

Synthèse de l'imitation vocale et représentation graphique des gestes phonétiques sous-jacents à la description du timbre de la guitare classique

Caroline Traube, Laboratoire d'Acoustique Musicale, Faculté de Musique, Université de Montréal, Canada
caroline.traube@umontreal.ca <http://www.musique.umontreal.ca/396.html>

Nicolas D'Alessandro, TCTS, Faculté Polytechnique de Mons, Ac. Universitaire Wallonie-Bruxelles, Belgique
nicolas.dalessandro@tcts.fpms.ac.be <http://tcts.fpms.ac.be/~dalessandro>

**Proceedings of the Conference on Interdisciplinary Musicology (CIM05)
Actes du Colloque interdisciplinaire de musicologie (CIM05)
Montréal (Québec) Canada, 10-12/03/2005**

Résumé

La guitare est un instrument qui permet la production d'une très vaste palette de timbres. Les guitaristes utilisent un large vocabulaire pour qualifier les timbres qu'ils produisent sur leur instrument. Ces adjectifs font référence aux différents domaines sensoriels : sombre, chocolaté, transparent, métallique, brillant, etc. Dans un contexte pédagogique, la signification de ce vocabulaire ne semble pas toujours être perçue clairement. L'objectif de cette recherche est de concevoir un outil informatique produisant la synthèse de l'imitation vocale ainsi que la représentation graphique des gestes phonétiques sous-jacents à la description du timbre de la guitare classique, en fonction des paramètres du geste instrumental et basé sur des analogies entre les sons de guitare et les sons vocaux. À l'image de la méthode traditionnelle d'enseignement du tabla, qui s'effectue à l'aide d'onomatopées, l'imitation vocale des timbres de guitare pourrait fournir un langage commun aux interprètes de la guitare, complémentaire à l'imagerie mentale qu'ils utilisent couramment.

Introduction

La guitare est un instrument qui permet la production d'une très vaste palette de timbres. Le musicien peut interagir de façon complexe sur le mécanisme de production du son en variant notamment la position du point de pincage de la corde, l'inclinaison de la main et la zone de contact du doigt avec la corde. Les guitaristes utilisent un large vocabulaire pour qualifier les timbres qu'ils produisent sur leur instrument (Traube 2004). Ces adjectifs font référence aux différents domaines sensoriels : sombre, chocolaté, transparent, métallique, brillant, etc. Dans un contexte pédagogique, la signification de ce vocabulaire ne semble pas toujours être perçue clairement. Il a été suggéré qu'un sous-ensemble de ces adjectifs font référence à des gestes phonétiques : ouvert, fermé, rond, ovale, mince, creux, nasal, etc. (Traube & Depalle 2004a). En effet, la phonétique qualifie les voyelles en des termes similaires. Parmi les traits distinctifs des voyelles du français par exemple, on retrouve la labialité (les lèvres sont plus écartées pour le /i/ et plus arrondies pour le /o/), l'aperture (la bouche est plus ouverte pour le /a/ et plus fermée pour le /i/) et la nasalité (Abry 1980).

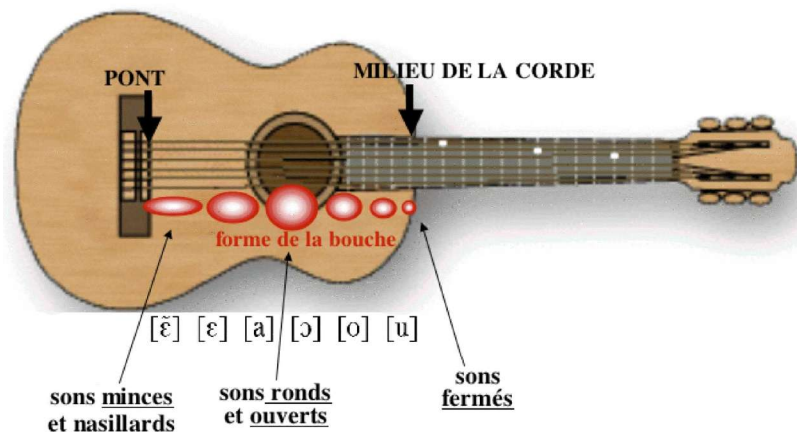


Figure 1. Gestes phonétiques sous-jacents à la description du timbre de la guitare classique et voyelles correspondant à l'imitation vocale de sons de guitare joués à différentes distances du chevalet.

