Originally published in *New Music and Aesthetics in the 21st Century, Vol. 4: Electronics in New Music*,
Claus-Steffen Mahnkopf, Frank Cox, and Wolfram Schurig, editors.
Wolke Verlag, Hofheim ISBN-10: 3-936000-15-8 and ISBN-13: 978-3-936000-15-3

Gesture in the Mouth and Throat: Electroacoustic Approaches to Vocalization in *Rally*

Chris Mercer

Vocalization as Gesture

In “The Listening Imagination,” composer Denis Smalley offers the following definition of *utterance* as a special case of physical gesture:

> Gesture and utterance are not distinctly differentiated fields. At a psychological level they can be synonymous because both are articulated through the energy-motion trajectory, and both are concerned with proprioceptive perception. Hence the term *vocal gesture*, which expresses this shared indicative significance. However, the spectro-morphologies of language-utterance and paralinguistic utterance sound very different from instrumental gesture…The fact that the sounds of utterance are generated within the body, and that they are the essential vehicle of personal expression and communication, make utterance intimate and emotionally charged.[1]

The term *vocal gesture* acknowledges that vocalization is both related to and distinct from musically or sonically significant physical gesture as usually understood, i.e. as a function of the arms and legs and, in some respects, as a translation of the choreographic into the sonic. However, whereas the more outwardly kinetic gesture seems to pass smoothly into the realm of musical figure—where it takes on an immediate secondary significance—the inner physical gesture in the mouth and throat has already taken on secondary significance, namely that of linguistic communication, before it passes into musical figure. Of course, all forms of physical gesture can be used to communicate, and that becomes part of their musical power. But vocal gesture communicates with a degree of exactitude and primacy that marks it with *functionality*. In a music that seeks to problematize the morphology of sound itself and to dismantle both the semantics and the mechanics of physical gesture and its figural derivatives, the everydayness of vocalization is a stumbling block. The communicative/expressive function of utterance tends to blur apprehension of both its purely sonic and purely physical-gestural aspects, making it difficult to treat them as raw materials in composition. The reduced listening that may be possible in degrees with other sound sources is bound to fail in the face of recognizable human or animal vocalization.

Perhaps a true return to first principles is both impossible and undesirable in the case of vocalization. Can we really reconceive the vocal tract as merely a pneumatic device with which to produce sound? Yet without entirely denying its built-in expressivity, composers may wish to examine the gesturality of the mouth and throat without succumbing uncritically to all the extra-musical (and extra-sonic) associations attached to it. The most obvious of these is language itself. The first task is to find ways

[1] Denis Smalley. "The Listening Imagination: Listening in the Electroacoustic Era," *Contemporary Music Review*, vol 13, Part Two (1996), p. 86.

to drain the text out of speech. Smalley's definition also includes paralinguistic vocal phenomena. All manner of grunt, screech, click, cluck, and even expressively loaded respiration serve communicative functions. Limiting one's sound sources to textless vocalization does not solve the problem of association. Furthermore, even animal vocalizations are powerfully evocative; some express (ambiguous) emotion, some merely cue environment or location, but they are always unmistakably vocal. Synthetic voices also speak and express—however blandly or intentionally robotic, they emote and project personalities or, at least, we instinctively project personalities onto them. And just as a pair of eyes is instantly recognizable, even magnetic, in an otherwise abstract painting, so the slightest trace of the workings of the vocal tract will poke through even the densest sonic surface. The ear is so attuned to vocalization that its characteristics will be heard when they appear, even incidentally, in both non-vocal organic sources (stomach growling) or inanimate objects (bubble popping).

Rally for 2-Channel Tape (1994-2005) attempts to render the linguistic paralinguistic, the paralinguistic musical, the synthetic expressive, and the inanimate organic. Every sound in the piece is to *vocalize*, while no sound is to *verbalize*. To achieve this goal, the work combines spoken word recordings, extended vocal techniques, animal sounds, acoustic objects, and voice synthesis. The sound sources are shaped via analog tape manipulation, digital editing, and extensive digital signal processing into an array of unique sonic creatures (or herds of creatures) with distinctive behaviors and implied anatomies and the ability (and desire) to hybridize with one another.

