VOCAL SYNTHESIS AND GRAPHICAL REPRESENTATION OF THE PHONETIC GESTURES UNDERLYING GUITAR TIMBRE DESCRIPTION

Caroline Traube

Lab. d'informatique, acoustique et musique
Faculté de musique, Université de Montréal
Montréal, Québec, Canada
caroline.traube@umontreal.ca

Nicolas D'Alessandro

TCTS Lab
Faculté Polytechnique de Mons
Ac. Universitaire Wallonie-Bruxelles, Belgium
nicolas.dalessandro@tcts.fpms.ac.be

ABSTRACT

The guitar is an instrument that gives the player great control over timbre. Different plucking techniques involve varying the finger position along the string, the inclination between the finger and the string, the inclination between the hand and the string and the degree of relaxation of the plucking finger. Guitarists perceive subtle variations of these parameters and they have developed a very rich vocabulary to describe the brightness, the colour, the shape and the texture of the sounds they produce on their instrument. *Dark*, *bright*, *chocolatey*, *transparent*, *muddy*, *wooly*, *glassy*, *buttery*, and *metallic* are just a few of those adjectives. The aim of this research is to conceive a computer tool producing the synthesis of the vocal imitation as well as the graphical representation of phonetic gestures underlying the description of the timbre of the classical guitar, as a function of the instrumental gesture parameters (mainly the plucking angle and distance from the bridge) and based on perceptual analogies between guitar and speech sounds. Similarly to the traditional teaching of tabla which uses onomatopeia to designate the different strokes, vocal imitation of guitar timbres could provide a common language to guitar performers, complementary to the mental imagery they commonly use to communicate about timbre, in a pedagogical context for example.

1. BACKGROUND

1.1. Guitar timbre semantics

The guitar is an instrument allowing the production of a vast palette of timbres. The performer can interact in complex ways with the sound production mechanism. The location of the plucking point

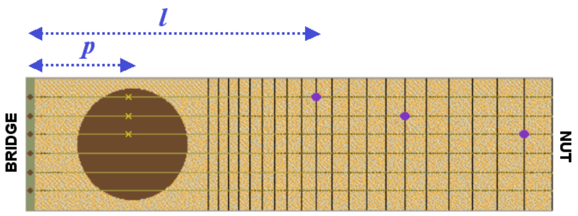

Figure 1: *Plucking point at a distance p from the bridge and fingering point at distance l from the bridge on a guitar neck.*

has a major influence on timbre. Plucking a string close to the bridge produces a tone that is brighter and sharper, generally richer in high-frequency components. When playing near or over the fingerboard, that is closer to the midpoint of the string, the tone is rounder, mellower, less rich in high-frequency components. This correlation between plucking position and brightness, acknowledged by most guitarists, coarsely describes the timbral palette of the instrument. In addition to brightness, guitarists seem to be sensitive to a vocal quality of guitar tones.

Guitarists use a large vocabulary to qualify timbres they produce on their instrument. These adjectives refer to the different sensory domains. Figure 2 presents about 60 of these adjectives, organized into clusters of synonyms [1]. In the lower left-hand corner of the map, sounds are usually obtained by plucking the string closer to its middle. In the upper right-hand corner, sounds are usually obtained by plucking the string closer to the bridge.

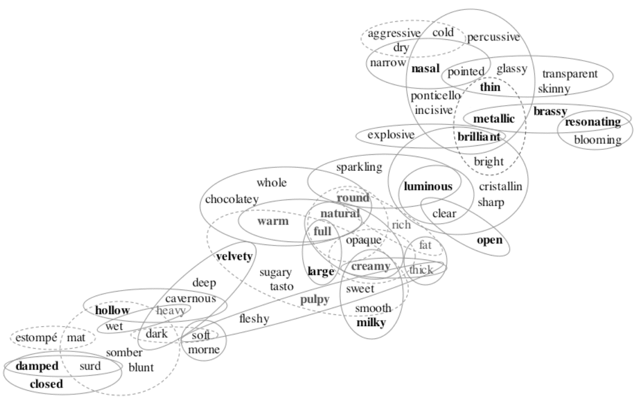

Figure 2: *Guitar timbre qualifying adjectives organized into clusters of synonyms [1].*

1.2. Vocal imitation of guitar tones

In a study (reported in [2]) aiming to investigate perceptual analogies between guitar tones and speech sounds, nine French-speaking musicians from Montreal were asked to vocally imitate guitar tones. The stimuli were recordings of a professional guitarist playing the same melody with four different timbres which were described by the guitarist as *ponticello*, *brassy*, *round* and *tasto* (Figure 3).

