

A LA RECHERCHE DE LA VOIX-PROTEE : PENSER LA VOIX DE SYNTHÈSE AUJOURD'HUI

Anastasia Georgaki

Music Department, Ionian University; 49100 Corfu, GREECE
georgaki@ionio.gr

Résumé

Cet article présente une méthodologie, basée sur le modèle aristotélicien des quatre causes, pour penser la voix de synthèse dans le cadre de l'informatique musicale.

Ce thème est développé en cinq parties allant de sa définition à ses buts et exigences.

Dans le premier paragraphe, nous essayons de donner diverses interprétations du terme voix de synthèse, afin de mettre l'accent sur sa spécificité, considérant ce terme du point de vue scientifique et du point de vue musical (la voix de synthèse revêt trois visages dans les créations de la musique contemporaine). Dans le deuxième paragraphe, nous exposerons les techniques et les modèles existants et leur performabilité. Le troisième paragraphe décrit les causes efficaces de la voix de synthèse en mettant l'accent sur les moteurs des scientifiques et de certains compositeurs qui les poussent à commencer un projet sur la voix de synthèse. Le quatrième paragraphe, celui des causes finales, explique les buts ultimes des scientifiques et des compositeurs pour la synthèse de la voix. Enfin, dans le dernier paragraphe, il est donné quelques éléments d'une structure représentationnelle satisfaisant aux exigences établies pour la conception d'un synthétiseur vocal, devenant un outil accessible aux compositeurs et aux musiciens qui n'ont pas nécessairement une formation scientifique dans la synthèse des sons vocaux, et pour dégager de nouvelles perspectives dans le domaine de la performance «sensorielle» de demain.

1. Introduction

Cette recherche sur la voix de synthèse a commencé au début des années 90, dans une approche plutôt musicologique que scientifique, partant d'une simple et naïve curiosité concernant l'incapacité des synthétiseurs numériques du commerce (Yamaha, Korg, Roland, etc) à produire des voix de synthèse fiables. Pendant un séjour à l'Ircam, cette curiosité nous a conduit à approfondir la recherche sur les synthétiseurs fabriqués dans des circuits de recherche d'informatique musicale, non seulement en France, mais partout dans le monde. Leur applicabilité en musique contemporaine dans des œuvres du vaste répertoire de la musique assistée par ordinateur devenait un champ d'exploration des plus attirants.

Dans cette voie, nous avons choisi d'une part d'étudier quelques modèles de synthèse de la voix adaptée aux besoins des chanteurs ou des musiciens (*Chant, Musse, Spasm, FM*, etc..) et de discuter de leurs avantages et inconvénients afin de mettre l'accent sur la recherche actuelle de la voix de synthèse. D'autre part, étant donné que la recherche en Informatique musicale, surtout à l'Ircam, est intimement liée avec la création musicale, nous avons fait l'inventaire du répertoire des œuvres qui ont fait usage de tel genre des voix. [Georgaki, 1998a].

Par cet article, nous souhaitons redéfinir "les raisons d'être" de la voix de synthèse, ses acquis et ses perspectives dans le monde de l'informatique musicale.

Préalablement à la discussion sur l'utilité d'avoir des voix de synthèse, ses incidences sur la création musicale et lutherie instrumentale, nous allons définir ce terme général - *la voix de synthèse* - qui inclut toutes les formes possibles que l'on peut engendrer à travers la synthèse pour une approche artistique et abstraite de la voix.

Afin de parvenir à ce résultat, nous avons préféré adapter nos questions sur le classique modèle aristotélicien des quatre causes¹{Bambrough 1963, *Metaphysics, I, 3, 983a*).

Dans la conception d'Aristote [Lalande, 1993] tout développement présuppose un but selon lequel *l'ousia*² progresse en se déployant du possible au réel. C'est *l'entéléchie*³ qui caractérise la matière. Aristote estime que lorsqu'on veut rendre compte d'un objet ou d'un événement, qu'il soit naturel ou artificiel, on doit prendre en considération quatre facteurs ou plutôt assigner au développement quatre causes :

- *La cause matérielle* qui nous donne des informations sur la structure de la matière de l'objet.
- *La cause formelle* qui définit la forme de l'objet.
- *La cause efficiente ou motrice*: qui donne des informations sur les moteurs et les personnes qui mettent en marche l'idée de construction de l'objet.
- *La cause finale* qui nous donne des informations sur le but final du développement de l'objet (à quoi cela peut-il servir?).

2. Du bruit blanc aux voyelles informatiques. Les causes matérielles

Les causes matérielles de la voix de synthèse nous fournissent des informations sur la structure de la matière du signal de synthèse.

