes du texte chanté selon des seuils déterminés par le compositeur. Ensuite, elles sont manipulées dans des manières et des espaces différents : les consonnes sont dirigées vers le granulateur *sogs~*¹² qui les rend moins éphémères, et sont diffusées derrière le public ; les voyelles sont dirigées vers un système de multidelays qui les rend plus élastiques, et sont diffusées au fond de la scène.

¹¹ PARRA, Hèctor. Pour une approche créatrice des interrelations structurelles entre les espaces acoustiques et visuels. p. 102.

¹² *Ibidem.* p. 100.

The image shows a musical score for a vocal line. At the top left, there is a box with '♩=54' and the instruction 'Nervoso - exagérer les consonnes - non vibr.'. The score is in 2/4 time and features a variety of dynamics: *pp*, *mp*, *ppp*, *mf*, *p*, *mp*, *pp*, and *mp*. Performance markings include 'gliss.', 'vibr.', 'expressivo', 'nv.', and 'trans.'. Rhythmic notations like '3:2' and '7:8' are present. The lyrics are: 'Ver-bracht ins Ge-lä-[n]-de mit der un-tri-glistischen Spur:'. A circled '1' is at the beginning, and a circled '4' is under the first measure.

Figure 9 : extrait de *Strette* de Parra.

Au fur et à mesure, la rupture spatiale et qualitative entre voyelles et consonnes disparaît. Les paroles deviennent plus intelligibles, le son de la voix plus concret, en analogie avec la notion de pénétrer dans l'univers de Celan. Au fil de ce processus, Parra introduit d'autres traitements électroniques : un résonateur en temps réel et une bande de la voix granulée et détimbrée du soprano. En employant ces traitements, qui rendent la voix grinçante et métallique, le compositeur souhaite mettre en relief l'incapacité de parler, mais veut aussi exposer une voix synthétique permettant de parler de manière impersonnelle et protégée.

Par la suite, cette voix synthétique engendre un « langage artificiel »¹³ généré par l'ordinateur en imitation au langage déstructuré et reconstruit de Celan. C'est à ce moment que le compositeur trouve les paroles de Celan plus réelles que son expérience, le langage permettant de franchir l'Holocauste en gardant une distance émotionnelle. Il a choisi de remplacer entièrement la voix de la chanteuse avec le langage synthétisé par l'ordinateur afin d'exagérer la facticité du texte.

Finalement, les paroles du début du poème réapparaissent, mais elles semblent plus sincères et tangibles au compositeur. Le langage synthétisé de l'ordinateur disparaît et la voix seule émerge pour la première fois ; par ailleurs, l'écriture vocale devient lyrique et accessible, en contraste avec l'écriture découpée et introvertie du début de l'oeuvre. Ainsi, Parra souhaite exprimer la réconciliation entre le présent et le passé, entre l'abstraction et la réalité, ce qui reste néanmoins insuffisant pour décrire la réalité de l'Holocauste¹⁴.

¹³ *Ibidem.* p. 107.

¹⁴ *Ibidem.* p. 109.

I.2.3 POETRY FOR // DARK -/ DOLLS – MARCO SUAREZ, CHRISTOPHE TARKOS

Contrairement à *Strette*, *Poetry for // dark -/ dolls*, de Marco Suarez-Cifuentes a été conçu sans texte. Ce choix est idoine car la pièce trouve sa genèse dans des gestes vocaux purs ; plaquer un texte ou une série de phonèmes sur ces gestes pourrait les rendre moins plastiques et peut-être même contrarier leur nature. Par ailleurs, *Poetry for // dark -/ dolls* est d'une grande complexité, et l'ajout d'une autre couche de sens – en l'occurrence le texte – ne s'avèrerait pas forcément nécessaire. Après avoir composé la pièce, Suarez a pourtant décidé d'utiliser un texte de Christophe Tarkos, et ce malgré les considérations exprimées ci-dessus. Nous tenterons ici d'expliquer le travail de Suarez et éclairerons ses choix quant à l'emploi de textes.

Poetry for // dark -/ dolls a été composé pour la chanteuse Valérie Philippin et une quintette instrumental constituée de flûte, clarinette basse, violon, violoncelle et contrebasse. La création a eu lieu dans l'espace de diffusion de l'IRCAM où la scène était divisée en deux par une barrière semi-perméable : la chanteuse était située d'un côté de la scène (Zone A), la quintette de l'autre (Zone B) (Fig. 10). Avant l'exécution, les sons joués par la quintette avaient été enregistrés, échantillonnés et granulés en Max/MSP par *FTM-gabor-Musaicing* pour créer une « archive sous forme de dictionnaire »¹⁵ des grains instrumentaux de 25 millisecondes. Le même système avait été appliqué à la quintette vocale. Pendant l'exécution, les sons joués par la quintette ont été analysés en temps réel avec le même logiciel. Ensuite, l'ordinateur a cherché les échantillons du « dictionnaire » vocal qui correspondent le mieux aux sons joués dans la zone A, et les a diffusés dans la zone B où se trouvait la chanteuse. La même opération a eu lieu de l'autre côté : le chant de la zone B a été restitué par les échantillons du « dictionnaire » instrumental et diffusé dans la zone A. Le public écouta la pièce deux fois, de part et d'autre de la scène. Par ce dispositif, Suarez souhaitait créer un jeu, un système interrogeant la question du manque et de la complémentarité, et par lequel l'intégralité de la pièce pouvait être recomposée par la mémoire de l'auditeur.¹⁶

¹⁵ SUAREZ-CIFUENTES, Marco Antonio. *Poetry for// dark-/ dolls pour voix, flûte, clarinette basse, violon, violoncelle, contrebasse et dispositif informatique temps reel*. p. 5.

¹⁶ *Ibidem*. p. 3.

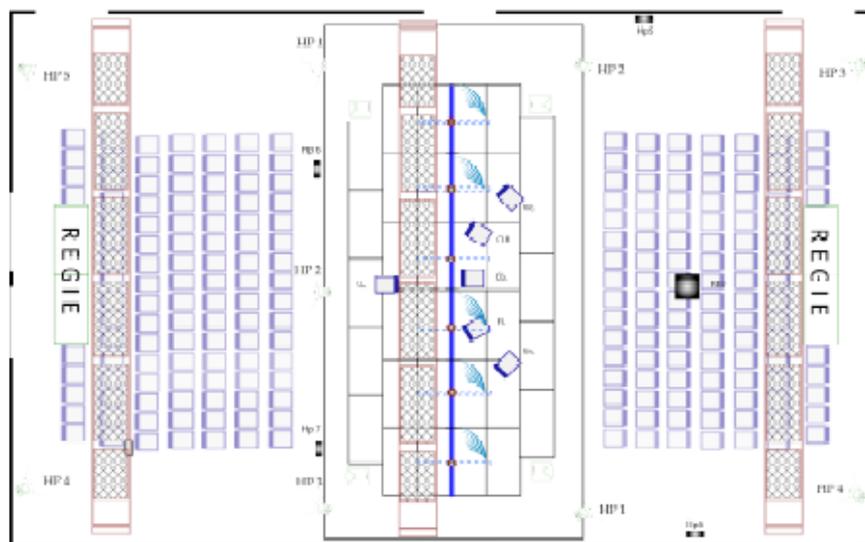


Figure 10 : Plan réalisé par Suarez pour la création de *Poetry for // dark-/ dolls* dans l'espace de projection de l'IRCAM¹⁷.

Bien¹⁷ que le concept de Suarez donne des résultats sonore séduisants, le compositeur s'est rendu compte que la pièce était trop longue (20 minutes pour chaque côté, 40 minutes en tout) et trop complexe pour être facilement suivi par le public. Il a donc décidé d'ajouter un texte sur certains passages vocaux afin d'aider l'auditeur à mieux se situer dans la pièce.

Le texte est composé de deux extraits du poème *Noir* (voir Annexe 1) de Christophe Tarkos. Dans ce poème, nous constatons encore que le sens n'existe que dans l'ordre, la fluidité et la sonorité des mots. Les mots et les phonèmes se répètent au fil du poème d'une manière découpée, chaque fois dans des contextes différents, ce qui tend à exagérer leurs sons et réduire leurs significations.

Ce texte sert particulièrement les fins du compositeur, car il forme un « flux continu autour du mot « noir » »¹⁸, qui lui permet de souligner des passages importants et marquer la mémoire de l'auditeur avec le mot « noir », tout en gardant la fluidité originelle de l'œuvre.

¹⁷ *Ibidem*. p. 3.

¹⁸ SUARUEZ-CIFUENTES, Marco Antonio. « Poetry for// dark-/ dolls », in *Saison 2008-2009, Tremplin-Cursus 2, programme*. p. 7.

Au final, la texture du poème réfléchissait les éléments granulaires et répétitifs de *Poetry for // dark -/ dolls*, et il a décidé de donner un rôle plus important au texte.¹⁹ Dans la version finale de la pièce nous trouvons des mots, des phrases entières de *Noir*, incrustés entre des séries des phonèmes bruts (Fig. 11). Elles représentent d'avantage des modes d'articulations que des structures narratives de la pièce. Suarez a réalisé son idée initiale, celle d'une écriture vocale libérée des contraintes textuelles, tout en intégrant la poésie de Tarkos dans sa création.



Figure 11 : extraits de *Poetry for // dark -/ dolls* de Suarez

I.2.4 *THEMA (HOMMAGIO A JOYCE)* – LUCIANO BERIO, JAMES JOYCE

Dans *Thema*, Luciano Berio utilise le début de l'onzième chapitre de *Ulysses* de James Joyce (voir Annexe 2). Tout comme les poèmes de Tarkos et de Celan, le texte de Joyce met en relief la sonorité et non pas la sémantique des mots. L'extrait d'*Ulysses* est une série de phrases et des cris disloqués, à travers lesquels sont racontés les événements et les personnages situés dans un bar. En employant cette écriture, Joyce souhaitait construire une polyphonie²⁰ des motifs qui se répètent, se développent, s'imitent – qui créé une version littéraire d'une fugue en canon.²¹ La musicalité du texte – marquée par

¹⁹ SUARUEZ-CIFUENTES, Marco Antonio. Entretien informel.

²⁰ BERIO, Luciano. « Poésie et musique – une expérience », in *Contrechamps*, N° 1. p. 27.

²¹ *Ibidem*. p. 27.

l'emploi persistant d'onomatopées, allitérations et références aux modes de jeu instrumentaux – paraît sans nul doute être une volonté de l'auteur.

C'est probablement pour cette raison que Berio a choisi d'utiliser ce texte, que lui avait fait découvrir Umberto Eco, à l'occasion d'une émission radiophonique. Ainsi en 1958, à l'aube de la musique mixte et avant la composition de ses pièces principales pour la voix, Berio a tenté « ...de vérifier expérimentalement une possibilité nouvelle de rencontre entre la lecture d'un texte poétique et la musique, sans pour autant que l'union doive nécessairement se résoudre au bénéfice d'un des deux systèmes expressifs : en essayant, plutôt, de rendre la parole capable d'assimiler et de conditionner complètement le fait musical. »²². Ses efforts, menés au Studio de Phonologie Musicale de la RAI en collaboration avec Eco, n'étaient que partiellement réussis. C'est alors que Berio a décidé de composer une pièce sur bande en quatre pistes.

La profondeur et la rigueur du travail de composition de Berio, sur le plan musical comme sur les plans littéraire et électronique, sont véritablement remarquables. Le compositeur a d'abord commencer ses expériences avec le texte et l'électronique : il a enregistré la lecture du texte par Cathy Berberian et l'a superposé sur elle même, deux fois, en variant régulièrement les rapports de temps et de dynamique, afin de faire émerger « spontanément »²³ les éléments linguistiques intrinsèques au texte (Fig. 12). Berio a appliqué ce même système à la traduction française du texte ; néanmoins, la lecture a été enregistrée avec une voix d'homme et une voix de femme afin d'imiter la diversité des timbres et la discontinuité de la langue anglaise. La version italienne, étant encore plus structurée que la version française, a nécessité l'emploi de trois types de voix différentes. Enfin, Berio a combiné les trois langues selon la ressemblance des éléments sonores de chaque langue : « Il s'agit d'une série d'échanges entre une langue et l'autre, qui s'effectue sur des points fixes et déterminés... par des critères de ressemblance ou de contraste. »²⁴ (Fig. 13).