The sound sources in *Rally* are:

-Extended vocal techniques, recorded and manipulated on consumer cassette decks

-Extended vocal techniques, recorded and manipulated on handheld cassette recorders

-Extended vocal techniques, recorded and manipulated on a 4-track cassette recorder

-Seaweed, recorded and manipulated on a 4-Track cassette recorder

-Speakers, including Noam Chomsky, Harry Partch, Henry Cowell, Langston Hughes, William Carlos Williams, and a Southern Baptist preacher recorded off the radio

-Growling stomach, recorded digitally

-Growling cat, recorded digitally

-Pigs, chickens, and a frog, from an effects library

-A maraca, recorded digitally

-A doorstop box (homemade instrument consisting of a wooded box with doorstops mounted on top—opening and closing the lid simulates vowel formants), recorded digitally

-Fonction d'onde formatique (FOF) synthesis

Approaches to Textless Vocalization

Throughout the 20th century, composers sought new approaches to text setting and vocalization, often attempting to separate out the sonic and semantic components of language or to appropriate for compositional purposes paralinguistic vocalization. Beginning with extreme word splitting strategies (with decidedly unspeechlike rhythms), texts were increasingly treated as the raw material of a plastic language art. Eventually, composers began to atomize texts into their component phonemes whose relationship to the source text is frequently rendered unrecognizable. In such cases, the original unprocessed text may or may not appear in the work. Where it does not appear in the work, the question arises as to what relationship its semantic content has to the pure phonemic content of the resulting composition. Although it is certainly significant to the composer that a given text was used to generate sonic-phonemic material, the listener is left out of the process, and hence its meaning, unless there is an explanatory program note. At the surface level, the phonemes, whatever their origins, are only phonemes.

Nevertheless, one can understand the modernist-constructivist desire to conceive of the phoneme as an atomic unit of speech-sound and to apply to it all the deterministic organizational principles that had been applied to other musical units/parameters. The reductive processing of extant text is one possibility; another is the creation of pseudo-linguistic materials from the ground up. Milton Babbitt's *Phonemena* for soprano and tape is an example: Babbitt obviates the text-relation problem entirely by using raw phonemes as material for serial organization. A "language" emerges that is entirely the result of the work's deterministic procedures; it is neither text-setting, nor sonic poetry, but acts as the illumination of serial processes running parallel to the parametric control of purely musical materials.

In some cases, the use of actual text may have a talismanic role in the composer's work process, despite its obfuscation in the resulting composition. Brian Ferneyhough's *Time and Motion Study III* uses source texts processed beyond recognition. Ferneyhough acknowledges and plays upon this fact:

> …this splitting apart of the syntactic and semantic dimensions of speech activity is deployed at a further axis. Affectively operative syntax (processually combinatorial successions of sound) placed against semantic emptiness; valid semanticity (texts in Latin, German, and English), so processually undermined as to lose all claim to independent existence.[2]

The source text is so thoroughly ameliorated that the phonemic material on the sonic surface is often savagely primal and nakedly emotional: strings of detached fricatives, long rolled R's, ingressive breathing, stretched unvoiced consonants. The sound world approaches that of animal vocalization, as though we have traced linguistic semanticity back to its roots in the world of expressive pre-language. It is ironic and poignant that multi-lingual source material processed by a highly organized determinism should cast a light on the instinctual physicality at the heart of the linguistic sign.

[2] Brian Ferneyhough. *Collected Writings* (Amsterdam: Harwood Academic Publishers. 1995), p. 115.

In electroacoustic music, work with vocalization at the sonic level can be carried out via direct voice synthesis. John Chowning's *Phoné* works with FM models of vowel formants voiced by strangely disembodied synthetic choirs. There is no text associated with these synthetic vocalists, so they occupy an electronic netherworld of both linguistic emptiness and unnatural timbre while still retaining the baseline sonic signature (filter properties) of human vowelness. In a sense, the work is an exploration of the vowel as a signal of a given disposition of the oral cavity. And furthermore, the blatantly synthetic quality of the voice is a signal of the technological triumph of having reduced the disposition of the oral cavity to the relevant formant filter values. Synthesis itself, then, is the true subject of the work, even as pure *vocal gesture* is both its source material and the measure of its virtuosity.

