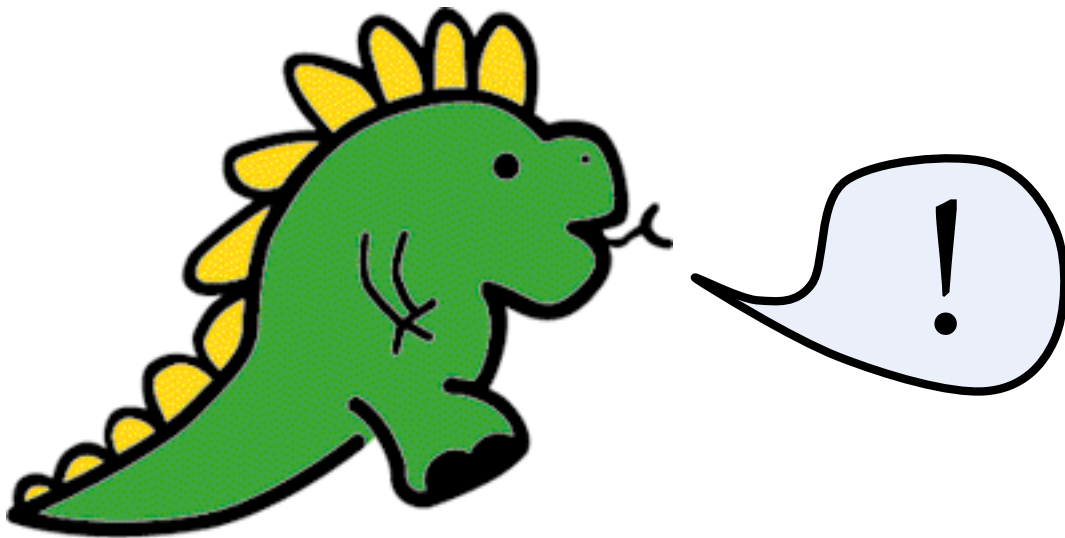# Synthasaurus:

An Animal Vocalization Synthesizer

Robert Martino
Master's Project
Music Technology Program
Advisor: Gary Kendall
June 6, 2000

Introduction

A compelling area of exploration in the domain of physical modeling and vocal synthesis is the production of non-human, expressive, animal-like vocalizations. Animal sounds can convey a wide variety of emotional states, and synthesizing life-like vocalizations would allow for interesting applications in the world of video games, film, music, and artificial intelligence systems. This paper describes Synthasaurus, a synthesis engine prototype developed in Opcode MAX/MSP, which enables one to create emotive animal-like calls, and provides enough flexibility to synthesize a variety of organisms that can resemble different mammals, birds, reptiles and amphibians. Alien, robotic, and other imaginary creatures can also be conceived. Synthasaurus builds on research and technology developed for human speech synthesis, with special kinds of control added for creating more animal-like sounds.

Basics of Animal Communication
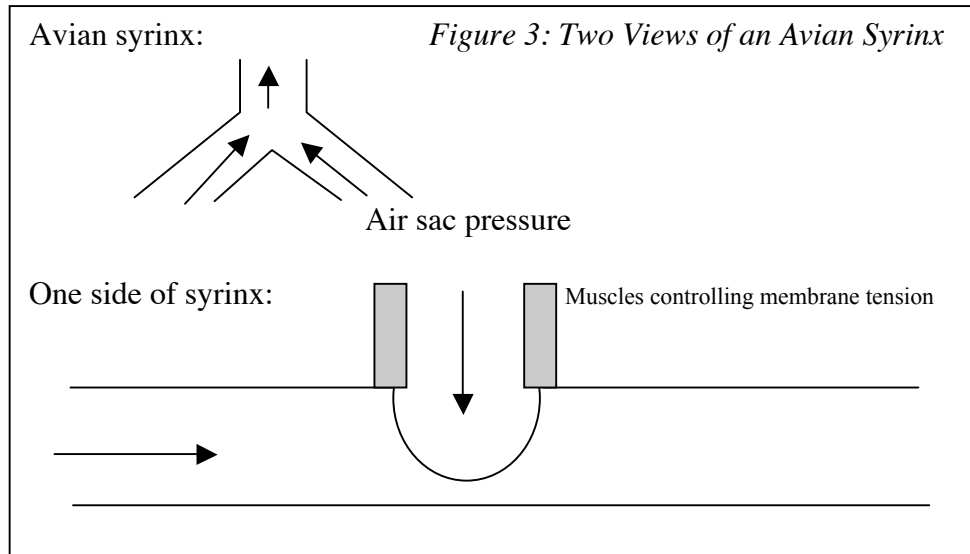
Animal Vocal Systems