²² *Ibidem.* p. 25.

²³ *Ibidem.* p. 28.

²⁴ *Ibidem.* p. 29.

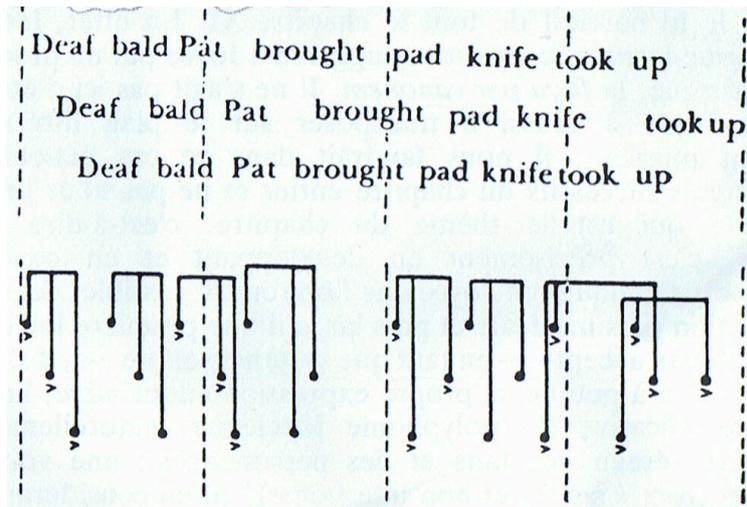
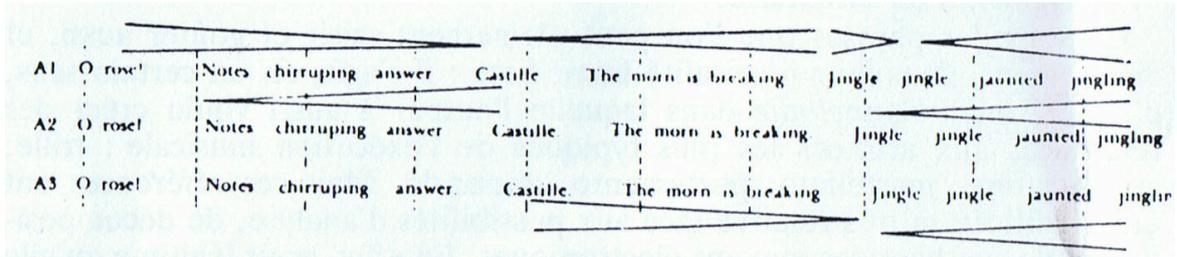


Figure 12 : Variations de vitesse et de dynamique appliquées aux enregistrements superposés de la lecture du texte²⁵.

Cette étude fut pour Berio un point de départ pour l'élaboration d'*Omaggio a Joyce* ; en suivant les conclusions tirées sur l'accentuation, le timbre et la métrique de la voix de chaque langue, le compositeur a pu aborder le développement musical du texte. Afin de mettre en valeur la richesse phonétique, rythmique et timbrale de la lecture anglaise de Cathy Berberian, Berio ne prépare l'entrée du traitement électronique qu'après une minute et cinquante secondes de lecture non-traitée.²⁶

Les transformations de la bande sont introduites progressivement à partir de 1'57''. Ici, la « continuité » mentionnée par Berio devient évidente : les sons [β] sont exagérés, dans leur lecture, à partir du trente-troisième vers (« Pearls : when she. Liszt's rhapsodies. Hissss. »²⁷), et le dernier [β], du mot « Hissss », est allongé, établissant ainsi un point commun à partir duquel surgissent les manipulations électroniques. Les moyens

²⁵ *Ibidem.* p. 28.

²⁶ STOIANOVA, Ivanka. « *Omaggio a Joyce* », in *Luciano Berio : Chemins en musique.* p. 52.

²⁷ JOYCE, James. *Ulysses.* p. 198

de traitement électroacoustique sont dès lors très présents, leur but étant de déconstruire et recomposer la partie vocale afin d'augmenter les possibilités de couleur et d'articulation proposées par la voix.

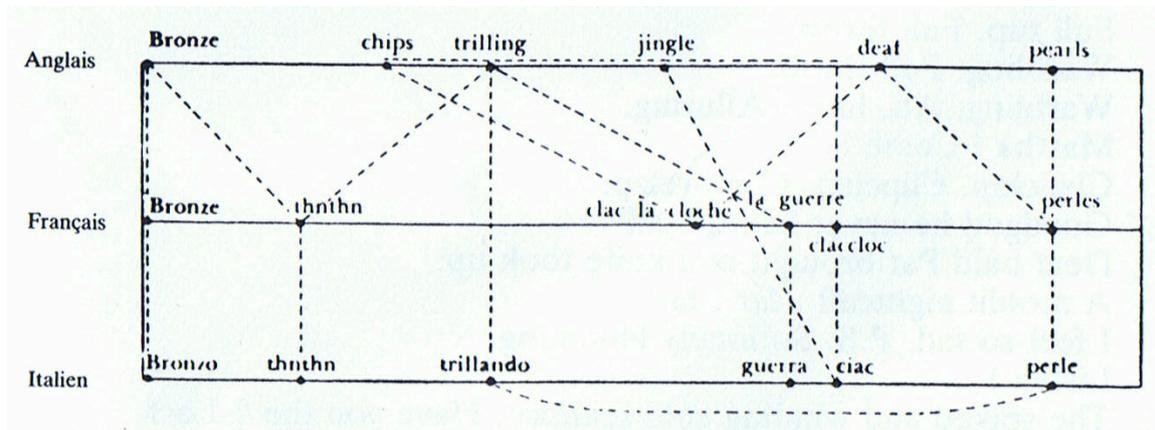


Figure 13 : Points d'échange entre les langues²⁸.

Afin d'atteindre cet objectif, Berio a classé les paroles du texte selon leurs « couleurs vocales »²⁹ – déterminées par le placement des formants des voyelles (Fig. 14). Ces composantes sont soumises à une constante variation de vitesse et de dynamique, afin de les éloigner graduellement des capacités du mécanisme vocal. Nous entendons, par exemple, des combinaisons des phonèmes articulées sur divers degrés de vitesse que la voix humaine ne peut pas facilement produire, comme l'enchaînement rapide des consonnes occlusives voisées et non-voisées³⁰. Ces nouvelles possibilités vocales sont atteintes au moyen de l'évolution progressive d'un son à l'autre comme le passage du son [β] – qui ressemble au bruit blanc et donc contient la plupart de fréquences – vers le [ʃ] et le [z], ou vers le [f] et le [v]. Ainsi le compositeur met en œuvre « la continuité essentielle »³¹ qui se trouve au cœur de son travail compositionnel.

En outre des éléments continus, Berio a incorporé des éléments périodiques afin de tempérer l'opposition entre les voyelles et les consonnes ; la rythmisation régulière des dits éléments permet de passer du son au bruit d'une manière plus aisée. A cette fin, le compositeur a fait appel aux traductions italiennes et françaises, qui sont davantage

²⁸ BERIO, Luciano. « Poésie et musique – une expérience », in Contrechamps, N° 1. p. 29.

²⁹ *Ibidem.* p. 31

³⁰ *Ibidem.* p. 32

³¹ *Ibidem.* p. 29

rythmiques et structurées. Il a choisi d'employer l'allitération française : « Petites ripas, il picore les petites ripas d'un pouce rêche, petites ripas »³² car elle est précise et reconnaissable, et la phrase italienne : « *morbida parola* » car elle met en relief la périodicité du [r] roulé.³³

(team)	(tip)	(tape)		(time)	
steelyringing	chips	fade	sail	by	
Peep	picking	awave	veil	Idolores	
leave	tink	Waves		cried	
sweetheart	pity			dyng	
feel	jingle			bright	
	listen			I	
	liszt			thigh	
	hiss			night	
				spiked	
				winding	
				silent	
(never)	(ever)	(tap)	(far)	(ton)	(talk)
breast	heard	satin	bronze	thnthnthn	more
never	answer	rang	stars	not	call
deaf	have	clacked	garter	love	morn
sad	pearls	smack	throb	tup	warm
when		pat	ah	come	war
		pad	Martha		all
		plash	far		lost
					saw
					call
					roar
(town)	(tone)	(took)	(tool)	(few)	
rebound	gold	lood	blue	blew	
now	note	could	bloom	fluted	
avowal	rose	full	who's	lure	
jaunted	chords	good	bloo	alluring	
	horn	took	bloomed		
	hawhorn		moonlit		
	so lonely		blooming		
	cold				

Figure 14 : Les mots du texte de Joyce arrangés selon les points de résonance de l'appareil vocal³⁴.

*Omaggio à Joyce*³⁴ est donc constitué des composantes discontinues issues de la langue anglaise et accentuée par leur superposition, des composantes continues issues des transformations électroniques progressives, et des composantes périodiques issues de l'emploi des langues latines. La fluidité du passage entre ces composantes et leur

³² *Ibidem.* p. 32

³³ *Ibidem.* p. 32

³⁴ *Ibidem.* p. 31

enchevêtrement naturel avec la voix nue, manifeste une intégration profonde des éléments littéraire et musicaux.

II . LE TRAITEMENT ÉLECTRONIQUE

En s'appuyant sur les recherches présentées dans la première section de ce travail, nous croyons important d'isoler et d'explorer les possibilités de traitement de chaque type de matériau vocal mentionné ci-dessus.

II.2. LA SEPARATION DES SONS

En premier lieu nous abordons quelques questions concernant l'isolation des sons harmoniques et inharmoniques. Cette séparation, comment peut-elle être effectuée avec Max/MSP? Quels objets nous seront utiles et comment seront-ils intégrés dans un patch ? Dans les limites de notre travail, nous avons choisi de discuter deux objets : *yin~* et *zerox~* (zéro cross).

II.2.1 *YIN~*

Yin~ (Fig. 15), développé par Alain de Cheveigne de l'ircam et Hideki Kawahara de l'université de Wakayama au Japon, est un estimateur de la fréquence fondamentale, de l'amplitude et de la périodicité d'un signal musical ou parlé. L'algorithme s'appuie notamment sur *la fonction de l'autocorrélation (ACF)*³⁵ mais révèle un taux d'erreur considérablement plus bas que celui d'autres méthodes d'estimation employant l'ACF, grâce a une série de cinq algorithmes supplémentaires qui corrigent certaines erreurs spécifiques à l'ACF.

