VOX V  by Trevor Wishart. The analysis of an electroacoustic tape piece.

Tom Williams

Context
Vox is a work for four amplified voices with electronics and tape in six independent movements (anything between one and all six can be in a concert), and totalling over an hour's worth of music. The cycle was composed over an eight year period between 1980 and 1988 (Wishart composed only a few additional pieces in this period) and is a development from such earlier works as Anticredos for amplified voices and the tape piece Red Bird. All these works explore the voice as a "multimedia" instrument through language, semantic and phonetic, dramatic intent; metaphor, transformation and the voice as a sonic generator conceptually similar to the analogue synthesizer. Vox is a synthesis of the electronic, the ethnic (with the incorporation of other world vocal techniques), the popular (especially in movement VI) the avant garde as well as polymetric/polyrhythmic techniques more usual in medieval musics.

Journal of Electroacoustic Music, Vol. 7, p. 6
Vox, or movements thereof, have been performed throughout the world with Vox V receiving international prizes and considerable acclaim. It is available on two major recording labels (Virgin and Wergo)—remarkable for a tape piece that is little more than five years old.

**Vox V: the graphic score**

In order to come to a deeper analytical understanding of this piece it has been necessary to take it out of time, to capture the phenomenological ‘now’ by giving some kind of visual representation of the piece over time. The graphic scoring of the work is different to a conventional Western music score in that it only allows us to survey the music not to reconstruct it. In that sense ‘map’ is a more appropriate term: a map’s sole purpose is to guide us through an area. Furthermore, unlike conventional, acoustic composition, where the composer begins from the score and consequently writes his/her music out of time, which determines, by this note by note dissection, the score and consequently writes his/her music out of time. The graphic representation of the piece allows us to understand the music and its evolution throughout the piece, allowing us to see the movement of the piece through an area of space. The graphic score allows us to see the piece as a whole and to understand the movements of the piece as it evolves through time. The graphic score is a representation of the piece and allows us to understand the movement of the piece through space and time.

**Vox V: scored**

Wishart, in some of his vocal writing (including Vox), has developed a notation for the timbral continuum. This is based on a mixture of mechanical instruction (how to make the sound) and the final sound as object— as in Anticredos, or Vox I. However, Vox V, a piece with no live musician, requires no musical artifact. With no performer, the presentation of the work no longer requires the symbols of conventional musical notation. They are simply redundant. The composer needs no middle party to give a presentation of the work and, as this is unnecessary, there is no symbolic convention.

Many composers of electroacoustic tape music construct a ‘diffusion score’. This is solely for the benefit of the sound projectionist in a concert situation. Certain key points within the electroacoustic piece need to be prescribed in order that the work be dynamically shaped and spatially diffused throughout the auditorium—ideally ea (electroacoustic) performances have multiple loudspeakers with individual dynamic control around the performance space. These ‘scores’ are usually very limited in their information. Wishart has made a diffusion score of Vox V, (see example). It itemises thirteen key points within the piece largely by verbal description and timing. Below it gives decibel indications for the projectionist; OdB is the optimum performance level, and from there the tape is cut or boosted by ‘riding the faders’ (volume control). Apart from the final section, there are no diffusion indications, for the piece is quadrophonic and therefore has a strong integral spatial field. At the final ‘thunderclap/rain’ section the rain and thunder can be projected upwards, as well as surrounding—adding to the impression of the ‘real’ landscape intended by Wishart.

**Vox V: mapped**

I have attempted to give a graphic representation, a mapping of events with the scaling of time on the horizontal axis. This mapping is a guide to the ordering of events and the type of gestures found therein. It is an extension of the Wishart diffusion score but now giving more detail. The graphic visualization of the sounds and their morphosis is an attempt to impart the gestalt. To show how the gestures move and evolve through time.

**Structure and electraonic music**

An ea piece dominant in the timbral continuum will not fall into any conventional forms, as there is no pitch hierarchy and therefore no ‘meaning’, no tonal tendency, no deviation from expectation; for now, deviation is the norm (see L. Meyer - Emotion and Meaning in Music). When ea pieces do attempt to refer to conventional forms the result can have a simplicity of structural content that can seem crass, let alone unsatisfactory—especially with ABA type structures.


An analysis of Trevor Wishart’s Vox V
Vox V  genesis and the poetic content
The piece is composed by taking a number of concrete (natural) sound sources and redefining them within the digital domain. It can be termed 'computer music' as the piece is generated solely within the computer environment. Wishart also uses the term ‘Sonic Art’ as this removes all musical preconceptions and has resonances of the artist as sculptor modelling sound.