The main instrumental gesture parameter that was varied to obtain the different timbres was the plucking position p (from very close to the bridge to over the finger board). The angle of the

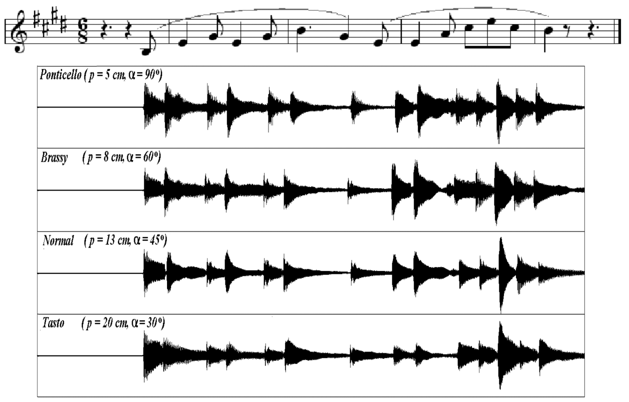

Figure 3: *Time-domain representation of the 14 tones of a melody played with four different timbres. The melody is an excerpt from the piece* L'encouragement *for two guitars by Fernando Sor (1778-1839).*

	Ponticello	Brassy	Round	Tasto
# 1	tɛ̃	tœ̃	ta	tø
# 2	tɛ̃-ti	d[ɛ̃-ɑ̃]	bɑ̃	bwɔ̃
# 3	kɛ̃	pɑ̃	dɔ	bɑ̃
# 4	kɛ̃	tɛ-tɔ̃	tɔ	dɑ̃
# 5	[k-t]ai	[d-p]aw	dɑ-dɔ	dɑ̃
# 6	kɛ̃	gœ̃	tɔ	dø
# 7	dɛ̃-kɛ̃	t[ɑ̃-ɔ̃]	dɔ̃-tɔ̃	gu-du
# 8	kɛ̃	tsɑ̃-pɑ̃	dɔ-tɔ	θɔ̃
# 9	kɛ̃	tɛ	ta	bu

Table 1: *IPA Phonetic transcription of the vocal imitation by 9 participants of guitar tones corresponding to different timbres [2].*

plucking finger also differed, positioned closer to a perpendicular to the strings for brighter timbres. This correlation between the two gesture parameters was necessary to preserve the naturalness of the plucking techniques. The *ponticello* timbre was played 5 cm from the bridge with fingers perpendicular to the strings (thus a 90^o angle between the fingers and the string). The *brassy* timbre was obtained by plucking the string 8 cm from the bridge with a 60^o angle. The *round* timbre was obtained by plucking the string 13 cm (close to the tone-hole) from the bridge with a 45^o angle. The *tasto* timbre was obtained by plucking the string 20 cm for the bridge with a 30^o angle.

Table 1 presents the phonetic transcription of vocal imitations of these four timbres by the nine participants in the study. Generally, guitar tones were vocally imitated by a syllable made of a consonant for the attack, followed by a vowel for the resonating part of the sound.

1.2.1. Plosive consonant as attack

As illustrated on Figure 4, the angle between the finger and the string seems to modify the nature of the perceived attack, which will be harder if the string is plucked perpendicularly, evoking the velar plosive consonant [k], and softer if the finger is inclined on the string, evoking the dental plosive consonant [d]. The voice

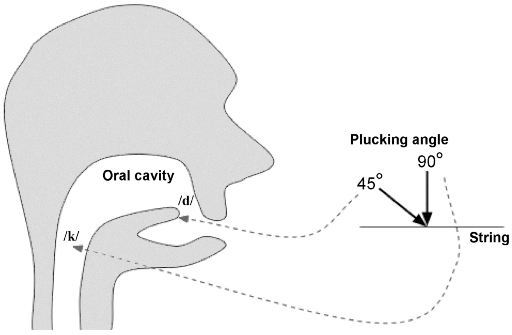

Figure 4: *Articulation point of the consonant evoked by the attack of the guitar tone as a function of plucking angle.*