³⁵ La corrélation croisée d'un signal par lui-même.

yin~

pitch analysis

by Norbert Schnell ATR - IRCAM Centre Pompidou

real-time implementation of the YIN algorithm by Alain Cheveigné et Hideki Kawahara

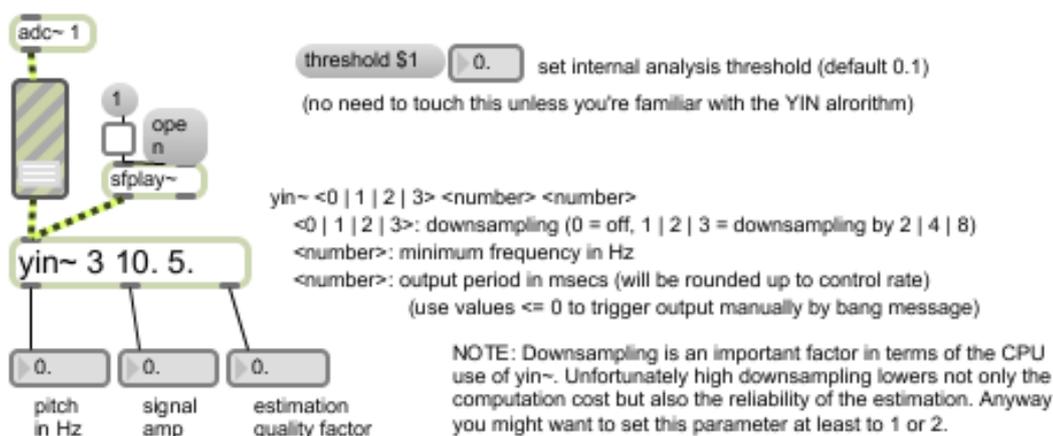


Figure 15 : l'objet yin~

La fonction de l'autocorrélation permet de détecter des éléments sinusoïdaux à l'intérieur d'un signal turbulent. D'abord, le signal est décalé et multiplié par le signal d'origine à chaque échantillon. Pour chaque signal résultant, les valeurs d'échantillons sont additionnées. Ces sommes sont en suite utilisées pour construire un nouveau signal dont les éléments cycliques sont bien plus évidents (Fig 16). A partir de ce dernier signal on peut estimer le niveau de périodicité et la fréquence fondamentale, même si elle n'est pas présente dans le signal d'origine. Bien que l'ACF soit l'algorithme le plus approprié pour ce rôle, il peut mal fonctionner lorsque certains paramètres du signal analysé sont trop bas, trop élevés ou trop variables. Pour cette raison, chacun des cinq algorithmes additionnels qui constituent *yin~*³⁶ intègrent les dits paramètres d'une façon plus fiable ; les lacunes importantes de l'ACF sont ainsi comblées.

³⁶ Difference function, cumulative mean normalized difference function, absolute threshold, parabolic interpolation, best local estimate. de CHEVEIGNÉ, Alain, KAWAHARA, Hideki. « YIN, a fundamental frequency estimator for speech and music » in *Journal of the Acoustical Society of America*. p. 1919-1921

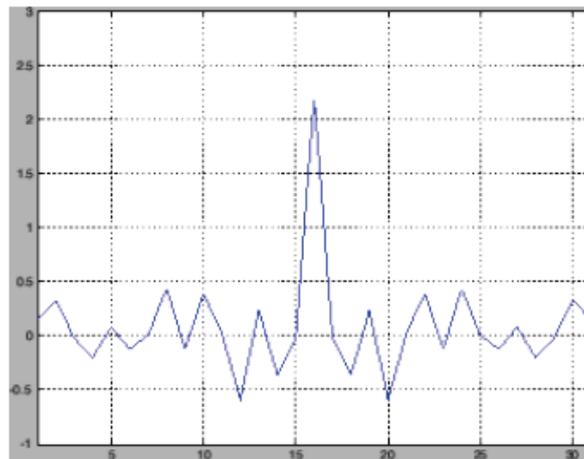


Figure 16 : le résultat de l'autocorrélation d'un signal bruiteux

Notre travail s'intéresse à la valeur rendue par la troisième sortie de *yin* : le facteur d'harmonicité. Celui est représenté par un nombre en point flottant entre 0 et 1, 0 étant le plus bruiteux et 1 le plus sinusoïdal. Nous avons observé que le facteur d'harmonicité est au delà de 0,9 pour une voyelle ou une consonne nasale chantée par une voix de femme et moins de 0,5 pour une consonne fricative. Les consonnes occlusives révèlent un résultat plus complexe : un facteur proche à 0 lors de l'occlusion, et entre 0,4 et 0,7 lors de la désocclusion. Nous constatons à travers cette dernière analyse que le taux de précision de *yin* est suffisamment élevé pour obtenir des données pour des étapes très courtes de la production sonore. Cependant, ce chiffre peut monter à 0,8 ou même 0,9 si la partie voisée de la désocclusion est légèrement plus forte ou tenue.

En s'appuyant sur les chiffres ci-dessus, nous fixons un seuil autour de 0,75. Dans notre patch Max/MSP nous pouvons employer les objets < ou > et l'objet *gate* afin de conduire les sons qui dépassent ce seuil vers une sortie et ceux qui ne l'atteignent pas vers l'autre. Ainsi nous distinguons les voyelles et nasales des consonnes occlusives et fricatives.

Il nous faut toutefois signaler que nous avons obtenu les chiffres ci-dessus en expérimentant avec des sons isolés, articulés avec une grande précision et ce, hors d'un contexte musical ou poétique. Au sein d'une œuvre, une chanteuse peut se trouver dans l'impossibilité d'articuler avec une telle exactitude sans compromettre la musicalité ou le sens de son geste. Nous avons trouvé que le facteur d'harmonicité d'une consonne occlusive peut monter à 0,9 dès que la désocclusion est légèrement trop forte ou tenue.

Egalement, une voyelle chantée rapidement dans une phrase composée de plusieurs consonnes peut ne pas atteindre le seuil fixé. Même une variation de hauteur ou d'embouchure suffit pour changer radicalement l'harmonicité mesurée par *yin*, bien qu'une telle variation ne perturbera point l'auditeur. Dans tous les cas mentionnés, non seulement le tri de sons est moins efficace, mais aussi le son à la sortie est perturbé par des cliques. La précision de *yin* est donc à double tranchant. Nous concluons que l'emploi réussi de *yin* pour la distinction des voyelles et des consonnes est favorisé par une écriture vocale qui permet d'exagérer les composants linguistiques – ce qui implique une certaine liberté rythmique.

II.2.2 ZEROX~

L'autre objet utile à notre travail, *zerox~* (zéro-cross), mesure le taux de changement de signe du signal, ou *zero-crossing rate* (ZCR). Nous avons expliqué dans la section I.1.1.1, un son non-voisé a un spectre turbulent et peu régulier. Par conséquent, le signal non-voisé oscille plus souvent entre des limites plus extrêmes que celui d'un son voisé, et son ZCR est remarquablement plus élevé. Nous observons que *zerox~* nous estime entre 8 et 35 changements de signe du signal pour la plupart de consonnes non voisées, et entre 1 et 4 pour les voyelles et les consonnes voisées. Il nous suffit donc de fixer un seuil alentour de 5 *zero-crossings* afin de trier les sons voisés et non-voisés (Fig. 17).

L'intégration de *zerox~* dans un patch max est simple. Comme il s'agit d'un objet MSP, le signal peut être trié à la sortie de *zerox~* sans passer par le domaine du control. La réalisation d'un tel traitement uniquement dans le domaine du signal est plus efficace et plus fiable. Cependant, cette efficacité est limitée au tri de signaux qui peuvent être tenus – comme les voyelles, les consonnes nasales et les consonnes fricatives. Ainsi, en employant cet outil nous facilitons le fonctionnement de notre patch, mais nous privons du traitement des consonnes occlusives et restreignons notre écriture à des contrastes plutôt grossiers.

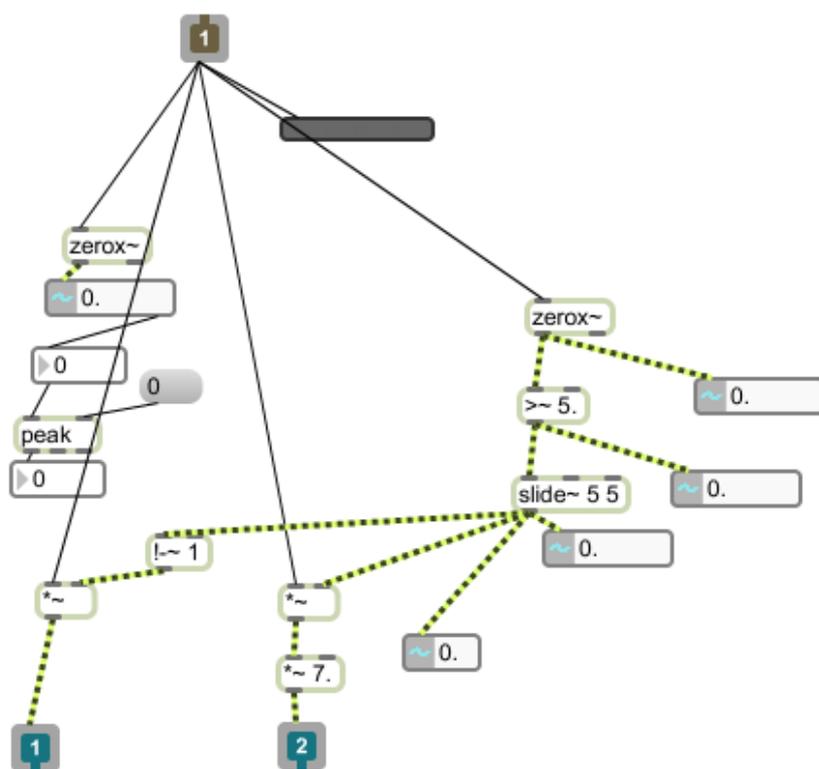


Figure 17 : le tri des consonnes et des voyelles effectué avec *zerox~*

II.2.3 CAUTIONS D'EMPLOI

Il est important de remarquer que la séparation des sons harmoniques et inharmoniques en appliquant les méthodes décrites dessus peut produire des cliques perturbants. Ceux-ci se manifestent quand le son est brusquement coupé et redémarré lors de l'alternance rapide des voyelles/voisés et consonnes/non-voisés qui se trouve naturellement dans la langue parlée. Uniquement dans la première syllabe du mot « s'asseoir » par exemple, nous entendons un ou plusieurs cliques quand : le premier « β » termine ; le « a » est prononcé ; le « a » termine ; le prochain « β » est prononcé. Afin que notre musique soit écoutable, il nous faut alors mitiger le passage d'un type de son à un autre.

Certains objets Max/MSP peuvent nous servir à ce but, notamment, *slide* (ou *slide~*), *adsr~* et *zl*. *Slide* lisse le passage entre chacune des données saisies par *yin~* ou *zerox~* (l'estimation d'harmonicité ou le ZCR) afin que les sauts brusques entre des valeurs soit plus graduels. Egalement, l'objet *zl* avec l'argument *modemedian* permet de

trouver la valeur médiane d'une liste et donc d'éviter les perturbations créées par des valeurs excentrées. L'objet *adsr~*, en comparaison avec *zl* et *slide*, n'amende pas les données utilisées pour trier le signal, mais traite directement les signaux déjà triés : nous pouvons alors établir un temps de rampe au début et à la fin du signal. En employant ces trois objets, seuls ou en combinaison, nous assurons l'application réussite de *yin~* et *zerox~* dans le contexte de notre travail.

II.3. LE TRAITEMENT DES CONSONNES – LA GRANULATION

Nous avons mentionné dans la section I.x que nous trouvons souhaitable de traiter les consonnes avec des outils s'appuyant sur des paramètres temporels. Dans cette sous-section nous discuterons un type de traitement, que nous trouvons utile à notre travail de création : la granulation.

Nous avons traité les consonnes avec trois granulateurs différents : *rogs~*, *sogs~*³⁷ et *munger~*³⁸. Tout les trois sont contrôlés par les mêmes paramètres principaux : la taille de grain, le taux de la variation de la taille de grain, la transposition de hauteur, le taux de la variation de cette transposition et la direction de lecture. Comme ces attributs de base sont trouvables dans la plupart de granulateurs, nous nous sommes intéressés dans les qualités spécifiques à chacun des outils nommés qui peuvent servir à notre travail de consonnes.

II.3.1 *Rogs~* et *sogs~*

D'un premier regard, nous remarquons que *rogs~* (Fig. 18) et *sogs~* (Fig. 19), qui ont été créés par le même auteur, lisent un fichier depuis un buffer dont l'utilisateur peut sélectionner la position ou le passage de lecture. Les deux objets permettent aussi de déterminer le nombre de sorties, et donc de spatialiser la granulation sur plusieurs canaux. Nous trouvons cette caractéristique avantageuse car la répartition directe du son traité nous semble plus convaincante que la multiple diffusion d'un son stéréo effectuée par un spatialisateur externe.