Un son de guitare peut ainsi être imité vocalement par une syllabe constituée d'une consonne pour l'attaque, suivie d'une voyelle pour la partie résonante du son. Le tableau 1 présente la transcription phonétique de l'imitation vocale par 9 participants de sons de guitare correspondant à différents timbres: *ponticello* (joué très près du chevalet), *cuivré* (joué entre le chevalet et la rosace), *rond* (joué près de la rosace) et *tasto* (joué près du manche). Ces analogies vocales ne sont pourtant pas consciemment exploitées par les guitaristes sous la forme d'onomatopées, comme cela se fait pour l'enseignement traditionnel du tabla par exemple.

	Ponticello	Cuivré	Rond	Tasto
# 1	tē	tœ	ta	tø
# 2	tē-ti	d[ē-ā]	bā	bwō
# 3	kē	pā	dɔ	bā
# 4	kē	tε-tō	tɔ	dā
# 5	[k-t]ai	[d-p]aw	da-dɔ	dā
# 6	kē	gœ	tɔ	dø
# 7	dē-kē	t[ā-ō]	dō-tō	gu-du
# 8	kē	tsā-pā	dɔ-tɔ	θō
# 9	kē	tε	ta	bu

Tableau 1. Transcription phonétique de l'imitation vocale par 9 participants de sons de guitare correspondant à différents timbres: *ponticello* – joué très près du chevalet, *cuivré* – joué entre le chevalet et la rosace, *rond* – joué près de la rosace et *tasto* – joué près du manche (Traube & Depalle 2004a).

Des analogies entre des sons produits par une guitare et des sons vocaux peuvent être identifiées dans le domaine spectral. Dans le cas de la voix, la glotte produit un signal harmonique à large bande dont certaines zones du spectre sont favorisées (formants) ou atténuées (anti-formants) par les cavités orale et nasale résonnant de manière préférentielle à certaines fréquences. Dans le cas de la guitare, le spectre d'amplitude d'une corde pincée possède des maxima et des minima à certaines fréquences suivant une enveloppe spectrale semblable à celle des filtres en peigne. On a proposé de considérer les maxima de cette enveloppe comme des zones formantiques (Traube & Depalle 2004b). Considérant une fréquence fondamentale f_0 et un point de pincage relatif R , la fréquence du premier maximum de cette enveloppe spectrale théorique est calculée par:

$$F_1 = f_0 / 2 R$$

Les autres maxima apparaissent à des fréquences qui sont des multiples entiers impairs de F_1 . Ainsi, $F_2 = 3 F_1$, $F_3 = 5 F_1$, $F_4 = 7 F_1$, etc. La figure 2 illustre le cas d'une corde accordée à une fréquence fondamentale de 100 Hz et pincée à un cinquième de sa longueur. L'amplitude des harmoniques dont l'ordre est un multiple entier de 5 est nulle. Le premier maximum apparaît à $F_1 = f_0 / 2 R = 5 f_0 / 2 = 250$ Hz. Les autres maxima apparaissent à 750 Hz, 1250 Hz, etc.

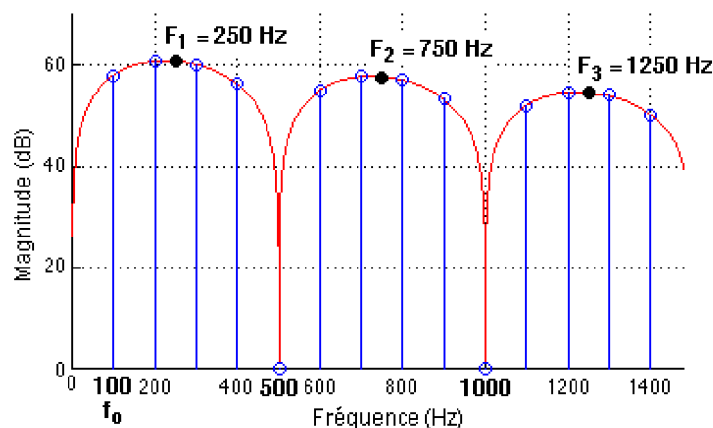


Figure 2. Spectre d'amplitude théorique d'un son de corde idéale accordée à 100 Hz pincée à un cinquième de sa longueur (point de pincage relatif $R = 1/5$).

