Encounters in the Republic of Heaven[1]

Trevor Wishart

Received 23rd September 2019
Revision 31st May 2020

Encounters is an exploration of the music inherent in everyday speech. Unlike *Globalalia*, however, the aim of this project was to capture the musical features of speech at the level of the spoken phrase, its melody and rhythm, and the sonority of individual speaking voices, that indefinable yet recognisable something that enables us to distinguish one person from another. I especially wanted to capture both the diversity of human expression, and the sense of an entire community of speakers, a poetic snapshot of the diversity of human life.

As the piece would attempt to encapsulate this community of speech, I decided to work in 8-channel sound-surround, so that the audience would be embraced within this community. I had worked in 4-channel sound-surround in the past (*Anticredos* and *Vox*), but when the 4-channel analogue tape-format died in the late 1980s I found it increasingly difficult to perform any of my 4-channel analogue works, and became cautious about embracing multichannel formats. Hence my works over the previous 20 years had been in stereo format, which could be expanded in performance by appropriate diffusion over multi-loudspeaker systems (see the chapter on *Diffusion*).

For *Encounters* I both modified my compositional tools to operate in (any) multichannel format, and developed new tools specifically for this work e.g. multichannel sound-spatialisation software. Also, the award of the *Giga-Herz Grand Prize* in 2008 enabled me to re-equip my studio to work in 8-channel sound, so I was able to experience directly the sound-surround output of the piece. This was essential to composing the work, as electro-acoustic composition relies crucially on the direct feedback of listening to the created sound output.

Origins of the Project: The Source Material

The composition of *Encounters* posed quite different problems from *Globalalia*. To capture a diversity of human speech across an entire community I would need to:

[1] The text by Trevor Wishart that follows is the only non-original contribution in this issue of «Music/Technology». In fact, it has already appeared in a collection of writings edited by the composer himself entitled *Sound Composition*. We want to thank Trevor Wishart warmly for allowing us to republish here.

Musica/Tecnologia, 14 (2020), ISSN online 1974-0050, ISSN print 1974-0042, pp. 9-33
DOI: 10.13128/music_tec-12344 – http://www.fupress.com/mt

- find people who would agree to have their voice recorded;
- make many recordings, to capture a wide range of vocal styles and colours;
- capture natural speech (i.e. not the voices of actors, media presenters, or interviewees).

The original idea was to make the piece around my home county of Yorkshire, in the North of England, where I might at least be accepted as a (possibly slightly eccentric) local. But even here, it's not possible to simply wander up to someone in a pub and, out of the blue, ask to record their voice. And to get a representative cross-section of vocal 'colours', you would need to make many more recordings than would finally be used, involving lots of recording time and the logistics and expense of recording equipment, materials and travel, all of which required funding of some kind. I couldn't see any practical way to proceed, so I demoted the project to my wish-list.

Then, in 2006, the 3-year post of Composer-in-Residence based at the University of Durham was announced. Funded partly by the University and partly by the Arts Council, this job involved obligations both to teach students and to work within the local community, but its primary purpose was to provide the opportunity for a composer to pursue musical composition. For very particular reasons the Durham department was also keen to appoint an electro-acoustic musician. This seemed the ideal opportunity to realise my project in the North, particularly as work in the local community would be an essential part of it. So I applied and was fortunately appointed to the post.

The first few months work involved making contacts with local community and arts organisations through which I might meet people willing to take part in the project. As human voices alter significantly with age (particularly during childhood, but also in later life), I aimed to record a cross-section of voices, both male and female, across the widest range of ages. Schools and drop-in centres for the elderly were obvious places to investigate - recording working-age adults was more of a problem. Initially unsure of how the speech sounds would eventually be organised, I also recorded speech in the street and in markets and the crowd at the Newcastle football stadium, together with the voices of amateur choirs to whom I provided a skeletal structure for improvising with different speech colours. As the project developed, however, I decided to generate the massed voices from the voices of the individuals I had recorded.

To capture 'natural' speech patterns I wanted to avoid recording in a staged situation – a recording studio or an interview. Instead I went to people in schools, meeting centres, homes, pubs and clubs and encouraged them to talk freely with almost no intervention from me. Often it was only some way into the session that speakers began to relax into a more freely flowing style. In schools I quickly discovered that children had much more to say when talking amongst themselves in a group, so I usually recorded with 2 or 3 kids together in a small room. I also recorded types of everyday speech 'performance', in particular the sales patter of a travelling Butcher selling meat from a lorry in Chester-le-Street market.

Selection and Preparation of the Sources

Having gathered a large volume of recordings I now needed to select what I needed for the piece. There were five preparatory tasks:

- deciding how to put together these more clearly recognisable materials as music;
- selecting appropriate voices;
- selecting appropriate narratives of a suitable duration to use;
- cleaning the sources;
- extracting the musical information needed to make the piece.

By extracting the syllabic content of speech in *Globalalia* I had removed both the narrative content and the sense of individual personalities delivering the words. In the new piece, using entire spoken phrases, the subject matter being talked about could no longer be ignored - I would need to run with the narrative threads. More importantly, in *Encounters* the individual speakers were immediately apparent. The sound material could no longer be treated simply as a set of 'musical instruments' - the personal embodiment of the speaker had to be respected. This meant that the type of sound-transformation techniques employed would need to be restricted so as not to 'do violence' to the materials. I decided therefore to make a sequence of individual sound-portraits (or portraits of groups of children), preserving a narrative in most of them. These would be set amongst two other types of materials, the first using voices organised en masse in various ways, and the second using sounds abstracted from the speaking voices and deployed more like musical instruments.

From the many recordings made, I selected a set of voices with contrasting colours and styles of delivery. There needed to be a balance of male and female, and of ages (the very young, teenagers, adults, elderly people) but, most of all, voices with their own special features, so that sound-portraits with distinctive musical and sonic characteristics could be devised. Speaking voices can often be typed almost like singing voices - James Bell (*The Soldier's* Tale) was a bass and Alan Sambrook (*The Bellydancer)* a tenor (a fact which later becomes apparent when his voice is made to 'sing' at the end of the piece). Sometimes the use of the vocal range is important - Joyce Dent's voice (*The Dancer's Tale*) plays across a wide range of pitches, which had strong harmonic implications. Other voices have noticeable sound markers, Kathleen Teward's (*The Farmer's Tale*) a distinctive cross-break leap (often by the interval of a 5th), Edna Gallagher's (*The Budgie*) a sub-audio rasp.

Next, particular materials had to be chosen from each speaker. For a narrative we usually want to elaborate the story and fill it with descriptive detail. From a musical perspective, however, a limited set of sound-material (which can then be further developed, as sound) is usually preferable. Taking into account the total scale of the work, this suggested that each portrait should be around 2 or 3 minutes in length, and should perhaps focus on key phrases or sounds within the narrative. I had up to 2 hours of recordings from each speaker, and the large amount of material collected made the task of selecting a short, interesting narrative easier. Some materials could

be excluded for technical reasons (e.g. too noisy, too quiet, too many people talking at once), some for lack of 'naturalness' (e.g. too stilted, too unnaturally hesitant in front of the mike), some for reasons of narrative content (e.g. too repetitive, too predictable), and some for musical reasons (e.g. too similar in sound to already selected materials).

The final stage of this preliminary selection used standard sound-editing procedures for narrative or documentary, removing hesitations, repetitions and vocal glossalalia ('um', coughs, breath sounds etc.) then condensing narratives to their sparest form (removing narrative repetition, and so on). This material was then separated into short vocal phrases, the musical 'units' of the piece. When the portraits were later assembled, these narratives would be slightly rearranged from a musical/poetic perspective, for example repeating key phrases or interjections ('uh, huh'); smaller or greater adjustments towards a more metronomic pulse (done largely by adjusting the time-gaps between successive phrases, rather than by altering the phrases themselves); gathering together hesitations, repetitions and glossalalia for their expressive content (*The Dancer's Tale*), and so on.