³⁷ Développés par Norbert Schnell à l'IRCAM. Font partie de la bibliothèque *IMTR*.

³⁸ Développé par Dan Trueman à la Princeton University, Princeton. Fait partie de la bibliothèque *PeRColate*.

rogs~

Romain's granular synthesis

by Norbert Schnell, IRCAM - Centre Pompidou

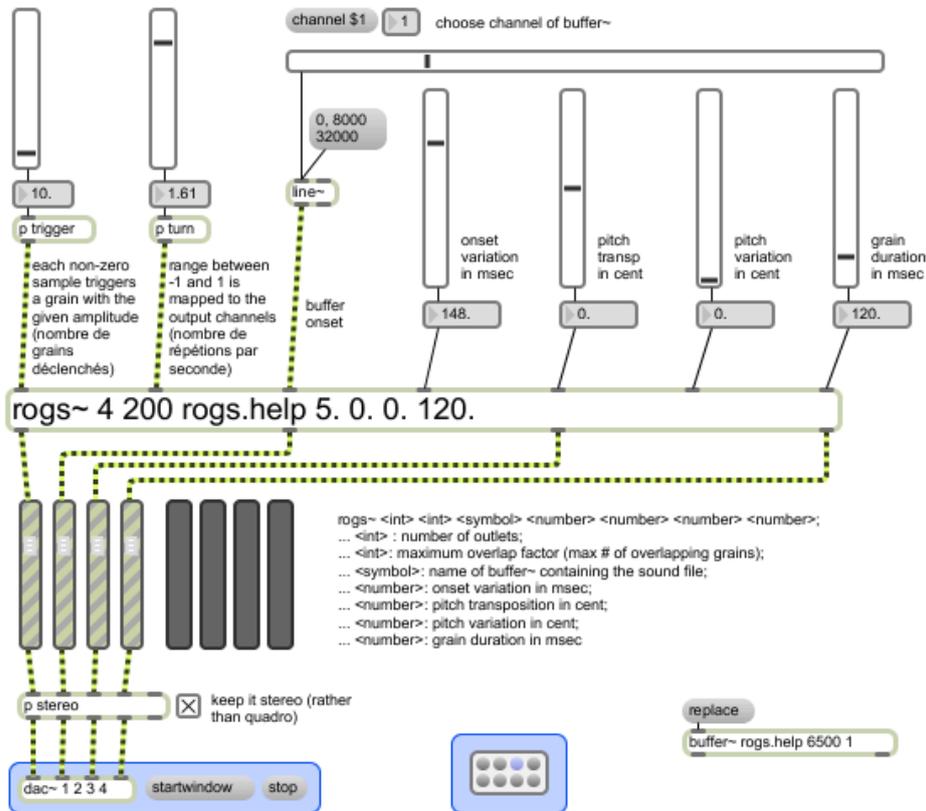


Figure 18 : l'objet rogs~

En outre des paramètres standards mentionné dessus, *rogs~* peut recevoir deux autres types d'information : le nombre de grains à déclencher et la vitesse de répétition des échantillons granulés. Le premier gère la densité de la granulation, le deuxième gère l'effet de boucle et – d'une manière plus limitée – la séparation des grains. Ce premier paramètre peut rendre des effets intéressants : en fixant une basse quantité de grains déclenchés nous pouvons obtenir une séquence d'événements courts dont le développement est imprévisible ; nous pouvons également passer d'une texture raréfiée à une texture dense avec une grande liberté.

sogs~

smooth overlap granular synthesis

by Norbert Schnell, IRCAM - Centre Pompidou

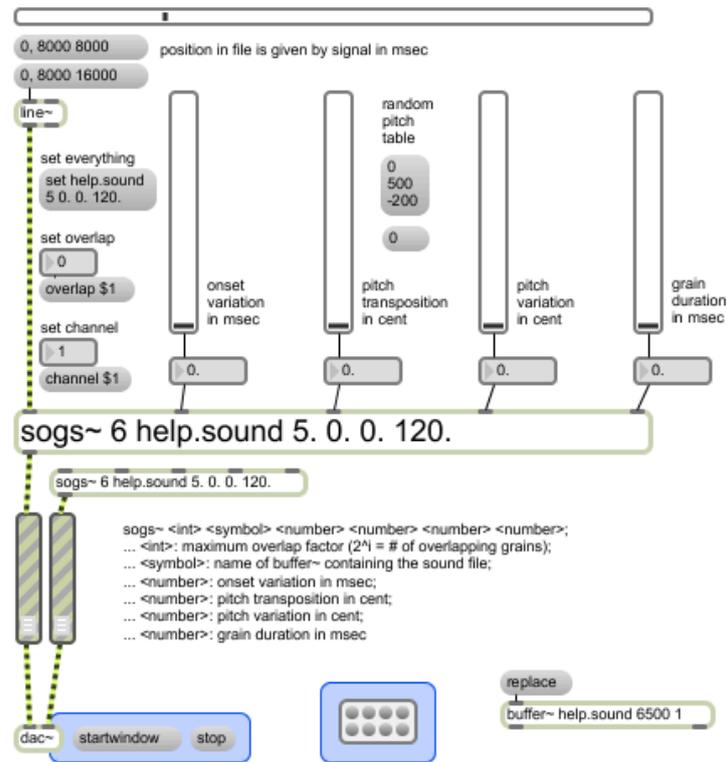


Figure 19 : l'objet sogs~

Sogs~ se distingue de *rogs~* par l'emploi de la technique *PSOLA*³⁹ qui divise le signal en sections courtes superposées et les étire ou les compresse (Fig. 20). Le résultat est une texture continue et réverbérée, colorée par une granulosité plus ou moins marquée selon la taille de la superposition sélectionnée par l'utilisateur.

39 Pitch Synchronous Overlap Add Method. Cette technique est employée notamment pour transposé des signaux sans changer leur durée. LEMMETTY, Sami. *Review of Speech Synthesis Technology*. p. 34-36.

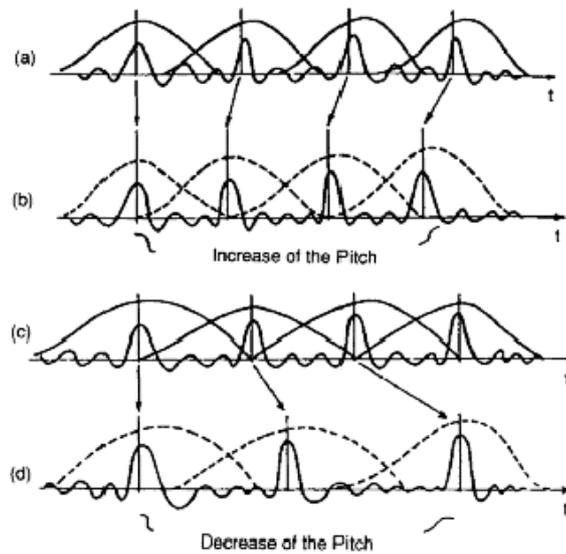


Figure 20 : schema du fonctionnement de l'algorithme *PSOLA*

II.3.2 *MUNGER~*

En comparaison avec *rogs~* et *sogs~*, *munger~* (Fig. 21) n'emploie pas un objet buffer mais lit directement depuis une source (fichier son ou *adc~*) dont le contenu est retenu par une ligne en retard⁴⁰. Cette méthode permet d'établir une interaction plus évidente avec la chanteuse mais rend le contrôle des matériaux traités moins rigoureux. Néanmoins, nous pouvons régler la durée de la ligne en retard (*delaylength_ms*), c'est à dire combien de millisecondes du signal d'origine seront gardées et traitées. Ce simple paramètre permet de varier énormément le rendu, entre des textures dispersées et inintelligibles (retard minimal), et des textures presque polyphoniques (retard longue, au delà de 5000 ms). Par ailleurs, la durée de retard n'affecte pas forcément la qualité de granulation, qui est toujours gérée par l'argument « taille de grain »⁴¹ ; ainsi, nous pouvons fixer une durée de retard assez longue tout en gardant le sentiment d'une granulation précise.

Parmi les autres caractéristiques spécifiques à *munger~*, nous trouvons intéressant le nombre de voix (*voices*) et la séparation des grains (*grain separation*). Le nombre de voix, comme le nombre des grains déclenché dans *rogs~*, permet de gérer la densité de la

⁴⁰ Le ligne en retard fonction comme un buffer interne.

⁴¹ La taille de grain est limitée à 1/3 de la durée de la ligne en retard.

granulation. Dans *munger~*, ce paramètre agit sur la totalité de la durée déterminée par *delaylength* et non seulement sur des positions fixées dans le buffer. En sélectionnant peu de voix nous nous trouverons alors avec une « phrase », découpée d'une manière imprévisible ; en sélectionnant plusieurs voix nous entendrons une superposition des dites phrases. Nous jugeons les sonorités issues de gestion de ce paramètre dans *munger~* plus fascinantes que ceux issues de *rogs~*, mais les processus de raréfaction et densification plus convaincants quand ils sont effectués avec *rogs~*.

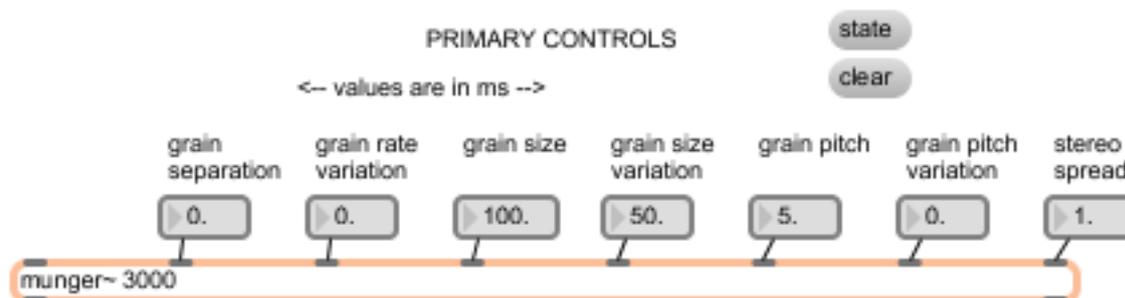


Figure 21 : l'objet *munger~* : paramètres principaux

Le paramètre *grain separation* détermine le nombre de millisecondes entre chaque grain et permet de gérer la densité. En variant le temps de séparation entre 0 ms et 500 ms, nous nous trouvons avec une masse continue plus ou moins épaisse ; en sélectionnant un temps de séparation au delà de 500 ms, nous entendons une série d'impulsions plus ou moins rapides. Ainsi, ce paramètre agit sur la densité ou sur le rythme de la granulation selon la valeur employée. Nous pouvons également employer cet outil afin de créer le sentiment d'accélération ou de ralenti.

II.4. LE TRAITEMENT DES VOYELLES

Or nous revenons à l'idée de réaliser le traitement des voyelles avec des outils s'appuyant sur des paramètres spectraux du signal vocal. Les dits outils comprennent, mais ne sont pas limités aux harmonisateurs, aux modulateurs de fréquence, aux filtres et aux résonateurs. A force de récurrence dans ces manipulations dans des contextes populaires et commerciaux, nous croyons que leur emploi peut créer chez l'auditeur des associations qui nuisent à l'intérêt musical de notre œuvre. Nous mènerons donc notre travail avec les outils en cherchant à éviter, dans une certaine mesure, ces connotations.

Afin de favoriser une mise en œuvre plus originale de ces traitements, nous avons choisi d'employer l'objet *psychoirtrist* de la bibliothèque *IMTR* dans notre travail. Dans cette sous-section nous discuterons alors du fonctionnement et de quelques applications rendues possibles par cet outil.

