TIMBRAL ANALOGIES BETWEEN VOWELS AND PLUCKED STRING TONES

Caroline Traube, Philippe Depalle

Sound Processing and Control Lab (SPCL), Faculty of Music, McGill University
555 Sherbrooke Street West, Montreal, Quebec, Canada H3A 1E3
caroline.traube@umontreal.ca, depalle@music.mcgill.ca

ABSTRACT

Classical guitarists vary plucking position to achieve different timbres from nasal and metallic —closer to the bridge —to round and mellow —closer to the middle of the string. An interesting set of timbre descriptors commonly used by guitarists seem to refer to phonetic gestures : *thin*, *nasal*, *round*, *open*, ... The magnitude spectrum of guitar tones being comb-filter shaped, we propose to see the local maxima of that comb filter structure as vocal formants. When guitarists talk about a guitar sound as being *round*, it would mean that it sounds like a *round-shaped-mouth* sound such as the vowel /O/. Although the acoustical systems of the guitar and of the voice mechanism are structurally different, we highlight the fact that guitar tones and a particular set of vowels display similar formant regions. We also investigate the possibility of applying some distinctive features of speech sounds to guitar sounds.

1. INTRODUCTION

Among the instrumental gesture parameters that contribute to the timbre of a guitar sound, the location of the plucking point along the string has a major influence. 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, 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 is well-known and acknowledged by most guitarists and coarsely describes the timbral palette of the instrument. Beyond brightness, guitarists seem to be sensitive to the vocal aspect of guitar tones. In this paper, we attempt to identify the acoustical basis to some perceptual analogies between guitar and vocal sounds and show that some of the dimensions of the timbre subspace for guitar can be borrowed from the phonetics domain. In sections 2 and 3, we present the comb filtering effect associated with the plucking excitation, then we identify the local maxima of the comb filter as formants. In section 4, we compare guitar and voice acoustical systems and formant structures. Finally, in section 5, we show how and why some distinctive features of speech sounds can be applied to guitar sounds.

2. PLUCKING A STRING AND COMB FILTERING

The plucking excitation initiates wave components travelling independently in opposite directions along the string. The resultant motion consists of two bends, one moving clockwise and the other counter-clockwise around a parallelogram [1]. In the ideal cases, the output from the string (force at the bridge) lacks those harmonics that have a node at the plucking point. The amplitude C_n of the nth mode of the displacement of an ideal vibrating string of length

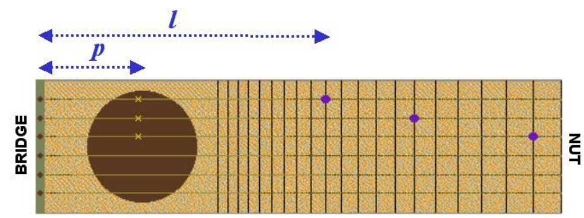


Figure 1: *Plucking point at a distance p from the bridge and fingering point at distance l from the bridge on a guitar neck.*

l plucked at a distance p from the bridge with an initial vertical displacement h is given by :

$$C_n(h,R) = \frac{2h}{n^2\pi^2 R(1-R)} \sin(n\pi R) \tag{1}$$

where $R = p/l$ is the relative plucking position, defined as the fraction of the string length from the point where the string was plucked to the bridge [2], as illustrated on Figure 1.

The digital signal processing interpretation of the physical phenomenon is the following: in a simple digital physical model of a plucked-string instrument, the resonant modes translate into an all-pole structure (i.e. the harmonic structure of the signal), while the initial conditions (a triangular shape for the string and a zero-velocity at all points) result in a FIR comb filter structure of the type $y[n] = x[n] - x[n-d]$ (i.e. the spectral envelope structure of the signal). As the fundamental period T_o corresponds to the time the wave needs to travel along a distance that is two times the vibrating length of the string ($2l$), the relation between the comb filter delay D and the relative plucking position R is [3]:

$$\frac{D}{T_o} = \frac{2p}{2l} = R \tag{2}$$

The comb filtering effect is illustrated on Figure 2 for a recorded guitar tone plucked 12 cm away from the bridge on a 58 cm open A-string (fundamental frequency = 110 Hz). The relative plucking position R is approximately 1/5 (12 cm / 58 cm = 1.483).