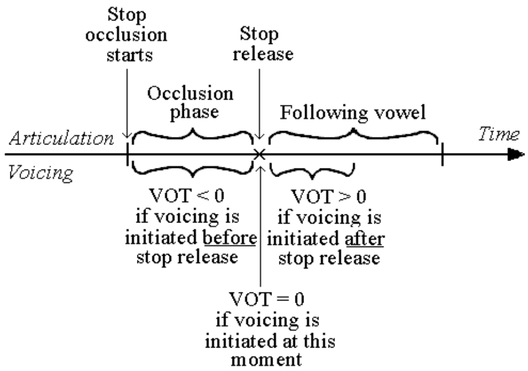

Figure 5: *Voice onset time (VOT) is a parameter of speech that designates the time interval between consonant onset and the onset of periodical vocal folds vibration.*

onset time (VOT - see Figure 5) of plosive consonants depends on the place of articulation and is shorter for bilabial plosive [p] and longer for velar plosive [k]. The greater relative strength of the attack for tones plucked close to the bridge might be responsible for a backward masking effect, evoking a consonant with a longer VOT. Also, VOT of unvoiced plosives [p, t, k] is longer than that of voiced plosives [b, d, g][1]. Hence, articulation (legato or staccato) determines if the consonant is respectively voiced (ex: [da-ga-da]) or unvoiced (ex: [ta-ka-ta].). A staccato articulation evokes unvoiced stops, such as [k], which are characterized by a prolonged VOT clearly interrupting the harmonic glottal source, whereas a legato articulation (where a note resonates until a new note is played) evokes voiced stops, such as [g], which have smaller VOT.

1.2.2. Vowel as resonance

Figure 6 illustrates the overall trend in the choice of vowels for the resonant portion of the guitar tone. For plucking positions ranging from near the bridge to closer to the midpoint of the string, we obtain the sequence [ɛ̃], [ɛ], [a], [ɔ], [o], [u]. The approximate shape of the mouth forming these vowels is also shown on Fig-

[1]As the VOT is increased in incremental steps, perception rapidly changes from a voiced plosive consonant to an unvoiced plosive consonant at an interval of 20-40 ms in American English.

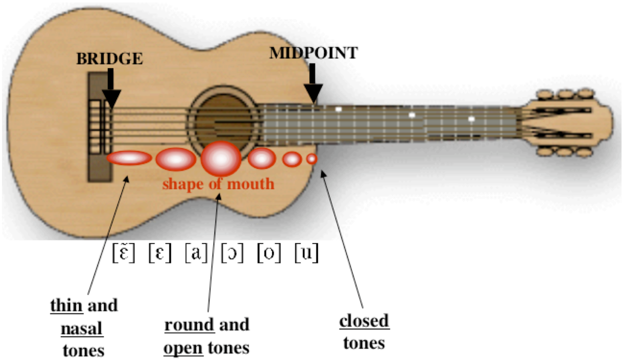

Figure 6: *Phonetic gestures underlying guitar timbre description and vowels corresponding to the vocal imitation of guitar sounds played at different distances from the bridge along the string [1].*

ure 6. When plucked close to the bridge, the string produces a sound that is associated with a thin-shaped mouth. Moving closer to the tone-hole, the mouth seems to open up to a round shape. Then, from the tone-hole to the midpoint of the string, the mouth closes progressively while keeping a more or less round shape. At midpoint, the guitar sound lacks all even harmonics. In fact, perceptually, this sound is generally described as hollow and some guitarists will call it a *bassoon* sound or a *pipe* sound.

1.3. Phonetic gestures underlying guitar timbre description

The results of the experiment reported in Table 1 support the idea that perceptual dimensions of the guitar timbre space can be borrowed from phonetics. It was suggested that a subset of adjectives qualifying the timbre of guitar tones actually refer to phonetic gestures: *open*, *closed*, *open*, *oval*, *thin*, *hollow*, *nasal*, etc. [2]. In fact, phonetics qualifies vowels in similar terms. The distinctive features of French vowels for example include *labiality* (lips are more spread apart for [i] and more rounded for [o]), *aperture* (mouth is more open for [a] and more closed for [u]) and *nasality* [3]. Hence, qualifying a guitar sound as *round* would mean that it sounds like a *round-shaped-mouth* vowel such as [o].