The poetic precept from which Wishart began work is a conceptual image of Shiva - the Hindu god of creation and destruction - creating the world through ejections from the 'supervoice'. This image then requires a number of 'found objects' of the real world which are recognizable, thus referential and therefore, intrinsically metaphorical; furthermore, they need to be sonically maleable. These 'sonic transformations', into urban sounds - social (crowds) or manmade (machine, bell) or the non-urban sounds of nature, are, as Wishart states, on a poetic level, an exploration of the range of human experience and expression. Moreover, as objects, they need an aesthetic value - the listener needs to found an interest in them for no more reason than that the sound is interesting. The voice, or poetically Shiva's 'supervoice', is the central concrete source. Most of the other sound sources will then be transformations from the voice into whatever, with many of the seemingly non vocal sounds having a vocal root. But the characteristic of bell or gun, the two most distinct examples, are imposed by the computer processing the vocal source.

The work is working on two distinct levels: the sonic or musical discursive, which would be categorized as 'abstract' in Simon Emmerson's article in The Language of Electroacoustic Music and against the 'abstracted' - sounds taken from and referential to the real world, and with symbolic intent. Emmerson cites Wishart's earlier tape piece Red Bird as an example and suggests that the combination of abstract and abstracted sounds produce a surreality/dreamscape.

Vox V  as metaphor
To hear the work is to hear sounds of the real world. At every stage of the piece sound-images are apparent. The image of Shiva ejecting the world from his mouth is always there. Yet, the work is not a series of consecutive sounds stemming from a vocal utterance, but rather a flux of transformations with strong associative reference, where, usually, a vocal utterance is the trigger for the transformation. The metaphor then is as much in the transformation as in the original sound-object. It is, as Wishart states, with regard to the much earlier piece Red Bird, 'a rich metaphorical field of discourse'. On this point Wishart continues, in 'On Sonic Art' (p91): 'Using concrete metaphors (rather than text) we are not 'telling a story' in the usual sense, but unfolding structures and relationships in time'. He also writes, 'we must both be sonically and metaphorically articulate'. This underpins the point that it functions, as discussed in relation to Emmerson, on two aesthetically independent levels. Wishart, in his article 'Sound Symbols and Landscapes', refers to the depth of meanings within transformation: 'If this transformation is set within a whole matrix of related and transforming images the metaphorical implications become increasingly refined and ramified'(p54). Later, he points out that 'the listener may of course deny or blank out the metaphorical implications but this is possible with all other art forms which use metaphor'.

So multiple and transforming sound symbols are, as Wishart proposes, 'giving a depth of meaning'. But is this the case? Suzanne Langer in her book 'Philosophy in a New Key, A Study in the Symbolism of Reason, Rite, and Art' states that a symbol conveys a concept and that a symbol is 'the essential act of thought' for 'the human brain is constantly carrying on a process of symbolic transformation' (p43); it is 'a typically overt activity ... the sheer expression of ideas'. This is a cognitive, conscious activity but this activity is also deeper than the cognitive and, as Langer says, accounts for 'ritual, art, laughter, weeping, speech, superstition and scientific genius'. She later writes (p89) 'this appreciation of forms is the primitive root of all abstraction'. [The notion of 'forms' is Kantian and can be seen as 'categories of understanding'.] With this in mind we can approach Vox V and see levels of sound symbolism from the verbalized to the primitive and inarticulatory. If this was not the case then its symbols would be translatable into another medium like language. The syntactic level of Vox V is completely unique and untranslatable to any other work of art. The ordering of the material, its grammatical structure, has, to borrow Langer's words, a 'further source of signification ... it has a symbolic mission ... a fragmentary connotation ... to make a complex term'. It is not possible to test this argument as it is impossible to divorce the metaphor from the 'musical' - that is the sonically organized. However, when hearing the work, the sound-image is self apparent and therefore is symbolic - metaphorical. By this hearing in, we 'appreciate forms' on a 'primitive' level as well as a surface level. It is an apprehension as when viewing a painting where there are a myriad of relationships but no 'vocabulary'; it is 'presentational symbolism', as Langer writes.

Sound Sources
This list is 'as heard', not as gathered from Wishart, and is arrived at by listening in real time and at half speed as well as considering the diffusion score.

