



HAL
open science

Les instruments chanteurs

Christophe d'Alessandro, Samuel Delalez, Boris Doval, Lionel Feugère, Olivier Perrotin

► **To cite this version:**

Christophe d'Alessandro, Samuel Delalez, Boris Doval, Lionel Feugère, Olivier Perrotin. Les instruments chanteurs. *Acoustique et Techniques*: trimestriel d'information des professionnels de l'acoustique, 2019, 89, pp.36-43. hal-02025861

HAL Id: hal-02025861

<https://hal.science/hal-02025861>

Submitted on 19 Feb 2019

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

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

Les instruments chanteurs

**Christophe d'Alessandro,
Samuel Delalez,
Boris Doval,
Lionel Feugère,
Olivier Perrotin**

LAM * Lutherie-Acoustique-Musique
Institut Jean le Rond d'Alembert
Sorbonne Université - CNRS, UMR 7190
4, place Jussieu (case 162)
75252 Paris CEDEX 05

E-mail :
christophe.dalessandro@sorbonne-universite.fr

Ces travaux ont été réalisés pendant que
les auteurs travaillaient au :
LIMSI-CNRS
Campus universitaire bât 508
Rue John von Neumann
91405 Orsay CEDEX

Résumé

Les instruments chanteurs sont nés de la rencontre entre synthèse vocale et nouvelles interfaces pour l'interaction humain-machine. La voix n'est pas un instrument de musique, car il n'y a pas d'objet externe mis en jeu par les membres ou par le souffle. La synthèse numérique permet pour la première fois une coupure entre le sujet et sa voix, en construisant des instruments chanteurs manipulés par les mains, les pieds, ou par toutes sortes d'interfaces humain-machine. Cependant, les possibilités de contrôle des instruments chanteurs sont encore limitées à certains aspects, car une transposition des gestes internes du chant effectués par l'appareil vocal en gestes externes des membres ne va pas de soi. Certains gestes sont analogues, d'autres sont médiatisés par un espace perceptif. Les travaux menés sur trois instruments chanteurs sont présentés : le contrôle de l'intonation par un stylet sur une tablette graphique et des gestes d'écriture ; le contrôle des voyelles et de la qualité vocale sur une surface ; le contrôle bimanuel de l'articulation consonantique ; le contrôle syllabique du rythme. Les modèles de synthèse sous-jacents utilisent soit la simulation du modèle source-filtre, soit la modification d'échantillons préenregistrés et étiquetés. Le contrôle des instruments chanteurs est multimodal, impliquant l'ouïe ainsi que la vue, le toucher, et la kinesthésie. Cette combinaison sensorielle et motrice permet dans certains cas de rendre les instruments chanteurs plus justes et précis que la voix, la vue privilégiant les aspects mélodiques et l'audition les aspects rythmiques. Miroir de la voix, l'instrument chanteur autorise toutes sortes de spéculations : musicales bien sûr, avec le Chorus Digitalis, chœur de voix de synthèse, mais aussi pour l'analyse des pratiques vocales, pour l'éducation ou la rééducation, en renforçant l'apprentissage de gestes vocaux par des traces visuelles, des gestes manuels ou corporels. Le statut symbolique de la voix est également affecté par la possibilité de contrôler et de produire le son vocal en dehors du corps : corps augmenté, mise en scène de l'expression vocale, double de la voix, jeu de la voix d'un ou d'une autre.

Abstract

Singing instruments are the result of the encounter between voice synthesis and new interfaces for human-computer interaction. Voice is not a musical instrument, since it does not involve an external object stimulated by limbs or breathe. In contrast, digital synthesis allows for the first time to separate the subject from its voice, by building singing instruments manipulated by hands, feet, or any human-computer interface. However, possibilities for singing instruments control are still limited to some aspects, as the transposition from internal singing gestures realised by the vocal apparatus to external gestures is not trivial. Some gestures are analogous, while others are transposed in perceptive spaces. Related work realised on three singing instruments is introduced: the intonation control by a stylus on a graphic tablet and writing gestures; the vocalic and voice quality controls on a surface; the bi-manual control of consonantal articulation; the rhythmic control of syllables. The underlying voice production models use either the simulation of a source-filter model, or the modification of pre-recorded and labelled samples. The control of singing instruments is multi-modal, involving hearing, sight, touch, and kinaesthesia. In some extent, this sensorimotor combination allows the singing instrument to be more accurate and precise than natural voice: the sight favouring melodic aspects while hearing being more related to rhythmic aspects. Mirror of voice, the singing instrument allows any kind of speculation: indubitably musical with Chorus Digitalis, a choir of synthesised voices, but also for the analysis of vocal practices, for education or re-education by strengthening the learning of vocal gestures through uses of visual traces, and manual and corporal gestures. Finally, the symbolic status of voice is also affected by the possibility to produce a vocal sound from outside the body: augmented body, staging of vocal expression, voice double, play of someone else's voice.

La synthèse ou l'analyse-synthèse de la voix chantée est un sujet de recherche déjà ancien. La voix chantée a été synthétisée dès les premières tentatives de synthèse numérique de la parole dans les années 1960. Le propos de cet article est de présenter le principe de fonctionnement des instruments chanteurs que nous avons développés au cours des dernières

années, et de discuter des applications et de l'avenir de cette approche. Comme la place manque pour détailler tous ces travaux, une liste de publications en fin d'article permettra au lecteur curieux d'approfondir le sujet. Les instruments ont été utilisés dans divers concerts et des exemples sonores se trouvent facilement sur la toile.