Next began the processes of cleaning and cataloguing these materials. The decision to record in relaxed settings, like the home or pub, or in small groups (the children), in order to obtain 'natural' speech, created its own technical challenges. These materials had now to be cleaned up. To achieve this I gathered together existing CDP tools for editing, filtering, masking with silence, and spectral subtraction (of the noise floor) into an integrated package (the *Cleaning Kit*) and developed new tools, in particular a means to remove one of two pitches, when two voices are speaking (vowels) at the same time. Using these tools I painstakingly went through all selected materials, systematically eliminating blemishes, wherever possible.

Organising the Sources

To enable me to organise these source phrases I greatly expanded the *Properties File* tool developed for *Globalalia*. Clicking on the name of a property file now displayed it as a graphic table from which the listed sounds could be played. In addition there were graphical means to enter rhythmic patterns and motivic shapes, a mechanism to hear pitch motif and harmonic field properties and, most importantly, means to compare the property values and to extract statistics about them.

I used a large set of properties, not knowing initially what might be most important.

(1) Information about the origin and type of the speech material

- src source recording from which the phrase was taken
- who the individual recorded
- sex gender of the person
- age child, teen, adult, elderly

(2) Information about the musical content of the speech

- motif the sequence of pitches defined by the speech line (if any)
- HF the Harmonic Field defined by the pitches of the speech line
- tonic any implied tonic of the pitch line
- rhythm the rhythmic pattern of the speech e.g. "5:4", "fast6:8"
- MM the tempo of the speech, as a metronome mark
- rcode the rhythmic pattern of the speech, graphically encoded
- quality The quality of the voice - bright, light, soft, laughed, hoarse etc

(3) Information about the narrative content

- text the words spoken in the phrase
- keywords words emphasized by stress, or repetition within or between speakers e.g. 'never-ever', 'remember' etc.
- markers verbal markers e.g. 'you know', 'of course', 'aye'.
- hes-rep hesitations and repetitions e.g. 'but-uh', 'uh-aye', 'uh-uh'

(4) And

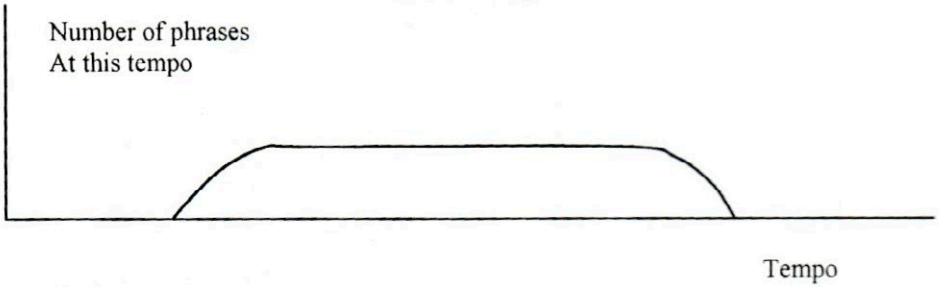
- ideas ideas about how to use the particular phrase.

Classifying the pitch properties of speech was not as straightforward as this might suggest. A speech phrase doesn't initially appear to have a melodic shape, partly because our attention is focused on the meaning of the speech and this is conveyed mainly by changes in the spectrum (the change from vowel to consonant to vowel) rather than by the pitch. Also the pitch is neither steady (it slides about) nor focused on any familiar scale. However, if we play a recording of a spoken phrase repetitively we usually hear a melody emerge – repetition reinforces our perception of the limited set of pitches which the speech-line uses, its harmonic field. We also tend to unconsciously approximate these pitches to the familiar tones and intervals of musical scales we know. For the "Speech Harmony" section I wanted to work with many speech phrases at the same time, so I needed their harmonic fields to fall on the same tuned scale system. I therefore approximated the pitches of the speech to the nearest tones of the tempered-scale when creating the "motif" and "HF" properties[2].

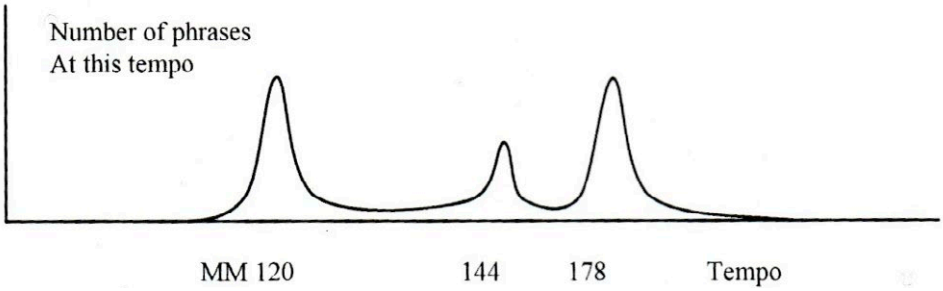
Speech has distinctive event-groupings and accents (the metrical feet of poetry) similar to musical groupings, and speech tempo is consistent within a small range but rather fluid, unlike the regularity of dance music. 'Rhythm' and 'rcode' captured the rhythmic pattern of the spoken phrase as closely as possible, and 'MM' recorded its tempo.

[2] This was helped by the fluidity of the pitch. I often came across the situation where, on a first analysis, the notes of a phrase fell in one tonality, say A minor, especially when heard against other A minor material, but the same phrase could as easily be heard in Bb minor, when set against other Bb minor materials.

More important were the Statistical Tools which helped me to determine which Harmonic Fields, which Tempi, or which words (word-starts etc) were most common amongst my collected materials. For example, although it's clear that speech tempo falls within a fairly narrow range, I expected different tempi to lie smoothly across this range, perhaps like this:



When I plotted the tempi of my collection of phrases, however, I came up with a tempo plot like this:



MM 178 is close enough to MM 180 to be one and a half times MM 120, the ratio of, say, quavers to triplet quavers, and this immediately suggested a way to organise the rhythm of the speech material (used in the central interlude of Act 1)[3].

Statistical analysis of the text (set to search for sub-phrases, common words, word-starts, rhymes and assonance) uncovered common words and syllables (like 'remember', 'different' and 'li') used in constructing the 'Clouds of Speech' central interlude of Act 2.

Later, in order to organise the speech-phrases in terms of their implied harmonies (central interlude of Act 3), I needed to find appropriate vocal phrases. As much speech falls in a narrow range, many phrases generate a chromatic cluster of small intervals as their harmonic field. I didn't feel that these clusters would produce a sufficiently interesting musical result so, using the search tools, I looked for phrases which had

[3] It's interesting that MM 120 corresponds roughly to double a fast walking pace, and is the tempo of much late C20, early C21 dance-music, while MM 144 is a typical Allegro in Western classical music.

larger intervals (a minor 3rd or greater) in their harmonic fields, ascertained which of these (with specific pitches) were the most common and, given these, decided which might be combined to produce interesting chord formations.

This is one example of how the process of cataloguing materials, particularly in its early stages, constantly threw up new ideas or insights about how to organise the piece, on the small or large scale. Hence the "ideas" property was added to the table to enable me to keep track of these thoughts, and cross-reference between the ideas themselves and their associated sounds.

After the first year of establishing contacts and making recordings, this cleaning and cataloguing stage of the work occupied me for the following 18 months.