II.4.1 *PSYCHOIRTRIST*

*Psychoirtrist*⁴², développé à l'IRCAM par Norbert Schnell, s'appuie sur l'algorithme *yin* (voir la section 2.1) pour identifier la fréquence fondamentale et les partiels d'un son, il est constitué de trois modules multicanaux : l'harmonisateur/transpositeur, le délai et le *pitch modulation* (Fig. 22). L'association de ces trois modules et l'incorporation de la spatialisation sur plusieurs canaux est particulièrement adapté à la création de traitements hétéroclites.

L'harmonisateur/transpositeur est composé de deux parties: 1) un objet *kslider* – un interface utilisateur sous la forme d'un clavier – qui rend la hauteur et la vélocité du touche appuyé à la sortie, et 2) une série de *sliders* qui détermine le taux de transposition en cents, jusqu'à une octave au dessus et en dessous de leur hauteur originelle. En activant la fonction polyphonique de *kslider*, il est possible de synthétiser une voix pour chaque touche appuyée sur le clavier et construire ainsi une harmonie complexe. En réglant les *sliders*, nous pouvons corrélérer une transposition à chaque point de diffusion. Une deuxième série de *sliders* détermine le taux de transposition des formants en cents ; l'utilisateur pouvant également varier l'enveloppe spectrale sur chaque canal.

Bien que l'effet de chœur généré par ce module déforme la voix d'une manière connotée et inesthétique, son emploi assumé peut produire des résultats intéressants et inhabituels. Nous pouvons envisager, par exemple, d'utiliser *kslider* pour la construction d'une harmonie laide et épaisse, qui est enlevée subitement afin de mettre en valeur la richesse spectrale de la voix nue. En réalisant des transpositions contrastantes de hauteurs et de formants sur des différents points de diffusion, nous pouvons également imaginer la création d'espaces sonores distincts sur une seule scène.

⁴² Pitch synchronous yin-based choral harmoniser.

psychoirtrist~

pitch synchronous yin-based choral harmonizer

by Norbert Schnell ATR - IRCAM Centre Pompidou

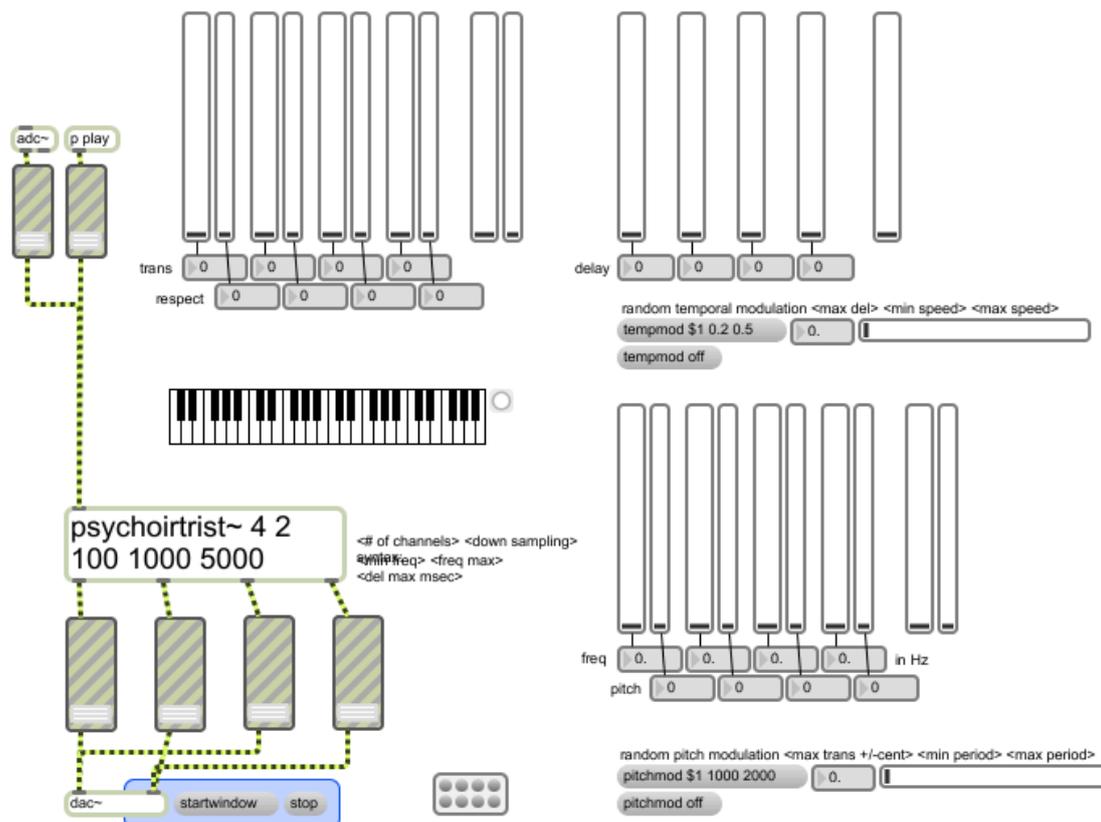


Figure 22 : les modules de l'objet *psychoirtrist~*

Le module de délai, lui aussi composé d'une série de *sliders*, permet d'ajouter un retard variant de 1 et 1000 ms à chaque canal ou à l'intégralité des canaux. Un tel traitement peut avoir diverses fonctions : nous pouvons l'employer afin de créer l'impression d'une espace plus ou moins grand, de souligner les différences entre les traitements effectués sur chaque canal, de mettre en relief certaines zones de l'espace de diffusion ou encore de créer des effets secondaires comme le déphasage. Ce module comprend aussi une modulation aléatoire de délai qui prend comme argument le délai maximal ainsi que la vitesse minimale et maximale.

Le dernier module permet de gérer la fréquence (hertz) et la hauteur (midicents) de chaque canal, et de générer une modulation aléatoire de fréquence. Comme dans le module de transposition, l'emploi subtil de ce traitement peut s'enchevêtrer avec la voix afin de

produire des colorations microtonales et de rendre la spatialisation plus dynamique. Cependant, nous ne conseillons d'utiliser la modulation de fréquence que dans un cadre assez restreint car sa présence devient vite envahissante.

II.5. CONCLUSIONS

Ayant exposé les outils Max/MSP que nous trouvons propices à notre travail, nous croyons important d'envisager notre future oeuvre en fonction des qualités spécifiques aux dits outils. En choisissant de séparer les voyelles et les consonnes avec *yin~* ou *zerox~* par exemple, nous déterminons aussi si notre écriture, ou notre choix de texte, doit privilégier la clarté de l'articulation de phonèmes, ou le contraste entre les sons voisés et non-voisés.

Egalement, en sélectionnant les traitements des consonnes et des voyelles, nous déterminons quels paramètres caractériseront la partie électronique et son interaction avec la voix. Le choix de granuler les consonnes avec *rogs~* ou *sogs~* favorise un traitement continu mais réglementé, ainsi qu'une spatialisation fine; l'emploi de *munger~* favorise un traitement imprécis mais libre, qui est idiosyncratique et qui interagit plus directement avec l'interprète. Nous constatons aussi que le traitement des voyelles avec *psychoirtrist~* encourage des transformations subtiles s'appuyant sur une spatialisation soignée. En fonction de ces observations, nous pouvons former le discours et les matériaux compositionnels qui se prêteraient à l'interaction avec l'électronique.

III. REALISATION D'AUTO-DETERMINATION POUR VOIX DE FEMME ET DISPOSITIF ELECTRONIQUE EN TEMPS REEL

En s'appuyant sur les recherches et les analyses exposées dans la première section de ce mémoire, nous avons composé *Auto-détermination* pour voix de femme et dispositif électronique en temps réel, qui a été créé par Armelle Orioux le 24 juin, 2010.

La pièce a été élaborée à partir de l'idée d'une interaction voix-électronique basée sur le contenu spectrale de la voix. Nous avons pensé qu'en employant les variations d'harmonicité intrinsèques à la langue parlée pour générer les composantes électroniques de la pièce, nous lierons infailliblement le texte, la musique et l'interprétation.

Suivant cette idée, nous avons enregistré divers gestes vocaux, enchaînements de phonèmes et lectures de texte ; nous avons également échantillonné des pièces vocales que nous trouvons intéressantes, comme *Sequenza III* de Berio et *Monomanies* d'Aperghis. Par la suite, nous avons testé le fonctionnement et l'efficacité d'une panoplie d'outils électroniques sur les fichiers obtenus ; les notions acquises de cette étude ont guidé notre travail compositionnelle préliminaire.

III.1. LE TEXTE

En s'appuyant sur les résultats de l'étude discutée ci-dessous, nous avons cherché un texte dont les éléments sonores et phonétiques se prêteraient facilement au traitement électronique. Le langage répétitif du poète surréaliste Gherasim Luca⁴³ nous a frappé comme étant particulièrement adapté à nos fins. A partir de son recueil de poèmes intitulé *Héros-Limite*, nous avons sélectionné quelques poèmes dont le traitement de phonèmes nous a évoqué des idées musicales ; parmi ces poèmes, nous avons trouvé la dimension humaine et théâtrale de *Auto-détermination* (publié en 1953) particulièrement attirante (voir Annexe 3).

⁴³ Poète Roumain né en 1913 à Bucharest. Il fait partie du groupe dadaïste roumain avec Tzara, Brancusi, etc. A cause de la deuxième guerre mondiale et l'invasion soviétique, il s'installe à Paris à 1952. Il a collaboré souvent avec A. Breton, G. Deleuze, P. Celan, M. Ernst, etc. La grande partie de son œuvre a été écrite en français. VELTER, André. Préface à *Héros-Limite*. p. I-VIII.

III.1.1 L'INTERPRETATION DU TEXTE

Nous considérons *Auto-détermination* comme ayant trois sections de caractères contrastantes. La première, qui dure du début du poème jusqu'à la ligne 17, est marquée par l'emploi des consonnes [l], [d], [m] et par la combinaison de [s] et [w] (autrement dit « soi », comme dans le mot « soir »). Etant percussives et éphémères, ces trois premières consonnes forment des structures rythmiques en fonction de l'ordre des mots et ponctuent l'écoulement de texte. Au fur et à mesure, les points de ponctuation se raréfient, les enchainements des mots s'allongent et la juxtaposition des sons de consonnes devient moins organisée, ce que nous apercevons comme un processus d'intensification. Nous trouvons ainsi que la prosodie de cette première section crée un sentiment de précipitation et d'impétuosité, éventuellement d'agitation.

la manière de
la manière de ma de maman
la manière de maman de s'asseoir
sa manie de s'asseoir sans moi
sa manie de soie sa manière de oie
oie oie oie le soir
de s'asseoir le soir sans moi
la manie de la manière chez maman
la manie de soi
le soir là
de s'asseoir là
de s'asseoir oui! de s'asseoir non! le soir là
là où la manière de s'asseoir chez soi sans moi
s'asseoir à la manière de
à la manière d'une oie en soie
elle est la soie en soi oui! oui et non!
la manie et la manière de maman de s'asseoir chez soi
sans moi

La deuxième section dure de la ligne 18 jusqu'à la ligne 26. Ici nous observons la prédominance des consonnes fricatives [β] et [ʃ], et la réduction des consonnes de la première section : [l], [m] et [d]. Par ailleurs, les voyelles de la deuxième section semblent plus présentes grâce au contraste avec les consonnes fricatives et à l'introduction du phonème « ô ». Les sons percussifs et ponctuels de la première section

sont ainsi remplacés par des sons continus. Nous trouvons ici que la prosodie nécessite une déclamation plus lente, accélérant au fur et à mesure avec la présence amplifiée des consonnes [l][m][d] et avec l'emploi plus fréquent des marques d'exclamation vers la fin de la section.

s'asseoir chez soi chérie! chez soi et toute seule chérie!
le soir à la manière d'un cheval
s'asseoir à la manière d'un cheval et d'un loup
d'un châte-loup ô chérie!
ô ma chaloupe de soie! ô! oui! s'asseoir non!
s'asseoir le soir et toute seule chez soi ô! non et non!
manière de s'asseoir sans moi chez soi
sans moi sans chez ô chérie!
c'est une manière chérie!