Il s'agit d'un signal sonore qui - dans une des ses formes - est déterminé par le générateur et le résonateur. Le générateur est constitué par un oscillateur digital et un générateur de bruit (qui produit des impulsions ou du bruit blanc) et simule le fonctionnement des cordes vocales. Le résonateur est déterminé par des filtres spéciaux qui simulent le fonctionnement du conduit vocal humain (cavités buccales).

Il s'agit donc d'une voix *informatique* qui est décrite par des paramètres différents comme : la hauteur fondamentale, l'amplitude, les formants, l'amplitude, la largeur de bande et les trajectoires des formants, le vibrato, des variations aléatoires du vibrato, la forme du spectre général, l'intensité sonore, l'évolution temporelle du spectre, etc.. [Rodet, Depalle, 1985; Sundberg 1989].

On tentera de donner ici deux interprétations de voix de synthèse : la première concerne la recherche scientifique et la deuxième la création musicale (1980-1995).

2.1. La voix de synthèse "scientifique"

¹Selon Aristote dans le cas d'une statue, sa cause matérielle est le bronze, sa cause formelle est ce qu'elle représente, sa cause efficiente est le statuaire et sa cause finale est le but pour lequel on l'a sculpté.

²Ousia: mot grec dérivé de *ousa* (participe présent féminin du verbe *einai -être*). Le sens philosophique du terme (à la fois ce qui est, et ce qui fait de ce qui est (est ce qu'il est) peut être traduit en français par les termes substance, essence ou quiddité.

³*Entéléchie* (étym. qui a sa fin en soi) caractère de ce qui est achevé, déploiement de l'essence, terme créé par Aristote.

Ainsi, par une première approche de définition, nous pourrions dire que la *voix de synthèse est une sorte du signal sonore qui dérive d'un modèle de simulation qui condense, dans un ensemble de structures paramétrées, l'information dont on dispose a priori grâce à l'analyse.*

Afin de mieux adapter ce terme à la perception humaine, une deuxième approche pourrait être la suivante : *la voix de synthèse est représentée par un son synthétique qui garde suffisamment des caractéristiques pour que l'oreille l'aperçoive comme quelque chose d'apparenté à la voix. Au delà, tout est permis.*

Il faut aussi souligner que le terme utilisé *la voix de synthèse* est lié plus aux interactions de *la voix chantée* qu'à celles de la voix parlée et la recherche menée autour d'elle est au service de l'art et de la création musicale; même quelques voyelles chantées, bien décrites et formulées, suffisent à la définir.

Comme il a déjà été question de l'analyse, il faut souligner que souvent dans le cas de la voix on pense *analyse-synthèse* car pour l'instant on ne dispose pas de toutes les informations pour construire un modèle physique du fonctionnement de l'appareil phonatoire (cordes vocales et conduit vocal) [Depalle 1995]. C'est pourquoi, très souvent, on utilise un modèle, on prélève certaines données mesurables pour extrapoler ou pour simuler. A partir de cette analyse détaillée, on peut avancer à la synthèse du son, soit la resynthèse pour simuler le son original en contrôlant tous ses paramètres, soit la synthèse de nouveaux sons déclinant du modèle initial.

2.2. La voix de synthèse dans un contexte "musical"

D'après notre recherche [Georgaki, 1998a], nous pouvons constater que la voix de synthèse dans le monde des compositeurs n'a pas la même définition que la voix de synthèse vue par les scientifiques. Par contre, elle revêt trois aspects différents :

- la voix de synthèse *in abstracto*
- la voix de synthèse qui dérive de l'*analyse-resynthèse* d'une voix existante
- la voix qui dérive de la synthèse des formants en temps réel par un système interactif.

On se trouve donc confrontés à trois types de *voix de synthèse* :

- Le premier type, *la voix de synthèse in abstracto*, basé sur des modèles préconçus de la voix qui combinent différents paramètres dérivés de l'analyse physique de certaines voix, la variété des jeux vocaux, la quantité des associations de timbres et la sophistication de contrôle, par ex. modèle FM, *Phoné* (1981) de J.Chowning.
- Le deuxième type résulte généralement de l'analyse-synthèse, en nous obligeant à passer par les étapes d'échantillonnage, d'analyse et de remodelage d'une voix naturelle, par ex. *Mortuos plango, vivos voco* (1981) de J.Harvey, Voix de *Farinelli*;
- Le troisième, quant à lui, se situe entre le traitement et la synthèse de la voix en temps réel, [Puckette, 1991], la voix jouant non seulement la source de production synthétique mais, aussi, l'outil de contrôle pour la synthèse d'autres sons. Le problème ici, est qu'on ne peut pas intervenir sur tous les paramètres qui dérivent de l'analyse mais uniquement sur la fréquence, l'amplitude et la largeur de bande des formants, par ex. *En Echo* (1994) de Ph. Manoury.