Ainsi, quand la corde est pincée proche du chevalet, les zones formantiques sont plus espacées, évoquant les voyelles telles que le /i/. Quand la corde est pincée proche du milieu de la corde, les zones formantiques sont plus rapprochées et plus étroites, évoquant des voyelles rondes et plus fermées, telles que le /o/.

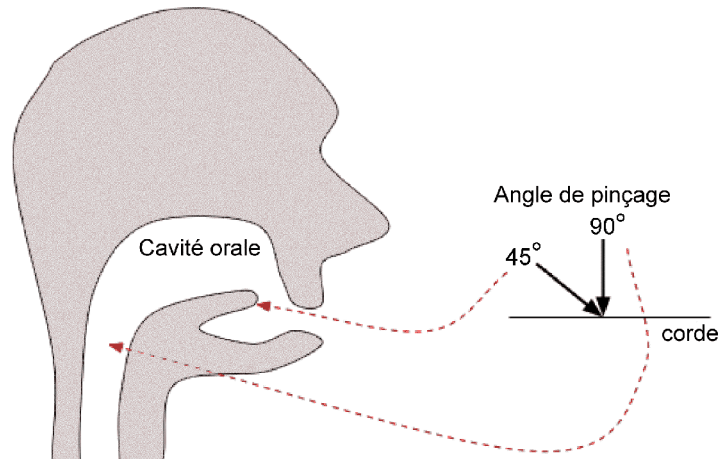


Figure 3. Point d'articulation de la consonne évoquée par l'attaque du son de guitare en fonction de l'angle de pincage.

L'angle du doigt par rapport à la corde modifie quant à lui la nature de l'attaque qui sera perçue plus dure si la corde est pincée perpendiculairement, évoquant la consonne gutturale /k/, et plus molle si le doigt est incliné sur la corde, évoquant la consonne labiale /p/. L'articulation, legato ou staccato, détermine si la consonne est respectivement voisée (ex: /da-ga-da/) ou non-voisée (ex: /ta-ka-ta/).

Fonctionnement de l'application

À partir de paramètres du geste instrumental tels que la position du point de pincage, l'inclinaison de la main droite et l'articulation (legato ou staccato) et de paramètres musicaux tels que la hauteur et la durée de la note (cf. Figure 4), une imitation vocale des sons de guitare (correspondant aux phonèmes perceptivement similaires) est synthétisée par concaténation de diphtonges suivant la technique de synthèse *MBROLA* (Dutoit 1997). La représentation graphique de la forme approximative du conduit vocal est produite en fonction des traits distinctifs d'*ouverture* et d'*acuité*, tels que définis dans Slawson (1985). L'outil graphique fournit également l'évolution en fonction du temps de deux autres traits distinctifs, le *relâchement* et la nature de la consonne (D'Alessandro 2004).

