Words, Movement and Timbre

Alex McLean and Geraint Wiggins
Center for Cognition, Computation and Culture
Department of Computing
Goldsmiths College
{ma503am,g.wiggins}@gold.ac.uk

Abstract

Phonetic symbols describe movements of the vocal tract, tongue and lips, and are combined into complex movements forming the words of language. In music, *vocables* are words that describe musical sounds, by relating vocal movements to articulations of a musical instrument. We posit that vocable words allow the composers and listeners to engage closely with dimensions of timbre, and that vocables could see greater use in electronic music interfaces. A preliminary system for controlling percussive physical modelling synthesis with textual words is introduced, with particular application in expressive specification of timbre during computer music performances.

Keywords: Vocable synthesis, timbre

1. Introduction

A human speaker can produce timbre of subtlety and range at great speed, through movements of their vocal tract, tongue and lips. Perhaps even more impressive is the ability of a listener to derive the state of a speaker's vocal tract from the sound produced, evident in their reconstruction of the phonemes, words and phrases that a speaker intends to communicate [1].

Perception of speech is distinct from auditory perception. This is made clear by *sine wave speech* [2], where the sound signal of speech is reduced to the variance of just four sine waves. The frequency and amplitude of three sine waves are mapped from the lowest three formant frequencies, and the fourth sine wave from a fricative formant. The result is a bistable illusion, where a human subject perceives the sound as a kind of formless burbling until they are primed with the original speech signal – they then perceive the sine waves as intelligible speech. Distinct speech perception is further demonstrated by the *McGurk-MacDonald effect* [3], where for example *seeing* lip movements for the word 'ga' while *hearing* the word 'ba' causes the listener to *experience* the word 'da', produced by an imagined articulation halfway between 'ga' and 'ba'. Here the phoneme 'da' exists only as a speech percept influenced by both visual and auditory stimuli.

Here we focus on vocal speech, but it is important to note that the vocal tract is not unique in its expressive encoding of symbols with movement. Deaf culture has produced sign languages where symbols are represented by movement of the hands and face, yet otherwise exhibit all the features of a spoken human language, including grammar, pragmatics and metaphor [4]. It may seem odd to mention languages for the Deaf in a paper about musical interfaces, but we do so to support our placement of movement at the heart of the phonetic categories which support language. On this basis we argue that the symbolic classification of movement is a general function of the brain, an ability useful in musical expression, as we will see in the next section.

2. Vocables – musical words

In musical tradition, vocable words are those used to describe an articulation of a musical instrument. An instructor may use their voice to describe the sound their student should try to make on their violin, perhaps by singing the pitch contour while using a particular consonant-vowel pattern to indicate a particular bowing technique. Over time the student will be able to perceive the phonetics of their instructor's voice as sound categories of their instrument.

A formal system of vocables for use in the scores of western classical music has been proposed and used by Donald Martino [5] but has so far not found wider use. However many of the oldest musical cultures have well-developed formal systems in widespread use. Indian classical music has the *bol* syllables [6] where, for example, *ṭe* represents a non-resonating stroke with the 1st finger on the centre of the dāhinā (right hand) drum. In the Scottish Highlands we find *canntaireachd* of the bagpipes [7], for example *hiaradalla* represents an echo of the D note in the McArthur canntaireachd dialect. In China the delicate finger techniques of the *guqin* (*Chinese zither*) is notated using the *chien-tzû*. For example *ch'üan-fu* indicates that the index, middle and ring finger each pull a different string with a light touch, making the three strings produce one sound 'melting' together.

In her doctoral thesis "Non-lexical vocables in Scottish traditional music" [8], Chambers terms formalised vocables

as culturally *jelled* and ad-hoc vocables as *improvisatory*, acknowledging that the line between the two is sometimes blurred. Chambers goes on to term onomatopoeic vocables as *imitative* and more arbitrarily assigned vocables as *associative*. At this point we hit upon major perceptual issues, as Chambers reports: "Occasionally a piper will say that a vocable is imitative (indigenous evaluation) when analysis seems to indicate that it is actually associative (analytic evaluation) because he has connected the vocable with the specific musical detail for so long that he can no longer divorce the two in his mind" [8, p. 13]. In other words, a vocable may appear to mimic an instrumental sound on the perceptual level, when on the level of the sound signal it does not. This seems to be true in the general context of onomatopoeia – for example where a native English speaker hears a hen say "cluck", his German neighbour may perceive the same sound as "tock". Research into tabla bols have however found them to be genuinely imitative, sharing audio features with the instrumental sounds they represent, identifiable even by naive listeners [9]. Further correlation has been found between the words commonly used to describe vowel-like quality of guitar sounds and the associated mouth shapes [10].