1.4. Analogies at the spectral level

Analogies between guitar tones and vowels can be found in the spectral domain. In the case of voice, the frequency-content of the glottal harmonic source is either enhanced (formants) or attenuated (anti-formants) by the resonances and anti-resonances of the oral and nasal cavities. In the case of the guitar, as shown on Figure 7, the spectral envelope of a ideal plucked-string tone is comb-filter-shaped with its first local maximum F_{C_1} located at half the ratio of the fundamental frequency f_o over the relative plucking position along the string R.

$$F_{C_1} = \frac{f_o}{2R} = \frac{lf_o}{2p} \tag{1}$$

The other local maxima in the magnitude spectrum are odd integer multiples of F_{C_1}. Hence, $F_{C_2} = 3F_{C_1}, F_{C_3} = 5F_{C_1}$, etc. Figure 7 illustrates the case of a string tuned to 100 Hz fundamental frequency and plucked at one fifth of its length. Harmonics whose order is an integer multiple of 5 are cancelled. The first maximum appears at $F_{C_1} = f_o/2R = 5f_o/2 = 250$ Hz. The other maxima appear at 750 Hz, 1250 Hz, etc.

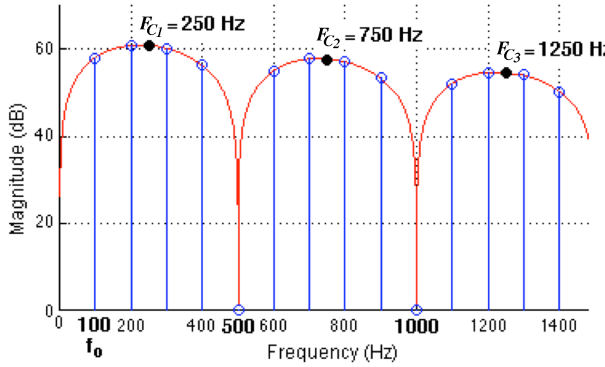

Figure 7: *Theoretical magnitude spectrum of an ideal string tuned to 100 Hz and plucked at one fifth of its length (relative plucking position $R = 1/5$).*

It was proposed to consider the frequencies F_{C_n} of the maxima of this comb-filter envelope as the central frequencies of *comb filter formants* [4], taking the literal definition of a formant, that is a frequency range in which amplitudes of spectral components are enhanced. In fact, when the string is plucked closer to the bridge, these *comb filter formants* are further apart, evoking vowels such as [i]. When the string is plucked closer to the middle of the string, the *comb filter formants* are narrower and closer to each other, evoking rounder and closer vowels such as [o].

1.5. Applying distinctive features of speech to guitar tones

The distinctive feature theory, proposed by Jakobson, Fant and Halle in 1951 and then later revised and refined by Chomsky and Halle in 1968 [5], codifies certain long-standing observations of phoneticians by hypothesizing that many sounds of speech can be placed in categories based on the presence or absence of certain distinctive features: whether the mouth is open, whether there is a narrowing of the vocal tract at a particular place, whether a consonant is aspirated. Jakobson, Fant and Halle detected twelve inherent distinctive features in the languages of the world. In his book *Sound Color*, Slawson selected three of the features related to vowels (*compactness*, *acuteness* and *laxness*) as candidates from which to derive dimensions of sound colour. Figure 8 displays the equal-value contours for these features in a (F_1, F_2) plane, representing the second formant frequency F_2 as a function of the first formant frequency F_1 for various English vowels.

According to Slawson, OPENNESS (replacing the term COMPACTNESS given in [5]) is named for the tube shape with which it is correlated. The approximate acoustic correlate of OPENNESS is the frequency of the first resonance. ACUTENESS reflects its connotation of high or bright sound. It increases with increasing frequency of the second resonance. Finally, LAXNESS is said to correspond to a relatively relaxed state of the articulatory musculature. The equal LAXNESS contours are closed curves on the (F_1, F_2) plane centered on the maximally LAX point (or minimally tense point) point, corresponding to the neutral vowel [ɜ] (**ne**) [2]. This central

[2]Instead of using the International Phonetic Alphabet, Slawson decided to adopt a two-letter convention that he believes to be more evocative of

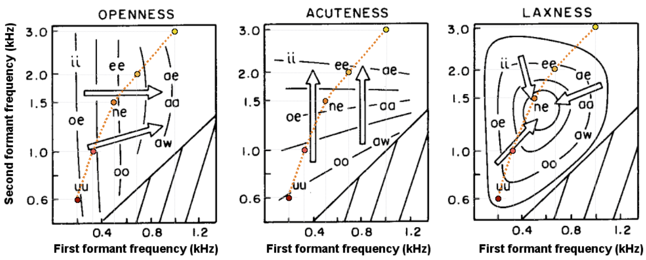

Figure 8: Equal-value contours for three distinctive features of speech in the (F_1, F_2) plane (from [6]) with superimposed *guitar vowels* trajectory (dotted line) corresponding to the relationship $F_2 = 3\,F_1$.

point corresponds to the formant values that would arise, in theory, from the vocal mechanism in the position to which it is automatically brought just before beginning to speak [7].