Sound productions systems in typical vertebrates (mammals, anurans, and birds) share a similar basic mechanism. Air flowing through a tube causes one or more membranes in the path of the flow to vibrate. These vibrations can then be modified (such as through a resonating chamber) and are then coupled to the propagating medium (Bradbury 1998).

Trachea   **Muscles**   *Figure 1: Mammalian Larynx*

Air flow

Glottis
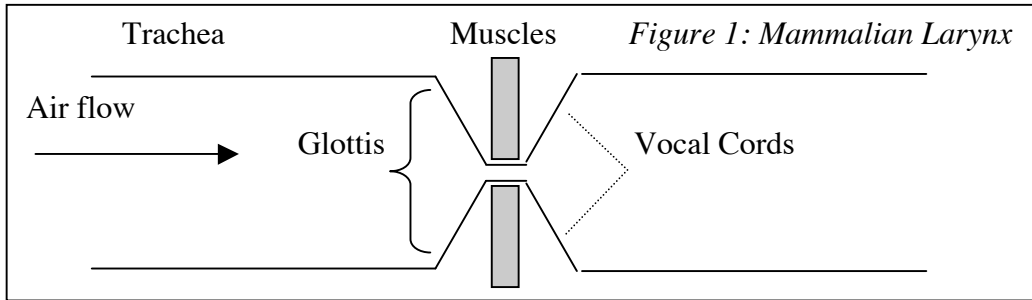
Vocal Cords

In the larynx of mammals (figure 1), two vocal cords (which make up the glottis) block the airflow from the respiratory system. Enough air pressure will push the glottis open, releasing a burst of air, and a Bernoulli force is generated which pushes the vocal cords back together. The result is a series of periodic air fronts of a non-sinusoidal nature. This harmonically rich signal can then be filtered via a resonant chamber that ends with the mouth and nose.

Trachea   **Cartilage**   *Figure 2: Anuran Larynx*

Air Flow

Vocal Cord
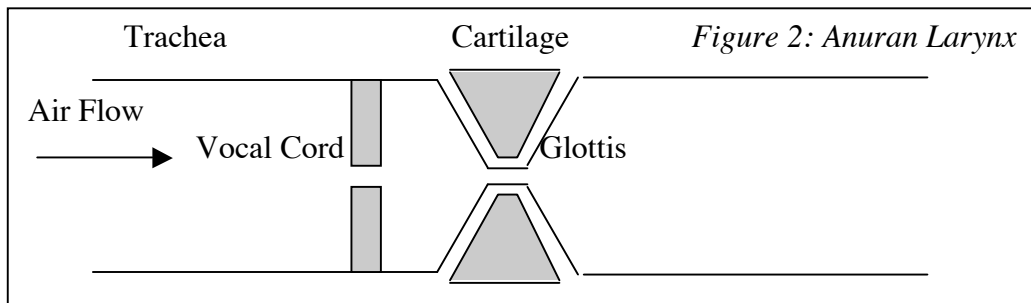
Glottis

Anurans (frogs and toads) also have a larynx system (figure 2), but in this case a second pair of membranes upstream from the glottis can oscillate at a frequency independent from the glottis. Thus amplitude modulation occurs. Air then passes into a throat sac rather than escaping through a mouth or nose, and this air can also be recycled back into the lungs.