La notion de voix de synthèse recouvre trois aspects différents :

- Les vocodeurs, ou systèmes d'analyse-synthèse. Dans ce cas, la voix d'un chanteur est analysée, transformée puis resynthétisée. Les vocodeurs sont utilisés pour corriger certains aspects du son (justesse, timbre) sans que cela soit perçu explicitement ; pour ajouter des effets sonores audibles pour le timbre, l'intonation, le rythme ; pour étendre les possibilités vocales. Les vocodeurs sont omniprésents dans la musique enregistrée et diffusée. Ils sont presque systématiquement intégrés dans la chaîne de production de la musique vocale en studio.
- Les systèmes de synthèse à partir du texte produisent la voix chantée synthétique à partir de la partition et des paroles. De nombreuses productions musicales font appel à ces technologies, intégrées à une station de travail audionumérique, et fonctionnant en temps différé, comme un éditeur de texte. Un exemple est le système Vocaloid de Yamaha (contraction de Vocal Android), qui a connu un immense succès populaire au Japon.
- Les instruments chanteurs solistes sont des instruments numériques pour la synthèse du chant en temps-réel. L'effet de « chœur » est présent dans les synthétiseurs musicaux depuis longtemps. Mais le réalisme étant limité, ces chœurs de voix ne sont employés que comme des effets d'orchestraux, et non comme des voix solistes. La recherche sur les instruments chanteurs s'est développée depuis une douzaine d'années aboutissant à plusieurs systèmes qui permettent de jouer des voix solistes expressives et de belle qualité. Ces instruments étendent la palette sonore des instruments numériques.

La synthèse du chant a très tôt été employée dans divers styles de musique. La voix "robotique" est un lieu commun dans les musiques pop/rock ou électroniques, depuis les vocodeurs utilisés par les groupes des années 1970 (par exemple le groupe allemand Kraftwerk), jusqu'aux avatars chanteurs comme Atsune Miku (Yamaha), en passant par l'effet Auto-Tune (Antares), ingrédient indispensable et marqueur stylistique de divers genres musicaux. Le cinéma a fait appel à la voix synthétique ou transformée par analyse-synthèse, pour incarner des robots (« 2001, Une odyssée de l'espace », Stanley Kubrick, 1968), ou pour recréer des voix aux possibilités très étendues (voix de castrat dans « Farinelli », Gérard Corbiau, 1994). De façon plus créative ou expérimentale, toutes les sortes de synthèse vocale ont été employées dans la musique contemporaine, instrumentale ou électroacoustique.

Une voix déplacée

La voix n'est pas un « instrument » de musique, au sens d'un artefact mis en vibration par les membres ou par le souffle. Les organes vocaux sont internes, en grande partie invisibles, et contrôlés de façon complexe par plusieurs ensembles musculaires (respiration, phonation, articulation). Le contrôle vocal est donc par nature intéroceptif, alors qu'il est davantage kinesthésique et extéroceptif pour les instruments de musique. Ce contrôle passe par l'oreille, mais la modalité visuelle ne joue aucun rôle pour la voix.

L'avènement de la synthèse numérique a permis pour la première fois le rendu d'un son indéniablement vocal par un dispositif instrumental externe, déplacé et mis à distance de l'appareil vocal. Il y a une coupure entre le sujet et sa voix ou la voix. Les « instruments chanteurs » sont « manœuvrés » par les mains, les pieds, par le truchement de toutes sortes de capteurs ou d'interfaces humain-machine. Cette mise à distance pose la question du contrôle du chant dans des termes tout à fait différents de ceux du contrôle d'un instrument acoustique. Il faut en effet s'abstraire de l'espace physique, utiliser des espaces intermédiaires et mettre en place des espaces de contrôle [1]. Les avancées sont suffisantes pour présenter aujourd'hui des instruments chanteurs effectivement joués en concert, qui ouvrent également des perspectives pédagogiques, voire thérapeutiques.

Une triple tâche sémiotique

Le chant est le point de rencontre entre organe vocal, langage, musique et expression. Le langage et la musique utilisent des systèmes de signes (grammaire et solfège), culturellement déterminés, qui peuvent aussi exister en l'absence de la voix (musique instrumentale, langage gestuel). Ces signes et leurs structures sont des moyens de représentation de ce qui est dit ou joué, ils portent la signification du texte et de la partition.

Mais la voix seule, événement vocal pur, exprimée sans parole ou sans référence à un système musical est déjà du chant. La musicalité, ou l'expression, est proche du geste et du mouvement corporel, en suivant la célèbre formule « la musique est la science des mouvements bien ordonnés » (*musica est scientia bene movendi*) [2], ce qui reflète bien l'étymologie d'« émotion ». La variation expressive qui peut être symbolique (attitudes, affects typiques) ou non, résiste à la description structurale, au sens linguistique ou musical du terme. L'expression, toujours située et incarnée, relève de l'interprétation (herméneutique), des moyens de séduire ou convaincre (rhétorique), en bref de la manière de dire ou de jouer. L'expression est essentielle en situation de communication pour donner le sens musical, au-delà de la signification du texte. Les gestes du chant effectuent donc plusieurs tâches simultanément, suivant la triple contrainte linguistique, solfégique et expressive :

Tâche linguistique : L'unité minimale pour rendre le chant intelligible, semble être la syllabe. En effet, le phonème est une abstraction linguistique, mais la réalisation effective d'un geste vocal est toujours une syllabe. Sur et par la syllabe peuvent se construire la double articulation phonème/mot, la prosodie, et les niveaux linguistiques supérieurs. La contrainte linguistique est donc de produire des syllabes intelligibles, éventuellement par combinaisons de phonèmes.