The trajectories that we plotted with a dotted line on top of Slawson's equal-value contours for distinctive features of speech in Figure 8 correspond to the relationship

$$F_2 = 3\,F_1 \tag{2}$$

which is found for the first two local maxima of a comb filter [3]. In that way, one can see which "vowels" may be obtained by varying the plucking position from the middle of the string (**uu** region) to the bridge (**ee** region or further up depending on fundamental frequency of the string). For a given string, the absolute plucking position p will determine the vowel colour, regardless of the note that is played, since F_1 can also be expressed as the ratio $c/4p$ (speed of sound over 4 times the plucking distance) which does not depend on f_o. As a result, vowel colour is maintained for any note on a given string, except for relative plucking position $R = 1/2$ which is the case of an odd-harmonic only spectrum, perceived as a distinct timbre.

Str. #	First formant frequency for $p = 12$ cm and $l = 60$ cm	Sound color	IPA symbol
6	$(30 \times 83)/12 = 207.5$ Hz	**uu**	u (b*oo*t)
5	$(30 \times 110)/12 = 275$ Hz	**oe**	ø (b*ö*se)
4	$(30 \times 146)/12 = 365$ Hz	**oo**	o (b*oa*t)
3	$(30 \times 202)/12 = 505$ Hz	**ne**	ɜ (th*e*)
2	$(30 \times 248)/12 = 602$ Hz	**ee**	e (b*ai*t)
1	$(30 \times 330)/12 = 825$ Hz	**ae**	æ (b*a*t)

This table gives the first formant frequency calculated with Eq. (1) for the six strings of a guitar tuned with the standard tuning (EADGBE), together with the closest sound colour and corresponding IPA symbol, for a string length l = 60 cm and a plucking position p = 12 cm from the bridge.

Applying the three selected distinctive features of speech to guitar sounds, we could infer that the adjectives *closed*, *round*, *large*, *open* indicate different degrees of OPENNESS. The adjectives *thin* and *round* would be opposites along the ACUTENESS dimension. A *warm* or *chocolatey* sound would likely be associated with the maximally LAX point. In fact, a *warm* sound likely evokes the sound that one makes while exhaling warm air, usually with the vocal tract in a neutral position. Finally, a *hollow* or *cavernous* sound would actually sound like the [u] vowel produced as the mouth forms a hollow cavity.

most English speakers' phonetic intuitions.

[3]The curve $F_2 = 3\,F_1$ is not a straight line because the F_2 axis is not linear.

2. APPLICATION

In this section, we present the computer tool we have developed for the synthesis of the vocal imitation as well as the graphical representation of phonetic gestures underlying the description of the timbre of the classical guitar, as a function of the instrumental gesture parameters and based on perceptual analogies between guitar and speech sounds.

2.1. Graphical interface

As shown on Figure 9, the graphical interface of the computer application, developed in Max/MSP, allows the user to set instrumental gesture parameters such as the chosen string (by clicking on the string on the picture) and finger, the plucking distance from the bridge, the angle between the finger and the string and the articulation (legato or staccato), and of musical parameters such as the tempo, the duration of the note, as well as its pitch (by clicking on the staff).

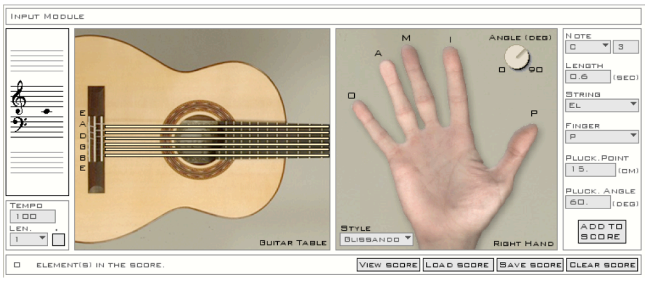

Figure 9: *Graphical interface for the setting of instrumental gesture and score parameters.*

2.2. Text score

All the variable parameters that can be set with the graphical interface can also be stored in a text score, as shown on Figure 10.