Le deuxième et le troisième type de synthèse qui tirent leurs données de l'analyse, soit en temps différé, soit en temps réel, utilisent *des voix déjà existantes* dont on veut changer la forme par une simple manipulation des paramètres. En revanche le premier type de synthèse, la synthèse *in abstracto*, par lequel on construit *des voix inouïes* qui n'ont jamais existé dans la nature, est considéré comme un nouvel

instrument qui pose beaucoup de problèmes esthétiques (absence d'âme, voix qui n'appartient à personne, etc.).

Ayant défini les trois types de synthèse de voix comme ils se présentent dans les créations musicales, nous pourrions remarquer que la voix *de synthèse in abstracto*, appliquée dans les œuvres des années '80 laisse de plus en plus la place à la *voix par analyse-resynthèse* qui s'inscrit, elle, dans le cadre du traitement sophistiqué du signal vocal [Georgaki, 1998a].

3. De la MUS(S)E qui CHANT(e) : Les multiples visages de la voix de synthèse dans la recherche musicale

En ce qui concerne les *causes formelles* définissant la forme de la voix de synthèse, nous mettons l'accent sur la question : comment est-elle produite et sous quelle forme ?

3.1. Les techniques d'analyse-synthèse, les règles et le contrôle forment le modèle

La forme dépend de la technique de synthèse et des règles appropriées appliquées, à la fois, selon le choix des chercheurs, adaptées ou non, aux besoins des musiciens [Georgaki,1998]. Par exemple, la voix de synthèse issue du modèle *Chant* [Rodet,1984] a une identité sonore différente de celle issue du modèle *Musse* [Sundberg, 1989] ou du modèle *Spasm* [Cook, 1993], ou le modèle FM [Chowning,1981] etc.

Plus précisément, une des caractéristiques primordiales du signal vocal est *la nature du contrôle de la fréquence fondamentale* [Rodet 1985] et *le contrôle des formants*. Ainsi le processus qui amène à la synthèse d'une phrase chantée est basé sur deux axes : d'un côté sur *la recherche en micro-échelle* du signal vocal (les déviations aléatoires et périodiques du signal vocal, les trajectoires des formants, etc.) et, deuxièmement, sur *la recherche en macro-échelle des systèmes de règles* pour la performance vocale et le phrasage [Berndtsson 1995] et des systèmes de règles pour la composition [Rodet, 1986].

On peut donc suggérer qu'il y a une multitude de techniques utilisées comme processus de base pour la synthèse du signal vocal (technique additive, LPC, Source-filtre, FM, Modèles physiques) dont l'efficacité dépend de plusieurs facteurs.⁴

La deuxième part du processus de synthèse est *la recherche du contrôle* du modèle de synthèse. En contrôlant les paramètres d'entrée des algorithmes de synthèse, l'utilisateur peut espérer obtenir les résultats perceptifs qu'il recherche. Le contrôle pertinent rend un système de synthèse *flexible* et intéressant pour le musicien, puisqu'en modifiant seulement un faible nombre de paramètres, il peut obtenir un grand registre de sonorités.

La technique d'analyse, de synthèse, le contrôle et les règles appropriées constituent le corps principal du modèle de synthèse et déterminent le résultat sonore.

⁴- Le coût de calcul informatique des algorithmes de synthèse, mesuré en nombre d'opérations informatiques effectuées par échantillon du son synthétisé.

- La combinaison de la réduction du coût des machines utilisées
- L'augmentation du nombre d'instruments synthétisés
- Les temps des calculs réduit.

La technique et le contrôle de la voix sont décrits par des modèles cognitifs de la voix qui sont les suivants.

Modèles de synthèse	Techniques de synthèse	Centres de recherche
Chant (1978)	Fof (Modèle spectrale)	IRCAM, Paris
Music V (1982)	Additive(mod. spectral)	Univ. de Luminy, Marseille
Musse (1975)	Synthèse par formants	KTH, Stockholm
Musique 10 (1981)	FM (spectrale)	CCRMA, Stanford Univ.
Soufil (1991)	LPC/Fof	IRCAM, Paris
ICMS (1975)	LPC (spectrale)	CSC, Padova
Spasm/Singer (1993)	Modèle Physique	CCRMA, Stanford Univ.