A third distinction can be made between vocable words in *written* and *spoken* form. A reader of a vocable applies *paralinguistic* phrasing not derived directly from the text, but nonetheless with great musical importance. Conversely a transcriber may resolve ambiguity in a spoken vocable, by writing a precise interpretation of what was intended. This issue is of course common to all symbolic notation systems. We can say however that to some degree a written vocable may capture the *essence* of a sound.

3. Vocables as Musical Interface

Electronic music allows sound synthesis free from the constraints of physical instruments. However this freedom presents the problem of how to create new interfaces to control the new synthesis parameters. Fruitful research into 'tangible' physical interfaces to synthesisers is ongoing, but vocable words offer an alternative approach, which a few have explored.

From the early 1980s David Evan Jones has played on the boundary between auditory and speech perception in his music, for example by using the CHANT software [11] to apply vowel like quality to instrumental sounds. In "Speech Extrapolated" he describes how he leads the listener to perceiving non-speech as speech, and vice-versa [12]. Speech synthesis has also featured as a source of musical timbre in electronic dance music, for example the largely unintelligible singing synthesis software written by Chris Jeffs for use in his compositions under the 'cylob' moniker [13]. General use tools are rare, but pioneering research into voice-control of synthesisers with vocable words is provided by Jordi Janer, where syllables are sung into a microphone and the sound signal analysed and mapped to instrumental parameters [14].

Of course innovation in vocable expression continues outside of electronic music. Luciano Berio's *Sequenza III* is a vocal piece featuring mutterings, clicks and shouts of the female voice, notated with a unique system of symbols including vocables. The manipulation of the voice is taken to different extremes in *human beatboxing* where extended vocal techniques are employed to produce convincing impersonations of drum machines and bass lines [15]. Beatbox rhythms may be notated with a system of vocables called *standard beatbox notation* [16].

4. Vocables as metaphor

Our particular interest in vocables is that sounds and words may be perceptually related through imagined, simulated movement. This argument is supported by work in other fields. In neuropsychology, research by V.S. Ramachandran into synaesthesia has found 'cross-wiring' between sensory and/or conceptual maps of the brain to be an evolutionary trait, particularly common in artists [17]. Synaesthetic cross-wiring is characterised as an extreme case of the normal brain's ability to draw metaphors. Ramachandran posits that cross-wiring could even have provided the original evolutionary stepping stone to language itself.

From cognitive science, Peter Gärdenfors proposes the *theory of conceptual spaces* [18], placing a level of conceptual representation *between* the symbolic level (of computation) and sub-symbolic level (as commonly modelled by artificial neural networks). This level represents concepts in *geometry*, where distance represents dissimilarity, sets of dimensions form conceptual domains and shapes represent properties of concepts. This accords well with Ramachandran's account, indeed Gärdenfors characterises metaphor in much the same way, working it into a theory of semantics. The application of the theory of conceptual spaces to music and creativity is covered in greater detail in an earlier co-written paper [19].

These are powerful ideas that could have great consequences for future artistic practice. Indeed, this cross-wiring could provide a neural basis for exactly the kind of cross-domain mapping that our research into *vocable synthesis* aims to exploit.

5. Vocable Synthesis

We term *vocable synthesis* as the process where vocable words are specified in written form, which are mapped to articulations of a physical model of a (real or imagined) musical instrument. The use of physical modelling synthesis allows us to posit that a listener can perceive time variance of audio features as physical movement. The musician is then describing movement with words, which the listener experiences through the medium of timbre.

Vocable synthesis was introduced in an earlier paper [20] and artwork[21], where Karplus-Strong percussive synthesis is controlled by vocable words. Our current system models a drum head using a 2D waveguide mesh [22] in a triangular geometry for maximum accuracy [23]. The drum head is excited through interaction with a drumstick, using a mass-spring model [24]. This model gives greater control over a broader range of timbre than our previous work. The drum head has parameters to control the *tension* and *dampening* of the surface, and the drumstick has parameters to control its *stiffness* and *mass*. The drumstick is thrown against the drum head with parameters controlling the *downward velocity*, *starting x/y position* and the *angle and velocity of travel* across the drum skin.