Tâche solfégique : Un solfège, explicite et formalisé ou non, existe pour la plupart des langages musicaux traditionnels. L'unité minimale est alors la note, soit une hauteur mélodique (suivant une échelle mélodique), une valeur de durée, éventuellement des nuances d'intensité ou de timbre. La contrainte solfégique est donc de produire des hauteurs précises avec des durées contrôlées.

Tâche expressive : L'expressivité relève principalement de la variation du phonostyle, par rapport aux contraintes linguistique et musicale. Cette variation se manifeste à toutes les échelles (micro-mélodie, ornements, autant que mouvement sur une phrase entière), en particulier dans la prosodie et le timbre vocal.

Il faut nuancer ce tableau, car le chant n'a pas forcément besoin d'être intelligible : scat, glossolalies, langues inconnes ou imaginaires, proto-langues comme dans le cas du « *human beatbox* » qui est constitué d'un lexique de symboles sonores. De même, un solfège n'est nullement obligatoire, en distinguant les « musiques de notes » qui s'appuient sur un solfège, des « musiques de son » qui peuvent volontiers s'en passer.

Cependant, afin de pouvoir évaluer les résultats obtenus, il nous a semblé important dans le développement de nos recherches sur les instruments chanteurs de maintenir un paradigme d'imitation, en étant capable de jouer des notes de musiques justes et des paroles intelligibles. L'imitation est la meilleure épreuve de la qualité sonore et de la finesse de contrôle. Mais bien sûr, pour la création musicale, le modèle peut être étendu en offrant de nouvelles possibilités de contrôle et des gestes et des sons qui se démarquent de l'imitation vocale, ne sont plus contraints par l'intelligibilité ou le solfège.

Le geste vocal

Dans l'espace physique, le geste vocal comprend des gestes respiratoires (thorax, abdomen), des gestes laryngés, des gestes articulatoires (lèvres, voile du palais, langue, mâchoires). Certains gestes co-verbaux sont sans effet sonore, comme les gestes oro-faciaux non articulatoires (yeux, sourcils), les gestes des membres et du corps. Les gestes produisant du son se divisent en :

Geste articulatoire : Le geste articulatoire correspond à une fermeture/ouverture plus ou moins complète dans le conduit vocal, depuis la glotte jusqu'aux lèvres. L'intéroception de ce geste est une contraction, ou un mouvement des articulatoires, coordonné avec le contrôle du souffle (par contraction abdominale ou thoracique) et l'activité glottique. Le geste articulatoire est rythmé par l'oscillation mandibulaire.

Geste phonatoire : Le geste intonatif, mélodique ou phonatoire, correspond à l'abduction/adduction et à la tension du larynx, mis en jeu par un ensemble complexe de muscles et cartilages coordonnés avec la respiration. L'intéroception de ce geste est une contraction laryngée, ainsi que le contrôle du souffle par contraction abdominale ou thoracique.

Le rythme vocal est construit sur la syllabe, qui correspond au geste articulatoire d'ouverture/fermeture de la mandibule, et sur le geste phonatoire et respiratoire. Le geste expressif est également porté par le geste articulatoire et/ou le geste phonatoire : il correspond à la conduite et à la variation des mouvements articulatoires et laryngés. La plupart des instruments chanteurs ne prennent pas en compte la respiration, à l'exception de SkeezeVox [3] qui utilise un soufflet d'accordéon.

Cantor Digitalis

Cantor Digitalis est un instrument chanteur construit sur un synthétiseur vocal paramétrique [4]. Les gestes de contrôle sur une tablette graphique utilisent les deux mains : main préférée avec un stylet, et la main non-préférée avec les doigts. La synthèse paramétrique permet de calculer les formants, l'intonation, et un modèle de la source vocale. L'instrument produit des voyelles (voisées ou non) et toutes sortes de sons dérivés du modèle vocalique, associés à de multiples gestes expressifs. La figure 1 présente le schéma de principe de cet instrument, ainsi que son contrôle. Les deux panneaux du haut montrent les interfaces de contrôle gestuel, ainsi que l'interface graphique sur un ordinateur. À partir de ces contrôles, des paramètres acoustiques sont calculés par règles. Ces paramètres pilotent le synthétiseur à formant (panneau du bas) qui calcule le signal acoustique.

Écriture mélodique

La chironomie, ou externalisation par les mouvements de la main du geste phonatoire passe par la construction d'un espace perceptif, l'espace mélodique. Cet espace perceptif, à une dimension, peut se déployer dans un espace à trois dimensions (geste libre), à deux dimensions ou à une dimension. Le geste chironomique (spatial) diffère beaucoup du geste laryngé (tension glottique).

Le geste d'écriture semble un excellent compromis entre liberté expressive et facilité d'apprentissage [5,6] : le clavier (même MPE, *Multi Polyphonic Expression* [7]) favorisant les notes discrètes, et la précision étant difficile avec le geste libre du type Theremin [8].

Dans le Cantor Digitalis c'est l'image d'un clavier qui représente la hauteur mélodique (voir l'architecture de l'instrument sur la figure 1). Le geste d'écriture semble particulièrement facile à utiliser pour le contrôle mélodique. En effet, c'est un geste acquis dès l'enfance, à l'aide de la main préférée. La justesse et la précision du contrôle mélodique par un stylet sont parfois meilleures qu'avec la voix [9]. L'audition, la vision, la kinesthésie coopèrent dans cette tâche gestuelle de visée de cible, alors que le chant naturel ne peut compter que sur l'intéroception et l'audition [10].