The Form of the piece

The final form of the piece gradually crystallised around the extracted narratives in the 2nd year. I envisaged the piece falling into four separate Acts, each presenting four sound-portraits projected in front of the audience in wide stereo.

These would be embedded in an 8-channel sound-surround sea of vocal or voice-derived materials. Acts 1 and 3 (the start and mid-point of the work) would start with introductions using the "Voicewind" sound generated from a very dense mass of speaking voices, and the entire work would end with the voices dissolving back into this sound.

The first 3 acts would also contain a central interlude using all the voices in sound-surround, each organised using a different common property of the speech:

- for Act 1 the rhythm of the spoken phrases;
- for Act 2 the sonority of words or syllables;
- for Act 3 the harmonic fields of the spoken phrases.

The finales of the first 3 acts would seek to draw out key musical elements from the voice-portraits within the act, and develop them into a more abstract, purely musical world. This helped to determine the final order of the portraits (which also took into account the balance of vocal quality, gender, age, musical treatment and narrative content) as the sound materials from the portraits within an act would determine the musical nature of these finales.

Act 4, starting as Act 2 with 2 portraits, would merge together the interlude and finale, drawing on and developing 8-channel materials from the earlier acts, and leading to the final transformation of the spoken voices into song, before the voices remerged into the Voicewind.

The portraits themselves were each to be approached differently, depending on the sonic character of the vocal material. The technical "abstraction" of vocal-character (one of the technical goals of the piece) proved elusive, too complex a web of favoured rhythms, pitch-contours, accentuation-tics, glossalalia, verbal-markers and so on to be extractable using computer algorithms, and I had to fall back on musical intuition.

The final form of the piece looks like this.

ACT 1

Introduction:	Voicewind
Portraits:	The fisherman's tale
	Children's stories I (mainly boys)
Interlude:	Speech waltzing
Portraits:	The budgie
	The bellydancer
Act 1 Finale:	

ACT 2

Portraits:	Teens gossip
	The dancer's tale
Interlude:	Clouds of speech
Portraits:	Travelling butcher
	Children's stories II (mainly girls)
Act 2 Finale:	

ACT 3

Introduction:	Voicewind
Portraits:	The farmer's tale
	Heathcliffe, come here!
Interlude:	Speech harmony
Portraits:	The stories I hear
	The soldier's tale
Act 3 Finale:	

ACT 4

Portraits:	The poet's tale
	The best thing that ever happened to me
Interlude/Finale:	
Coda:	Voicewind

A Sound-Surround Working Environment

For Encounters, a new environment had to be built for mixing sounds with more than 2 channels. I had developed a way to specify the signal routing from an input sound (with any number of channels) to an output space (of any number of channels), and a mixing process allowing multichannel files to be mixed together in an entirely flexible manner. To avoid having to write out the complicated-looking text-files controlling this process, I developed a high-level tool which would alter lines of the mix text at the click of a button e.g. to change the output level on all channels at once, or copy the signal routing of some channels to others, or from one file to a file with a different number of input channels. I also added a means to alternate easily between building musical phrases in the mix environment and building sounds themselves.

Algorithms to generate convincing rotation around or panning across the ring of loudspeakers were developed. Panning the signal rapidly across a changing (random) permutation of all the speakers produced a strange sense of everywhereness which can be heard with some of the "It's a bloke" transformations in the Finale of Act 1. Once a multichannel sound-image has been made it is possible to put the entire frame in motion e.g. spinning the whole 8-channel image clockwise, a procedure I'd speculated about in a chapter in "On Sonic Art" in the 1980s. And I added wrap-panning, taking a stereo image, itself already in motion across the stereo panorama, and moving it through the 8-channel space, e.g. out of a mono-condensate to a stereo display at the front, wrapping it around the listener to a stereo-display and thence a mono merge at the rear.

Adding reverberation to signals in a multichannel space also raised new issues. In stereo, reverberation is produced by very many echoes of a sound returning from different locations in the stereo space, creating the sense of sound being reflected around a room. With a multichannel space one could either create these reflections around all the channels, so the sound appears to resonate around the entire surround space, or e.g. make 8 signals in the 8-different channels reverberate from a restricted (e.g. stereo-width) space around their current location retaining a strong sense of the original spatial layout (channel-centred reverberation).

In addition, processes such as texture-generation and brassage, which originally took mono input sounds and generated stereo output, were extended to generate output in a multichannel space.

The musical sections

It's not possible to describe in detail every compositional process used in *Encounters*. I will outline some key features in each section. As a general observation, each process applied would typically vary over time. And several such processes were often progressively overlaid. For example we might begin by applying a time-varying time-stretch to our source; we then apply a low pass filter to the resulting output, increasing the filtering progressively; we then gradually shift the pitch of the end of the sound, upwards;

we then introduce slightly randomised tremolo to the end of the sound, gradually increasing the tremolo depth from zero, and progressively changing the tremolo frequency; we then dovetail the end of the sound (its level falls away to zero), and so on.

Introductions, Interludes & Coda

Introductions Acts 1 and 3: Coda Act 4: Voicewind

(*All examples from 8-channel sections are stereo reductions from the 8-channel originals, so cannot always convey the exact musical output composed; the audio examples are available in the data set at the following URL: https://zenodo.org/record/4240436).*

If many different vocal phrases are overlaid in a texture, the sound space becomes increasingly crowded – we hear human voices, but not individual speech. When the density approaches a thousand vocal phrases per second the sense of human voices is almost completely lost, and by ten thousand voices per second we hear only a wall of noise, the average colour of all the voices, spread around the 8 outputs. A stereo sound which presents such a space-filling image can be made to shudder across the stereo space[4]. Treating each adjacent pair of the 8 output channels as 8 stereo pairs, independent shudders were applied to each pair. The resulting sound appears like the noise of a gale force wind blowing around our ears in a storm. At a lower density, where human voices can still be heard, the shuddering effect is not apparent (we seem to hear only random fluctuations of loudness of individual voices in a very dense mix), so we can make a transition from a mass of speaking voices to shuddering wind by gradually increasing the density.

Interlude Act 1: Speech Waltzing

Collecting together spoken phrases at MM 120 and synchronising their attacks as closely as possible produced only a sense of a crowd of speakers – the underlying rhythm of the phrases wasn't apparent. The phrases were therefore subtly retimed[5] so that attacks were precisely synchronised, whereupon the many voices locked into a clearly pulsed rhythm. This approach was also applied to the MM 178 materials.

The opening of the waltz uses such rhythmically retimed phrases but each word in each phrase is sent to the next loudspeaker in the ring, the phrase thus stepping rapidly around the ring (stepped-rotation) **(Sound Example 117)**. In the main materials of

[4] A random frequency tremulation is applied to a stereo sound such that the loudness peak of any one tremulation is (randomly) different in the two channels. If the sound is a space-filling image (not one image to the left, and another to the right) the tremulations seem to move randomly around the space.

[5] After experimenting with various alternatives the simplest approach proved best – adding tiny slivers of silence in the troughs between syllables, or removing tiny slivers of sound from the same place.

the waltz, whole vocal phrases from different voices enter on successive loudspeakers around the ring, but the phrases themselves do not move. As the phrases had different pitch-ranges, lengths, qualities, and emphases, each had to be carefully chosen to tell effectively against the previous and following overlapping phrases. In the overall form of the interlude, sections in 2-time (MM 120) and 3-time (MM 178) alternate.