A la ligne 26 nous observons une reprise brève qui amène la conclusion du poème. Le contenu de cette dernière section est une synthèse des matériaux exposés auparavant : des enchaînements de mots provenant de la première partie (« manie de la manière de ») et des sons utilisés dans la deuxième partie (alternance de [s] et [ʃ]).

une manie de
une manie de la manière de
manière de s'asseoir chez soi sans chaise
s'asseoir s'asseoir sans chaise c'est ça!
c'est une manière de s'asseoir sans chaise

Notre attention est aussi attirée par des mots pôles spécifiques à chaque section : « maman » dans la première et « chérie » dans la deuxième. Etant des noms propres, ceux sont aussi des formes d'adresse ; ils impliquent une interaction humaine entre le narrateur du texte et d'autres personnages. Ainsi, nous trouvons que ces mots sont des potentielles porteuses de signification.

Sous l'angle de la dramaturgie, nous interprétons le texte comme étant le monologue insensé d'un personnage troublé. Au regard de cette interprétation, nous remarquons dans le texte une indignation croissante qui mène à un sentiment de frénésie à la fin de la première section. Nous imaginons qu'après avoir été porté par sa passion, notre personnage se fatigue et recommence son discours d'une manière plus calme mais toujours méfiant ; ainsi le caractère de la deuxième section nous semble doux. A

l'arrivée de la troisième section, nous retrouvons l'agitation et la frénésie du début mélangées, teintées par la fatigue et de frustration. L'avant-dernière ligne nous semble être une révélation cathartique qui permet au personnage de réconcilier ses inquiétudes. Nous trouvons le caractère de la dernière ligne calme et résolue.

III.1.2 CONCEPTION DE LA PIÈCE *AUTO-DETERMINATION*

En s'appuyant sur l'interprétation ci-dessous, nous établissons quelques notions générales qui guideront le déroulement de notre pièce:

- 1^{ère} section :
 - la musique est plutôt rapide et accélérera d'une manière extrême
 - l'écriture du rythme est très précise afin de mettre en relief la prosodie
 - le traitement électronique a le rôle de renforcer le caractère percussif et la brièveté des consonnes
 - l'électronique n'est pas très présente au début de la section afin d'assurer que la déclamation précipitée de la chanteuse soit comprise ; néanmoins, elle se densifie relativement à l'intensité de la partie vocale
- 2^{ème} section :
 - la musique est extrêmement lente afin de permettre le temps nécessaire pour sculpter les sons continus de cette section ; elle accélérera vers la fin de la section pour arriver à la vitesse du début
 - l'écriture est moins rigoureuse
 - le traitement électronique est constitué d'outils spectraux pour traiter les voyelles et d'outils temporels pour modifier les consonnes fricatives
 - l'électronique est assez présente afin que les traitements en temps réel soient perceptibles
- 3^{ème} section :
 - la musique reprend à la vitesse du début ou à une vitesse légèrement plus élevée
 - la section est allongée par des boucles afin que toutes les sections soient de mêmes proportions
 - le traitement accentue de nouveau les consonnes et la prosodie

- la chanteuse et l'électronique ont des rôles d'égale importance, et ce, en alternance;
- la chanteuse termine seule
- général :
 - le déroulement de la pièce suit nos impressions de caractère du personnage

A partir de ces idées de base nous avons commencé d'élaborer l'écriture vocale et l'emploi d'outils électroniques.

III.2. L'ECRITURE VOCALE

Dans cette section nous expliquerons les décisions pratiques et esthétiques selon lesquelles nous avons concrétisé l'écriture.

III.2.1 GENERAL

Nous l'avons jugé important d'exagérer les consonnes afin de faciliter leur traitement électronique, de mettre en valeur les qualités sonores du texte et de créer un langage spécifique à cette pièce. Dans l'écriture, ce choix se manifeste à travers la répétition des sons de consonnes ponctuelles (Fig. 23), la prolongation des sons de consonnes continues (Fig. 24) et le raccourci de certaines voyelles. Nous avons trouvé, conséquemment à ces manipulations, que l'énoncé du texte évoque l'impression d'un bégaiement, d'une incapacité à s'exprimer ou à aboutir un discours. Cette image correspond bien au personnage frénétique que nous avons envisagé pour l'interprète.

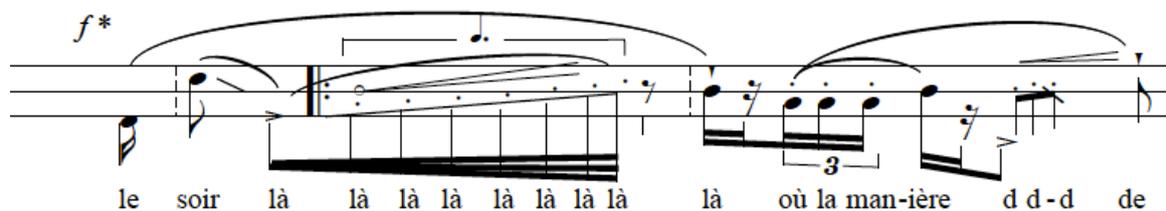


Figure 23 : la répétition des consonnes courtes

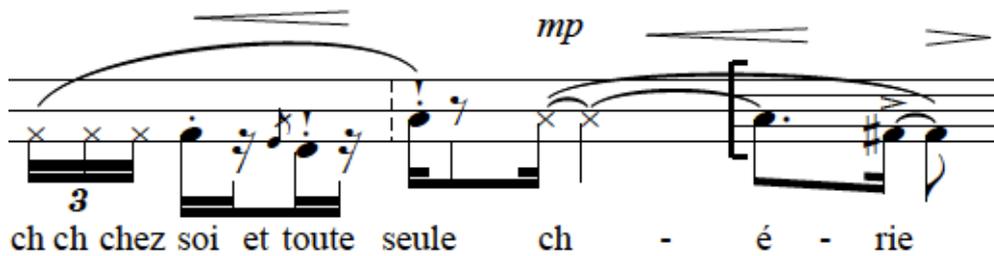


Figure 24 : la prolongation des consonnes continues

Comme nous l'avons mentionné antérieurement, la voix chantée se distingue de la voix parlée par son taux élevé de sons harmoniques, principalement des voyelles ; ainsi, l'exagération des consonnes courtes ou non-voisées privilège un style vocal peu lyrique et plutôt déclamatoire. D'après cette observation, nous avons estimé accessoire la notation d'hauteurs précises. Nous imaginons également qu'un travail rigoureux sur les hauteurs réduirait la force du rythme et de la gestualité, paramètres fondamentaux de la pièce. Nous avons ainsi noté la pièce d'une manière approximative sur des portées de trois lignes, à l'exception de certain mots clés contenant des sons harmoniques ; ceux-ci – notés sur des portées standards – aident à structurer la progression de la pièce.

III.2.2 CONSTRUCTION DE GESTES

Les gestes et cellules rythmiques exposés au début de la pièce se reproduisent, se transforment et s'enchaînent au fil de l'œuvre, comme les cellules de mots du poème. Nous avons dérivé des critères pour le traitement de chaque phonème, cellule ou mot, à selon ses attributs sonore :

- La consonne [d], qui est particulièrement percutante (Fig. 25)
 - détermine le placement de l'accent dans la plupart des phrases, et le placement des barres de mesure
 - est parlée avec la voix de poitrine et articulée d'une manière la plus percussive possible
 - peut être répétée pour la rendre plus présent ou pour varier le rythme de cellule
- Les cellules qui précèdent le [d] accentuée (Fig. 25)
 - forment des anacruses qui précipitent vers le temps fort

- sont articulées par les consonnes [l], [m] et [d]-faible (à distinguer de *d* accentuée mentionné ci-dessus) ; ces consonnes sont courtes et ponctuelles comme indiquent les points de piqués, et peuvent être répétées

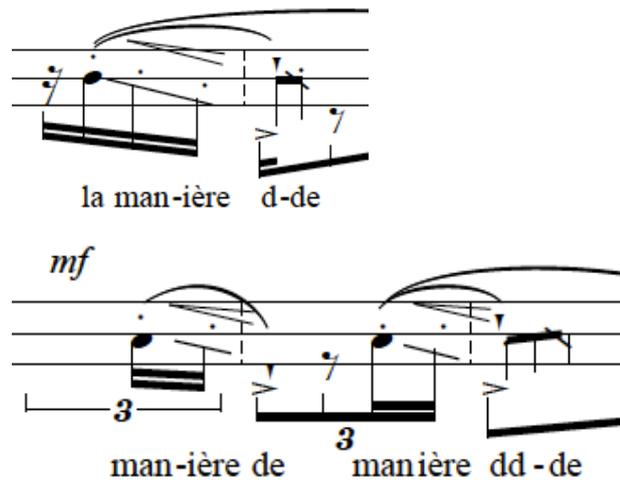


Figure 25 : l'accentuation de [d]

- La consonne [w] (comme *oi* dans « soir ») qui est diphtonguée et qui donc subit un changement interne de fréquences et de formants
 - est tenue, en contraste avec les consonnes citées ci-dessus
 - est chantée en glissant rapidement pour mettre en valeur sa diversité de fréquence
 - est transformée par la chanteuse, ce qui modifie le placement de formants
- Les consonnes fricatives *s* et *ʃ*
 - sont tenues d'une manière exagérée, afin de donner l'impression que les voyelles qui suivent sont retardées
 - sont transformées par la chanteuse, ce qui modifie le placement de formants en réarticulant ou non
- Les mots « oui », « non » et « ô » (Fig. 26)
 - sont souvent chantés et tenus sans trop varier la hauteur, pour les distinguer d'autres sons et pour marquer leur importance discursive
 - « oui » peut être séparé en deux sons : [w] et [i] ; le [w] est parfois transformée par les changements de résonance effectués par la chanteuse

- « non » est accentué à la fin du mot et peut être séparé en deux sons : [n] et [on] ; [n] est souvent tenue sur une note grave et [on] est créée d'une manière percussive
- « ô » est chanté et tenu sur une hauteur plutôt aigu ; ceci est le son le plus stable dans la pièce

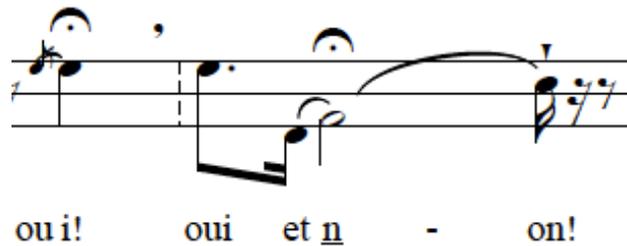
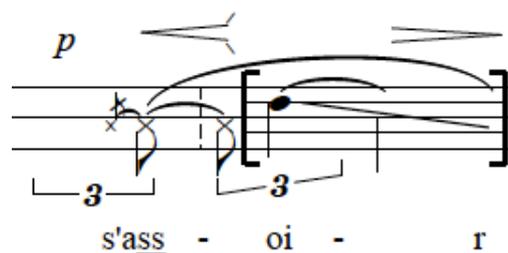


Figure 26

- Les mots « maman », « chérie », « s'asseoir » (Fig. 26)
 - Sont chantés et notés sur des portées de cinq lignes
 - « maman » et « chérie » sont chantés sur l'intervalle d'une tierce mineur descendant, en imitant le caractère plaignant d'un enfant ; la tierce est transposée au fil de l'œuvre
 - « s'asseoir » est séparé en deux parties : la partie non-voisée [βaβ] et la partie voisée [war] ; la partie voisée est chantée en glissant, en commençant sur une hauteur aiguë qui monte au fur et à mesure en fonction de l'intensité du discours



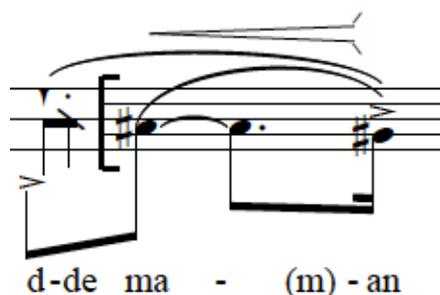


Figure 27 : traitement de “s’asseoir” et “maman”

Dans la dernière section, les gestes exposés depuis le début de la pièce sont complètement déconstruits ; nous n’employons que des phonèmes isolés et nous permettons à la chanteuse de les chanter seuls ou en combinaison dans l’ordre qu’elle choisit.