Le geste vocalique

Le contrôle physique des voyelles passe par la conformation du conduit vocal, forme qui dépend de la position des articulatoires. Pour le contrôle temps-réel, les voyelles peuvent se répartir dans un espace perceptif décrit par deux dimensions acoustiques, le triangle vocalique, défini par la fréquence centrale des deux premiers formants spectraux. C'est ce mode de contrôle qui est proposé dans le Cantor Digitalis à l'aide de la main non-préférée (voir Figure 1, en haut à gauche). Les transitions entre voyelles sont assez réalistes, même si l'espace est acoustique plutôt qu'articulatoire. Un ensemble de voyelles cibles permet de reconstituer par interpolation tout l'espace vocalique. Si on le souhaite, les deux dimensions de l'espace vocalique peuvent se projeter sur un seul axe de la tablette, en fixant un parcours particulier des voyelles dans le plan.

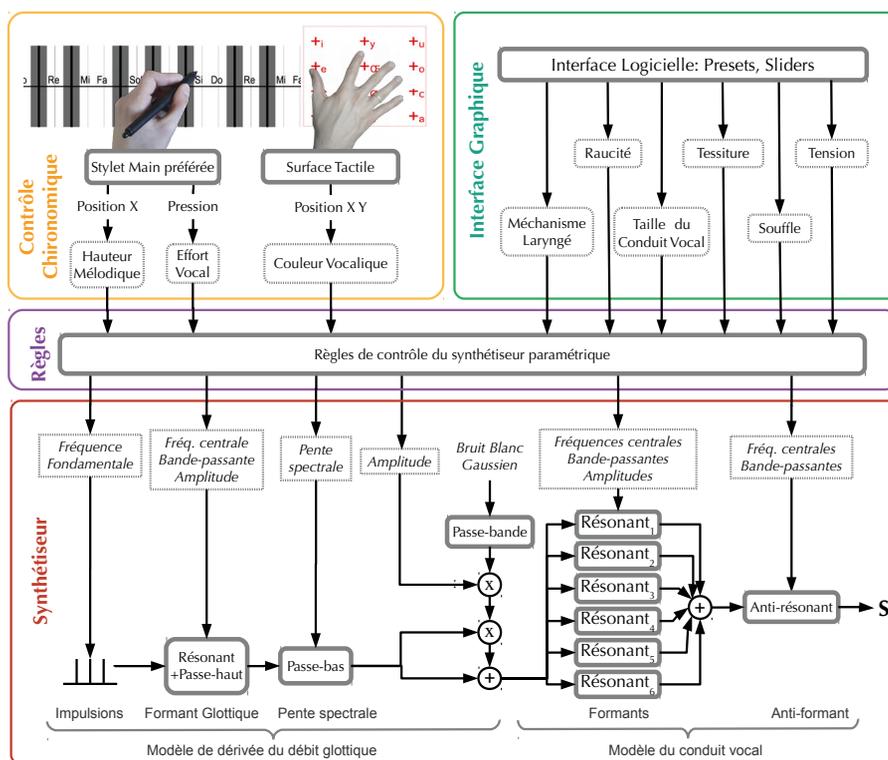


Fig. 1: Cantor Digitalis. Principe du contrôle bimanuel et interface de contrôle (en haut). Schéma de principe du synthétiseur paramétrique (en bas)



Fig. 2 : Choral à quatre voix harmonisé par Jean-Sébastien Bach. Chanté sur les voyelles /a/ et /u/

La qualité vocale

La variation de qualité vocale peut être réduite à deux dimensions principales, l'effort et la tension. L'effort vocal se manifeste par la pression subglottique, la résistance glottique, la contraction supraglottique. La tension et le relâchement de la voix résultent de gestes d'adduction/abduction des plis vocaux. En passant par un espace perceptif, le contrôle gestuel de l'effort ou de la tension peut être associé à une dimension spatiale (par exemple la dimension verticale sur la tablette) ou bien, de façon plus analogue au geste physique, à la pression sur le stylet. Un autre aspect de la qualité vocale est l'allongement du conduit vocal, dont la perception peut être contrôlée dans un espace à une dimension, par exemple sur un des axes de la tablette graphique.

L'expression : un exemple choral

Les figures 2 et 3 montrent la partition d'un choral à quatre voix, chanté par le Chorus Digitalis (un chœur de Cantor Digitalis) et les tracés graphiques des quatre voix. Le choral est chanté sur les voyelles /a/ et /u/, représentées sur l'axe Y, avec le temps sur l'axe X et la hauteur mélodique sur l'axe Z. La première phrase du choral est reprise en /a/, puis en /u/ ce qui correspond aux deux premiers ensembles, puis la seconde phrase alterne des /a/ et /u/ et continue sur /u/, ce qui donne le troisième ensemble de notes. Le style gestuel (et donc l'expression) de chaque musicien et musicienne ressort nettement des tracés de la figure 3, alors que les contenus mélodiques et rythmiques sont comparables pour les quatre voix. On observera : le vibrato, la forme et la netteté des transitions entre notes, la justesse, la simplicité ou la complexité des gestes.