Figure 3: Two Views of an Avian Syrinx

Avian syrinx:

Air sac pressure

One side of syrinx:

Muscles controlling membrane tension

Birds have a bronchial-tracheal junction called a syrinx (figure 3), whereby two bronchial paths join with a single trachea. Membranes either in the trachea or on each side of the bronchial passages vibrate when air passes over them. The tension of these membranes can be modified to modulate frequency and amplitude of sounds. When these membranes occur in the two bronchial passageways, they can sometimes be controlled independently, thus creating two independently controlled sounds. (Bradbury 1998)

Communicating Animal Emotion

While one important goal of this synthesis model was the ability to create sounds with physically realistic timbres, another is to communicate emotion, possibly evoking a particular "mood". Despite the variety of animal species and sound production systems they employ, there are some generalizations that have been made as far as understanding the intention or emotional message of an animal's auditory signals. This kind of

information would be helpful in relating emotional states of an organism to the physical properties of sounds it might make in those contexts. Darwin suggested that the size of an animal determines the pitch of it's voice, and that larger individuals are generally more dominant than smaller ones (Darwin 1965). Using this reasoning he argued that aggressive vocalizations tend to be characterized by lower pitch, and submissive vocalizations are relatively higher.

Morton (1992) developed a more comprehensive model that related the structure of many mammal and bird vocalizations to motivational states, which he called the Motivational/Structural rules. As an animal gets more aggressive, its vocalizations tend to become more broadband (harsh) and lower in pitch, and as an animal becomes more fearful, the pitch of its vocalization tends to rise and become tonal. Combinations of various degrees of aggression and fear reflect more ambiguous motivational states that combine sonic properties from both (figure 4). Each block represents a basic sonogram, with thickness of the line representing bandwidth, and height of the figure representing frequency. Arrows suggests shapes that can vary in pitch, and dotted lines represent degrees of change in slope. Tones in the upper left corner of the chart show non-aggressive, friendly sounds that are tonal and vary in pitch. Fear is indicated by increasingly higher pitch. Aggression is expressed through harsher sounds that are lower in frequency, and can be mixed with fear characteristics. The "neutral" chevron shape in the middle can express a sense of general alarm or excitement, and depending on the frequency and length is characteristic of a "bark" like sound in many species (Morton 1992).
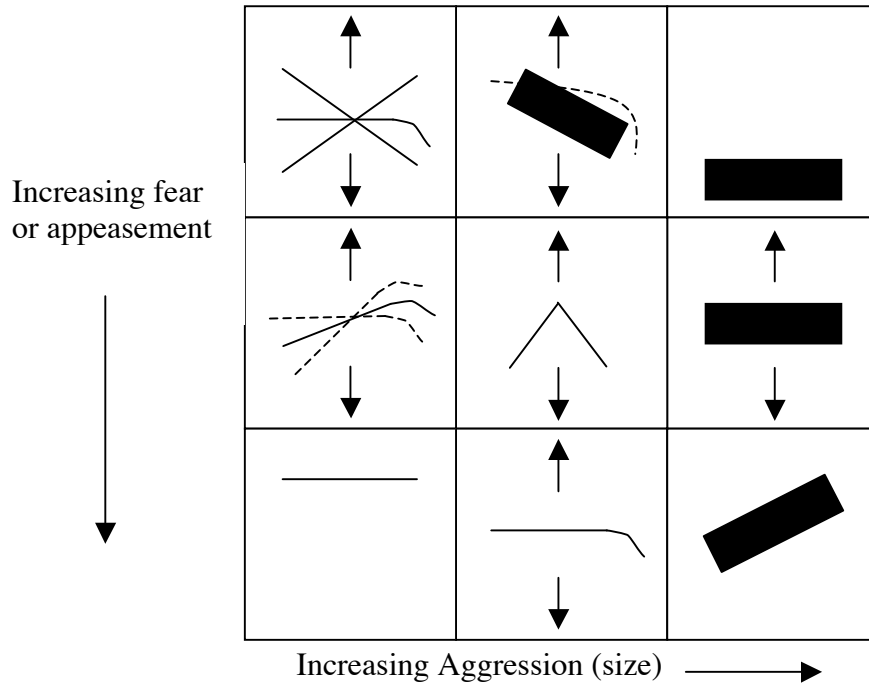
Figure 4: Morton's Motivational/Structural Rules
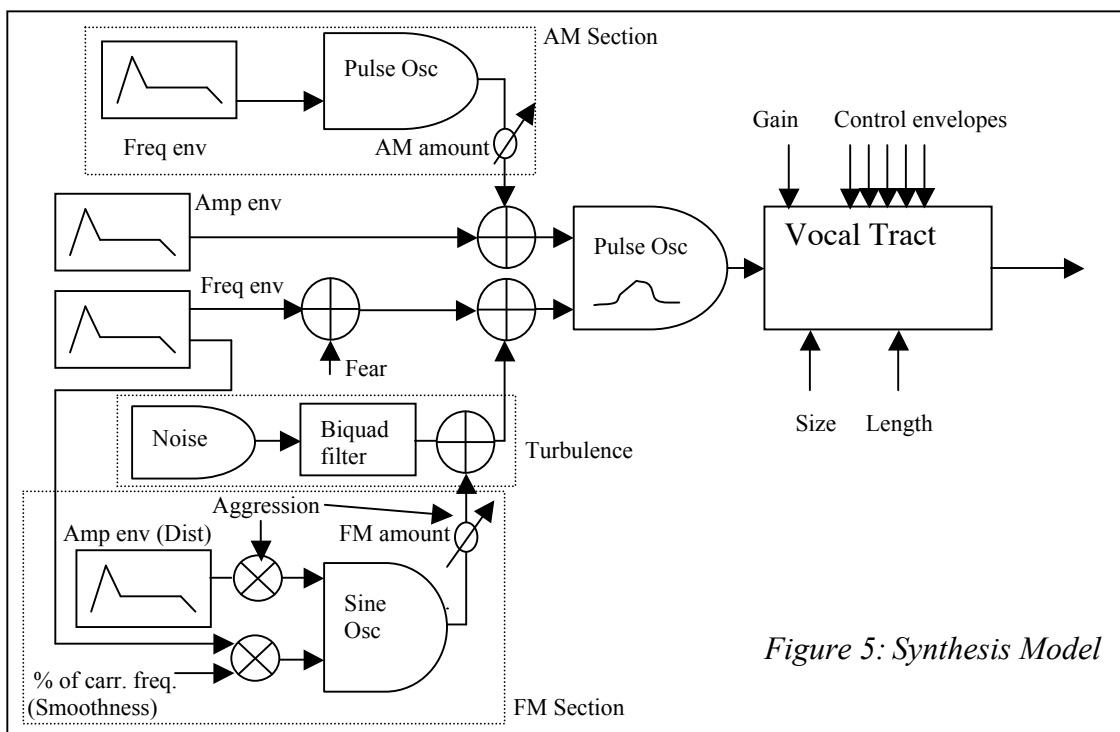
The Synthesis Model