3. COMB FILTER FORMANTS

Considering a string of length l plucked at a distance p from the bridge and resonating at fundamental frequency f_o, the frequency F_1 of the first local maximum in the comb-filter shaped magnitude spectrum equals the inverse of twice the delay D from Eq. 2:

$$F_1 = \frac{1}{2RT_o} = \frac{f_o}{2R} = \frac{lf_o}{2p} \tag{3}$$

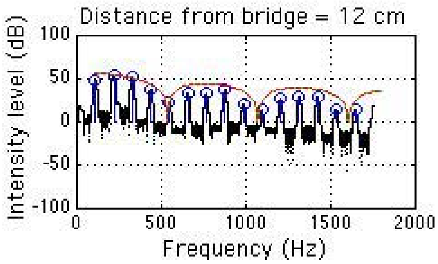


Figure 2: *Magnitude spectrum of a guitar tone, plucked at 12 cm from the bridge on a 58 cm string.*

The other local maxima ($F_2, F_3, ...$) in the magnitude spectrum are odd integer multiples of F_1. We propose to see those frequencies F_n as the central frequencies of *comb filter formants*. Here we consider the literal definition of a formant, i.e. a frequency range in which amplitudes of spectral components are enhanced. In most cases, formant regions are due to resonances but in the present case, the local maxima do not correspond to resonances per se but rather to anti-notches. An interesting fact is

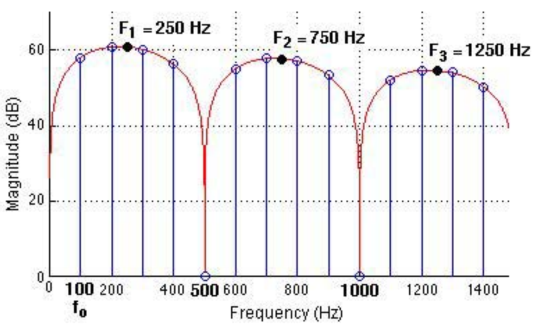


Figure 3: *Comb-filter shaped magnitude spectrum of a plucked-string tone with fundamental frequency of 100 Hz and relative plucking position of 1/5. Zeros occur at integer multiples of 5 × 100 = 500 Hz and local maxima occur at odd integer multiples of 250 Hz.*

that the formant frequencies F_n are constant for a given absolute plucking position p on a given string, regardless of the note being played. For example, to play a note that is, say, a third higher than the open string frequency, the vibrating length of the string is shortened by a (5:4) factor by pressing the string with a finger against the corresponding fret. More generally, calling α the transposition ratio, the fundamental frequency f_o is multiplied by the ratio α while the string length l is divided by the same ratio α. By a simple inspection of Eq. 3 giving F_1 as a function of f_o and l, we can see that the α's will cancel each other. This all makes sense considering that the product lf_o is a constant for a given string and equals half the speed of sound c (being equal to the square root of the ratio of string tension and linear mass density). Therefore, F_1 can also be expressed as the ratio $c/4p$.

Using Eq. 3, we can determine the usual ranges of frequencies for the first formant frequencies F_1 of the different strings of a guitar. We will consider a range of absolute plucking position p going from 3 to 30 cm from the bridge on 60 cm strings tuned with the standard tuning. We find that the range for F_1 goes from f_o (at 30 cm from the bridge) up to $10f_o$ (at 3 cm from the bridge).

Str. #	Note name	Tuning frequency	Range for first formant frequency
6	E (Mi$_2$)	83 Hz	[83 → 830] Hz
5	A (La$_2$)	110 Hz	[110 → 1100] Hz
4	D (Ré$_3$)	146 Hz	[146 → 1460] Hz
3	G (Sol$_3$)	202 Hz	[202 → 2020] Hz
2	B (Si$_3$)	248 Hz	[248 → 2480] Hz
1	E (Mi$_4$)	330 Hz	[330 → 3300] Hz

Finally, since we have talked about the comb filter local maxima as being formant regions, we will specify the bandwidth of those formants. The magnitude spectrum being proportional to a sine function, we conclude that the 3dB-bandwidth equals F_1 (frequency range corresponding to a $[\pi/4, 3\pi/4]$ phase range). For example, the bandwidth is 250 Hz for the case illustrated on Figure 3. This is wider than the bandwidth of a usual vowel formant.