III.3. TRAITEMENT ELECTRONIQUE DANS *AUTO-DETERMINATION*

Afin de choisir les outils les plus appropriés à notre travail, nous avons appliqué divers outils aux échantillons d’une lecture d’*auto-détermination* que nous avons enregistré. Le contenu inharmonique a été traité avec les granulateurs *rogs~*, *sogs~* et *munger~*, et avec les multidelays de la bibliothèque *jamoma* ; nous avons trouvé que *munger~* nous propose les résultats les plus convaincants et permet de développer une grande interactivité. Le contenu harmonique a été traité avec les harmoniseurs *psych~* et *psychoirtrist~* de la bibliothèque *IMTR*, et avec les résonateurs de la bibliothèque *CNMAT* ; nous avons choisi d’utiliser *psychoirtrist~* car l’objet nous permet d’effectuer un traitement raffiné sur plusieurs canaux. Par ailleurs, nous avons constaté que les traitements appliqués à l’intégralité du contenu vocal ne sont pas suffisamment fiables et qu’ainsi, les sons harmoniques et inharmoniques doivent être triés à l’entrée du patch avant d’être traités. Ce tri a été tenté avec les objets *zerox~* et *yin~* ; dans la version finale de la pièce, nous employons *zerox~*, que nous trouvons plus performant.

D’après notre expérimentation, nous avons pu aussi observé que le traitement en temps réel est largement moins efficace sur la récitation rapide du texte : la granulation rend des résultats discontinus et peu subtils ; les outils développés pour la séparation des consonnes et des voyelles révèlent un taux d’erreur trop élevé. En conséquence, nous

avons décidé que les traitements en temps réel seront appliqués à une bande dans la première et dernière section, et à la voix directe dans la section centrale. Plusieurs fichiers son, composés d'un collage de consonnes et de mots enregistrés par Armelle Orioux, ont été créés à cette fin.

Finalement, nous avons décidé d'employer un contrôleur wii afin de mettre en place une interaction directe entre l'interprète et l'électronique. En ajoutant cet outil, nous espérons produire une partie électronique taillée, qui garde un côté aléatoire et qui peut être adaptée à différents interprètes.

III.3.1 TRAITEMENT DE CONSONNES

III.3.1.1 *MUNGER~*

Nous avons trouvé trois paramètres de *munger~* particulièrement fructueux: la durée de la ligne à retard qui forme le buffer interne (*delaylength*), le nombre de lignes en retard (*voices*), et la séparation de grains (*grain separation*). Dans la première section, nous avons envisagé une partie électronique peu présente qui se densifie au fil de la section selon le taux de frénésie de la chanteuse ; les paramètres *voices* et *grain separation* nous ont servi pour la réalisation de ce processus à partir du chiffre 3 (voir Annexe 4 – la partition de « Auto-détermination »). Ici, un fichier constitué des phonèmes et des mots exposés depuis le début de la pièce est déclenché par la chanteuse (déclenchement A). *Munger~* lit ce fichier en boucle et le traite selon les réglages suivants :

- la taille de grains est fixée à 200 ms ; le *delaylength* est fixé à 10,000 afin que l'auditeur puisse reconnaître le contenu de fichier
- la séparation de grains commence à 2000 ms et diminue à 20 ms d'une manière irrégulière sur une période de 25 secondes, selon une fonction gérée par l'objet *clocker* ; ceci génère un passage non-linéaire entre une texture très raréfiée et une texture dense.
- le nombre de voix commence à 1 et augmente à 10 sur une période de 16 secondes ; ceci aussi contribue à la densification de la texture

Ainsi, l'électronique est clairement présente au chiffre 4. A partir de ce moment, la partie vocale est très chaotique est assez libre ; l'électronique est donc gérée par le wii de la chanteuse : l'axe *roll*⁴⁴ (voir annexe 5) contrôle la séparation de grains – et donc la densité, entre 1 et 500 ms ; l'axe *pitch*⁴⁵ permet à varier légèrement la transposition de hauteur entre 0,9 et 1,1. Grâce à cette configuration, la chanteuse peut doser la présence de la partie électronique en fonction de son interprétation : l'électronique est plus manifeste quand la voix est faible, et vice versa. Quand la chanteuse arrive à la fin de 4, elle s'assure que l'électronique soit à son état la plus dense afin de le couper d'une manière très évidente avant le chiffre 5. Le choix d'intégrer le wii dans cette partie de la pièce est issu de notre précieuse collaboration avec Armelle Orioux.

Dans la deuxième section, la musique est suffisamment lente pour effectuer les traitements en temps réel et pour les contrôler avec le wii D'abord le son est filtré à l'entrée du patch avec *zerox~* afin d'isoler les consonnes [s] et [ʃ]. En suite, les consonnes sont soumises à une granulation, s'appuyant principalement sur le paramètre *delaylength* pour varier les grains et sculpter le son. Nous employons les réglages suivants :

- le nombre de voix est fixé à 11 et la séparation des grains est fixée à 100 ms afin que la granulation soit suffisamment dense pour que les variations internes soient perçues
- le *delaylength* varie de 200 à 3000 ms et est contrôlé par l'axe *roll* du wii
- la variation de hauteur varie de 0,3 à 2 et est contrôlé par l'axe *pitch* du wii

Ainsi, dans cette section, le traitement électronique est directement lié et largement dépendent des mouvements de la chanteuse.

Dans la dernière section nous réintroduisons des fichiers son : six fichiers composés de trois ou quatre sons de consonnes dans l'ordre dans lequel ils apparaissent dans le texte ; quatre enregistrements de Armelle Orioux chantant les quatre lignes de texte qui précèdent la ligne ultime (lignes 26 à 29). Les six fichiers sont déclenchés consécutivement par la chanteuse ; chacun est lu en boucle par *munger~* et est réglé ainsi : la taille de grain est fixée à 100 ms, le *delaylength* est fixé à 10,000 ms et la séparation de grains est fixée à 30 ms. Les quatre fichiers enregistrés sont aussi

44 la valeur entre 0 et 1 qui correspond à l'angle de torsion horaire-antihoraire par rapport au plan horizontale, en étant la valeur 0,5 quand la manette est en position horizontale (voir annexe 2).

45 La valeur entre 0 et 1 qui correspond à l'angle d'inclinaison verticale par rapport au plan horizontale, en étant la valeur 0,5 quand la manette est en position horizontale (voir annexe 2).

déclenchés consécutivement par la chanteuse, mais sont diffusés sans être traités en même temps que les fichiers granulés.

III.3.1.2 LA SPATIALISATION

Tout le matériau granulé est canalisé vers le spatialisateur *ambipanning~*⁴⁶ avant d'être diffusé. Cet objet reprend le signal stéréo généré par *munger~* et le diffuse sur six points de diffusion équidistants, placés selon la forme d'un cercle. Afin de rendre la spatialisation dynamique nous avons créé un sous-patch qui décale l'azimut de chaque point de diffusion par $\pm 30^\circ$ en fonction de niveau sonore de la voix à l'entrée du patch. Néanmoins, cette donnée varie d'une manière trop extrême quand la voix passe de grave à l'aigu, ou d'un son non-voisé à un son voisé ; ainsi le déplacement des points de diffusion est trop rapide et donc une source de cliques. En essayant d'atténuer ce mouvement, nous avons trouvé le résultat irrégulier et souvent inaudible. Finalement, nous avons appliqué une fonction qui déplace graduellement les sources sonores (les deux sorties de *munger~*) du centre jusqu'à la circonférence du cercle et vice versa, et qui est gouvernée par l'objet *clocker*.

III.3.2 LE TRAITEMENT DES VOYELLES

Le traitement de voyelles avec *psychoirtrist~* est appliqué directement à la voix à l'entrée du patch. En souhaitant éviter des transformations trop évidentes ou trop connotées, nous avons choisi de colorer les voyelles en effectuant de légères transpositions (entre 1/5 et 1/4 de ton) sur six canaux. Quatre parmi les six transpositions sont aussi accompagnées par un délai minuscule de deux à cinq millisecondes. Ayant trouvé le résultat de ces traitements trop statique, nous avons créé un sous-patch similaire à celui cité dans le paragraphe précédent, qui varie l'intervalle de transposition en fonction du niveau sonore de la voix. Le rapport entre le niveau sonore et les variations

⁴⁶ Développé à l'Institute for Computer Music and Sound Technology à la Zurich University of the Arts.

est non-linéaire et non-uniforme à travers les six canaux. Ainsi nous générons une multiple-transposition variable qui rend la spatialisation dynamique et crée des effets de « phasing ».

III.4. CONCLUSIONS

Dans *Auto-détermination*, nous avons tenté de créer une relation permanente entre l'écriture vocale, le traitement électronique et le texte, dans laquelle chaque composant influence et est influencé par les autres composants. Certains aspects de cet objectif ont été atteints, d'autres restent encore à être développés et améliorés.

Nous trouvons la musique appropriée au texte : ni le déroulement de la pièce, ni l'écriture vocale ne pourraient être facilement associées à d'autres paroles. Néanmoins, la musique n'est pas secondaire au texte : son rapport avec le poème provient d'éléments sonores et musicaux intrinsèques, et non pas d'un figuralisme ou d'une structure imposée par une signification dérivée. La pensée qui amène à établir une telle liaison est issue de notre travail sur les traitements électroniques. Sans réfléchir sur les possibilités de traitement et d'interaction voix-électronique, et sans penser les matériaux vocaux sous l'angle du domaine du signal, nous n'aurions pas été capables de développer cette corrélation.

Egalement, nous trouvons réussi le développement d'une écriture vocale originale en fonction des méthodes de traitement électronique envisagées. En souhaitant s'assurer que les traitements des voyelles et des consonnes soient plus précises, nous avons employé des modes de production vocale, des registres, des rythmes et des traitements de hauteurs différents pour chaque type de son ; l'enchaînement de dits sons produit des gestes hétéroclites que nous n'aurions pas inventé sans avoir considéré le traitement électronique et son association avec le texte.

Cependant, l'interaction entre l'interprète et l'électronique nous semble insatisfaisante. Ayant envisagé une partie électronique adaptable qui réagit directement aux variations de contenu spectral de la voix, nous avons découvert au fil de notre travail qu'une interaction si étroite nécessite soit une simplification importante du matériau vocal, soit un travail approfondi de programmation s'appuyant sur la modélisation et la synthèse vocal. La distinction entre les signaux harmoniques et inharmoniques générés

par la voix – sur laquelle notre système d’interaction est fondé – reste trop théorique dans notre travail ; en réalité, la voix contient plusieurs gradations d’harmonicité qui varient en fonction du contexte linguistique et expressif. Par ailleurs, nous n’avons pas considéré l’effet des fluctuations du niveau sonore qui se produisent lors du changement de tessiture ; ainsi, les oscillations de registre présentes dans l’écriture génèrent des montées et des baisses soudaines de niveau qui peuvent provoquer des réactions imprévisibles. Afin de compenser les défauts ci-dessus et de resserrer l’interaction entre interprète et électronique, nous avons introduit le contrôleur wii.