In the text score, the data corresponds successively to the note to be played, the octave number, the duration in seconds, the chosen string (from 1 to 6) and finger (i, m or a), the plucking distance from the bridge in cm, the angle (from 0^o to 90^o) and the articulation (0 = glissando, 1 = legato, 2 = staccato).

The plucking positions of all playing fingers (p_t, p_i, p_m and p_a, related to the thumb, index, major and annular fingers respectively) are computed from the plucking angle α (0^o = parallel to strings, 90^o = perpendicular to strings), the spacing between fingers δ (1.5 to 2 cm) and a global plucking position p along the string, which is set to be equal to the plucking position of the index finger p_i.

$$p_t = p - \alpha/45 \tag{3}$$

$$p_i = p \tag{4}$$

$$p_m = p - \delta \tag{5}$$

$$p_a = p - 2\delta \tag{6}$$

```
Musical/Gestural Score:
note  B   3  0.3  3  i  13  45  1
note  E   3  0.6  2  m  13  45  1
note  G#  3  0.3  1  a  13  45  1
note  E   3  0.6  2  i  13  45  1
note  G#  3  0.3  1  m  13  45  1
note  B   4  0.9  1  a  13  45  1
note  G#  3  0.6  1  m  13  45  1
note  E   3  0.3  3  i  13  45  1
note  E   3  0.6  3  m  13  45  1
note  A   4  0.3  2  i  13  45  1
note  C#  4  0.3  1  m  13  45  1
note  E   4  0.3  1  i  13  45  1
note  C#  4  0.3  1  m  13  45  1
note  B   4  0.6  2  i  13  45  1
```

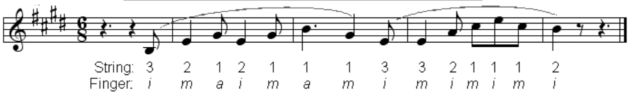

Figure 10: *A possible fingering for this melody and its corresponding text score. The strings are numbered from the highest to the lowest (1=E, 2=B, G=3, D=4, A=5, E=6). Index, major and annular fingers are notated i, m, a respectively.*

2.3. Synthesis of vocal imitation

On the basis of the instrumental gesture and musical parameters specified with the help of the graphical interface or within a text score, a vocal imitation of the guitar tones (corresponding to the phonemes perceptually similar) is synthesized using the MBROLA synthesis technique [8]. The speech synthesizer MBROLA produces the synthetic speech by concatenating elementary units of speech, called diphones extracted from a large database. The diphones are obtained from the segmentation of long sequences of speech pronounced by a person. In this application, the Max/MSP external object MaxMBROLA[4] is used.

2.3.1. Choice of consonant

The continuum of right hand angles is segmented in order to determine the appropriate consonant for the vocal imitation. Three ranges of angles (0^o-30^o, 30^o-60^o and 60^o-90^o) are assigned to the voiced plosives [b, d, g] (from the softest to the hardest) for the legato articulation and unvoiced plosives [p, t, k] for the staccato articulation.

2.3.2. Choice of vowel

Figure 11 illustrates the distribution, in a (F_1, F_2) plane, of the reference vowels used in [6] (Slawson's notation) on the left-side figures and of vowels from the MBROLA FR1 voice (SAMPA notation[5]) on the right-side figures with respect to the guitar timbre $F_2 = 3F_1$ trajectory (top figures). The distribution of MBROLA vowels is less homogeneous but nevertheless perceptually coherent, as illustrated by the vowel triangles in both cases.

Each region of the subspace $F_2 = 3F_1$ will be assigned to the phoneme (extracted from the MBROLA database) which is the closest in the (F_1, F_2) plane. The first comb filter maximum is calculated with

$$Fc_1 = \frac{c}{4p} \tag{7}$$

[4]MaxBROLA is an external object for Max/MSP 4.5 developed par the TCTS Lab (FPMs, Mons, Belgium) which allows the MBROLA synthesis engine to run in real time.

[5]SAMPA = Speech Assessment Method Phonetic Alphabet.

where p is the absolute plucking position, and c is the speed of sound characterizing the chosen string, which can be derived from the fundamental frequency of the open string f_o times twice its length l:

$$c = \frac{2l}{T_o} = 2lf_o \tag{8}$$

Then, the phoneme which is the closest to the point $(Fc_1, Fc_2) = (Fc_1, 3Fc_1)$ in the (F_1, F_2) plane is selected. For example, in the case of the first note of the melody presented on Figure 10, the fundamental of the string on which it is played is the G-string (f_o = 202 Hz). Considering a 60 cm long string, $c = 2lf_o = 2 \times 0.60 \times 202 = 242.4$ m/s and $Fc_1 = \frac{c}{4p} = \frac{242.4}{4\times0.13} = 466.15$ Hz. Then we obtain the point $(Fc_1, 3Fc_1) = (466, 1398)$ on the (F_1, F_2) plane. The closest phoneme is **ne** ([ɜ] in IPA or @ in SAMPA).