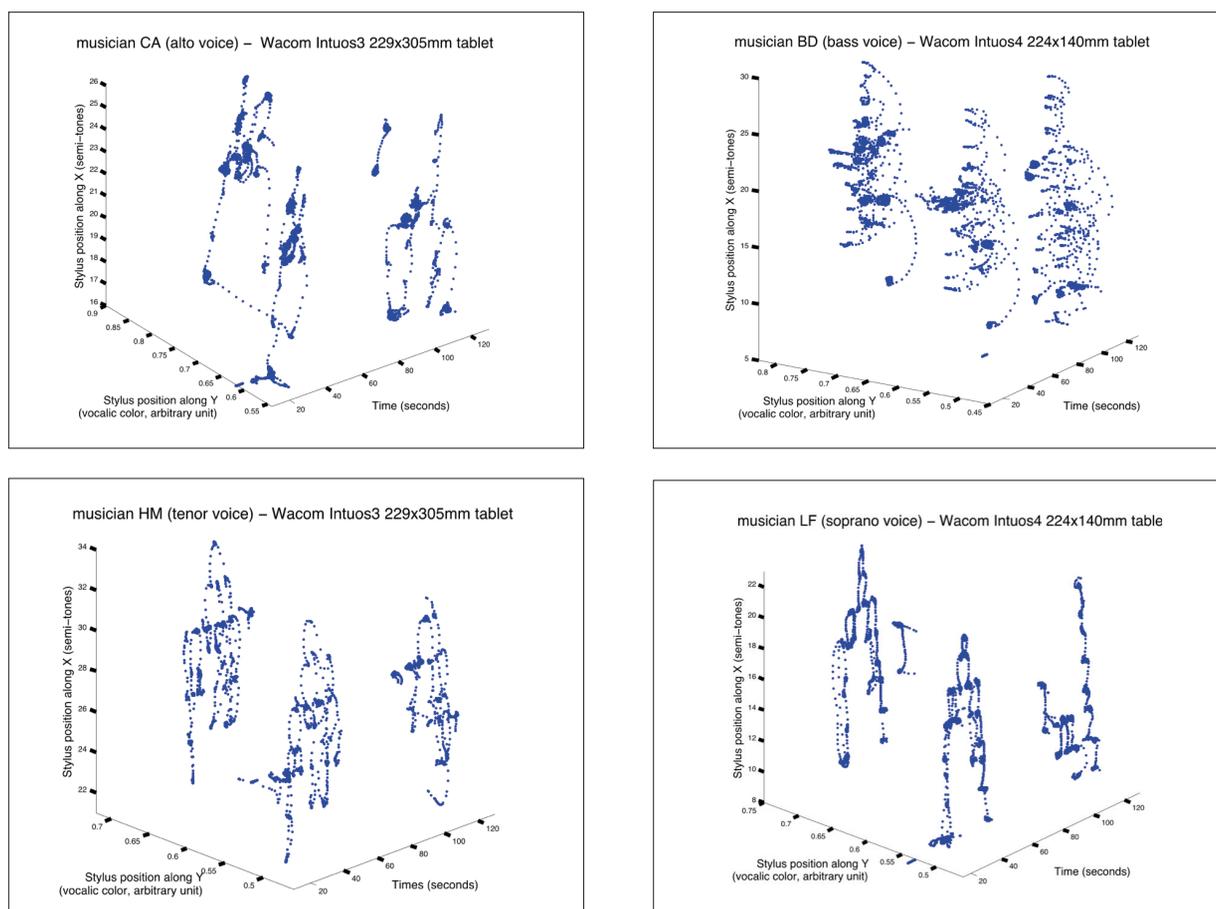


Fig. 3 : Traces des quatre voix chantant le choral. Le temps en abscisse, la fréquence mélodique en cote, la position de la voyelle en ordonnée

Cependant, il faut noter que l'effort vocal n'est pas rendu sur la figure, ce qui serait nécessaire pour préciser l'expression.

Digitartic

La durée du geste consonantique est en général plus brève que celle du geste vocalique dans la syllabe. Les dimensions de l'espace consonantique sont difficiles à évaluer, car les consonnes sont produites par des moyens très divers. On distingue au moins les dimensions de lieu d'articulation (labial, dental, palatal, vélaire etc.), de mode d'articulation (plosif, fricatif, nasal, liquide), de voisement (voisé, non voisé). En plus de ces dimensions intervient la coarticulation avec les voyelles, soit deux dimensions supplémentaires, et les différentes phases temporelles pour certaines consonnes (phases très courtes, quelques dizaines de millisecondes voire quelques millisecondes).

Le contrôle direct du geste consonantique pour tous les phonèmes semble problématique, sinon impossible. Il a été tenté pour la synthèse de parole, mais sans parvenir à une qualité suffisante, ni même à une intelligibilité acceptable [11]. Une extension du Cantor Digitalis, l'instrument Digitartic (voir l'architecture sur la Figure 4) est un synthétiseur paramétrique par règles [12]. Il permet le contrôle gestuel direct d'un sous-ensemble de consonnes.

Il est possible d'articuler un ensemble de syllabes avec une voyelle et une consonne initiale ou finale, mais pas de restituer un énoncé tant soit peu complexe. À l'usage, l'instrument s'est révélé difficile à maîtriser, le rendu de consonnes intelligibles étant difficile.

L'instrument Digitartic a montré les limites de la synthèse paramétrique pour la parole articulée, lorsque l'on cherche à contrôler de façon synchrone les variations rapides et détaillées de nombreux paramètres. D'une part l'espace des paramètres consonantiques a beaucoup de dimensions (GloveTalk utilisait les cinq doigts, le geste manuel en 3D et les pieds afin de multiplier les dimensions contrôlées [11]), d'autre part la dynamique des mouvements consonantiques est beaucoup plus rapide que celle du geste d'écriture.