4. ANALOGIES BETWEEN GUITAR TONES AND VOWELS

4.1. Comparing guitar and voice acoustical systems

It is well-known that the recognizable quality of vowel sounds is due to the existence of formant regions, which are frequency ranges where the sound is enhanced by the cavity resonances of the human vocal tract (among the oral, labial and nasal cavities).

In his book *Sound Color*[1], Slawson reviews the research that has investigated *vowel-like resonances in some musical instruments* [4]. He notes that *most musical instruments have sources that are driven by the resonance systems of the horns, strings, or membranes that make this instrument and that there is little in those systems, apparently that is vowel-like*. He refers to Jansson [5] who has compared the bow-string system to the vocal source and the resonance box to the vocal-tract filter system. The analogy was not very successful. So at the time Slawson has ruled out musical instruments as models for sound color because of the strong coupling between source and filter: *there may be some basis for studying the sound color of musical instruments if other decoupled resonance systems are discovered* (p.157 in [4]). In a source/filter

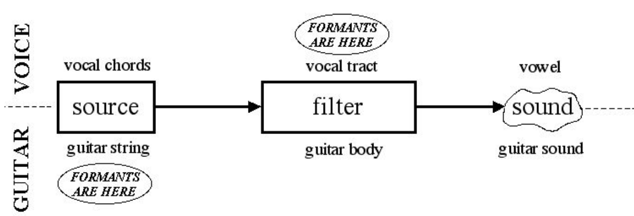


Figure 4: *The source/filter model of the vocal mechanism and of the guitar.*

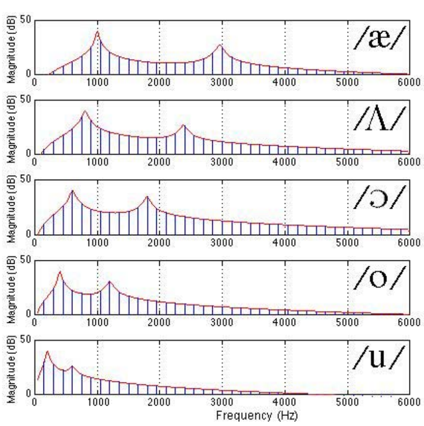
model of the vocal mechanism, the source is the glottal excitation produced by the vocal chords and the filter corresponds to the cascade of resonators formed by the vocal tract. The acoustical

[1]For Slawson, sound color is an aspect of timbre pertaining ot the steady-state portions of sounds.

system forming a guitar could also be decomposed into a source and a filter. The string couples via the bridge into the resonating body which is needed for coupling to the surrounding air. Spontaneously, we can say that the string is the source and the guitar body, the filter, expecting that, as it is the case for the vocal mechanism, the formant structure belongs to that filter. In fact, we found that the vowel-like formant structure is due to the localized plucking excitation point (resulting in a comb filter effect) rather than to the main resonances of the body which occur at quite low frequencies, around 100 or 200 Hz. On the other hand, the comb filter effect for a guitar string is characterized by odd formant frequencies ($F_2 = 3 \times F_1$, $F_3 = 5 \times F_1$, etc.). Vowels show similar patterns in their magnitude spectrum since the vocal tract is, in first approximation, a tube closed at one end, also favoring odd resonant frequencies. In conclusion, we believe it is not necessary to find strong similarities between the acoustical systems, i.e. between the causes of the sounds. It is enough to find similarities between the acoustical signatures of the sounds produced by those systems, i.e. between the effects, independently of the cause.