Figure 5: Synthesis Model

The basic structure of Synthasaurus (figure 5) is a glottal pulse oscillator (which can be modified in a variety of ways) which passes through a waveguide model of a vocal tract. This model most closely reflects a mammalian vocal tract, although anuran and bird like calls are also possible because of the amplitude and frequency modulation possibilities incorporated into the model. The waveguide model of the vocal tract is similar to the one used in Perry Cook's SPASM. In this implementation, a simplified, six-section straight tube is used (as opposed to the three-way system used in Cook's model with throat, mouth and nose passageways).

The glottal oscillator is a custom MSP object designed for this project (developed in C with the MSP Software Development Kit), which provides a "smoothed" curve pulse wave that can be lengthened or shortened with a slider for different timbral qualities (this can also be set to modulate randomly). The user can specify a pitch envelope for this oscillator, and a frequency range within which this envelope works (as well as a base frequency). The user can specify an overall amplitude envelope function.

This oscillator can be amplitude modulated with a relatively low frequency (0-100 Hz) oscillator of the same smooth pulse type. This not only enables one to simulate the glottis upstream from the vocal cords in anuran vocal tracts, but also provides an effective way to create rapid "stuttering" effects which help in the creation of purring and growling type sounds. The pulse width of this oscillator can be controlled, as well as the strength of the amplitude modulation (0-100%). A configurable envelope function controls the modulation frequency.