Tableau 1. Les synthétiseurs et les modèles vocaux

On peut donc soutenir que chacun des modèles vocaux étudiés a un certain caractère et un certain résultat sonore. Ceux-ci ne sont pas uniquement utilisés dans la composition assistée par ordinateur, mais également dans l'étude du fonctionnement de l'appareil vocal pendant le chant, comme outil d'éducation vocale des chanteurs etc..

3.2. Les différences entre les modèles

Essayons de mettre l'accent sur les différences de chaque modèle qui, du point de vue physiologique et cognitif, pourraient rapprocher les particularités du fonctionnement d'un être humain.

Nous avons choisi d'étudier les modèles principaux *Chant*, *Musse*, *Spasm* et dans quelques applications le *modèle FM* parce qu'ils sont encadrés par des règles musicales pour le contrôle du chant.

Selon les exemples sonores de simulation de la voix tirés de chaque modèle qu'on écoute, nous pourrions déduire que la voix de synthèse se différencie selon le modèle (cela dépend des règles, de la technique de synthèse et du contrôle).

Musse est plutôt encadré par des règles qui décrivent le fonctionnement de l'appareil vocal pendant le chant et il s'adresse aux phonéticiens, psychoacousticiens, aux chanteurs du bel canto mais est moins adapté aux besoins des musiciens. *Chant* est régi par des règles musicales qui s'adressent aux compositeurs, *Spasm* est plutôt un modèle, une aide à l'éducation vocale des chanteurs. [Sundberg, 1987]

On pourrait dire qu' entre *Chant* et *Musse* le modèle de base est assez similaire car ils sont tous deux des modèles *source-filtre* avec la même particularité : la fonction formantique. Cette fonction introduit un paramètre de plus qui permet de contrôler plus précisément l'étendue spectrale d'un formant. Leur différence réside dans le mode de conception: *Musse* est conçu plutôt dans l'esprit scientifique d'acquérir des connaissances par simulation-étude de la voix chantée et c'est pour cette raison qu'il y a de nombreuses règles pour changer de registres sur le vibrato , d'articulation, etc.. Plus précisément, la spécificité du programme *Musse* réside dans la possibilité qu'il offre à l'utilisateur de gérer un très grand nombre de règles déduites de l'étude de la voix chantée ; des règles qui ne sont pas limitées seulement à la synthèse des voyelles et des consonnes mais qui sont aussi appliquées à l'expression musicale. C'est donc un outil remarquable pour l'étude du chant, tant pour ses aspects strictement vocaux que pour ceux de l'interprétation. [Carlsson, Ternstrom, Sundberg, 1991]

De plus, la recherche sur la synthèse du chant vocalique par *Musse* s'étend de l'étude acoustique d'une seule voix jusqu'à la recherche sur le chant choral. Un autre avantage du système *Musse* réside sur le contrôle : la possibilité qu'il donne à

l'utilisateur de convertir automatiquement des fichiers de musique en chants basés sur des règles. Ces fichiers ne contiennent pas seulement des informations sur la musique, mais aussi sur les voyelles et les consonnes. *Musse*, par contre, est moins adapté aux besoins des compositeurs car son contrôle des paramètres pour aller au-delà de la voix est encore limité.

Dans le cas du modèle *Chant*, il y a certainement moins de règles implantées pour le contrôle de la voix chantée que dans le modèle *Musse*. Par exemple, il y a quelques règles pour changer de registres, des règles de relation entre timbre et amplitude sonore. Par contre l'aspect de la contrôlabilité par des compositeurs y est plus développé avec des accès de formants pour pouvoir introduire des fonctions d'évolution de paramètres⁵. De plus, c'est le seul système de synthèse sonore largement utilisé par les compositeurs.

Dans le cas du modèle *Spasm*, quoique basé sur le modèle physique, on peut facilement faire un lien entre le filtre utilisé dans *Chant* et *Musse*, qui modélise le conduit vocal. Il y a moins de règles de connaissance d'acoustique musicale liées au fonctionnement de la voix chantée.

Mais sur la première approche avec les modèles, nous nous apercevons que la différence réside avant tout dans le contrôle : *Spasm* a un côté beaucoup plus accessible que les autres modèles, parce qu'il y a un contrôle graphique direct, ou un contrôle sur la forme du conduit vocal qui correspond à une réalité physique et qu'on peut appréhender facilement. Ainsi, *Spasm* peut devenir accessible aux chanteurs et professeurs de chant qui veulent comprendre concrètement comment fonctionne le placement de leur voix, le changement des registres, et d'autres particularités de l'enseignement du bel-canto ou des techniques extra-européennes, par une interaction entre ordinateur et utilisateur. [Cook, 1993]

Enfin, *la synthèse par FM* est un quasi-modèle de production de sons vocaux qui n'a aucune connexion directe avec la perception et pas, non plus, avec le fonctionnement de l'instrument de musique (sauf peut-être pour les chants d'oiseaux). De plus, c'est un modèle abstrait qui permet de faire des sons avec des spectres assez riches, à partir de peu de paramètres et peu de moyens calculatoires. Donc, cette synthèse a l'efficacité de pouvoir synthétiser des sons avec des timbres très facilement variables, avec peu de paramètres et un faible coût algorithmique.