4.2. Comparing guitar and vowel sounds formant structures

The first two or three formants are generally sufficient to identify vowel sounds. We have synthesized vowel-like sounds with two formants located where the first two local maxima in a comb-filter structure would be, i.e. the second formant frequency equals three times the first formant frequency: $F_2 = 3 \times F_1$ with F_1 = 1000, 800, 600, 400 and 200 Hz. The five magnitude spectra are displayed on Figure 5. These are the formant patterns that we



	f	A	BW
F_1	1000	0	40
F_2	3000	-13	90
F_1	800	0	70
F_2	2400	-14	80
F_1	600	0	70
F_2	1800	-6	80
F_1	400	0	40
F_2	1200	-10	80
F_1	200	0	40
F_2	600	-18	60

Figure 5: *Two-formant spectral envelopes of five "guitar vowels". The table specifies the central frequency, the amplitude et the bandwidth of each formant.*

would obtain for a string with length l = 60 cm and fundamental frequency f_o = 200 Hz plucked successively at 6, 7.5, 10, 15 and 30 cm from the bridge (cf. Eq. 3). This sequence of sounds simulates the narrowing of the comb-filter structure when moving the plucking position from the bridge to the midpoint of the string. The synthesized sounds are perceived as close to /æ/ (as in b*a*t), /Λ/ (as in b*u*t), /O/ (as in b*ou*ght), /o/ (as in b*oa*t), /u/ (as in b*oo*t). At this point, we can already draw attention to the shape of the mouth forming those vowels, as illustrated on Figure 6. When plucked close to the bridge, the string produces a sound that is associated with a thin-shaped mouth. Moving closer to the tonehole, the mouth seems to open up to a round shape. Then, from the tonehole 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, the sound is generally described as hollow and some guitarists will call it a *bassoon* sound or a *pipe* sound. Note that the transitions from a thin-shaped mouth to a round-shaped mouth and then to a closed mouth are the same transitions that one would have to go through in a continuous way to imitate the sweeping flanging effect of a landing airplane, for example. As we noted at the end of section

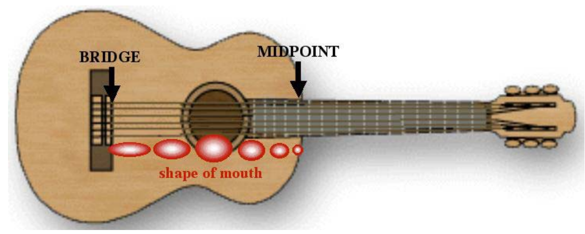


Figure 6: *Phonetic gestures associated with plucking positions.*

3, the bandwidth of the lobes of a comb filter is much larger than the bandwidth of vowel formants. The closer to the bridge the plucking position is, the wider the comb filter formants are. This could explain why guitar tones are perceived as more nasal when plucked closer to the bridge since one of the main effects of adding the nasal cavity to the main oral pathway is the broadening of all formant bandwidths and the flattening of spectral peaks [6].

5. APPLICATION OF THE DISTINCTIVE FEATURES OF SPEECH SOUNDS TO GUITAR SOUNDS

Professional guitarists perceive subtle variations in the tones they produce and they have developed a very rich vocabulary to describe those timbre nuances. An interesting set of timbre descriptors seem to refer to phonetic gestures such as, for example, *oval*, *round*, *thin*, *open* and *hollow*. Supported by the analogies that we found at the signal level, we will show, in this section, how and why the distinctive feature theory can be applied to guitar sounds.

The distinctive feature theory, proposed by Jakobson, Fant and Halle in 1951 and then later revised and refined by Chomsky and Halle in 1968 [6], 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 designated three of the features related to vowels (*compactness*, *acuteness* and *laxness*) as candidates from which to derive dimensions of sound color. Figure 7 displays the equal-value contours for those features in a (F_1, F_2) plane [2]. According to Slawson, OPENNESS (replacing the term

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

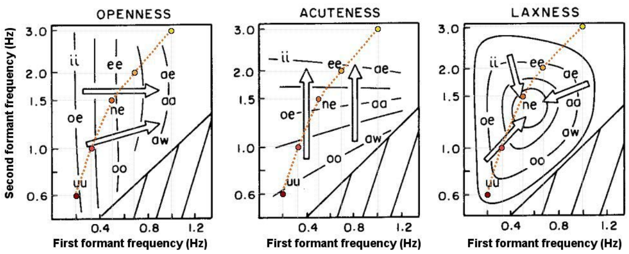


Figure 7: *Equal-value contours for three distinctive features of speech in the (F_1, F_2) plane [4] with superimposed "guitar vowels" trajectory (dotted line) corresponding to the relationship* $F_2 = 3 \times F_1$.