The 'Supervoice'
- vocalized: phonemes, consonants - usually as attacks for transformation as in 'Ko' for the bell/voice (2'40): the phoneme 'Ko' has an inherent bell characteristic.
  - ululation (1'16), an African vocal technique used in much of the Vox Cycle.
  - mutterings - like incantations, these may be manufactured from other vocal sources and not originally sampled muttering. [sampling: recording a sound into the digital domain]
- 'snaps' - see Wishart diffusion score - the characteristic of this sound when heard at half speed suggests it is a processed thunderclap, however it could be a vocal phoneme or even a computer generated compound of the two.
- crowd - crowd sounds, used in Vox IV, are here also. Some of the multiplex textures are created with the crowd sound source.
- unvoiced; breath - inhalation, and exhalation - dovetailing of Vox IV to V
- sibilants and their noise constituents.
- lips - flabby sounds generated

Apart from these abstracted sources other the natural utterances are:
- animal: human baby cry, horse neighing and crows calling.

Metaphorically, the baby is the only human sound. The final ‘living’ sound is the stridulence of a swarm of bees.

**Nature - elemental**
wind - there is a natural correspondence between this and breath,
thunderclap and subsequent rumbling - also the ‘snap’ sounds are frequency shifted thunder.
rain - with its granular characteristic, the ‘surge’ gestures may have rain as the original source. (it is possible to see some of the vocal sounds as watery.)

**Mechanical**
- there are no mechanical sounds! - examples of metaphorical ambiguity as well as the moulding of a distinct source (voice) into other apparently as distinct:
  - large gun at (4'08) is the voice - this becomes apparent when listened at slower speed.
  - bell - also voice generated by processing the voice, see when listened at slower speed.

**Spatial**
The work was conceived and is usually performed quadrophonically. For this a 4-track tape recorder is required with four loudspeakers placed in the four corners of the auditorium so the audience is in the centre of the square. Wishart does this not only because he sees this as being of musical interest - sounds moving right around the auditorium, but also as it allows a graphic kind of representation of the ‘supervoice’. All articulations can begin from centre front and from there move to the back to give the illusion of surrounding the audience. The most dramatic example is the opening. For the first minute where the landscape of crow/wind is heard it is all centre front but the very last transforming seconds at 1'12 - 16' the sounds swoops to the back speakers and from there the ululation vigorously moves through the whole space. Compact discs and the hi-fi market is stereo, and as a consequence Wishart mixed the work down onto two tracks for CD mastering. These important musical gestures are all but lost in the stereo version - this is also true of all the other Vox pieces, as they are all quadrophonically conceived - the scores of the various movements have spatial instructions. However, even in its two channel mode the work still maintains a clear spatial discourse (at its most persuasive over headphones).

**Overall Form:**
A 0' - 1'16 Landscape 1. wind/crows duration 76"
B 1'16 - 4'18 ea discourse duration 182"
C 4'18 - 6' (end) Landscape 2. thunder/rain duration 104"

**Episodes within B1:**
Five episodes:-
All are demarcated by the end of one sound continuum, each episode begins with a new vocal ejaculation. They are metamorphologically and gesturally defined - each has their own individual gestalt. There is a consequent, inherent forward propulsion. Even the landscapes are undergoing continual change. It is a journey for the listener moves through the landscape: there is a sense of becoming - not being.

Episodes:
1. 1'16 - 33 (17") double ululation one panned fully L (left) one R (right) and from there wide spatial sweeping and filter sweeps
2. 1'33 - 46 (13") consonant attacks and snap (thunderclap) to noise band (sibilance)
3. 1'46 - 2'00 (14") vocalized breath later mixed multiplex
4. 2'0 - 13 (13") 4 phonemes with transformations - use of crowd source
5. 2'13 - 40 (27") cyclic transformation voice - bees - voice, mono texture

Episodes within B2
Four episodes:-
1. 2'41 - 3'01 (20") voice/bell attacks with vocoder continuum.
2. 3'01 - 18 (17") utterances to multiplex (dovetailing to next episode)
3. 3'17 - 53 (36") vocal vocoder continuum to single cry
4. 3′53 - 4′19 (26") powerful climactic gestures with voice, horse, gun, baby, lips, crowd

Landscape 1.

0" Begins with a breath exhalation - this has two vocalised 'creaks' to characterize the voice - whether the auditor is conscious of the two is not important for the function is to make explicit the notion of human breath. The higher frequencies are slowly filtered out over sixteen seconds. It is a direct link with the previous Vox giving a smooth transition between the live and the tape.

4" From nothing emerges the wind, this continues for a minute where it is slowly subsumed into the overall texture. The wind has mixed into a rattling; this functions wholly within the landscape of wind, just as the vocalized 'creaks' do in the context of human breath. Wind by its energy makes other objects rattle and we intrinsically recognize this and therefore apprehend it as wind.