Le problème est qu'on ne peut pas faire d'analyse car la FM n'est pas une technique de synthèse linéaire et ne permet pas d'avoir d'extensions directes entre les paramètres et la perception.

Comme les modèles de synthèse des tons chantés n'ont pas le fonctionnement physique d'un instrument, il s'agit plutôt de reproduire quelque chose qui est proche de la voix, mais il est très difficile d'élaborer tout un système de manipulations de sons vocaux, sans avoir de règles appropriées.

L'intérêt de la voix de synthèse, en FM, est, donc, de produire un paquet d'harmoniques à forte amplitude et de simuler ainsi, assez facilement, les formants.

4. La simulation comme point de départ. Les causes efficientes

Parlons maintenant *des causes efficientes* qui donnent des informations sur les moteurs et les personnes qui mettent en marche l'idée de construction de l'objet.

D'une part, *les causes efficientes* de la voix de synthèse sont déterminées par la curiosité des scientifiques liée ou non, aux nécessités des compositeurs, ou des chanteurs, d'investiger, analyser, comprendre, reproduire et simuler ce genre de signaux sonores. Ainsi, le scientifique crée un modèle vocal, d'après lequel il est capable de faire des prédictions. L'artiste n'a pas cette obligation. D'autre part, les scientifiques ont plus de maturité sociale, car leur travail est toujours un travail collectif ; ils échangent des informations sur leurs progrès de recherche, afin d'enrichir leur champ de connaissances et d'arriver à leur objectif final.

Selon Depalle [Depalle 1995] les scientifiques veulent valider leur démarche, et montrer qu'elle est identique au départ. Une fois que l'on sait que cela fonctionne, on peut extrapoler. C'est à ce moment-là que les compositeurs peuvent donner leurs idées.

En ce qui concerne les compositeurs qui ont eu l'idée-motrice de mettre en marche un projet sur la voix de synthèse pour donner naissance à de nouveaux timbres (indispensables pour leur travail artistique)⁶, nous tenons ici à mentionner deux personnalités : John Chowning, compositeur et chercheur, qui par la technique de synthèse FM a essayé d'approcher la voix de synthèse dans une conception artistique et J. C. Risset qui a essayé aussi de reproduire quelques voyelles chantées par la technique de la synthèse additive sans poursuivre la simulation parfaite, voulant donner une autre dimension à la voix et l'utiliser dans un contexte musical pour exprimer l'au-delà (l'autre face, 1985).

La collaboration du chercheur Xavier Rodet avec le compositeur Gérard Bennet a également donné naissance aux premières voix synthétiques de l'Ircam [Bennet, 1981].

A part ces exemples rares de compositeurs-luthiers qui avaient un très bon niveau scientifique et musical, nous remarquons un fossé entre la recherche sur la voix de synthèse et son applicabilité à la création ou la lutherie instrumentale d'aujourd'hui. [Georgaki, 1998b].

Quelle est donc la cause finale de la voix de synthèse ? A quoi peut-elle servir dans la musique et dans la recherche contemporaine ? Ne se serait-on pas égaré tout en étant fasciné par la construction de nouveaux modèles vocaux ? La création artistique est-elle dirigée par des innovations scientifiques, immolée sur l'autel de la recherche musicale ?

5. A la recherche de la voix-Protée : les causes finales

Selon Pierre Lévy [Lévy 1987] si on reprend la théorie des quatre causes d'Aristote, une part importante de la création artistique contemporaine s'est concentrée sur des causes efficientes, matérielles et formelles, au détriment des causes finales.

Les causes finales de la voix de synthèse sont encore à l'ombre des causes efficientes, formelles et matérielles. Les scientifiques sont à la recherche d'un modèle

⁶Parlons donc des artistes «originaux» à l'esprit de Proust dans *le temps retrouvé* : des artistes-compositeurs des artistes -musiciens qui ne désirent pas utiliser les sons-clichés, des synthétiseurs existants, mais qui sont toujours à la recherche du «différent», sous l'expérience d'un monde vocal imaginaire, sous les mots, sous les voix et dont la capacité de gérer des voix pareils présuppose une forte connaissance de la voix et de ses particularités acoustiques;

parfait et flexible de la voix, en appliquant de nouveaux algorithmes qui donnent une solution à des problèmes temporels, en désirant souligner le domaine du contrôle et de la flexibilité du modèle pour que tout ceci soit adapté aux besoins du musicien et du compositeur.