A frequency modulation section provides for further signal modification. Low frequency modulation of the carrier waveform with a sine wave creates sidebands that

contribute to the "harshness" of the sound (which in turn often relates to the degree of aggressiveness in an animal call, as described by Morton). An envelope control is provided for controlling the depth of frequency modulation (which can be further strengthened by the "Aggression" parameter described later), as well as a slider for controlling FM frequency (which is calculated as a percentage of carrier frequency). Filtered noise can also be injected into the modulating oscillator's signal to simulate air turbulence in the vocal tract.

The vocal tract (figure 6) is also a custom MSP object developed for this project. It consists of a six section waveguide model, divided by junctions which reflect or transmit signal energy depending on the radius of each tract section, as described in Cook's model (Cook 1993). Envelopes can be defined to control the radii of the sections, and are input into the tract object as sample rate signals (for smooth sounding transitions).

Figure 6: Waveguide Model
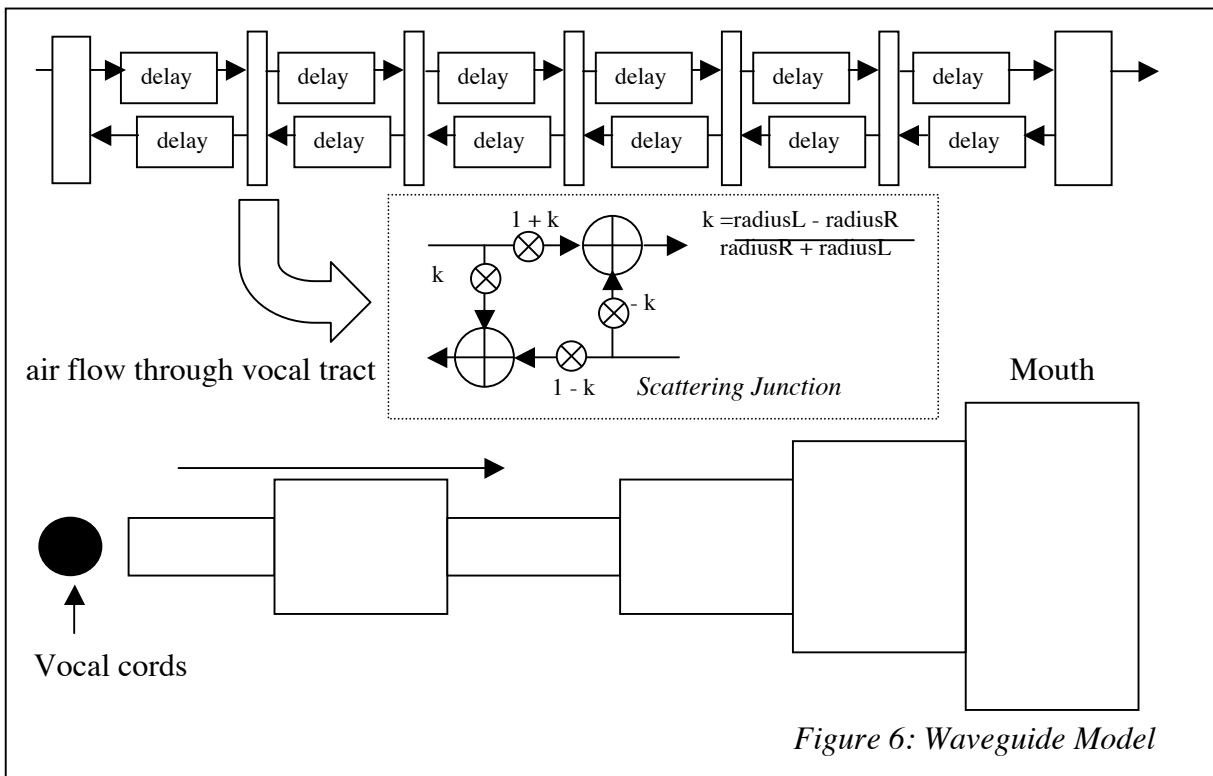
At the end of the tract model, where the "mouth" of the animal would be present, a simple crossover filter system controls the reflection characteristics of the vocal tract: higher frequencies escape the tract and lower frequencies are reflected back. The cutoff frequency of this filter can be controlled. By allowing more low frequencies to escape the tract, the impression of a larger tract (and thus larger animal) is created. The Vocal Tract Size slider represents this cutoff parameter. The delay time of the waveguide sections can also be increased to allow for the lengthening of the vocal tract.
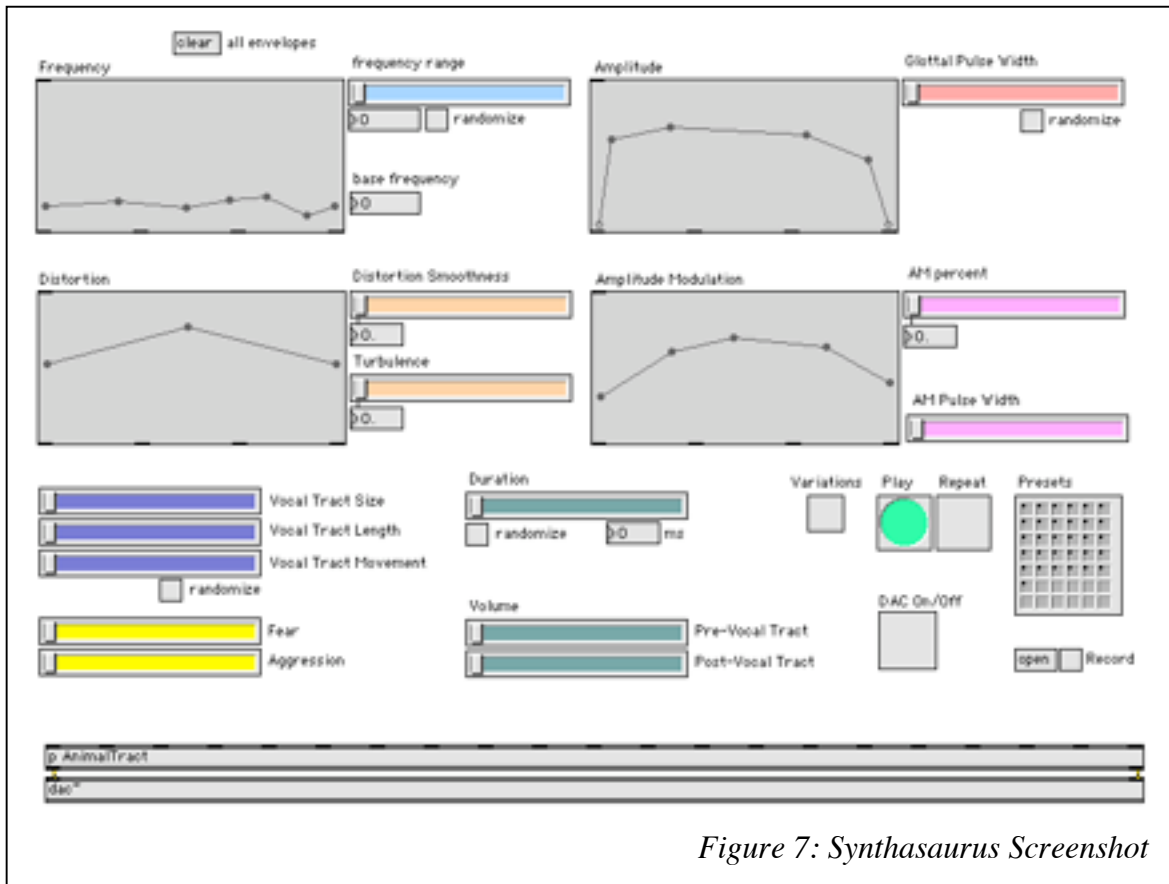
The User Interface
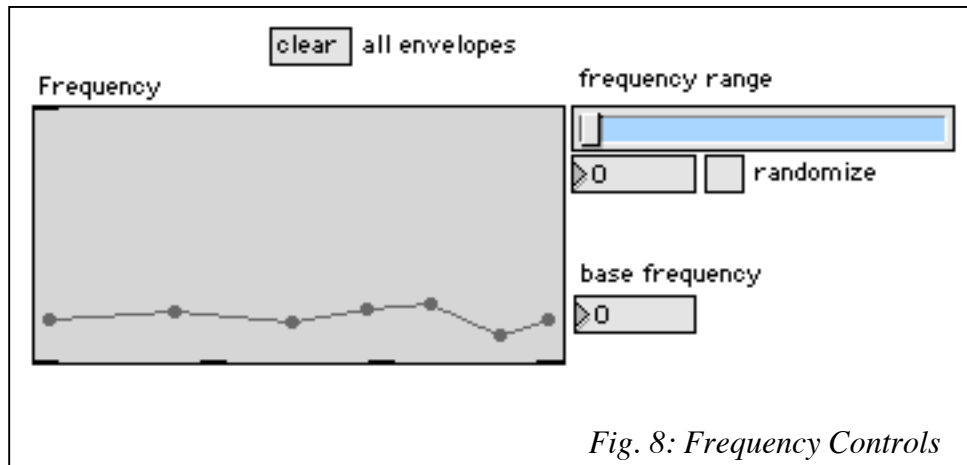