14" Crows slowly emerge from background to foreground by 50" they are loud, dominant and and no longer indentified within the 'real' world - from then onwards there is a surreality.

33" The third distinct, bubbly, texture running concurrently is 'unreal', abstract, for it has no apparent source or reference in the 'real' world. It is not electronic - no sounds are - and it most probably comes from some vocal utterance.

ca 1': beginning the merger of the crow and bubble textures with the wind - it is like a chemical reaction of three elements to produce a compound substance - it results in a fluidity of gesture that sweeps the piece into its ea discourse with the first vocal attack - ululation.

Landscape 1 sets up an imaginary lanscape from which the listener can contextualize the objects of ea discourse which are both abstract and abstracted (see Emmerson article in The Language of Electroacoustic Music) from the world - the surreality.

Landscape 2 is where the storm metaphorically breaks - perhaps an electric summer storm with much needed rain to quench the thirst of a dry world.

4'18 the first very 'electric' thunderclap is heard, as if it is overhead - there is filtering, the boosting and cutting of some frequency bands. Thunder was used in Vox I as an image of creation. It is one of the most powerful and emotive natural sounds. And, with the sound of the wind and rain, it is the sound of the birth and the death of the world. Is then the first crack of thunder the birth of the world from the 'supervoice' with the ea discourse as the moulding of the world? This is then one interpretation of the metaphoric. But as Wishart writes in On Sonic Art (p.91): 'Using concrete metaphors (rather than text) we are not 'telling a story' in the usual sense, but unfolding structures and relationships in time'. However, it should be remembered that Wishart's principle 'raison d’etre' is to produce structures in sound. And this is close to the semiotologist Jean-Jacques Nattiez's definition of music; the metaphoric through referentiality of any sound is of metaphoric import, but, not necessary in comprehending the work. (Comprehensibility is a Schoenberg term and is deemed an essential requirement if music is to succeed in its ability to communicate with the listener. To take this further Leonard Meyer argues that music has its own intrinsic meaning. To comprehend anything, meaning needs to be imparted.)

4'28 Thunderclap two with rumbles

5'21 Thunderclap three is lower and so apparently more distant and from here it rumbles on. Right from the first crack we have a voodoo type continuum, an 'unreal', abstract sound which takes nearly a minute to, in a sense, travel across the landscape (like a comet?) while all this time the rain becomes stronger and distinctly wetter. This sonic 'comet', I gather from Wishart, is a massively time-stretched vocal bell utterance. This I did not recognize in my own listening. By learning of this from the composer, it is no longer a phenomenological insight but, instead, similar to the analyst who much later discovers that a passage near the end of the work, he is analysing, has an augmented retrograde inversion of the original tone row. For neither is heard as such but both have an integral relationship to what has gone on before. Yet, without the early versions of the material the later radical transmogrifications would not be possible.

The work is then closed by the symbol of growth, rain. Or is it like Noah's flood where it is the drowning of the old evil world for a new beginning? A cleansing? The metaphoric readings are myriad.

The role of the Computer.

Vox V would not have been realisable using the classic electronic music studio resources. Red Bird is aesthetically in the same domain of metaphoric landscape with sonic structures as the compositional feature but composed at a time when he had only the resources of the analogue studio. The computer has allowed the original sound objects to be transformed in ways impossible with analogue technology.

Transformation

Transformation is an important feature of ea music for it has such intrinsic gestural content, and gesture is one of the most important elements within all music - gesture reflects the inner emotion of our lives and so has great emotional pull (see S. Langer). When pitch and rhythm are no longer apprehended then gesture in the sound continuum through change of one or more of the properties in that sound over time is an essential compositional device. Transformation in the work of Wishart from one sound symbol to another - from one overtly recognizable sound to another overtly recognizable sound - has important metaphoric ramifications. Wishart is looking for transformations that are illusive, that fool the ear, that work because as a sound gesture it has an
The phase vocoder can compute the smoothest journey between two sounds by analysing the sound into 513 windows in the frequency domain each pertaining to approximately 2 milliseconds of sound. To stretch a sound the program looks at a window and then creates an entirely new window exactly the same length as an original but now there are more. These new windows are interpolated between the original windows to ensure appropriate data. Vox V has perhaps the classic sonic example of this process in the transformation from the voice to the swarm of bees and the return to the voice at 2.13. These are two very distinct sounds. However, when listening to this transformation there is no true point of departure or return. It happens so smoothly that we are only left saying now it is a voice or now it is the bees. The difference is that in the analogue studio we would have two separate sounds and as one was faded out the other was faded in but they would be always two separate strands. With the computer it is one continuous sound that transmogrifies.