The rhythm is pushed forward by tutti moments (voices synchronised on all loudspeakers) emphasizing the rhythmic pulse **(Sound Example 118)**. Canon and echoes (synchronised to the pulse) are also used at some points. Occasional bursts of double-speed material, some randomly panned, animate the mix. Laughter and time-stretched exclamatory words in continuous-rotation (rather than stepping from channel to channel) - lubricate the flow. Towards the end, clipped syllables enter at double tempo on successive channels around the ring, a very rapid stepped-rotation **(Sound Example 119)**.

Interlude Act 2: Clouds of Speech

The Properties files were used to find common words and syllables, and relevant phrases containing these automatically retrieved. Words, and sub-phrases containing them, were edited out and grouped (e.g. remember, remembering, memory).

Words were chosen because they were used by many speakers and on many occasions, providing a wide variety of sources; because they contained different sounds to the words chosen in previous clouds (different vowels; use of sibilants or liquids); Sometimes for their meaning (remember", "never").

Event clouds were then assembled based on particular words or syllables. The "Remember" cloud assembles short clusters of very short phrases containing the word "remember etc." **(Sound Example 120)**; strong channel-centred-reverberation (see above) of these phrases, producing a different event **(Sound Example 121)**; the syllable "mem" repeated, rotating around the sound-surround space - several different "mem"s rotate together, but at different speeds, and with different rates of repetition, a prominent motif in the Finale of Act 4 **(Sound Example 122)**; a heavily shredded[6] version of this scattered over the sound-surround space **(Sound Example 123)**.

The "Something" cloud, eventually dissolving in its "s" sound, assembles short phrases starting with the word "something" to form a sound-surround texture - this accelerates, rising in pitch (sample-rate change), with its low frequencies gradually filtered away, and the whole image rotates at a faster and faster rate **(Sound Example 124)**; the reverberated end of this mixed to mono, tremulated rapidly and rotated rapidly, used as a separate strand; the same combined in canon, the two images rotating, in parallel, 180 degrees out of step **(Sound Example 125)**; time-stretched versions of the original phrases; aggregates of the "ss" sounds within the phrases, some time-stretched.

The "ll" cloud combines several strands of material each from an individual voice, using multichannel sequences of "li", "la" "lo" syllables, swapping from channel to

[6] Sound is cut into random length chunks, and the chunks rejoined in a random order. The resulting sound is the same length as the original, and can be shredded again, then again and so on.

channel; the same material with iterated repeats[7] of each syllable, each fading into its own reverberation.

The event based on "Different" works similarly but, as it progresses, focuses on the "diff" of "different ", then just the "ff", eventually using time-extended "ff" sounds with tremulations **(Sound Example 126)**. Similarly, the event based on "good" dissolves into a sea of "g" sounds.

The "never" event stresses the pitches of words or syllables through reverberation, or time-stretching, some extended events pulsing and/or rotating round the space. Others use very long spectral-time-stretches which have a "metallic" resonance **(Sound Example 127)**.

Interlude 3: Speech Harmony

Using the Properties Files statistics I searched for pitch-groups common to motifs in several different voices, and containing larger pitch intervals. In addition, as adult male voices are typically lower than adult female voices, and both usually lower than young children's voices, I needed to find the most common useful pitch-groups in each range, and decide how these might be combined into harmonies. Eventually a set of 6 harmonic fields were selected, together with the vocal phrases using their pitches. For the 7th field, the low C which ends the sequence, I took vocal lines which hovered around that pitch, and adjusted all the voice pitches closer to the C.

To focus attention on the pitch content, harmonic field filters[8] were constructed, tuned to all the pitches of a chosen set of motifs, and the motifs passed through the filter bank. The first harmonic area in the first half of the section mixes an array of the original voices, one on each channel **(Sound Example 128)** ; the same material mixed in mono and resonated through pitch-following filters **(Sound Example 129)**; the same mono material filtered in such a way that we hear pitch resonances, but not vocal features (like sibilants), all reverberated **(Sound Example 130)**. The mono variants are projected through every loudspeaker so the harmonic field is everywhere. In the second half of the Interlude, these mono materials, reverberated, are made to rotate slowly around the space.

The second harmonic area adds to the reverberant strand a randomised fluttering[9] over the 8 output channels. In the 2nd pass there are two versions of this fluttering

[7] Exact repetition (looping) of materials has a particular synthetic quality. The CDP process of Iteration adds slight randomly variations to the timing and pitch-shift of each repetition of the sound, producing an output more like a naturally-occurring iterated sound.

[8] I developed a method to harmonically-colour the massed material by using a bank of filters tuned to specific harmonic fields. By (gradually) varying the tightness (Q) of the filter, these massed sounds could be strongly or subtly harmonically coloured (or not) as the music progressed.

[9] Tremolo is applied to a multichannel sound such that the sound rises in level in each loudspeaker in turn. The sequence of channels used for this can be specified in a regular pattern or as a random permutation of all of them (followed by a different permutation, etc. – this ensures all output channels are used equally often).

material, one echoing the other, slightly delayed and 180 degrees apart (opposite) in the sound-surround space **(Sound Example 131:** *simulation of 8-channel effect***)**.

Some filtered materials, strongly disguising their vocal origin and each with different filter parameters, were combined by sequential cross-fading to create long, pitch-oscillating pedals. At the very end, the low C material of this type is transposed upwards by various octaves and 5ths and these versions combined so the sound appears to swell across the register, and subside into the bass once more. At the same time another strand on C, strongly filtered but still vocally recognisable, rotates around the space **(Sound Example 132)**.

The Portraits

(most of the portrait examples are from the stereo originals)

The Fisherman's Tale

Two key ideas here were capturing and time-extending the Northumbrian burr and developing a recurring motto phrase based on the word "herrn". The burr is a characteristic feature of Northumbrian speech, rolled softly on the arch of the tongue, a little further forward than the French "rr". This occurs in the word "herrn" (herring).

Grain-extension[10] was used to greatly extend the burred "rr" in several occurrences of the word "herrn". Shorter and longer extensions, and (static or time-changing) pitch-shifts up one octave, and down one and two octaves were made (without changing the vowel shape or rate of iteration of the "rr"). Combined, these form phrases which descend to the lowest octave using loudness enveloping to crescendo to the final "...rr-n" of the last event **(Sound Example 133)**. More or less reverberation could be added to the tail of this sound.

In parallel the syllable "he" of "herrn" was combined with a reversed copy of itself ("eh"), and the output repeated to produce an ululated version of the syllable. The process was applied to the syllable in four occurrences of the word "herrn" with different pitches, and the ululating lines combined into short phrases **(Sound Example 134)** used on their own or to precede the octave-descending phrase. This ululating material is decelerated as an anacrusis to the final strong cadence.

In addition "Cos you didn't mek nowt" and "Coming aboot (salmon)" are extended by rhythmically overlaid repetition. In the latter case, the phrases gradually fall in pitch (preserving the vocal vowel shapes). The phrase "years ago" is extended by texturing first "years", and then "ago", with no pitch-change of the sources, so that the 3 pitches of the phrase are foregrounded **(Sound Example 135)**. The "o" of ago is then time-stretched, so that its pitch is sustained. The phrase "was still gan" is filtered

[10] The same process was used in the "*Ma-rrr*" section of *Globalalia*.

with a pitch-following-filter[11] and transposed up two octaves to produce the seagull-like sounds which echo the pitch-contour of that vocal phrase **(Sound Example 136)**.

Childrens Voices I

This portrait combines story fragments told by several small boys, attempting to retain the overexcitement of many of the tellings. Various musical objects are created.

Back-to-back versions[12] were made of moments which step between clear pitches. These pitch-oscillating events were overlaid in various ways to make assemblies of characteristic harmony. The end of this material was then strongly reverberated, and this material used independently **(Sound Example 137)**.