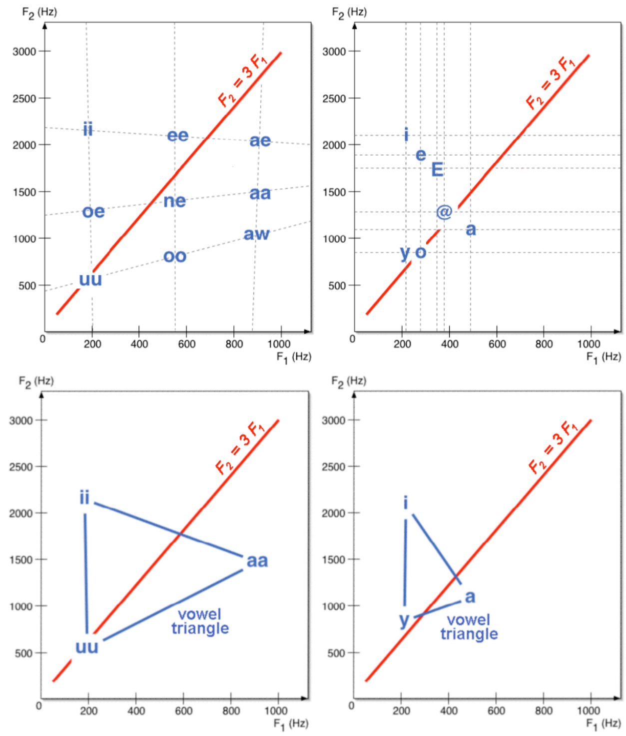

Figure 11: *Location, in a (F_1, F_2) plane, of vowels used in [6] (left) and of vowels from the MBROLA FR1 voice (right) with respect to the guitar timbre $F_2 = 3F_1$ trajectory (top figures) and their corresponding vowel triangles (bottom figures).*

2.4. Graphical representation of phonetic gestures

The graphical representation of the approximate shape of the vocal tract evoking the phonetic gestures underlying the description of guitar timbre is obtained from the parametrisation of the three distinctive features of speech selected by Slawson [6] to characterize sound colour: OPENNESS, ACUTENESS, LAXNESS.

Slawson's equal-value contours presented on Figure 8 are modelled as functions of F_1 and F_2 (derived from Eq. 1 and 2 respectively). As shown in Figure 13, in these simplified models of the

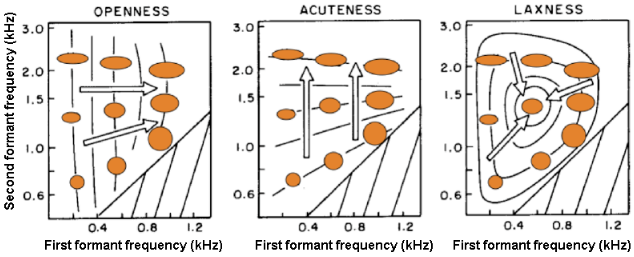

Figure 12: *Mouth shapes associated with vowel colours centred on the corresponding* (F_1, F_2) *points.*

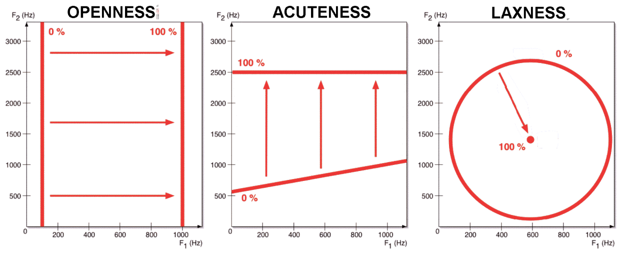

Figure 13: *Simplified models of the equal-value contours in the* (F_1, F_2) *plane.*

equal-value contours, the OPENNESS parameter P_O is directly proportional to F_1.

$$P_O = \frac{1}{9}F_1 - \frac{100}{9} \tag{9}$$

The ACUTENESS parameter P_A is mainly influenced by F_2.