Vokinesis

L'organisation du geste syllabique

Pour s'affranchir des problèmes de contrôle de la synthèse paramétrique, la synthèse par échantillonnage consiste non plus à calculer les paramètres d'un modèle de production du signal, mais à modifier des échantillons préenregistrés de signal, des phrases ou des textes. Les échantillons peuvent également provenir d'un synthétiseur à partir du texte qui calcule le signal au préalable, avant qu'il ne soit joué.

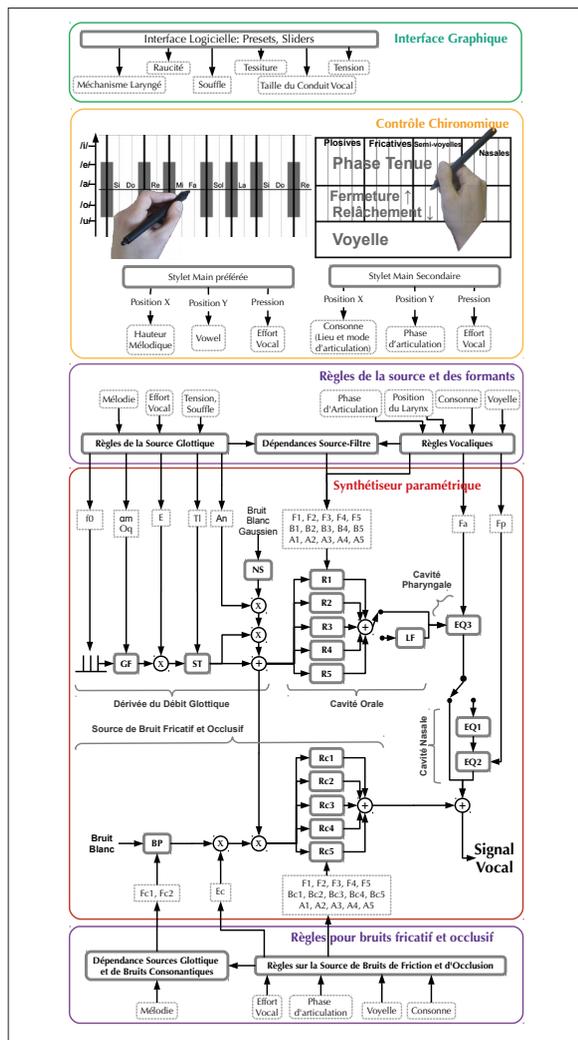


Fig. 4 : Digitartic. Principe du contrôle gestuel (en haut) et schéma du synthétiseur paramétrique (en bas)

Des systèmes de contrôle en temps-réel de la fréquence fondamentale et du rythme ont été développés et évalués. Les résultats montrent un contrôle mélodique excellent puisque pour l'imitation mélodique, le geste est perceptivement équivalent à la voix [13].

Le rythme de séquences vocales préenregistrées peut être contrôlé simplement en variant la vitesse de lecture, ou le débit. Il ne s'agit pas d'un contrôle à proprement parler rythmique, mais plutôt d'un contrôle du défilement du temps. Une autre façon de contrôler le temps est de piloter le défilement des échantillons (le pointeur temporel), ce qui permet aussi de lire le signal à l'envers ou de jouer à volonté n'importe quelle partie de la séquence. Ici le rythme produit correspond au mouvement rythmique du pointeur, mais pas forcément au rythme syllabique.

Contrôle du rythme syllabique

La syllabe est l'unité prosodique de base pour contrôler le rythme. Cela se reflète par exemple dans la poésie, dont la rythmique est construite sur le nombre de syllabes dans un vers. L'organisation du geste syllabique a été étudié par phonologie articulaire. La syllabe comprend

trois éléments : l'attaque (optionnelle, constituée d'une ou plusieurs consonnes), le noyau vocalique (obligatoire), et la coda (optionnelle, constituée d'une ou plusieurs consonnes). En suivant la théorie *frame-content* [14], le mouvement syllabique est constitué d'un cadre vocalique (frame) qui va de voyelle à voyelle, et qui correspond aux mouvements articulaires de grande amplitude (ouverture de la mandibule, avancée de la langue) et d'un contenu, qui correspond aux mouvements consonantiques plus rapides qui s'inscrivent dans le cadre.

Le geste syllabique implique donc au moins deux grands sous-ensembles, les gestes vocaliques et les gestes consonantiques, qui forment le mouvement rythmique des syllabes. Notons que cette théorie est proche de la description classique du rythme comme successions d'Arsis (levée du pied) et Thesis (pose du pied). Pour le contrôle du rythme syllabique, il faut au préalable repérer la position des phonèmes dans le signal, afin de déterminer les frontières syllabiques. Jouer un rythme revient à viser des cibles pour la levée et la pose des syllabes. Ce mouvement comprend deux phases pour chaque syllabe de la même façon qu'une note de musique comprend deux phases : la phase d'attaque (ou *Onset*) et la phase de relâchement (ou *Offset*). La première phase est « l'ouverture » de la syllabe, entre le point consonantique le plus fermé et le point de la voyelle correspondant à la pleine ouverture de la mandibule : cette phase inclut l'attaque et le début du noyau vocalique, c'est la posée de la syllabe ou Thesis. La seconde phase, de fermeture, va de l'ouverture de la mandibule dans le noyau vocalique à la fermeture dans la coda, c'est la transition, la levée de la syllabe ou Arsis. Les deux phases peuvent être contrôlées par divers dispositifs, un bouton *On/Off* ou des contrôleurs continus, comme des potentiomètres ou de pédales. Alors que dans l'instrument Digitartic tout le détail de la réalisation acoustique des phonèmes est contrôlé directement, dans l'instrument Vokinesis [15], seul le rythme syllabique est contrôlé, les détails acoustiques étant préenregistrés sous forme d'échantillons.