The "tha..t" at the end of "and that was that" was greatly time-stretched by roughness extension[13], then made to rise in pitch (sampling-rate change) **(Sound Example 138)**. Further variants were made, using spectral time-stretching, spectral-tracing[14], and adding tremolo to the stretched tail. This becomes a musical motto.

The clown story falls over itself in excitement, accumulating one bizarre event after another. This is exaggerated by the musical treatment, mixing a layer of variously time-stretched versions of the story with a layer of start-synchronised copies. The copies in this second layer are placed at different positions in the stereo space and each accelerates at a minutely different rate so that stereo echoes gradually emerge **(Sound Example 139)**. Each thread of the premix also gradually merges into a shredded[15] version of the text. Towards the end of the portrait the clown-story is harmonic-field-filtered[16] with increasingly tight filtering, so the pitches of its harmonic field begin to resonate strongly **(Sound Example 140)**.

In the escaped cow story, overexcited delivery of the word "pump"[17] produced "pfump" (an accidental vocal imitation of the cow's flatulence). "Pf" is developed into noisy extended aggregates which are later greatly extended in the Finale of Act 1 **(Sound Example 141)**.

The breaths in two very breathy voices are selected and some developed by time-stretching. In the Finale of Act 4 this material is further stretched, with time-reversals, becoming more like pneumatic machinery, then trembling and rotating round the space **(Sound Example 142)**.

[11] The Harmonic Field filter mentioned previously, now confined to a single pitchline, which follows the pitch of the vocal line being filtered.

[12] A time reversed copy of a sound is made, and the original sound edited onto the end of this.

[13] This is the grain-extension process, mentioned elsewhere, applied to the roughness in a vocal sound.

[14] *Only* the most prominent partials in the spectrum are retained, *on a window-by-window basis.* See my book *Audible Design* for more details.

[15] See Footnote 6.

[16] See footnote 8.

[17] North of England word for "fart".

The Budgie

The principal idea here was to capture and extend the grittiness of the voice. A roughness-extension procedure[18] was used to capture and extend grittiness fragments in the voice. Once captured and extended, the rasp-stream could be panned, and/or decelerated, either without changing the duration of the grains, or with time-stretch **(Sound Examples 143)**. The sibilant "s" was extended using a technique similar to roughness-extension[19]; the extended sounds pan across the space, sometimes with a slow slide in tessitura and/or a slow or rapid time-varying tremulation **(Sound Example 144)**. Individual vowels are emphasized by time-stretching and/or reverberation (e.g. the first "e" of "Peter"). Some syllables or words are stereo-bounce echoed with ("Blue") or without ("Peter") pitch-focusing filtering[20]. Very softly-spoken phrases are aggregated **(Sound Examples 145)**, the aggregates punctuating the main narrative.

The Bellydancer

This is the only section of the piece to use any sounds *not* derived from the voice recordings themselves. A small brass ensemble picks up prominent pitched-phrases in the speech-line as the story progresses, and these are folded into a simple rhythmic brass-band accompaniment **(Sound Example 146)**. The brass ensemble was enhanced through changing the spectrum of sounds and by mixing sounds together, to generate a larger (imaginary) ensemble. Other modifications were made for specific moments (reverb, excessively wide vibrato etc.).

The principal sound idea is the capture and extension of the rasp-sonority of 'bloke' in the motto phrase "It's a bloke". The "oke" of "it's a bloke", when first heard, has been shortened (by editing). It next occurs in its original form and later is extended using the roughness-extension process **(Sound Examples 147)**. This extended sound becomes important in the Finale of Act 1, where a sequence of exponentially-falling envelopes counted over wavesets[21] generate a bell-like sequence, itself then dissolved by spectral tracing **(Sound Examples 148)**. At the climax of the story, the original speech is enhanced by echoes, texturing, pitch-shifting by octaves, and time-stretching. Part of the reverberant echoes which spread in stereo behind the time-stretched "I've been over there…" are made by extracting the peaks from the vocal phrase and time stretching them 32 times **(Sound Examples 149)**. The very deep transformation uses multiple processes gradually applied to the source ("it's a bloke") including pitch-following filtering, time-stretching producing pulsations (a phasing artefact of spectral time-stretching) and low-pass filtering.

[18] See Footnote 13.

[19] This process searches for noise-bands, rather than iteratives, in the vocal stream, then extends them.

[20] See Footnote 11.

[21] Wavesets are delinated by zero-crossings in the signal. In a complex sound these are irregularly spaced in time. Hence the envelope-attack onset-times, based on counting these, are slightly irregular.

Teens Gossip

Here the aim was to capture the essence of gossip in a rhythmic musical device, hiding the text-content of the gossip. Examples of exclamatory speech were used as a foil to this rhythmic material. Using peak-extraction the peak of each syllable in the speech is grabbed and these peaks reassembled at a regular tempo[22]. This retains the pitch-contour, vowel sequence and changing expressivity of the speech, while losing the specific words used **(Sound Example 150)**. These rhythmic gossip-lines are articulated in various ways; call-and-response in stereo, repetitions of elements in the line, larger-than-life tremolo, texturing of several strands. One gossip phrase is time-stretched to twice its length (half its speed), a cadencing device just before the portrait ends. And at the end of the finale of Act 2, one line of this gossip material recurs in a repeating cycle, circling rapidly around the 8-channel space.

The Act begins with rising phrases from this rhythmic material, used as echoed snippets with reverberation. These recur in textures which are sometimes octave transposed to higher registers (and correspondingly speeded up). These animate the eight-channel space around the front-stereo "story-telling" **(Sound Example 151)**. The gossip material is interrupted by blocks of several exclamatory voices in which particular words can be heard. These blocks evolve e.g. rising in octave-shifts with no vowel change. Some exclamatory words (e.g. "whopping!") are vastly extended by time-stretching and panned across the stereo space.

The Dancer's Tale

In this portrait we work principally with the pitch implications of the wide-range vocal line. The hesitations and glossalalia in the voice were also collected together to form recurring assemblies of hesitation, resonated in their own harmonic field **(Sound Example 152)**.

To resonate the pitch of the vocal lines, two types of filter were used, a pitch-following filter (PFF) which is a time-varying filter following the changing pitch of the vocal line[23], and a harmonic-field filter (HFF) which is a bank of filters whose tunings do not change with time, but are tuned to all the pitches used in the vocal line (the Harmonic Field of the phrase)[24].

[22] Using a coarse time-window we track the fluctuations in the loudness of speech which correspond to the individual syllables within words. The envelope peaks tell us where the syllables lie, and this information can be used to slice out a specified-duration segment from each syllabic peak (adjusting the cut-points slightly in time to capture the start of the syllables if possible). These can then be reassembled at some specified tempo (or in a specified rhythmic pattern). The result is a rhythmic stream of events which captures both the melodic contour, vowel content and changing expressivity of the speech, whilst losing the specific verbal content.

[23] See Footnote 11.

[24] See Footnote 8.

By using a sequence of PFFed echoes of the vocal line, with increasing tightness (Q), and filtering with decreasing numbers of harmonics, the vocal line appears to dissolve into pure melody as we pass from one echo to the next **(Sound Example 153)**. The tail of the echoes can be extended with reverb or time-stretching, while adding time-varying vibrato and/or tremolo to it. An HFF based on the pitches in a particular vocal line can be applied to that entire line, or to a texture of sounds cut from the line (without changing the pitch, therefore preserving the harmonic field). The filter-resonated events can evolve, as different Q-valued versions merge into one another, perhaps with slight tremulations or randomised corrugation of the envelope[25] as they evolve. In this way the harmonic implications of speech lines can be extended in diverse ways **(Sound Example 154)**. Individual words or syllables (e.g. the "awe" of "awestruck") were highlighted by time-stretching and sustained PFF resonances. Later, these same resonances recur, with variation **(Sound Example 155)**. Each syllable of a word can be extended in different ways **(Sound Example 156)**.