La déroute arrive lorsqu'ils perçoivent que l'idée motrice de leur travail est *la simulation de la voix* sans savoir exactement pour quelle raison et comment elle sera utilisée cette dernière dans la musique contemporaine.

En outre, les compositeurs, souvent mal informés sur les nouvelles technologies, ayant des connaissances insuffisantes sur le fonctionnement de l'appareil vocal, sur les techniques d'analyse et de synthèse, sur les problèmes techniques sous-jacents, tendent à créer des voix *ex-machina* sans bien savoir ce qu'ils veulent exprimer exactement, désorientés par les informations inadéquates qu'ils obtiennent en peu de temps, et se laissant porter par le désir profond d'innover dans le domaine vocal. Il en résulte des phonèmes bizarres, des hybrides mythiques, des voix imaginaires qui n'arrivent même pas à articuler un seul mot.....

Nous pouvons donc soutenir que, l'objectif final des scientifiques, est la construction d'un modèle flexible de la voix, d'où on ne doit pas seulement tirer des informations sur la structure de la voix chantée, mais aussi simuler et engendrer de nouveaux sons vocaux. [Sundberg 1989 ; Cook 1993 ; Rodet 1984]

Spécialement à l'Ircam, les scientifiques visent, en premier lieu, *la simulation vocale* pour mieux comprendre le fonctionnement de l'appareil vocal et la structure phonétique des voyelles, et en second plan, *l'interpolation et l'extrapolation* de la voix pour enrichir la palette sonore et tester les limites de la voix (où une voyelle perd son caractère vocal) ; troisièmement, *l'accroissement des connaissances* relatives à la perception et à la production des sons, l'intérêt final étant d'aboutir à *la conception de systèmes*, à la fois généraux et modulaires, capables de structurer et de modifier les connaissances acquises par l'analyse.

De l'autre côté, les compositeurs ne s'intéressent pas du tout à la simulation vocale, utilisent les modèles qui leur donnent la possibilité de produire des voix de synthèse avec des orientations artistiques diverses :

- La première orientation de l'intérêt compositionnel dans l'analyse/ synthèse⁷ de la voix, a ses racines dans *les recherches linguistiques* des compositeurs. [Dodge,1985; Lansky1989; David Evan Jones, 1990]

- La seconde orientation des compositeurs, dans l'usage des systèmes d'analyse-synthèse de la voix, est la *la synthèse croisée*, qui permet aux résonances de la voix de s'imposer à la source de n'importe quel son instrumental. [Moorer1978 ; McNabb 1981 ; Manoury 1991]

- La troisième orientation est de simuler des voix que l'être humain ne peut produire et d'en avoir un meilleur contrôle, par rapport à la voix humaine, (jouer sur les harmoniques facilement) pour créer des images sonores vocales multi-texturales et polyphoniques. [Chowning, 1981; Barrière 1983]

- La quatrième orientation concerne l'intervention sur la transformation du timbre vocal par les techniques d'extrapolation (étirer le signal vocal à l'extrême),

⁷Comme on peut analyser l'évolution des résonances de la fréquence fondamentale dans le temps, on a les trois dimensions de la voix indépendamment pour les manipuler dans la resynthèse.

On peut par exemple changer le temps sans changer la hauteur, changer la hauteur sans changer la qualité timbrale, etc.

l'hybridation, l'interpolation et l'abstraction, où le signal perd son caractère vocal. [Barrière 1983 ; Harvey 1981, etc.]

Pour conclure, la comparaison entre le but ultime des compositeurs et des scientifiques concernant la voix de synthèse, on remarquera, dans la figure suivante, que la simulation intéresse les compositeurs à un degré limité (pourquoi simuler une voix, quand on a, à notre disposition, de belles voix humaines expressives ?); leur intérêt est plutôt centré sur l'extrapolation, l'hybridation, l'interpolation et l'abstraction du son, afin de créer de nouveaux timbres vocaux inouïs.

En contre partie, les scientifiques s'intéressent à la construction des synthétiseurs capables de simuler les voix chantées pour construire des outils utiles non seulement à la composition mais aussi à l'étude du chant classique, à la pédagogie vocale, à l'aide musicologique et à la performance.