Les gestes utilisés sont semblables ou analogues aux gestes de battue rythmique avec les mains ou avec les pieds. De façon analogique entre les mouvements rythmiques des membres reproduisent ceux de la mandibule. Les gestes d'écriture sur une tablette graphique sont utilisés pour le contrôle mélodique. L'architecture du système Vokinesis [15] est décrite dans la figure 5, page suivante.

Conclusion : qui chante?

Le contrôle de la synthèse numérique de la voix se démarque de celui des instruments de musique. Trois niveaux de transposition gestuelle se dessinent : au premier niveau, celui du piano numérique ou du saxophone MIDI, l'interface de contrôle de l'instrument acoustique est conservée ; au second niveau, celui de la modélisation physique des instruments monophoniques, comme les flûtes, les gestes de contrôle peuvent être transposés sur des contrôleurs MPE [7] (EigenHarp, Haken Continuum, Seaboard Roli, LinnStrument etc.) ; au troisième niveau, celui de la voix, des gestes linguistiques s'ajoutent aux gestes solfégiques et expressifs.

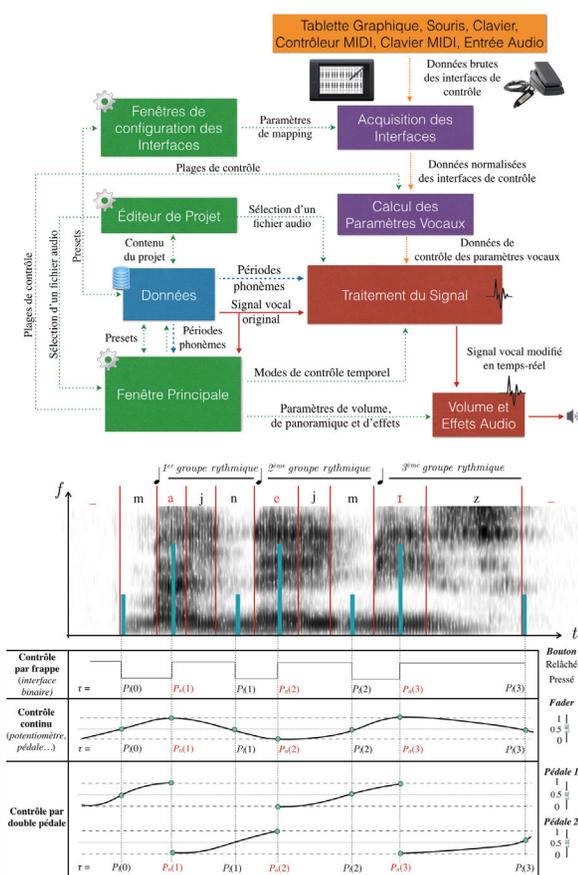


Fig. 5: Vokinesis (en haut). Principe du contrôle rythmique des syllabes (en bas)

L'externalisation des gestes de contrôle vocal fait appel à plusieurs niveaux d'abstraction, résumés dans le tableau 1. Le geste de contrôle peut être analogue au geste vocal : dans le cas du contrôle rythmique, la transposition manuelle ou pédestre du mouvement oscillatoire de la mandibule fonctionne bien, est facile à apprendre, et permet une grande précision ; la résistance du larynx et la pression subglottique, responsables de l'effort vocal, sont bien rendues par la pression du stylet sur la tablette.

Pour les gestes mélodiques, un niveau d'abstraction supplémentaire est nécessaire. Il faut passer par l'espace perceptif des hauteurs musicales, et projeter cet espace sur, par exemple, une surface. D'autres gestes, comme l'allongement du conduit vocal passent par un espace perceptif pour être rendus par des gestes qui n'ont que peu d'analogie avec le geste physique.

Enfin les gestes articulatoires, surtout les consonnes, semblent beaucoup trop rapides pour être contrôlés dans le détail. Il est possible de contrôler la lecture des échantillons, mais difficile actuellement de réaliser à volonté tout geste consonantique dans la temporalité de la synthèse.

Les gestes employés utilisent le retour haptique, du stylet sur une surface, des boutons ou des pédales. Comme pour les instruments acoustiques, la résistance opposée par des dispositifs physiques permet de caler les gestes et d'affiner leur précision. Le contrôle des instruments chanteurs est multimodal, impliquant l'ouïe, la vue, le toucher et la kinesthésie. Cette combinaison sensorielle et motrice permet dans certains cas de rendre les instruments chanteurs plus justes et plus précis que la voix. Il est possible d'ajouter une automatisation ou semi-automatisation du contrôle vocal : correction de justesse pour la mélodie [16], durée des transitions voyelles/consonnes par exemple.

Outre une réflexion sur la nature du contrôle gestuel, les instruments chanteurs offrent des possibilités musicales nouvelles. L'instrumentiste devient le chanteur d'une "voix", mais c'est une voix artificielle, qui n'est pas la sienne, aux possibilités extrêmes en ce qui concerne le registre, le timbre, la vitesse voire même le sens du temps. Les instruments autorisent l'imitation, jouer une voix qui ressemble à une voix (au système de reproduction sur haut-parleurs près), et l'extension de la voix. Cette extension provient de paramètres ou de gestes impossibles dans le cas naturel.