In addition, slightly time-stretched echoes of the (possibly PFFed) phrase, suggest the words themselves linger in the resonant space; sibilants and plosives ("s" and "p") were extracted and used to form textures, lingering around the original text **(Sound Example 157)**; the rhythm of "Bob and I go ballroom dancing" was emphasized by rhythmic repetitions, dissolving into filtered variants **(Sound Example 158)**.

The Travelling Butcher

The travelling butcher amplified his sales patter, powering it from a noisy generator attached to his lorry engine **(Sound Example 159)**. The first problem was to separate his voice from the rumble of the generator. The thin quality of the extracted voice, however, gave it a sonic character distinct from the other portrait voices, and the lorry rumble itself could be used as a distinct musical element.

First of all we use spectral subtraction[26] to remove the generator's rumble from the voice. As the rumble is very loud, the remnant voice has lost all its low-frequency components and is strongly coloured, as if heard through a poor-quality walkie-talkie **(Sound Example 160)**. The transformed voice was then cleaned with the Cleaning Kit (see "Cleaning"). To obtain the lorry, parts of the source where the voice was silent were separated out and joined together, then extended by repetitions to produce a sound of generator-rumble only. A pitch-focusing filter, of low Q and at the pitch

[25] A sound like laughter (ha-ha-ha-…) has a loudness contour loud-quiet-loud-quiet- etc. A rolled-rr sound has the same contour at a much faster speed. We can exaggerate this contour by making the area of quietness slightly longer and the loudness correspondingly shorter (we are not time-stretching the sound here, just altering the loudness envelope). This is envelope corrugation.

[26] A portion of the recording with generator sound only (no voice) is spectrally analysed and the maxima, in all spectral channels, are subtracted from the complete (generator + voice) recording. For a very low level noise floor, this is an excellent means to remove unwanted noise, but in this case the lorry is sufficiently loud that, on subtracting it from the voice, we lose most of the voice's low frequencies.

(and just 2 harmonics) of the most prominent pitch in the rumble, is applied to focus the sound **(Sound Example 161)**.

The rumble material was then used as a distinct musical strand, pitch-shifted up by 1, 2, 3 and 4 octaves (transposing the spectrum), and these transpositions gradually cross-faded so that the rumble rises in tessitura. The material of the highest transposition was then cut into segments, isolating individual 'plops' of the sound, and these used to form a texture (sounding much like the pre-textured source). The sounds within this were gradually pitch-shifted upwards (sample-rate change) and the texture then cross-faded into the end of the transposing rumble, to extend the tessitura of the complete sound upwards, with variations **(Sound Example 162)**.

The extracted voice is progressively processed, first by spectral-tracing, then spectral-blurring and low-pass filtering, losing clarity as it dissolves. This complete strand wanders slowly away from centre stage. A copy then enters at the centre, taking the opposite route, then a third. The three gradually merge to a mono moving line, and vibrato is added, the verbal content completely obscured. Finally the peaks of the obscured-voice are abstracted and retimed at approximately double the tempo of the transformed-lorry 'plops', just before the original voice-line recurs **(Sound Example 163)**.

In the second half of the portrait the materials begin again at their starting point, but they are progressively time-contracted, and mixed with accelerating vocal lines in an accelerating spin through the stereo space.

Childrens Voices II

This portrait combines story fragments told chiefly by girls. Harmonic implications of stories and phrases are extended. The various "Do you know?" phrases are used as a recurring refrain, sometimes clustered, leading to the girl's story towards the end of the portrait. Various speakings of "Humpty" and "Dumpty", from the telling of the nursery rhyme, are aggregated, and the aggregates progressively harmonic-field-filtered. The ends of the aggregates are reverberation-resonated. This resonated material is extended and used in its own right **(Sound Example 164)**.

The hamster-tale voice has characteristic rising intonation. This is emphasized in a texture made from these rising gestures. Towards the end this texture is 'plucked' by steep-falling loudness-envelopes (based on counting wavesets) **(Sound Example 165)**. The rising scale of the laugh in the "wishing star" story is time-extended by regular repetitions spread across the stereo space. An abstracted variant is made by harmonic-field-filtering (few harmonics and tight filtering), disguising vocal features.

The rising cry "I went out" at the end of the aeroplane/pushchair story is aggregated with various pitch-following-filtered versions of itself, retaining the characteristic harmonic field **(Sound Example 166)**.

The Farmer's Tale

This voice has very characteristic cross-break articulations. These were used to create a bagpipe-like instrument to accompany the story. The various vocal-breakings were first cut out from the source and then passed through pitch-following filters with the appropriate number of harmonics and Q-value, preserving the pitch-leaps in the voice. The ends of individual notes were extended or contracted by time-stretching, and more or less of the start-articulation included, depending on the note's position in the 'bagpipe' phrase. In this way enough individual bagpipe-articulated notes were generated to create a virtual instrument which could be used to accompany the original spoken voice **(Sound Example 167)**. These articulated notes are further developed in the Finale of Act 3 **(Sound Example 168)**.

In addition the phrase "round and round and round" is extended by rhythmic repetition, later combined with a similar treatment of "as you go round". The principal pitches of the two phrases lie on a pentatonic scale and this material, itself rotated in the stereo space[27], is used to accompany the storytelling **(Sound Example 169)**. The occasional 'fluttering' of syllables in the original voice is quietly extended, a sound feature used in the Finale of Act 5 **(Sound Example 170)**; a harmonic field resonance from a phrase lingers, and so on.

Heathcliffe, come here!

In this portrait several extended narratives are presented, all pushed nearer to the same regular moderato tempo, close to the original, with repetitions of individual phrases, spatial antiphony etc. used to underline this, increasingly so as the portrait progresses.. As the tales unfold, individual words or phrases are sonically developed, becoming recurring musical elements, articulating the musical flow.

Individual FOFs[28] within the phrase "Heathcliffe, come here!!" are captured and time-frozen on their particular pitch to generate several variants of the phrase. Variable vibrato is added to the sustained FOFs to give a more plausible sung quality. Further variants are made by panning across the stereo space. Later, some of these have part of their spectrum spectrally pitch-shifted up or down by octaves producing chords in the voice, or events in higher registers. All these variants become "recurring" musical elements in the portrait **(Sound Example 171)**.

[27] The spatial circulation is created by left→right→left→ etc. panning, slowing at the edges of the space, and an increase in level moving to the centre in one direction, but a decrease (plus some filtering out of high frequencies) when moving in the opposite direction, creating a sense of front and rear for the rotation. The level changes rotate slightly ahead of the spatial panning (so the sound is at its loudest just after leaving the loudspeaker, as it moves 'towards' us, rather than when it reaches the centre of the space) giving a more psychologically convincing illusion of rotation. Also, as no doppler shift is used, the illusion is that of observing rotation in a *magnified small space* rather than rotation around a large spatial arena.

[28] The voiced sounds in speech are produced by a sequence of small wave-packets or FOFs.

The mainly unvoiced "..cliffe" part of this phrase and, later, various speakings of the word "pencil" (with strong plosive and unvoiced elements) are elaborated through the repetition of tiny different-sized chunks within the sound, selected by counting specific numbers of zero-crossings (waveset repetition). "Pencil" is also elaborated by textured repetition, superimposing different speakings of the word, mixes of all of these with the waveset-repetitions, time-stretches of the tail of "pen", aggregates of "pe", and random shredding of the word itself **(Sound Examples 172)**.