compactness given in [6]) 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 in the (F_1, F_2) plane centered on the maximally LAX point (or minimally tense point). This central point correspond 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. This is the neutral position of vocal tract which can be best approximated by a single tube closed at one end. Since a tube of length L closed at one end can only resonate at frequencies for which L is an odd multiple of one quarter wavelength and since the average length of the vocal tract of males is about 17.5 cm, the resonances appear at approximately 500, 1500, 2500 Hz, etc. We determined that the formant pattern of this neutral vowel is actually found in the magnitude spectrum of a guitar tone produced by the third string ($f_o = 202$ Hz) when it is plucked 12 cm from the bridge (assuming the string open and 60 cm long). More precisely, from Eq. 3, we find that the first formant frequency equals $\frac{60\times 202}{2\times 12} = 505 \approx 500$ Hz. The second formant is centered approximately on $3 \times 500 = 1500$ Hz, the third on 2500 Hz, and so on.

Slawson claims that the dimensions of OPENNESS, ACUTENESS, and LAXNESS are fundamental biological features serving some pre-speech function that are part of the auditory processing of all sounds. This is in fact what we have noticed when examining the timbre descriptors commonly used by guitarists. Although the analogy with vowels does not tend to be conscious, a large set of adjectives seem to refer to phonetic gestures. Applying the distinctive features of speech, 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 could likely evoke the sound one makes when exhaling warm air, usually with the vocal tract in the neutral position. Finally, a *hollow* or *cavernous* sound would actually sound like the *u* vowel produced as the mouth forms a hollow cavity.

The trajectories that we plotted with a dotted line on top of Slawson's equal-value contours for distinctive features of speech in Figure 7 correspond to the relationship $F_2 = 3 \times F_1$, which is found for the first two local maxima of a comb filter. In that way we can see what are the achievable vowels with a guitar string when 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 color, independently of the note that is played. Knowing that, we can determine the set of vowels corresponding to plucking all strings 12 cm away from the bridge (if strings are 60 centimeter long). The table below gives the first formant frequency (calculated with Eq. 3) for each string together with the closest sound color and corresponding IPA symbol. Note that vowel color 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	ɘ (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)

6. CONCLUSION

In this paper, we attempted to identify the acoustical basis of the perceptual analogies between guitar tones and vowels by highlighting the fact that both types of signal display similar formant structures. Deriving timbre dimensions from phonetics, we also clarified the vocabulary used by guitarists to describe timbre.

7. ACKNOWLEDGMENTS

We would like to thank classical guitarist Peter McCutcheon, professor at Université de Montréal, and his students (in particular Alexandre Savard) for their contribution to this work through lengthy discussions about timbre and plucking techniques.

8. REFERENCES

[1] N.H. Fletcher and T.D. Rossing, *The Physics of Musical Instruments*, second edition, Springer-Verlag, 1998.

[2] C. Traube and J.O. Smith, "Estimating the Plucking Point on a Guitar String", Proc. Conf. on Digital Audio Effects (DAFX'00), Verona, Italy, pp. 153-158, December 2000.

[3] C. Traube and P. Depalle, "Extraction of the excitation point location on a string using weighted Least-Square Estimation of a Comb Filter Delay", Proc. Conf. on Digital Audio Effects (DAFX'03), pp. 188-191, London, 2003.

[4] W. Slawson, *Sound Color*, Univ. of California Press, 1985.

[5] E.V. Jansson, Analogies between bowed string instruments and the human voice: source-filter models, Speech Transmission Laboratory, Royal Institute of Technology (KTH), QPSR3/1966, MIT, Cambridge, 1666.

[6] R. Jakobson, C.G.M. Fant and M. Halle, Preliminaries of speech analysis. The distinctive Features and Their Correlates, The MIT Press, 1967.