Table 1. Mapping of consonants to mallet property (columns) and movement relative to drum head (rows).

	heavy stiff	heavy soft	light stiff	light soft
across	q	r	y	s
inward	c	m	f	w
outward	k	n	v	z
edge	x	d	t	b
middle	j	g	p	h/l

Table 2. Mapping of vowels to drum head tension (columns) and dampening (rows).

	tense	loose
wet	i	u
	a	
dry	e	o

Vocable words are composed from the 26 letters of the modern English alphabet. The consonants map to the drumstick and movement parameters, and vowels to the drum head parameters, shown in Tables 1 and 2. While this mapping is largely arbitrary, the consonant/vowel organisation is inspired by the International Phonetic Alphabet [25], where vowels map to the position of the tongue and pulmonic consonants to the place and manner of articulation.

As an example, the articulation *"Hit loose, dampened drum outwards with heavy stiff mallet, then hit the middle of the drum with a lighter mallet while tightening the skin slightly and finally hit the edge of the skin with the same light mallet while loosening and releasing the dampening"* is expressed with the single vocable word *"kopatu"*.

5.1. Vocable rhythms

Vocable rhythms are implemented using syntax derived by the Bol Processor [26]. A sequence of vocables are separated with white space, with rests denoted with hyphens:

```
ba da - bing - -
```

Once the user types in such a sequence, it is played on a loop until the next sequence is entered. Sequences can be grouped together into polyphony, by separating sequences with commas, and surrounding them with braces:

```
{ba da bing, pip rrrre}
```

As the sequences in this example are of different lengths, rests are automatically inserted to pad them out to the length of the lowest common multiplier. The resulting polymetric structure is:

ba - da - bing -
pip - - rrrre - -

Note that 'ba' and 'pip' co-occur on the same measure, requiring a polyphonic articulation as described in §5.2.2. If square brackets are used rather than braces, then the sequences are repeated in order to fit, like so:

ba da bing ba da bing
pip rrrre pip rrrre pip rrrre

The sequences may be nested to create complex polyrhythms from simple parts.

5.2. Vocable manipulation and analysis

Now we have described the representation and mapping of vocables in our system, we examine ways of analysing and manipulating vocable words.

5.2.1. Symbolic level

As written vocables represent sounds in symbolic form, we have a wide range of techniques from computer science available to us. For example we may analyse sequences of vocables using Markov models and other statistical techniques. An approach to modelling structures of vocable rhythms in order to generate rhythmic continuations is introduced in earlier work [27].

We may also use standard text manipulation techniques such as *regular expressions* (regex). Regexes are written in concise and flexible language allowing general purpose rule-based string matching [28]. We have embedded a regex parser in our system, allowing operations such as the following:

```
~%3=0 /[aeiou]/to/ fe be
```

This replaces the vowels of every third vocable with the string 'to', resulting in the following sequence:

```
fto be fe bto fe be
```

5.2.2. Geometrical level

Our vocables are direct mappings from the geometry of a drum and its articulation. It is therefore straightforward to move from a symbolic to geometric representation, in order to perform spatial analyses and manipulations.

Combining vocables in polyphonic synthesis is straightforward, and implemented in our current system as follows. As consonants control the movement and mallet material, we allow two consonants to be synthesised concurrently simply by using multiple mallets in our model. Currently we allow up to five active mallets per drum, allowing five consonants to be articulated at the same time. As vowels control the properties of a single drum head, we combine them simply by taking the mean average of the values they map to.

We may exploit both symbolic and geometric vocable representations in one operation. For example we could estimate the perceptual *similarity* of two vocable words of different lengths with an approach similar to the symbolic Levenshtein edit distance [29], with edits weighted by phoneme similarity on the geometrical level. Producing such an algorithm is left for future work.

6. Conclusion

The work of many of those cited in this paper, in particular Ramachandran, Gärdenfors and Patel, stands out by looking not for *oppositions* between geometric and symbolic representations, or between language and music sounds, but in the *comparisons* and *interactions* between them. In this spirit we have shown a musical interface that allows symbolic control of a geometrical model in a manner that we hope is well matched to human perception and production of instrumental sounds.

Work is ongoing to greater understand the perception of vocables through experiment, while exploring vocable synthesis through artistic practice, in particular live coding improvisation [30].

Video demonstrations, and GNU public licensed source code for the software described here is available on-line from; `http://yaxu.org/category/vocable/`.

References

[1] A. M. Liberman and I. G. Mattingly, "The motor theory of speech perception revised," *Cognition*, vol. 21, pp. 1–36, October 1985.