Creature Design

When treating the vocal tract as a source of gesturality, certain associations are inevitable. Maybe the closest thing to a "first principle" of vocalization is its association with some kind of emitting organism. Synthetic voices may be one level removed from this starting point, but they are still understood as *emulating* an organism, and hence the association is still active. Many sound-producing physical gestures can occur independently of human or animal agency: wind blowing through trees, ice cracking, waves breaking, rocks falling. Vocalization has the special quality of signaling not only human-animal agency, but also direct corporality, since the body is both the mechanical trigger and the acoustical *site* of the sound object. Rather than conceive of the vocal tract in purely mechanical terms, I have accepted its association with organism and its attendant expressivity, while attempting to blur and distort this field of association, focusing on basic functioning states, expressive ambiguities, and peripheral in-between states, and devising a set of sonic pseudo-organisms whose motivic-behavioral characteristics imply psychological as well as anatomical features.

As a model for sonic creature design, I have looked not at sound design per se, but at the work of stop-motion animator Ray Harryhausen. I am interested in the behavioral repertoire of the animated creature and its projection of psychological states through anatomical design. In Harryhausen's work, each animated creature has a distinct personality, articulated largely by the relationship of its physiognomy to its gesturality. Take the motivic unity of Medusa[3]: In addition to her head of snakes, her entire body is serpentine (a kind of snake take on a mermaid), her comportment is accordingly slithering, her tail rattles and her head jerks in response to stimuli, her face is angular, her skin leathery. Her "bite" is found in her eyes, which glow bluish green when she turns a man to stone. This serpentine design is entirely Harryhausen's invention, and it is psychologically and anatomically consistent.

The slightly rigid, jerky quality of stop-motion animation gives the visual result an eerie, preternatural presence against the live action elements that benefits the total illusion. The creatures' personalities derive from their stop-motion-ness—the technology imposes its limitations on the creature design in a way that the animator turns to his advantage. Harryhausen explains:

[3] In *Clash of the Titans* (1981), Columbia Pictures.

> Fantasy is essentially a dream world, an imaginative world, and I don't think you want it to be too real. You want an interpretation. And stop-motion, to me, gives that added value of a dream world that you can't catch if you try to make it too real.[4]

One excellent example of this is Talos[5], an enormous metallic statue that comes to life to attack a ship. In order to convey Talos' stiff-jointed, mechanical gait, Harryhausen exaggerated the stop-motion jerkiness of the animation, robbing the creature of any sense of fluidity or flexibility. When Talos' body fluids are drained out through his heel, he gropes ineffectually at his throat, unable to bend and strangely unaware of his own anatomy. Hence, the basic functioning state is projected into the extreme state, the death throes.

This willing embrace of artifice, its deliberate exposure and incorporation into the fabric of the art object, is a primary component of the creature design in *Rally*. The vocalizations should sound at once organic and artificial, natural and engineered. In order to treat the organism itself as a first principle—or a best attempt at reduction or deconstruction—of vocalization, it is necessary to impose a critical aura around it. The notion of "critical aura" may seem aesthetically antithetical to Harryhausen's "dream world," but the perceptual goal is similar: The organism's artificiality must be part of its personality and, hence, its gesturality.

In *Rally*, the emphasis is not on the extreme state—the death throes, the anguished cry, the fight-or-flight response—but on the basic functioning state. Put bluntly, the creatures in *Rally* are more likely to bark than to yelp, to howl than to shriek, to sigh than to pant or gasp. Since expressive functionality is an impediment to the apprehension of pure vocal gesture, evocation of extreme emotional states would work against the basic thesis of the piece. The few vocalizations that are expressively charged tend to be ambiguous or in-between, minor modifications of the basic functioning state.