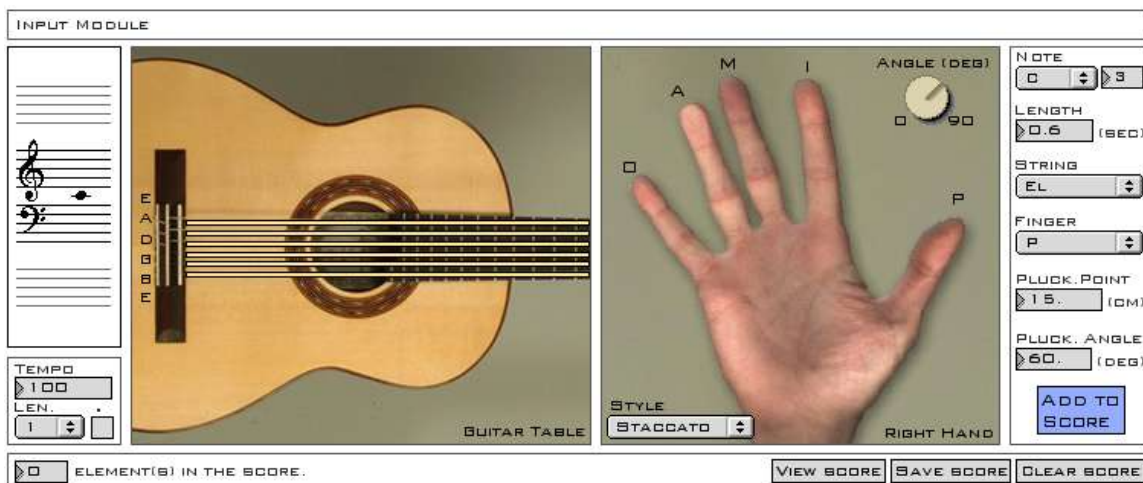


Figure 4. Illustration de l'interface de saisie des paramètres des gestes instrumentaux et de la partition: tempo, hauteur et durée de la note, position sur la table de la guitare (point de pincage et corde), doigt utilisé, orientation de la main et type de jeu. Ces paramètres peuvent ensuite être stockés dans une partition.

La saisie du point de pincage de chaque doigt est remplacée par une saisie plus intuitive en deux étapes. D'une part, l'interprète désigne un point de pincage général sur la table de la guitare. D'autre part, il choisit le doigt qui exécute le geste. Cette présentation nécessite un modèle simple de la main droite afin de calculer les valeurs réelles des points de pincage. Le point de pincage général est considéré comme celui ciblé par l'index.

$$p_p = p - \alpha/45$$

$$p_i = p$$

$$p_m = p - \delta$$

$$p_a = p - 2\delta$$

où p , p_p , p_i , p_m , p_a sont respectivement le point de pincage global, du pouce, de l'index, du majeur et de l'annulaire, α est l'angle d'inclinaison de la main (0° = parallèle aux cordes, 90° = perpendiculaire aux cordes) et δ est l'écartement entre les doigts (entre 1,5 cm et 2 cm).

Calcul des paramètres de l'imitation vocale

Le synthétiseur de parole MBROLA (Dutoit 1997) construit la parole synthétique en concaténant des unités élémentaires de parole, appelées diphones, extraites d'une base de données. Cette base de données est obtenue à partir de la segmentation de longues séquences de parole prononcées par une même personne. Pour attribuer le phonème le plus représentatif d'un timbre de guitare en fonction de la position du point de pincage, on segmente en régions la trajectoire

$$F_2 = 3 F_1$$

correspondant au sous-espace de timbres de la guitare dans un espace de voyelles dont les coordonnées (F_1, F_2) correspondent aux fréquences des premier et second formants (Traube 2004b). Chaque région du sous-espace se verra attribuer le phonème (extrait de la base de données MBROLA) qui lui est le plus proche dans le plan (F_1, F_2) . La figure 5 illustre le positionnement des voyelles utilisées dans (Slawson 1985) et des voyelles de la voix FR_1 de MBROLA par rapport à la trajectoire des timbres de la guitare. Nous constatons que la distribution des voyelles de MBROLA est moins homogène mais cependant cohérente (sur le plan perceptif) par rapport à la distribution utilisée dans (Slawson 1985).