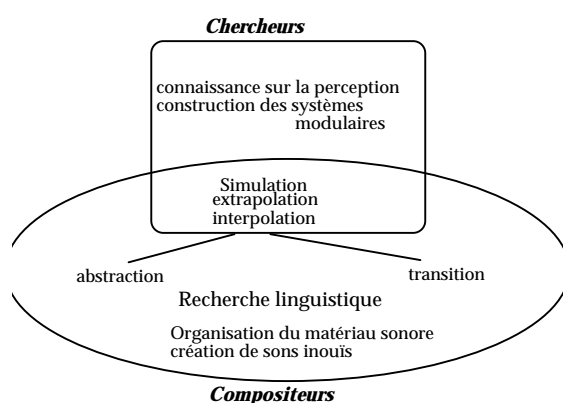


fig. 1. Les causes finales des compositeurs et des chercheurs pour la construction des voix de synthèse

En tant qu'accordéoniste, notre intérêt se porte plus spécialement à la performance musicale tout en rêvant d'un instrument qui pourrait gérer de nouveaux timbres ou former des phrase chantées synthétiques ; dans cet esprit, nous proposons un projet commun entre plusieurs pays (différentes langues et différentes techniques vocales) sur la conception d'un synthétiseur vocal cognitif universel qui pourrait condenser l'ensemble des techniques du chant et des langues de chaque pays. Cette opération pourrait être réalisée par des nouveaux modes de communication des chercheurs par Internet.

Ce synthétiseur donnerait la chance aux musiciens de connaître, après une formation spéciale, un nouvel instrument de musique, mettant la voix à la portée des mains.

L'idée donc de construire un synthétiseur vocal qui couvre toutes les gammes de la voix humaine, égalise les registres, a la possibilité de produire des voix douces ou fortes, condense dans à mémoire des voix caractéristiques d'autres civilisations, demande un contrôle extrêmement précis au niveau non seulement de l'étendue vocale, des créations de timbres vocaux (phonèmes, consonnes) mais aussi à leur adjonction. Dans ce cas-là, il est nécessaire d'étudier les modèles vocaux de chaque pays extra-européen et en fonction l'analyse, le compositeur disposera de données sur la structure des timbres qui lui permettront de gérer les paramètres avec ses propres structures et envies (voix byzantines, voix tibétaines, voix africaines,....).

Il est possible ainsi d'implanter de telles données sur des recherches appropriées dans le domaine de l'acoustique sur les styles vocaux de la musique occidentale (bel-canto, voix de «castrati») et extra-européenne (du chant sur les harmoniques, le chant byzantin, le chant mongolien, etc..)

Si l'on arrive à implanter dans un modèle intelligent de synthèse les règles des techniques vocales de la musique classique et contemporaine on peut imaginer obtenir *le meilleur des synthétiseurs vocaux*. De cette manière n'importe qui pourrait devenir le maître d'une «metavoix» qui change de visages comme *l'hypervoix* dans *le meilleur des mondes* de Aldous Huxley...

Il n'y a pas lieu ici de parler de la forme exacte de ce synthétiseur éventuel, nous nous limiterons donc à parler de nos attentes. Nous aimerions voir naître un synthétiseur qui donne un modèle plus concret pour la voix de synthèse dans le domaine musicale.

Conclusions

Cet article nous a donné l'opportunité de discuter les formes possibles de la voix de synthèse aujourd'hui.

De cette discussion se dégagent les constats suivants:

1. La voix de synthèse d'aujourd'hui revêt davantage un visage scientifique que musical. D'ailleurs la voix de synthèse n'est pas dans le centre d'intérêt des compositeurs. De nos jours, les outils du traitement sonore évoluent sur les mêmes principes, la même démarche méthodologique que les outils de la synthèse sonore et donnent la liberté désirée de créer des voix en partant d'une voix existante et de gérer leur paramètres
2. Enfin, dans la recherche musicale, il existe une multitude de modèles vocaux qui en sont au stade du développement, pouvant mener vers la simulation de la voix chantée et la construction d'outils flexibles donnant naissance à de nouveaux instruments électroniques et aux nouveautés dans la performance aussi bien que dans la composition musicale.

Références Bibliographiques

- Bambrough Renford(1963) *The philosophy of Aristotle*, Mentor, Penguin , New York.
- Barrière Jean-Baptiste (1983) *Chréode I.ICMC '83 proceedings*, New-York.
- Bennet Gerald(1981) *Singing Synthesis in Electronic Music. Research Aspects of Singing (34-50)*, The Royal Swedish Academy of Music, Stockholm, 1981.
- Berndtsson, Gunilla(1995) *Systems for synthesizing singing and for enhancing the acoustics of music rooms*. Dissertation, KTH, Department of Speech communication and Music Acoustics, Royal Institut of Technology, Stockholm.
- Carlson,G. and Ternström,S. and Sundberg, J. (1991) *A new digital system for singing synthesis allowing expressive control*.In *ICMC'91 proceedings*, Montréal,
- Chowning John (1981) *Computer Synthesis of Singing Voice*. In *ICMC '81 proceedings*, La Trobe University, Melbourne.
- Cook Perry (1993) *Spasm, a real-time Vocal Tract Physical Model Controller; and Singer the companion Software Synthesis System*. *Computer Music journal*, 17(1), MIT, Boston
- Depalle Philippe(1995) *Contribution orale, Ircam , 1995*.