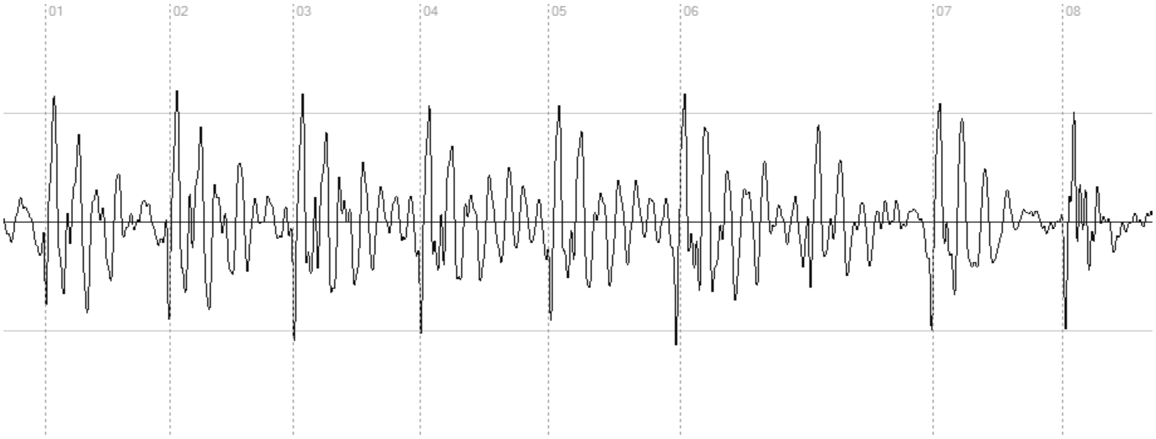
Vocal Fries and Synthetic Voices

One of the creatures found in *Rally* in various behavioral states and several distinct species was inspired by a remarkable use of vocal fry in Roger Reynolds' *Still (Voicespace I)*. "The austere text writhes in slow motion across the aspirate clicks of the performer's ingressive vocal fry,"[6] which is slowed to the point of revealing its individual glottal clicks. In the score, the composer has notated the timing of individual clicks in proportional notation with the text itself written over it. The performer mouths the words while performing the glottal clicks, which then act as sonic strobes in the oral cavity, revealing in sonar-like flashes the position of the mouth and hence sounding out the vowel formant. At the beginning of each section, the next line is read normally, so that the listener can grasp the vocal fry text as it occurs. This radical rethinking of the relationship between the source text and the mechanics of vocalization is a kind of lateral solution to the problem of deconstructive text processing discussed above.

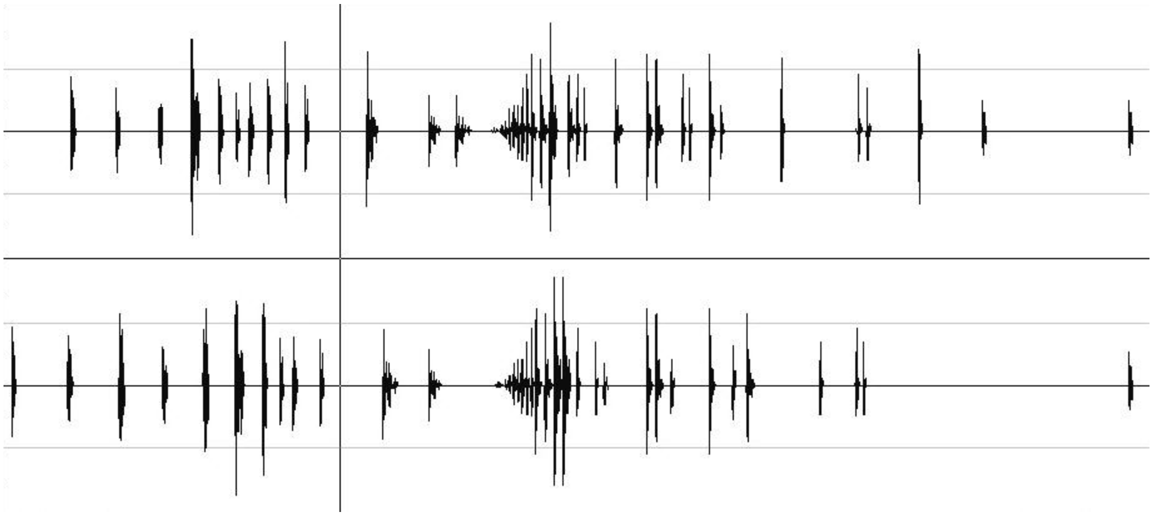
[4] *The Ray Harryhausen Chronicles* (2002), Columbia Pictures.
[5] In *Jason and the Argonauts* (1963), Columbia Pictures.
[6] Roger Reynolds. *Voicespaces*. CD Liner notes. (New York: Lovely Music, ltd. 1992), p. 3.