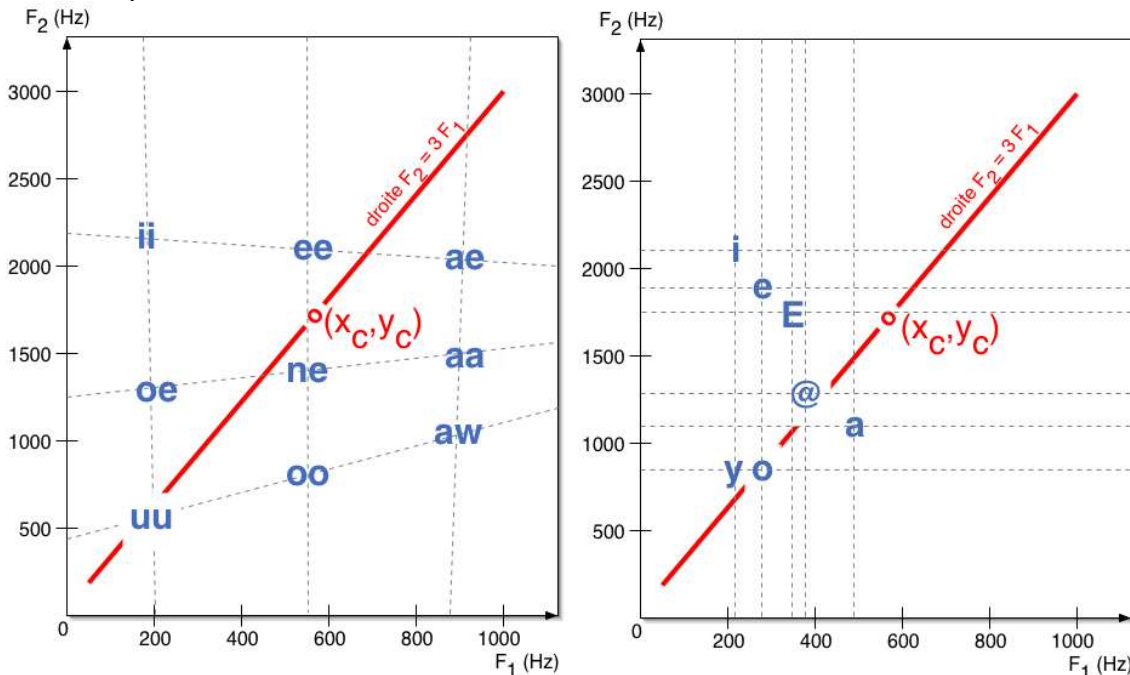


Figure 5. Comparaison entre le positionnement des voyelles dans le plan (F_1, F_2) utilisées dans (Slawson 1985) (à gauche) et des voyelles de la voix FR_1 de MBROLA (à droite) par rapport à la trajectoire des timbres de la guitare $F_2 = 3 F_1$.

Cette constatation est mise en évidence par la représentation du *triangle des voyelles* dans les deux situations, comme l'illustre la figure 6.

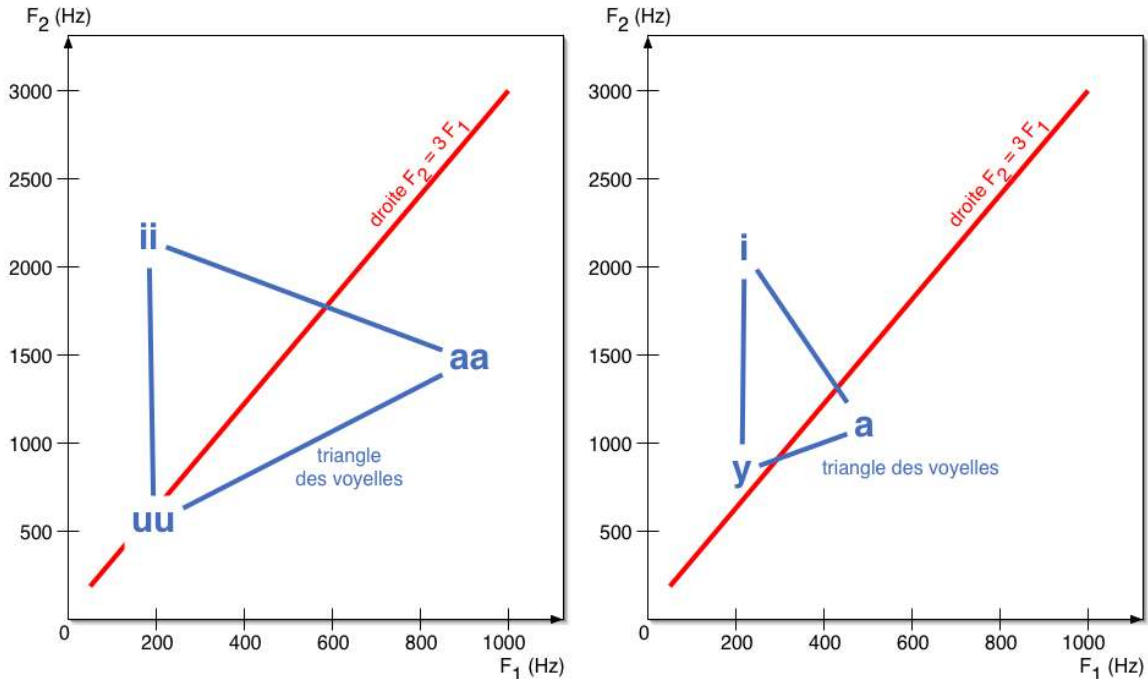


Figure 6. Comparaison entre le positionnement du *triangle des voyelles* dans le plan (F_1, F_2) utilisées dans (Slawson 1985) (à gauche) et du *triangle des voyelles* de la voix FR_1 de MBROLA (à droite) par rapport à la trajectoire des timbres de la guitare $F_2 = 3 F_1$.