[2] R. E. Remez, J. S. Pardo, R. L. Piorkowski, and P. E. Rubin, "On the bistability of sine wave analogues of speech," *Psychological Science*, vol. 12, no. 1, pp. 24–29, 2001.

[3] H. Mcgurk and J. W. Macdonald, "Hearing lips and seeing voices," *Nature*, vol. 264, no. 246-248, 1976.

[4] R. Sutton-Spence and B. Woll, *The Linguistics of British Sign Language: An Introduction*. Cambridge University Press, 1999.

[5] D. Martino, "Notation in general-articulation in particular," *Perspectives of New Music*, vol. 4, no. 2, pp. 47–58, 1966.

[6] J. Kippen, *The Tabla of Lucknow - A cultural analysis of a musical tradiation*. Cambridge University Press, 1988.

[7] J. F. Campbell, *Canntaireachd : articulate music*. Archibald Sinclair, 1880.

[8] C. K. Chambers, *Non-lexical vocables in Scottish traditional music*. PhD thesis, University of Edinburgh, 1980.

[9] A. D. Patel and J. R. Iversen, "Acoustic and perceptual comparison of speech and drum sounds in the north indian tabla tradition: An empirical study of sound symbolism," in *15th International Congress of Phonetic Sciences (ICPhS)*, 2003.

[10] C. Traube and N. D'Alessandro, "Vocal synthesis and graphical representation of the phonetic gestures underlying guitar timbre description," in *8th International Conference on Digital Audio Effects (DAFx'05)*, pp. 104–109, 2005.

[11] X. Rodet, Y. Potard, and J. B. Barriere, "The chant project: From the synthesis of the singing voice to synthesis in general," *Computer Music Journal*, vol. 8, no. 3, 1984.

[12] D. E. Jones, "Speech extrapolated," *Perspectives of New Music*, vol. 28, no. 1, pp. 112–142, 1990.

[13] C. Jeffs, "Cylob music system." `http://durftal.com/cms/cylobmusicsystem.html`, 2007.

[14] J. Janer, *Singing-driven Interfaces for Sound Synthesizers*. PhD thesis, Universitat Pompeu Fabra, Barcelona, 2008.

[15] D. Stowell, "Characteristics of the beatboxing vocal style," tech. rep., Queen Mary, University of London, 2008.

[16] Tyte, "Standard beatbox notation." online; `http://www.humanbeatbox.com/tips/p2_articleid/231`, 2008.

[17] V. S. Ramachandran and E. M. Hubbard, "Synaesthesia – a window into perception, thought and language," *Journal of Consciousness Studies*, vol. 8, no. 12, pp. 3–34, 2001.

[18] P. Gärdenfors, *Conceptual Spaces: The Geometry of Thought*. The MIT Press, March 2000.

[19] J. Forth, A. McLean, and G. Wiggins, "Musical creativity on the conceptual level," in *IJWCC 2008*, 2008.

[20] A. McLean and G. Wiggins, "Vocable synthesis," in *Proceedings of International Computer Music Conference 2008*, 2008.

[21] A. McLean, "Babble." online artwork; `http://project.arnolfini.org.uk/projects/2008/babble/`, 2008.

[22] S. Van Duyne and J. O. Smith, "The 2-d digital waveguide mesh," in *Applications of Signal Processing to Audio and Acoustics*, pp. 177–180, 1993.

[23] F. Fontana and D. Rocchesso, "Signal-theoretic characterization of waveguide mesh geometries for models of two-dimensional wave propagation in elastic media," *Speech and Audio Processing*, vol. 9, no. 2, pp. 152–161, 2001.

[24] J. A. Laird, *The Physical Modelling of Drums using Digital Waveguides*. PhD thesis, University of Bristol, 2001.

[25] P. Ladefoged, "The revised international phonetic alphabet," *Language*, vol. 66, no. 3, pp. 550–552, 1990.

[26] B. Bel, "Rationalizing musical time: syntactic and symbolic-numeric approaches," in *The Ratio Book* (C. Barlow, ed.), pp. 86–101, 2001.

[27] A. McLean, "Improvising with synthesised vocables, with analysis towards computational creativity," Master's thesis, Goldsmiths College, University of London, 2007.

[28] J. Friedl, *Mastering Regular Expressions*. O'Reilly Media, Inc., August 2006.

[29] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals.," *Soviet Physics Doklady.*, vol. 10, no. 8, pp. 707–710, 1966.

[30] N. Collins, A. McLean, J. Rohrhuber, and A. Ward, "Live coding in laptop performance," *Organised Sound*, vol. 8, no. 03, pp. 321–330, 2003.