Figure 7: Synthasaurus Screenshot

Fig. 8: Frequency Controls

Several presets are provided in this version Synthasaurus, which demonstrate its ability to create a variety of emotive sounds. The most compelling characteristic of a given sound in conveying emotion is the pitch envelope (figure 8), which is a good place to start in designing new sounds. Recordings of real animal calls in spectrograph format (frequency vs. time) are useful examples for designing pitch envelopes. Any of the envelopes on the screen can be edited by dragging existing points, clicking outside a point to add a new one, or shift-clicking to remove a point. A randomize feature is included in the frequency section for providing variation on each playback by offsetting the frequency envelope by a constrained random value.
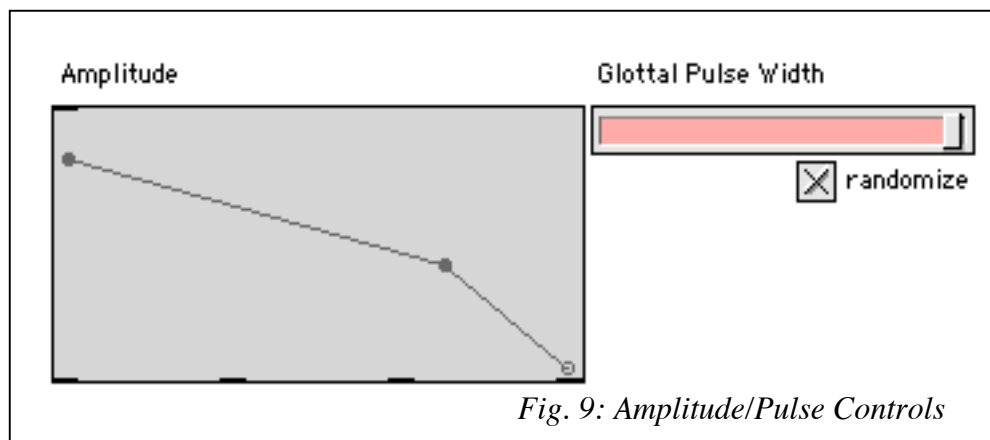

Fig. 9: Amplitude/Pulse Controls

An overall amplitude envelope (figure 9) provides overall volume contour for the sound, and sliders control pulse width (the narrower the pulse, the brighter the sound).