Major transformation points:
1. 1'06 -16 Crows/bubbles to sonic stream
2. 2'36 - 42 'snap' to ssss and then to voice
3. 3'45 - 59 in this episode are multiple transformations
4. 2'0 each of the 4 vocal phonemes is transformed, notably to horse and crowd
5. 2'14 - 40 voice-bees-voice - as an episode in itself
6. 3'52 multiplex continuum moving through to a single cry
7. 3'52 voice to horse
8. 3'57 baby to blowing lips.
9. 4'18 vox-horse-thunderclap

There are certainly more transformations within the piece but these seem the most significant. Others are more disguised as transformations and are part of the larger ea discourse.

**Stretching**

All the bell/voice sounds are the result of the analysis of the consonant vowel ‘ko’ which has a hard attack with an open vowel. Then certain of the upper partials of that sound object have been stretched inharmonically - in fractions rather than harmonic integers - and so creating the bell/voice. (This is also done using the window analysis technique of the phase vocoder.) It is somewhere in the middle ground where, at times, it is heard as a voice and at other times a bell. Our brains do not like the middle ground, we like to know what it is so we can make a decision either one way or the other (the survival instinct of friend or foe, to live a decision needs to be made very quickly). The repetitions of the bell/voice are all different but similar. The scalic, cascade upwards is probably not achieved by the transposition of the gesture but is rather an aural illusion as the same fundamental. It is still there but now certain upper partials are being highlighted and, in so doing, it gives the impression of transposition.

Bell/voice is first heard at 1.16 on the first attack of what I term ea discourse. There are no further attacks until the bell/voice episode of 1.41. After that it is heard twice more: at the first attack of 3.02 and again at 3.20. It has such a strong sonic characteristic that it acts as a central link throughout the piece. In the 2.40 episode it fulfills its compositional consequence.

With the phase vocoder it is also possible to stretch the sound over time. In an analogue environment the method to sustain a sound is to loop it - ultimately repetitive and static. The two phase vocoder passages: 1) 2.40 - 56", 2) 4.20 - 5'14, are voice stretched vastly. Others are more difficult to discern.

The final phase vocoder technique used in Vox V, but not in plain view, is the transposition of the sound object but with the retention of the original formants. Frequency shifting without this program will result in the speeded up record syndrome, particularly obvious with vocal formants transposed. The window analysis allows the frequency band over time to maintain its shape but now transposed. This allows radical pitch changes in the voice without it sounding like a tape recorded voice. Some of the sounds, notably the crowd source, have been frequency shifted non-analytically, but, those ‘classic’ tape techniques are not prevalent in the piece.

Most of the techniques are very much pioneered by Wishart when he was working on this piece at IRCAM, Paris. He wrote many of the processing programs using the phase vocoder analysis engine already developed at IRCAM. The use of these techniques have become, since this piece, prevalent as the powerful, number crunching computers have become more widely available in the domestic market, and a whole lot cheaper!

**References**

- Langer S. (1957): Philosophy in a New Key (Harvard University Press)
- Ferrara L.: Phenomenology as a Tool for Musical Analysis (Musical Quarterly LXX: 355 - 373)

Tom Williams is currently based at Boston University with a teaching assistantship in their studios whilst doing a DMA.
**VOX-5 diffusion score.**

This score is an approximate guide to diffusion.

If distant & close leaker pairs available at FRONT - START DISTANT & MOVE CLOSE

25secs

Im.05 Im.16 Im.33 Im.34

**WIND SOUNDS + CROWS**

CROWS BECOME VERY DENSE

VOCAL UTILIZATION

**SHAP**

**SPIT SYLLABLES**

(assesses moving around)

+2dB +2dB

**INPUT LEVEL** fade from zero 0dB

REAR OUTPUT

Keep rear leaks at -∞ until (no levels)

+6dB 0dB

1m58 2m13s. If WIDE & NARROW front leaker pairs available - move to narrow

2m39

3m.3s. 3m10s.

3m.52 3m.59

**Voice**

Various Crowds

Voice → Bees → Voice du → Bells

multi-voice Impressive 'screech'

K-crowd low drone

Voice to Horse

-4dB 0dB

quietly to make bees more 'realistic'

-3dB pull down these very bright sounds.

Voice out rear leaks (???)

+3dB +3dB

If diffuse leaker pairs available (e.g. upward facing side-pairs)

Use them (in addition to others) for this section as it evolves

3m59 4m15s

4m

5m

Frightened Crowds

Thunder Clap

Thunder clap 2

Final Thunder clap

+5dB

-5dB

OUT