- Depalle Philippe, G. Garcia, X. Rodet (1994) A virtual castrato. In *ICMC 1994 proceedings*, Aarhus, Denmark.
- Dodge Charles (1985) In celebration : The composition and its realization in synthetic speech. *Composers and the computer*, Inc. Curtis Roads ed., William Kaufman, Los Altos California.
- Dufourt Hugues(1991) *Musique, Pouvoir, écriture*, Ch. Bourgois, Paris.
- Gael Richard(1990) *Rules for fundamental frequency transition in singing synthesis*. Dept of Speech Communication and acoustics, Royal Institute of Technology, Stockholm.
- Georgaki Anastasia (1998a) *Problèmes techniques et enjeux esthétiques de la voix de synthèse dans la recherche et création musicales*. Thèse de Doctorat, EHESS/IRCAM, Paris.
- Georgaki Anastasia (1998b) Synthesis of the singing voice: links between research and creation. In the proceedings of the *First symposium on music and computers*, Ionian University, Corfu.
- Gibson, I.S., Howard, D.M., Tyrell, A.M.(1998) Real-time singing synthesis using a parallel processing system. *Proceedings of the IEE colloquium on Audio and music technology; the creative challenge of DSP*, IEE Digest 98/470, 8/1-8/6.
- Harvey Jonathan(1981) Mortuos Plango, Vivos Voco, A realisation at IRCAM, *Computer Music Journal 5*, MIT Press.
- Howell P. and N. Harvey(1975) Voice techniques. In *musical structure and Cognition*, ed. P. Howell, I Cross and R. West. New York Academic Press.
- Jones David Evan (1990) Speech extrapolated. In *Perspectives of new music*, vol.28(1).
- Lalande André (1993) *Vocabulaire technique et critique de la Philosophie*. Quadrige Presse Universitaires, Paris.
- Lansky Paul (1989) Compositional applications of linear predictive coding. *Current directions in computer music research* (pp. 5-8), ed. Max Mathews and John Pierce ed., Cambridge, Massachussets, MIT Press.
- Lomax Ken (1996) The development of a singing synthesizer. In *JIM'96 proceedings*.
- Manoury Philippe (1991) Les limites de la notion du timbre. *Le timbre, métaphore pour la composition*, textes réunis par Jean Baptiste -Barrière, Ircam et Ch. Bourgois éditeur, Paris.
- Paabon Peter(1994) A real-time singing voice analysis/ synthesis system. *ICMC 1994 Proceedings*.
- Pierucci p. , Paladin A. (1997) Singing voice analysis ans synthesis system through glottal excited formant resonators in *ICMC 97 Proceedings*, Thessaloniki.
- Potard Yves, Baisnée Pierre-François, Barrière Jean-Baptiste(1990) Méthodologie de synthèse du timbre : l'exemple des modèles de résonance. *Le timbre, métaphore pour la composition* (ed. Barrière) Ircam, Paris.
- Puckette Miller (1991) Music and speech synthesis using nonlinear distortion and amplitude modulation», *Journal of audio society*.
- Risset J.C.(1991) Timbre analysis by synthesis: representations , imitations , and variants for musical composition. In *Musical signals and representations*, De poli et al.
- Rodet X. et Al. (1984) The Chant project : From Synthesis of the singing voice to synthesis in general. *Computer Music Journal 8 (3)* (pp. 15-31), MIT Press.

- Rodet X., Depalle Ph., Poirot G.(1988) Diphone Sound Synthesis", *Int. Computer Music Conference*, Koeln, RFA.
- Sundberg John (1989) Synthesis of singing by rule. In *Current directions of computer music research*, ed. Max Mathews et John Pierce, MIT Press..
- Sundberg Johan(1989) *The science of singing voice*, Northern Illinois University Press,1987
- Tisato Graziano, Maccarini Andrea Ricci (1991) Analysis and synthesis of diphonic singing", *Bulletin d'Audiophonologie*, vol. VII no 5 et 6 - 619 -648, Ann. Sc. Aniv. Franche -Comté.