Le continuum d'angles d'inclinaison de la main droite est également segmenté afin de déterminer la consonne adéquate. L'association d'un nombre fini de consonnes au continuum d'inclinaisons se base sur l'analyse des résultats de l'étude rapportée dans (Traube & Depalle 2004a) où les sujets devaient fournir des imitations vocales de sons de guitare sous la forme d'onomatopées (cf. tableau 1). On adopte ici, pour un angle variant de 0 à 90°, la suite des consonnes voisées /b, d, g/ (de la plus molle à la plus dure) pour le jeu *legato* et la suite des consonnes non-voisées /p, t, k/ pour le jeu *staccato*.

Calcul des paramètres de la représentation graphique

La représentation graphique évoquant les gestes phonétiques sous-jacents à la description du timbre de la guitare est obtenue à partir du paramétrage de traits distinctifs de la parole que sont l'*ouverture*, l'*acuité* et le *relâchement*.

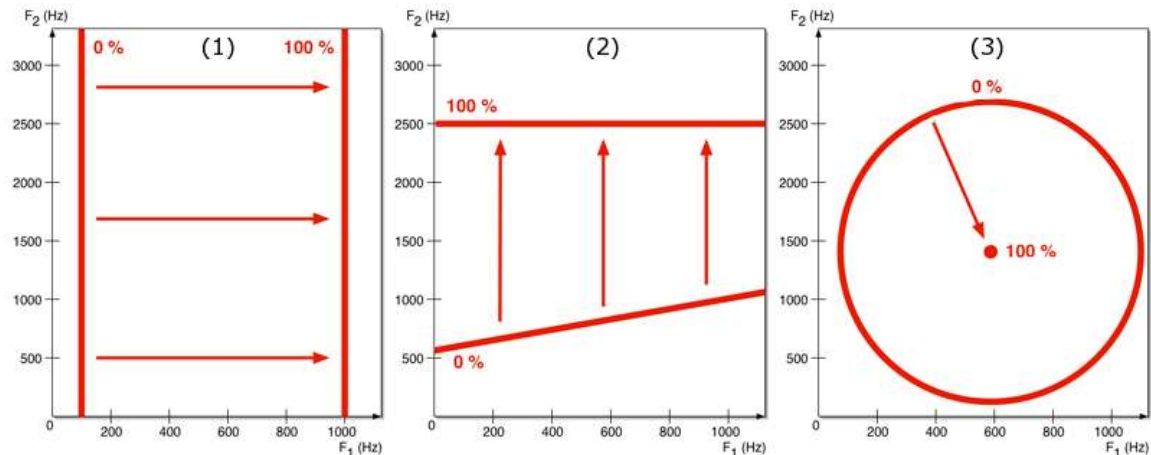


Figure 7. Paramétrage du plan (F_1, F_2) afin d'extraire une valeur en pourcents pour l'ouverture (1), l'acuité (2) et le relâchement (3).

La démarche consiste ici à paramétrer le plan (F_1, F_2) de manière à extraire des valeurs numériques (en pour cent) de ces trois traits distinctifs et ainsi piloter une représentation géométrique de la forme du conduit vocal. Ces paramètres sont calibrés sur le plan de manière à couvrir l'ensemble des possibilités gestuelles et phonétiques. Le paramétrage du plan est illustré à la figure 7.

Le paramètre d'ouverture P_o est directement proportionnel à F_1 .

$$P_o = \frac{1}{9}F_1 - \frac{100}{9}$$

Le paramètre d'acuité P_a est principalement influencé par F_2 .

$$P_a = \frac{100}{2500 - F_{2_{\min}}(F_1)}F_2 - \frac{250000}{2500 - F_{2_{\min}}(F_1)} + 100$$

$$\text{avec } F_{2_{\min}}(F_1) = \frac{4}{9}F_1 + \frac{5000}{9}$$

Le paramètre de relâchement P_r est conditionné par la distance entre la voyelle considérée et la *voyelle neutre* (Slawson 1985). Le point (600, 1400) est pris comme voyelle neutre (relâchement = 100%) et le point (100, 300) est pris comme voyelle la plus tendue (relâchement = 0%).