$$P_A = \frac{100}{2500 - F_{2_{min}}}F_2 - \frac{250000}{2500 - F_{2_{min}}} + 100 \tag{10}$$

where

$$F_{2_{min}} = \frac{4}{9}F_1 - \frac{5000}{9} \tag{11}$$

The LAXNESS parameter P_L is a function of the distance in the (F_1, F_2) plane between the chosen vowel and the neutral vowel. The (600, 1400) point is considered to be the neutral point (100% LAX) and the (100, 300) point is considered as the most tense vowel (0% LAX).

$$P_L = 100\left(1 - \frac{d}{d_{max}}\right) \tag{12}$$

with

$$d = \sqrt{(F_1 - 600)^2 + (F_2 - 1400)^2} \tag{13}$$

$$d_{max} = \sqrt{(100 - 600)^2 + (300 - 1400)^2} = 1200 \text{ Hz} \tag{14}$$

The mouth-like geometrical shape is realized with two $6th$-order cubic splines placed symetrically about a vertical axis. The OPENNESS parameter acts on the overall size of the shape (bigger if OPENNESS is increased). The ACUTENESS parameter acts on the ratio between vertical and horizontal dimensions, inducing a flattening of the shape when this parameter is increased. The LAXNESS parameter of the vowel and the hardness of the consonant are represented by symbolic modifications of the shape. LAXNESS being generally associated with warmth, lax vowels will be represented with a red-coloured mouth, and tense vowels by a blue-coloured mouth. The type of consonant influences the contour shape: rounded for soft consonant and broken for hard consonant. Figure 14 illustrates mouth-like shapes for different amounts of OPENNESS and ACUTENESS.

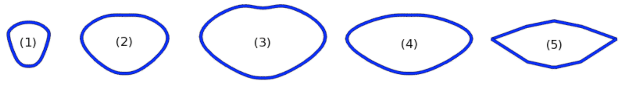

Figure 14: *Mouth-like shape produced by the application as a function of* OPENNESS *and* ACUTENESS *parameters: (1)* P_O *= 0%,* P_A *= 0%; (2)* P_O *= 50%,* P_A *= 0%; (3)* P_O *= 90%,* P_A *= 0%; (4)* P_O *= 90%,* P_A *= 50%; (5)* P_O *= 90%,* P_A *= 70% (hard consonant). In all cases,* P_L *= 0% (blue colour).*

3. CONCLUSIONS AND PERSPECTIVES

We have developed a computer tool producing the synthesis of the vocal imitation of guitar sounds, using the MBROLA technique, as well as the graphical representation of phonetic gestures underlying the description of guitar timbre nuances, as a function of instrumental gesture and score parameters and based on perceptual analogies between guitar and speech sounds. Similarly to the traditional teaching of tabla which uses onomatopeia to designate the different strokes, vocal imitation of guitar timbres could provide a common language to guitar performers, complementary to the mental imagery they commonly use to communicate about timbre, in a pedagogical context for example. We plan to investigate other phonemes (like French nasal vowels) and other MBROLA voices that might be closer to guitar timbres.

4. REFERENCES

[1] Caroline Traube, *An Interdisciplinary Study of the Timbre of the Classical Guitar*, Phd thesis, Faculty of Music, McGill University, 2004.

[2] Caroline Traube and Philippe Depalle, "Phonetic gestures underlying guitar timbre description," in *Proc. International Conference on Music Perception and Cognition*, Evanston (IL), USA, August 2004.

[3] C. Abry, L.-J. Boë, Corsi P., R. Descout, M. Gentil, and P. Graillot, *Labialité et phonétique.*, Université des langues et lettres de Grenoble, France, 1980.

[4] Caroline Traube and Philippe Depalle, "Timbral analogies between vowels and plucked string tones," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montréal, Québec, Canada, May 17-21 2004.

[5] Roman Jakobson, C. Gunnar M. Fant, and Morris Halle, *Preliminaries to Speech Analysis. The distinctive Features and Their Correlates*, The MIT Press, Massachusetts Institute of Technology, Cambridge, Massachussets, 1967.

[6] Wayne Slawson, *Sound Color*, University of California Press, Berkeley and Los Angeles, California, 1985.

[7] N. Chomsky and M. Halle, *The Sound Pattern of English*, Harper and Row, New York, 1968.

[8] Thierry Dutoit, *An Introduction to Text-to-Speech Synthesis*, Kluwer Academic Publishers, 1997.