In *Rally*, I chose to take this vocal fry “strobe” technique in another direction, robbing the vowels of their text articulation function, reducing them to pure formant filter play, and therefore severing the connection of the sonic to the semantic. The consonants have been removed and the vowels extracted from their original (word order) contexts. In some cases, two or more vowels have been inter-edited so that they interlock between channels and play out simultaneously. So while I strove to preserve the natural formation of individual vowels, I did not permit the perception of actual words. The resultant creature is of ambiguous morphology—the quasi-motoric rhythmic behavior of the glottal clicks signals a synthetic aspect; the faint trace of language in the vowel formants suggests a human presence; the lack of clearly recognizable linguistic structures and the strangely inhuman quality of the vocal fry itself evokes functional animal vocalization.

The creature was designed in the following manner: In collecting and examining spoken word recordings for possible source material, I came across several speeches by the linguist Noam Chomsky. (The irony of rendering a linguist’s speech non-linguistic was not lost on me.) Listening to his speaking voice, I noticed that it was, in effect, a continuously articulated vocal fry. A closer look at the waveform confirmed this, revealing clear strings of glottal clicks:



By placing spaces between glottal clicks, I created a controllable vocal fry over which vowels form clearly. Variations in pitch, glottal click rate, filter function, and editorial interlocking of simultaneous vowels gave the creature considerable plasticity of animation.



In order to extend this basic design, enhance the degree of parametric control, and develop additional species, I turned to FOF synthesis, which appears in two forms in the piece: raw click-formant simulation and analysis-driven resynthesis. The former is used only very rarely as an extreme instantiation of the synthetic voice, and its “clicking” quality is usually explicitly exposed, so that its role as a robotic analog to the vocal fry creature is clear. The latter, however, has a great range of behaviors, clearly related to or deriving from vocal fry, but capable of transformation and cross-synthesis. Since FOF synthesis derives vowel formant structures from successive sine tone bursts (glottal clicks), it is ideally suited to the extension of vocal fry material. Actual vocalization information was derived from spoken word recordings via Linear Prediction Coding (LPC) analysis. In Csound, the outputs of the LPC opcode were used to drive FOF synthesis parameters, resulting in a highly flexible vocal fry simulation instrument. The following instrument demonstrates this basic approach:

```
instr 1

ktimp linseg 0, p3, 7

krmsr, krmso, kerr, kcps lpread  ktimp, "wcw.lpc"

;ar fof  xamp,       xfund,      xform,     koct, kband, kris,   kdur,     kdec, iolaps, ifna, ifnb, itotdur

ar  fof  krmso*.2,  kcps*10,  kcps*.5,  1,     10,       .003,  kerr*2,  .007,  5000,   1,    2,     p3

aout lpreson ar

out aout

endin
```

The FOF generator is used as an LPC resonator, while simultaneously responding from within, as it were, to LPC data outputs. Thus, in addition to applying formant filters to simulated glottal clicks, this technique derives the rate and pitches of clicks themselves from the same speech analysis data that shapes the vowels.

Direct LPC techniques spawned several other distinct materials. When the unvoiced signal is used to filter white noise, the result is a raspy, respiratory speech sound, like a loud, hoarse whisper. This material is perhaps the closest thing to recognizably linguistic utterance in the piece and represents a sort of human extreme in the creature design. Nevertheless, its source text is never fully intelligible. Also, by feeding the voiced signal to simple sine and FM functions, I created a robotic complement to the pure FOF synthesis materials. Unlike the simple vowel formants of straight FOF, these LPC materials form complex strings of phonemes that elide into one another and have a genuinely speechlike cadence. Again, the text is obscured and we are left with a strangely unintelligible robot, earnestly but meaninglessly babbling in the rhythms of speech.