$$P_r = 100\left(1 - \frac{d}{d_{\max}}\right)$$

$$\text{avec } d = \sqrt{(F_1 - 600)^2 + (F_2 - 1400)^2}$$

$$\text{et } d_{\max} = \sqrt{(100 - 600)^2 + (300 - 1400)^2} = 1200 \text{ Hz}$$

La forme géométrique à contrôler est réalisée avec deux splines cubiques d'ordre 6 placés symétriquement autour d'un axe vertical. Le paramètre d'ouverture agit sur la dimension générale de la forme (agrandissement). Le paramètre d'acuité agit sur le rapport de la dimension verticale sur la dimension horizontale, créant ainsi un aplatissement de la forme. Le paramètre de relâchement de la voyelle et la dureté de consonne sont représentés par des modifications symboliques, sur la forme géométrique. Ainsi, le relâchement est lié à la couleur de l'ellipse par l'échelle de chaleur (rouge = relâché, bleu = tendu). Le type de consonne influence la forme du contour de l'ellipse (arrondi = consonne molle, brisé = consonne dure).

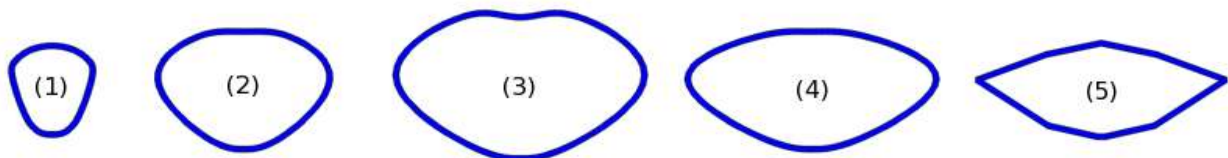


Figure 8. Évolution de la forme géométrique en fonction des paramètres extraits: (1) $P_o = 0\%$, $P_a = 0\%$; (2) $P_o = 50\%$, $P_a = 0\%$; (3) $P_o = 90\%$, $P_a = 0\%$; (4) $P_o = 90\%$, $P_a = 50\%$; (5) $P_o = 90\%$, $P_a = 70\%$ (consonne dure). Dans tous les cas, $P_r = 0\%$ (couleur bleue).

Conclusions et perspectives

Cette recherche se situe à l'intersection de plusieurs sphères de connaissances théoriques et pratiques: l'acoustique, le traitement du signal, la linguistique, la psychologie, l'interprétation et la pédagogie. Elle contribue à la création de nouveaux outils pour l'analyse de phénomènes liés à l'interprétation, tels que la description verbale des nuances de timbres sur un instrument. En effet, dans la plupart des cas, les pédagogues de la guitare abordent la notion de timbre au travers de leur propre vocabulaire issu de l'une ou l'autre école. À l'image de la méthode traditionnelle d'enseignement du tabla, qui s'effectue à l'aide d'onomatopées, l'imitation vocale des timbres de guitare pourrait fournir un langage commun aux interprètes de la guitare, complémentaire à l'imagerie mentale qu'ils utilisent couramment dans un contexte pédagogique.

Références

- Abry, C., L.-J. Boë, P. Corsi, R. Descout, M. Gentil, et P. Graillet, 1980. *Labialité et phonétique*. Université des langues et lettres de Grenoble.
- D'Alessandro, N., 2004. *Applications musicales et pédagogiques utilisant des synthétiseurs de parole et des analogies perceptives entre la guitare et la voix*, Travail de Fin d'Etudes, Faculté Polytechnique de Mons / Faculté de musique de l'Université de Montréal.
- Dutoit, T. 1997. *An Introduction to Text-to-Speech Synthesis*, Kluwer Academic Publishers.
- Traube, C., 2004. *A Interdisciplinary Study of the Timbre of th Classical Guitar*, PhD Thesis, McGill University.
- Traube, C. et P. Depalle, 2004a. Phonetic Gestures underlying Guitar Timbre Description, *Proceedings of the International Conference on Music Perception and Cognition (ICMPC8)*.
- Traube, C. et P. Depalle, 2004b. Timbral Analogies between Vowels and Plucked String Tones, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
- Slawson, W., 1985. *Sound Color*, University of California Press.