Externaliser la voix la désincarne, ou l'incarne différemment. Le statut symbolique de la voix est questionné, voire affecté par la possibilité de contrôler et de produire le son vocal en dehors du corps : corps augmenté, mise en scène de l'expression vocale, double de la voix, jeu de la voix d'un ou d'une autre. Cette coupure des représentations (sociales, corporelles, musicales) ouvre vers le jeu, l'apprentissage, l'analyse, une écoute différente. L'étrangeté de ce double artificiel de la voix peut jouer dans plusieurs sens : il dérange et attire, amuse et rebute.

Les instruments chanteurs sont susceptibles d'applications pédagogiques. L'instrument interne devenu externe, la voix qui s'expose permet l'analyse des pratiques vocales, pour l'éducation ou la rééducation.

	espace corporel	espace perceptif	espace instrumental
phonation	tension plis vocaux	mélodie	stylet, axe ou surface
	résistance glottique, pression	effort vocal	pression stylet
	résistance glottique	tension vocale	pression stylet
articulation	longueur conduit vocal	taille	stylet axe
	forme conduit vocal	voyelle	doigt surface/ stylet axe
	lieu d'articulation	consonne	stylet, surface
	mode d'articulation	consonne	dynamique stylet, espace
rythme	oscillation mandibulaire	arsis/ thesis	bouton/pédales

Tabl.1 : Espaces corporel, perceptif et instrumental pour les instruments chanteurs

Tout comme le renforcement multimodal permet une excellente justesse et précision, il ouvre des perspectives pour l'apprentissage des gestes vocaux en combinant perception et production de traces visuelles, de gestes manuels ou corporels et de sons vocaux.

Références bibliographiques

- [1] D. Arfib, J. Couturier, L. Kessous et V. Verfaillie, V. « Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces » *Organised Sound*, 7(2), 127-144, 2002
- [2] Augustin d'Hippone *De Musica* (Traité de la musique), Livre 1, Chap. III, ca 386-387
- [3] P. Cook et C. Leider « SqueezeVox : A New Controller for Vocal Synthesis Models », *Proc. ICMC 2000*, Berlin, Allemagne, 2000
- [4] L. Feugère, C. d'Alessandro, B. Doval et O. Perrotin, « Cantor Digitalis : Chironomic Parametric Synthesis of Singing », *J. Audio Speech Music Proc.* (2017) 2017: 2. <https://doi.org/10.1186/s13636-016-0098-5>, 2017
- [5] M. Wanderley, J.-P. Viollet, F. Isart et X. Rodet « On the choice of transducer technologies for specific musical functions ». *Proc. ICMC 2000*, Berlin, Allemagne, 2000
- [6] N. D'Alessandro, C. d'Alessandro, S. Le Beux et B. Doval « Real-time CALM Synthesizer New Approaches in Hands-Controlled Voice Synthesis », *Proc. NIME06*, Paris, France, pp. 266-271, 2006
- [7] G. Bevin (2018), « MIDI specifications for Multidimensional Polyphonic Expression (MPE) », <http://expressiveness.org/>, consulté le 21 mars 2018
- [8] C. d'Alessandro, N. D'Alessandro, S. Le Beux, J. Simko, F. Cetin et Hannes Pirker, « The Speech Conductor : Gestural Control of Speech Synthesis » *proc. eNTERFACE 2005*, Mons, Belgium, pp. 52-61, 2005
- [9] C. d'Alessandro, L. Feugère, S. Le Beux, O. Perrotin et A. Rilliard « Drawing melodies : evaluation of chironomic singing synthesis », *J. Acoust. Soc. Am.* 135 (6), 3601-3612, 2014
- [10] O. Perrotin et C. d'Alessandro « Seeing, listening, drawing : interferences between sensorimotor modalities in the use of a tablet musical interface », *ACM Trans. on Applied Perception*, 14(2), 2016
- [11] S. Fels et G. Hinton « Glove-Talk II - a neural-network interface which maps gestures to parallel formant speech synthesizer controls » *IEEE Trans. on Neural Networks*, 9(1):205-212, 1998
- [12] L. Feugère et C. d'Alessandro « Synthèse à partir du geste de la voix chantée : instruments Cantor Digitalis et Digitartic », *Traitement du Signal*, 32 (4), 417-442, 2015
- [13] C. d'Alessandro, A. Rilliard et S. Le Beux, « Chironomic stylization of intonation » *J. Acoust. Soc. Am.*, 129(3), 1594-1604, 2011
- [14] P. F. MacNeilage « The frame/content theory of evolution of speech production » *Behavioral and brain sciences*, 21(04), 499-511, 1998
- [15] S. Delalez et C. d'Alessandro « Vokinesis : syllabic control points for performative singing synthesis », *Proc. NIME 2017*, Copenhagen, Denmark 198-203, 2017
- [16] S. Delalez et C. d'Alessandro « Adjusting the Frame : Biphasic Performative Control of Speech Rhythm », *Proc. INTERSPEECH 2017*, Stockholm, Sweden, , 864-868, 2017
- [17] O. Perrorin et C. d'Alessandro « Dynamic Pitch Warping for Intonation Correction of Digital Musical Instruments », *ACM Trans. on Computer-Human Interaction* 23(3) : article 17, 2016