A patterned extension of the word "democracy" is created by chopping it into random chunks, then repeating these in a regular permutation sequence (defined by an English bell-ringing pattern). This is then cross-faded into a version of the same sequence of events, with each event pitch-focused-filtered at its own pitch **(Sound Examples 173)**. This is further developed in the Finale of Act 3 (hear **Sound Example 191)**.

Other words and events ("afternoon", "look", the laugh on "had ") are elaborated in various ways, "afternoon" being sung as a chord by superimposing pitch-shifted copies which retain the vowel-character of the original, then gradually adding a pulsating tremolo **(Sound Examples 174)**.

The Stories I Hear

Sylvia uses many different voices to tell her stories. I decided to use expressive snippets from the stories, juxtaposed with developments of various growls, yells and break-crossings from the voice. The phrases "then all of a sudden, it's dangerous..." and "the stories I hear" are repeated to give the portrait a song-like structure. The audience's "oo" and laugh responses were gathered into stereo groups, and become recurring musical elements **(Sound Examples 175)**.

The growling voice of "have you had a go(rr) an em", was developed, creating stereo mixes out of repetitions of part-phrases; textures of the growly element "had a" at regular intervals and increasing densities; and differently engineered roughness-extensions of "go(rr)" which are then rapidly panned. The last of these morphs briefly into a watery sound (using tiny displacements of the pitches of the grit particles during roughness-extension) **(Sound Examples 176)**.

For the high-pitched yell of "are you still using that thing", several individual FOFs[29] from the vowels are extracted. Then each is used as a very bright "instrument" playing at the original pitch-line of the voice (sometimes transposed down an octave). Several such instruments are combined into a squawking aggregate **(Sound Examples 177)**. Several variants are used. One spins round the stereo space, accelerating and rising in pitch as it fades.

In the yelled phrase with break-crossing "Three years on the bloody trot", individual syllables or words which jump the break in pitch (and some time-reversed versions) are aggregated into a stereo texture, preserving the pitches. A variant of this transposes the texture up an octave, preserving the vowel shapes **(Sound Examples 178)**.

[29] See footnote 27.

The Soldier's Tale

In this portrait, the sub-phrase "so close" is used as the source of both pitched and unpitched materials to be developed, especially during the Finale of Act 3. The "s" sounds from "so close" are extracted and assembled into a stereo event with a regular pulse, repeating like a shaken sand-filled rattle. This is made from the two syllables with their low frequencies filtered out, hence focusing on the sibilants. The resulting pulse streams are corrugated (see earlier) **(Sound Examples 179)**. In the finale of Act 3 this stereo materials returns, faster and more focused.

The falling minor 3rd of "so close" is developed. Very long reverberation added to the "o" of "close", greatly extends the lower pitch. This extended phrase is also transposed down a minor 3rd so the lower note of the original is in tune with the upper note of the transposed phrase – and also with the "o" of "ago" in "a long time ago". The rhythmic repetition of the two pitches (with randomised sequencing of pitches, and their accentuation) occurs half way through this portrait, becoming a key feature of the Finale of Act 3. This is made in the same way as the sibilant stream but with the sibilants filtered out, hence focusing on the pitched vowels. It is then corrugated **(Sound Examples 180)**. Towards the end of the portrait, the stream is pitch-shifted up an octave plus a 6th.

In addition, some phrases (e.g. "there was a ditch at the side") are echoed by soft brass-like sounds (pitch-following filtering of the vocal line), and a final variation of "a very red mark that's all" is transposed down an octave (without changing the vowels). The pitches of the vowels of this phrase are resonated by a harmonic field filter and extended in time and tessitura to generate the chord which swells over the end of the portrait.

The Poet's Tale

These rapidly delivered narratives go gradually "out of focus" as they are told, and sounds mentioned in stories (e.g. duck quacks, the toast) are elaborated. Each narrative is synchronised to the same regular pulse, slightly faster than the original. The stressed syllables in each word are slightly exaggerated (loudness contouring), then the phrases synchronised to the beat by subtly retiming the syllables. The rhythmic regularity is underlined by repeating elements and by echoes, playing across the stereo space, all synchronised with the pulse. As each narrative begins, two further copies of the story are introduced, each at a very slightly slower speed, so that the story becomes first reverberant, then echoed, then "muddied".

In addition, sibilants (especially the "ts" ending "pourin through mi texts") are extended, panned and trembled, articulating the rhythm, like percussion instruments **(Sound Examples 181)**. The phrase "travelling poet" becomes a repeating element with variations **(Sound Examples 182)** made variously by strongly reinforcing the "t" of "travelling" and the "p" of "poet"; isolating syllable peaks (with surrounding silence) then superimposing these on the original sound, emphasizing the rhythm of

the phrase; resonating the two pitches with a pitch-focusing filter using a high Q and few harmonics, to focus on and extend the pitches, which become pedal points for the tales. One of these pitches is echoed by similar treatment of "(a bit of a) do".

Sound events mentioned in the stories are picked up and elaborated. Keith's imitations of duck quacks are developed as an assembly of (human imitations of) duck quacks **(Sound Examples 183)**. The toast "To the Cosmos", related in the story, is echoed by a pre-recorded male chorus **(Sound Examples 184)**.

The Best Thing That Ever Happened To Me

The principal feature of this vocal material is the rising melodic shape of the ends of most phrases, some rising to the 5th of the scale, and some (as in "I just loved it") to the sharpened 4th. The various rising phrase-ends in the voice were collected together and used to form a texture. This was filtered through a harmonic field filter based on the (tempered scale) pitches of the vocal lines which rises to the 5th. The filtered version is added to the unfiltered version in the mix, and as it enters, the tuning of the texture is pushed upwards towards the 5th. Another process was used to extract all the vocal FOFs from one of these phrases, and the pitch-line of the phrase was pitch-tracked[30].Several versions of the pitch line were then resynthesized, each using a single FOF, and these enveloped with the loudness contour of the original voice. These string-like sounds were then textured together (heard most clearly towards the end of the portrait) **(Sound Examples 185)**.

The phrase "exciting life" also has strong pitch contours. The same "strings" procedure is applied to this phrase, and variants are constructed by forcing the final glissandi to rise to a note an octave (or 2 octaves) above the original goal pitch, or to rise from a register to the original pitches from an octave below. The various versions are combined to form a texture around the original vocal phrase **(Sound Examples 186)**.

The Finales

(Sound examples are stereo reductions)

Finale Act 1

In the Finale of Act 1 many sonic events from the Act are extended, elaborated and spatialised. Here are a few examples.

[30] Speaking (not whispering) voices, use pitch in the vowel sounds, and in many consonant sounds. The pitches are not restrained to any scale and generally slide about (rather than step carefully). This (real, sliding) pitch of speech can be tracked by first extracting the time-changing spectrum of the phrase. In each spectral window (time-frame) we search for peaks, and compare their frequencies. If these are (roughly)

The descending "herrn" phrase creates a recurring pedal, which now descends to deeper octaves and, at the very end of the finale, slowly pulsates.

The "f" of "pfump" is time-stretched by roughness-extension and rotates slowly around the space. Many "pf" sounds are massed in an extended sound-surround texture. On its final appearance, this gradually becomes very-dense **(Sound Example 141)** and crosses over into the Voicewind with which the Act had begun, and now ends. Another thread of this extended texture rises slowly in tessitura. In a further strand, the two (adjacent) pitches of the text, which tune with the "herrn" event, are focused by a harmonic field filter (the vocal character disappearing) and pulsed, the pulse rotating around the space.