Durant la composition de *Auto-détermination*, nous avons travaillé en proche collaboration avec la chanteuse Armelle Orioux. Cette collaboration fut essentielle à la réalisation du projet : grâce à l’ouverture d’esprit, l’expérience et le talent de Mme. Orioux, nous avons exploré des possibilités auparavant inimaginables pour nous.

Les recherches et les expériences exposées dans notre travail nous serviront comme un point de départ pour des futurs projets compositionnels d’une plus grande exigence. En s’appuyant sur les connaissances acquises à travers ce travail, nous pourrions envisager le développement d’outils électroniques et de systèmes interactives taillé à la voix humaine, ainsi que l’intégration des recherches phonologiques dans la pratique de la composition pour la voix.

CONCLUSIONS

Les styles, lyrique, sériel et avant garde ayant atteints un état de saturation, maints jeunes compositeurs sont à la recherche des nouvelles manières de composer pour la voix et d'incorporer des textes dans leurs musiques. Dans notre travail, nous avons abordé la problématique du rôle du texte dans la musique vocale sous l'angle du traitement électronique de la voix, en supposant que cette approche nous présentera des nouvelles perspectives jusque là inaccessibles. A travers notre quête, nous avons aussi pu toucher à la question de l'écriture vocale au sein d'un système interactif. Ici nous résumerons les matières abordées dans nos recherches et les déductions effectuées.

Dans une première étape, nous avons qualifié diverses caractéristiques de la voix hors d'un contexte stylistique ou esthétique. Ce travail fut révélateur : nous avons appris à considérer la voix comme un matériau – qui est exploitable dans le domaine musical ainsi que dans celui du traitement de signal –, plutôt que comme un genre⁴⁷. A partir de ce regard, nous avons pu focalisé notre travail sur des paramètres communs à la production vocale, au traitement électronique et à la littérature, afin de s'assurer, dès un stade initial, de l'intégration des dits domaines.

En parallèle, nous avons étudié les œuvres de trois compositeurs ayant incorporés musique et texte selon trois processus compositionnels différents. De cette étude, nous avons constaté le potentiel compositionnel des éléments sonores intrinsèques au texte et l'importance de travailler en alternance sur tous les domaines qui nous intéressent, jusqu'à ce que nous puissions les traiter simultanément.

Suivant les conclusions tirées de ces étapes préliminaires, nous avons pu sélectionner et évaluer les outils Max/MSP qui semblaient appropriés à nos objectifs. Nous avons observé que chaque outil possède deux ou trois paramètres qui le distinguent des autres outils ayant la même fonction. Ces paramètres ont été pris comme un point de départ pour l'élaboration de la partie vocale. Ainsi, nous avons appris qu'en privilégiant l'emploi d'un traitement électronique particulier, nous pouvons développer une écriture

⁴⁷ « La voix est un matériau qui, comme tous les matériaux, véhicule l'histoire des usages qui en ont été faits. Un genre est une convention sociale qui, comme toutes les conventions sociales, engendre des « horizons d'attente » chez le destinataire, le « consommateur de spectacle » ». BERIO, ECO. « Eco in Ascolto », p. 96.

vocale originale. La pièce *Auto-détermination* a été composée en s'appuyant sur les observations ci-dessus.

A travers ce travail de recherche et de création, nous avons tâché de montrer la possibilité d'une relation permanente entre musique et parole, entraînée par l'emploi des outils informatiques. Par ailleurs, nous avons pu déduire que la création d'un système incorporant le texte, l'écriture, le traitement électronique et l'interprétation, peut engendrer une grande quantité de nouvelles possibilités pour la composition mixte avec voix. Néanmoins, nous trouvons que dans un cadre moins restreint que celui exposé ici, un travail plus raffiné et approfondi peut être envisagé. Dans de futures recherches, des outils analysant le contenu d'un son ou des enchaînements des sons vocaux peuvent être développés, afin de faciliter une réalisation plus précise du système interactif proposé ici. En ce cas, le traitement électronique pourrait s'adapter à chaque interprétation et pourrait être appliqué à divers textes dans diverses langues. Le changement de contenu vocal peut en plus être utilisé pour déclencher des événements sonores. En développant notre travail dans une autre direction, nous pouvons imaginer remplacer certains composants ou certaines parties du spectre vocal par des sons instrumentaux. Ainsi, nous croyons que les extensions de nos recherches peuvent faire naître des systèmes d'interaction plus sophistiqués qui seront adaptés aux champs d'exploitation plus vastes.

BIBLIOGRAPHIE

BACHU R.G., KOPPARTHI S., ADAPA B., BARKANA B.D. «Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal ». *IEEE International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering, 2008*. (7 p.). Édition électronique : www.asee.org/documents/zones/zone1/.../ASEE12008_0044_paper.pdf, [Consulté le 28 août 2010].

BERIO, Luciano. « Poésie et musique – une expérience », in *Contrechamps*, N° 1, 1983. p. 24-35.

BERIO, Luciano. *Entretiens avec Rossana Dalmonte*. Traduit de l'italien et présenté par Martin KALTENECKER, Paris : Jean-Claude Lattés, coll. Musiques & Musiciens, 1983.

BERIO, Luciano, **ECO**, Umberto. « Eco in ascolto – Entretien avec Luciano Berio » in *Les Cahiers de l'IRCAM (Recherche et Musique)*, N° 6, Paris : Éditions Ircam – Centre Pompidou, 1994. p. 95-106.

BOSSIS, Bruno. « La voix des sirènes : *Thema – Omaggio a Joyce* de Luciano Berio » in *Le Modèle vocal : La musique, la voix, la langue*. Sous la direction de Bruno BOSSIS, Jean-Paul OLIVE, Marie-Noëlle MASSON. Rennes : Presses Universitaires de Rennes, 2007, p. 23-32.

BOSSIS, Bruno. *La Voix et la machine*. Rennes : Presses Universitaires de Rennes, 2005.

de CHEVEIGNÉ, Alain, **KAWAHARA**, Hideki. « YIN, a fundamental frequency estimator for speech and music » in *Journal of the Acoustical Society of America*, Vol. 111, N° 4, 2002, p. 1917-1931. Édition électronique : recherche.ircam.fr/equipes/pcm/cheveign/pss/2002_JASA_YIN.pdf, [Consulté le 21 août, 2010].

COOK, Perry R. « Singing Voice Synthesis : History, Current Work, and Future Directions » in *Computer Music Journal*, Vol. 20, n° 3, 1996, p. 38-46.

CURTIS, Susan R., **OPPENHEIM**, Alan V., **LIM**, Jae S. « Reconstruction of Two-Dimensional Signals from Threshold Crossings ». *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '85*, Tampa, Florida. (4 p.).
Édition électronique : <www.rle.mit.edu/dspg/documents/recon2dim_1985.pdf>,
[Consulté le 28 août, 2010].

DELALANDE, François. « L'Omaggio a Joyce de Luciano Berio », in *Musique en jeu*, N° 15, Paris : Éditions Seuil, 1974, p. 45-54.

HUCKVALE, Mark. « Lecture 2-4 : Acoustics of Fricatives » *Seminar Phonetic Science : Acoustics of Speech and Hearing*. University College of London, London, 2009. (7 p).
Édition électronique : <www.phon.ucl.ac.uk/courses/spsci/acoustics/lect2-4.pdf>,
[consulté le 3 août, 2010].

HUCKVALE, Mark. « Lecture 2-5 : Formant Transition » *Seminar Phonetic Science : Acoustics of Speech and Hearing*. University College of London, London, 2009. (8 p).
Édition électronique : <www.phon.ucl.ac.uk/courses/spsci/acoustics/lect2-4.pdf>,
[consulté le 3 août, 2010].

HUCKVALE, Mark. « Lecture 2-6 : Plosives and Nasals » *Seminar Phonetic Science : Acoustics of Speech and Hearing*. University College of London, London, 2009. (8 p).
Édition électronique : <www.phon.ucl.ac.uk/courses/spsci/acoustics/lect2-4.pdf>,
[consulté le 3 août, 2010].

JOYCE, James. *Ulysses*. Mineola, New York : Dover Publications, 2002. Extrait de chapitre 11, « Les Sirènes », p 196-198.

LEMMETTY, Sami. *Review of Speech Synthesis Technology*. Mémoire de Master, Department of Electrical and Communications Engineering. Mémoire soutenu à la Helsinki University of Technology, Helsinki, 1999. (113 p.). Édition électronique : <<http://www.acoustics.hut.fi/~slemmet/dippa/thesis.pdf>>, [consulté le 31 août, 2010].

LELONG, Guy. « Le stade de l'intégration » in *Les Cahiers de l'IRCAM (Recherche et Musique)*, N° 6, Paris : Éditions Ircam – Centre Pompidou, 1994. p. 26-52.

LOIZILLON, Guillaume. « Synthèse sonore et modèle vocal » in *Le Modèle vocal : La musique, la voix, la langue*. Sous la direction de Bruno BOSSIS, Jean-Paul OLIVE, Marie-Noëlle MASSON. Rennes : Presses Universitaires de Rennes, 2007, p. 33-40.

LUCA, Gherasim. « Auto-détermination », in *Héros-Limite*. Paris : Gallimard, 2001. p. 45-46.

OSMOND-SMITH, David. *Berio*. Oxford, New York : Oxford University Press, coll. Oxford Studies of Composers, 1991.

PARRA-ESTEVE, Hèctor. *Pour une approche créatrice des interrelations structurelles entre les espaces acoustiques et visuels*. Mémoire de DEA, Arts de la Scène et du Spectacle - Option Musique Esthétique, Technologie et Création Artistique. Mémoire soutenu à l'Université Paris 8, Vincennes-St. Denis, 2005. (133 p.).

PARRA ESTEVE, Hèctor. Entretien informel, réalisé le 20 juillet 2010.

PARRA ESTEVE, Hèctor. *Strette pour soprano, électronique, vidéo en temps reel et lumières*. Barcelone : Tritó Edicions, 2003. (13 min.).

RODET, Xavier, **POTARD**, Yves, **BARRIERE**, Jean-Baptiste. « The CHANT Project : From the Synthesis of the Singing Voice to Synthesis in General » in *Computer Music Journal*, Vol. 8, N° 3, 1984, p. 15-31.

ROTHENBERG, Martin. « The Source-Filter Model Lives » *Voice Foundation 37th Annual Symposium*, The Westin, Philadelphia, Etats-Unis. (9p). Édition électronique : <www.rothenberg.org/source-filter.../Source-Filter-Lives-paper-as-presented5.pdf> [Consulté le 12 août 2010].

STOIANOVA, Ivanka. « *Omaggion a Joyce* », in *Luciano Berio : Chemins en musique*. Paris : La Revue Musicale N° 375-376-377, 1985, p. 49-55.

SUARUEZ-CIFUENTES, Marco Antonio. « Poetry for// dark-/ dolls », in *Saison 2008-2009, Tremplin-Cursus 2, programme*. Paris : Ircam – Centre Pompidou, 2009, p. 6-8.

SUARUEZ-CIFUENTES, Marco Antonio. Entretien informel, réalisé le 5 décembre, 2009.

SUARUEZ-CIFUENTES, Marco Antonio. *Poetry for// dark-/ dolls pour voix, flûte, clarinette basse, violon, violoncelle, contrebasse et dispositif informatique temps reel*. Inédité, 2008. (40 min.).

SWERDLIN, Yoni, **SMITH**, John, **WOLFE**, Joe. *Glottal Signal, Vocal Tract Resonances and Output Sound*. University of South Wales, School of Physics, Sydney. Enligne : <<http://www.phys.unsw.edu.au/jw/glottis-vocal-tract-voice.html>>, [Consulté le 10 décembre 2009].

WOLFE, Joe, **GARNIER**, Maëva, **SMITH**, John. *Voice Acoustics : an introduction*. University of South Wales, School of Physics, Sydney. Enligne : <<http://www.phys.unsw.edu.au/jw/voice.html>>, [Consulté le 10 décembre 2009].