Lo-Tech Materials

I have written elsewhere on the aesthetics of lo-tech electronics.[7] Many of the oldest source materials in *Rally* were originally conceived, recorded, and manipulated on consumer tape recorders. The lo-tech aesthetic takes on special significance here, because the tape machines applied the first round of processing techniques, thereby shaping the vocal materials, in some cases, even as they were laid to tape. Hence, the lo-tech project of uncovering the distinct personalities of consumer devices projects itself onto the psychology of the resultant creature design. Not only do particular extended vocal techniques form the basis for a given sonic organism, but so too do the quirks, tics, and defects of the specific device onto which it was recorded or manipulated. I developed vocal techniques specifically intended to play to the sonic traits or manipulative capacities of certain tape machines.

The Harryhausen dream world mentioned above is especially relevant to these lo-tech materials. Here the dreamlike quality comes in the form of tape hiss, the inherently limited bandwidth of a built-in microphone, the warble—deliberately induced or otherwise—of the reels, or the weirdly imposed glissando of a poorly designed pause button. Basically, the tape machine acts as a transformative aperture through which vocalizations pass, and this filter of flawed technology imparts a quality of affectedness or artifice onto the sonic result, a sense of “found object,” as though the materials were artifacts from some documentary recording whose origin and purpose one could never guess.

In addition, many of the lo-tech materials in *Rally* were elaborated at the time of their conception into full-blown compositional statements. Materials recorded and manipulated on handheld tape recorder, for example, were often multi-tracked, in a sense, via a disabled erase head. They became miniatures with their own formal structures and distinct textures, and these too play a role in the subsequent creature design. Frequently, materials from one tape medium, such as the handheld tape recorder, were further developed in another, such cassette decks or 4-track machines. Thus emerges a history of preliminary compositions, ultimately subsumed by the final form of the work.

[7] Chris Mercer. “Composing Algorithms, Composing with Algorithms: A Critical Assessment” in Claus-Steffen Mahnkopf, ed., The Foundations of Contemporary Composition (=New Music and Aesthetics in the 21st Century, Vol. 3) (Hofheim: Wolke, 2004), pp. 166-168.

Just as specific sonic features of lo-tech devices influenced, on the input side, the choice and structuring of vocalizations, so too the application of DSP techniques was tailored to the special characteristics of resultant lo-tech compositions. There is a multi-layered interaction, then, between voice and tape machine, one tape machine and another, and tape materials and DSP techniques, with each step in the chain influenced by a desire to discover and develop *personality traits*—psychological, comportmental, physiological—in the end result. This process of interaction, and hopefully fusion, between the human voice and a series of processing machines might be seen as the polar opposite of the some of the work's other human-machine interactions, such as the "robotic" LPC resynthesis or, at the extreme, pure FOF synthesis, in which the machine simply imitates the human.

Animals, Organic and Inanimate Objects

One might view sub-linguistic human vocalizations and purely synthetic vocalizations as the extreme ends of the human-machine spectrum in *Rally*. In between, there are numerous animal sounds and sounds derived from organic and inanimate sources. Many of the animal vocalizations in the piece act either as connective tissue, variation, or textural enhancement. The work's sole use of a frog, for example, serves as an intermediate form of the glottal click creature, represented by FOF synthesis at the machine end and by actual edited human speech on the other. Farm animals add organic textural complexity, via convolution, to speech-derived materials (preacher-pig, Chomsky-chicken). The most prominent use of animal vocalization is the growling cat, transformed and featured as a distinct creature. Its function is, in fact, unequivocally animalistic, though its felineness is somewhat obscured by the degree of transformation. Also, because it contains strong vowel formants, it is used, via LPC analysis, to drive FOF-based glottal clicks as discussed above, providing a special instance of that creature.

Animal vocalizations are useful in that, short of extreme states, they tend to emote ambiguously. There is an organic purity about these sounds in that, as Wishart notes, not only their emotionality, but also their intentionality, is ambiguous.