The rising tail of the transformed "And that was tha…..aat" is greatly time-stretched, with tremolo added, rotating around the space, leading to an explosive end. Sometimes 3 (or more) rotations chase one another, spatially 120 degrees out of step **(Sound Example 187)**.

The "…oke" of "It's a bloke" is vastly time-stretched by roughness-extension. In one variant this sound is harmonic-field-filtered, and rapidly zigzag-panned[31] over the 8 channels, creating an illusion of "everywhereness". In another, waveset-counted steeply-descending loudness envelopes transform the sound to a stream of 'plucked' pitches **(Sound Example 148)**. The tail of this sound is elaborated variously, in the strand in the example by gradual spectral tracing crossing into a watery texture of the traced elements, whilst in another strand, pitch-rising and pitch-falling copies are superimposed on the original. It is combined with a slowly rotating extension of the extended "grit" sound from the Budgie Tale, gradually pitch-resonated through tight filtering.

Finale Act 2

The Finale begins with the fluttering material with which the entire Act begins, and ends with a long extension (by repetition) of one of the "Gossip" phrases, now rotating around the surround space and cadencing inconclusively, as in its original appearance. Harmonic field extensions of materials previously used in the Act are extended in time and animated in the 8 channels space, juxtaposed and variously transposed, to form a harmonically evolving web. The butcher's voice contrasts with this tempered scale material as it is approximately a quarter-tone out of tune, and spectral hold[32] is

whole number multiples of the lowest peak's frequency, then the spectrum is said to be 'harmonic' and corresponds to a clear pitch. Complications occur with sounds which are pitched but noisy (e.g."zz"), unpitched elements (e.g.``s``) and silences. So we need to deal with these (e.g. by interpolating between the real pitches on either side of these pitch-free events, or by inserting markers to indicate that no pitch is present) when generating a pitch-line equivalent of the speech phrase.

[31] Zigzagging reads a sound file back and forth at random (or in a user-specified way). Zigzag panning pans each zig or zag between output channels in a randomly permuted order. With rapid panning the sound seems to be everywhere at once.

[32] The spectrum in a single analysis window is sustained. With noisy, inharmonic or modified spectra (like the Butcher's voice) this produces a sustained inharmonic sound (we hear several unrelated pitches

applied to some of its syllables producing strange inharmonic resonances. Meanwhile the high-pitch fluttering, derived from the lorry rumble, rotates.

Finale Act 3

This finale is in ternary form. In the first and third parts, the rhythmic pulsed materials derived from "so close", from The Soldier's Tale, are developed. The "s" based material is now faster, and more regularly and incisively pulsed. It is introduced by a quickly rotating and tessitura-descending crescendo onto an explosive attack, a stack[33] of reverberated versions of one impulse **(Sound Example 188)**. This then tremolos into the tempo of the Finale, and recurs later (or is implied) as the Finale proceeds. The pulsating "ss" then wrap-pans[34] across the space, eventually leading to successive impulses being assigned to random permutations of the output channels. Other strands of this material are fed through pitch-focused filters where the Q value slowly increases, gradually focusing each strand onto a different pitch or pitch-pair. Meanwhile, the two-pitch stream also derived from "so close" is also panned around random permutations of the channels, at its original pitches, and a minor third lower.

"Heathcliffe" derived events are extended in time (FOF-extension or long reverb), some rotated with a slowly rising and falling pitch like a siren, others rising in pitch and pulsating at the basic tempo of the music, and so on. The "Three years" material from "The Stories I Hear" is time-extended, while the "are you still using that thing" event is elaborated by layering variously time-stretched copies animated by their own stereo delays. This creates short and long versions of a recurring phrase with a vibra-toed middle note. The "gorr" material becomes an extended aggregate with a long derived rasp at the end, juxtaposed with the above **(Sound Example 189)**. Horn-like vocal sounds, which step upward, derive from the voice in "The Farmer's Tale", **(Sound Example 190)** while deep pedals from the "Herrn" event recur beneath all this material .

The central section develops the "democracy" bell-ringing event from "Heathcliffe, Come Here". As the pattern repeats, the word chunks are increasingly pitch-focus-filtered at their own pitches, creating a bell-like resonance. The sequence is punctuated by increasingly deep stacks of these pitch-resonated sounds, deeper "bells" **(Sound Example 191:** *stereo reduction and contraction***)**. The horn-like vocals float above, some direct extensions of the voice in "The Farmer's Tale", some pitch-following filtered and more "instrumental" but with voice-like vibrato at their tails while other sounds deriving from this voice read back-and-forth across the break in the voice producing a rapid soft ululation **(Sound Example 192)**.

not lying on the tempered scale).

[33] Sample-rate-changed transpositions of the source (lower sounds being longer) are superimposed on one another. If the attacks of all copies of an attack-resonance event are precisely synchronised, the stacked events merge into a single new sound (we do not hear the result as a chord of several sounds).

[34] See section *A Sound-Surround Working Environment* earlier.

Finale Act 4

The finale of Act4 brings together and reworks many of the materials previously heard in the Finales of Acts 1, 2 and 3, leading to a tolling of the biggest of the "democracy" bells, each toll anticipated by a long anacrusis (a very long time-stretch of the opening of that same sound). Over this the speaking voices begin to sing. Some voices extend single notes by long reverberation; others use vowel-preserving spectral pitch-shifting on a time-stretched note. But most are extended by freezing the voice on a specific FOF impulse within a vowel, and using this to recreate a vocal line of designed pitch[35]. Time-varying vibrato of various (slightly randomised) depths and frequencies is added to the recreated line to produce a more plausible singing quality **(Sound Examples 193)**.

At the end of the piece, there are two strands of many voices, each strand filling all 8 channels in the surround-sound space. One of these strands appears to dissolve the voices in a patina of high frequency mouth-crackling. This is itself made up of two 8-channel mixes, each with 1 voice placed on each of the 8 channels. Both strands were fluttered[36], causing fragments of each voice in turn to be heard in its own channel. The entire sound-surround image was then set in rotation at an accelerating rate. Waveset omission[37] was then progressively applied to the sound end, giving it an increasingly gritty quality. The tail of this sound was passed through a pitched filterbank, a slowly rising bank of 4ths, and the output progressively high-pass filtered (so we hear only unvoiced or apparently unvoiced constituents) and reverberation gradually introduced as it fades to nothing.

The other 8-channel strand is slowly dissolved, through increasing density and shuddering, into the Voicewind sound, mirroring the opening of the entire piece.

[35] The voiced sounds in speech are produced by a sequence of small wave-packets or FOFs. The speed at which these FOFs come past determines the pitch of the voice, whilst the particular shape of the packet defines the sonority we hear e.g. an "aa" as opposed to an "oo". In the real world, when the pitch goes up, the FOFs are truncated (rather than squeezed) without altering too much their basic shape. We can emulate this by the simpler process of slightly overlapping the FOFS, again *without* changing the length of each packet. (If we overlap the packets too much we produce a kind of reverberation effect as the sound is effectively echoing itself). In the real world, when the pitch goes down, the FOFs are extended (rather than stretched) continuing their basic shape in some coherent way. We can emulate this simply by spacing out the FOFs, leaving silence in between them. Despite the artificiality of this approach, the results (vocal sounds transposed down in pitch while preserving their vowel character) are convincing.

[36] Tremolo is applied to a multichannel sound such that the sound rises in level in each loudspeaker in turn. The sequence of loudspeakers used for this can be specified in a regular pattern or as a random permutation of all of them (followed by a different permutation, etc. – this ensures all output channels are used equally often).

[37] A proportion of wavesets (delineated by counting signal zero-crossings) are replaced by silence.