> …a certain level of fear, aggression, and sexuality may produce a particular level of arousal and a particular articulation of the internal state causing the vocal apparatus to emit sounds of a particular form. As another organism of the same species will recognize these sound-objects as if it itself had emitted them, they may be taken to symbolise the particular state of the first organism. However, we cannot therefore assume that the emitter intended this symbolisation. Apart from the bringing into action of the vocal apparatus as a whole, the resulting evolution of the sound-objects may have been substantially involuntary, a direct utterance.[8]

Wishart refers here to extreme states, but the basic functioning state is perhaps of even more ambiguous expression and intention. In this respect, the origins of these animal vocalizations are truly distinct from the more engineered, manipulated, and goal-directed source production techniques found elsewhere in the work—human vocalizations prepared specifically to play to machine idiosyncrasies, carefully programmed synthetic

[8] Trevor Wishart. *On Sonic Art*, (Amsterdam: Harwood Academic Publishers, 1996), p. 255.

voices, imposition of extracted vowel formants onto white noise. The animals behave naturally, without artifice, and this is part of their importance as connective tissue.

There is one more general class of sounds in *Rally*: non-vocal acoustic sound sources. All but one of these—the maraca—closely resemble vocalization; and the maraca is in fact made to resemble it, when convolved with the Chomsky vocal fry (Chomaca). In addition to the maraca, there is a homemade instrument, the doorstop box. Two doorstops are screwed into the lid of a small wooden box. The doorstops are bowed as the lid is opened and closed, producing vowel-like filter effects over the strangely plaintive moan of the bowing. The result is surprisingly animal-like and serves as a compelling link between the growling creature and the vocal fry creature.

Finally, there are the growling stomach and the seaweed. The voice-like quality of a growling stomach is well known—we've all heard our stomachs chastising us at some point. In addition to the presence of discernable phonemic material, the growling stomach suggests the movement of viscous liquids. The seaweed consisted of strands of rubbery bulbs filled with water. I recorded the bulbs bursting and being dragged across concrete. This was, in fact, one of the work's lo-tech sources, and some tape manipulation was applied at the recording stage. The distinctive filter-sweep of the bursting bulbs suggested vowel formants, and this was exaggerated by the warbling tape manipulation. The squishy vocalizations of the bursting bulbs and the viscous vocalizations of the growling stomach complemented each other perfectly, and together they form a mass of pulsating, organic material, a sort of articulate bog in which more clearly defined creatures writhe, interact, and hybridize, not only with one another, but with the bog itself.

Cross Synthesis, Group Behavior, and Perspective

Formally, *Rally* is propelled by two concerns: the ongoing hybridization of creatures, to the point of near-continuous morphing; and interrelated shifts of perspective and creature behavior, alternating between solos, dialogues, and group vocalization, while the position of the listener in relation to these behaviors is in constant flux.

Nearly every significant creature hybridizes with every other creature at some point in the work. Some of the DSP techniques I have described (LPC/FOF) have cross-synthesis built into them. These are augmented with convolution, FFT cross-synthesis, and simple mixing and cross-fading techniques. It was important that the creatures be in constant morphological flux, always imparting qualities onto one another, and that this flux be perceived as an *urge*, a compulsion, something akin to a reproductive drive, only it does not seek propagation, but rather an endless exchange of physiognomic-behavioral traits.

This discourse of morphological exchange is viewed from various distances. In extreme close-up, the listener is placed within the throat of the emitting organism. The distant zoom-out reveals herds of like creatures, group vocalizations that merge into mass textures. At intermediate distances, the listener may be placed next to a creature or within a menagerie of disparate creatures, constantly in motion. In the work's climactic moments, the listener views a distant herd from within a menagerie, while occasionally moving into the throat of a passing creature. For instance, imagine a zoo situated atop a hill. Within the zoo, one is surrounded by a great variety of vocalizations, while in the distance, perhaps in a valley below the hill, one hears the bayings of a herd of like creatures. Occasionally, a nearby creature envelops the listener's entire head in its mouth, while simultaneously vocalizing. This is the strange perspective, a sort of perspectival hybridization, that I hoped to achieve in the last few minutes of *Rally*.

So, from a starting point of close examination of the vocal tract as a source and site of physical gesture, *Rally* generates an instrumentarium of vocalizing organisms—on the assumption that such an association is inevitable—using hybridization as its primary mode of transformation, and situating this population of sonic creatures within a constantly shifting set of spaces and perspectives. Without reducing the vocal tract to a mere mechanical device, the piece attempts to explore the mouth and throat as sources of meaningful organic sonic phenomena, without resorting to the use of text, overt expressivity (extreme states), or obviously familiar paralinguistic signification.