The frequency modulation section (figure 10) is useful for creating some aggressive distortion in the signal. At a low enough "smoothness" setting, frequency modulation becomes audible and is useful for bird like calls. The turbulence setting adds a degree of "breathiness" to the signal. Amplitude modulation (figure 11) enables one to create some interesting audible "pulsing" or "stuttering" effects, useful for feline purring and growling simulations.
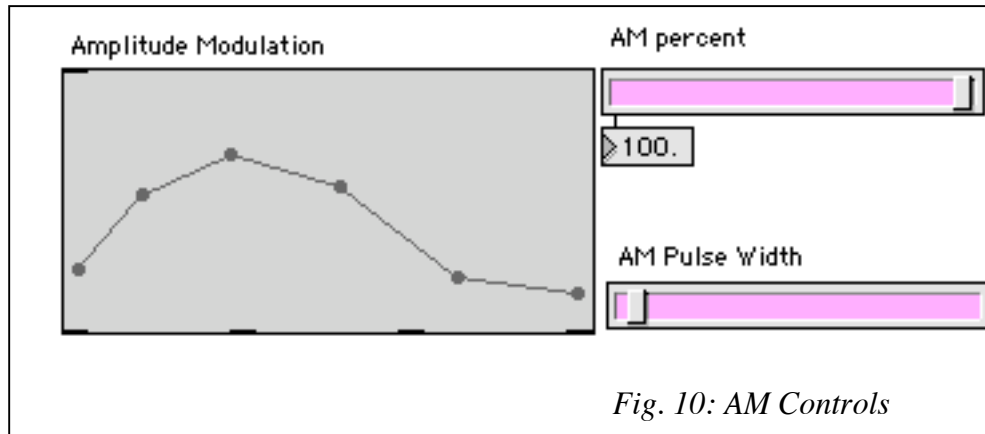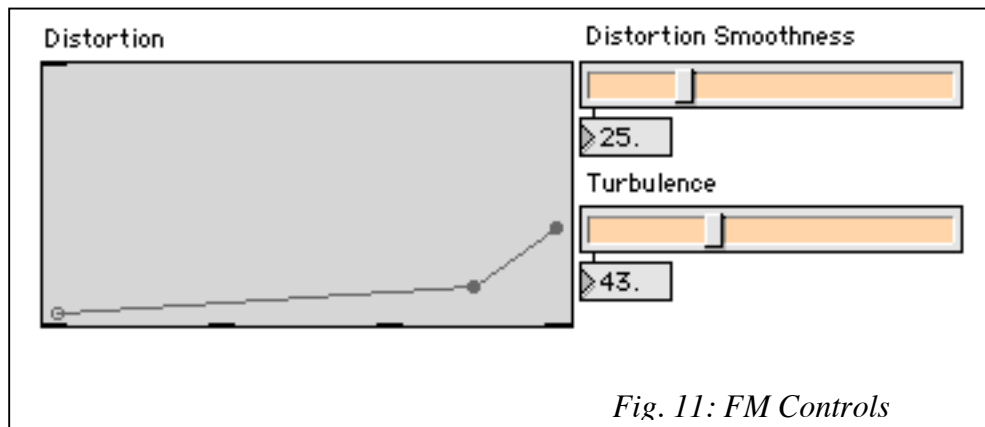
Fig. 10: AM Controls

Fig. 11: FM Controls

In the vocal tract section (figure 12), the size of the animal and articulation of the vocalization are easily specified. The Size slider controls the cutoff frequency of the

crossover filter at the mouth, and the Length increases the amount of delay in the waveguide model. The "Vocal Tract Movement" slider moves through a series of predefined vocal section movements, generally with the movement occurring more towards the base of the larynx when the slider is to the left and more towards the mouth on the right side. These variations of vocal tract movement create different kinds of articulations and formants during the course of a sound, sometimes effective in simulating a primitive "talking" effect. Enabling the Randomize feature causes a different tract movement to occur on each play occurrence.

Fig. 12: Vocal Model Controls
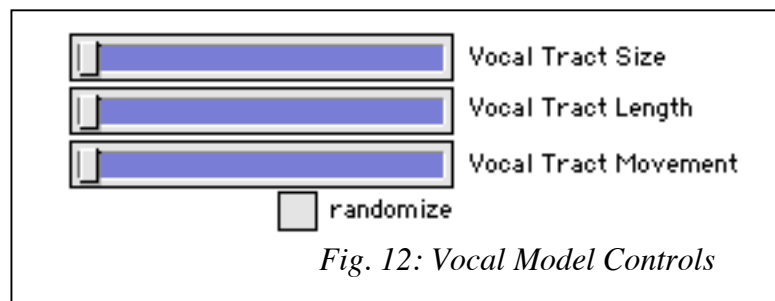
Sometimes the vocal tract model can overload due to its feedback nature. Thus two volume controls are provided (figure 14), one for pre-vocal tract gain and one for post-vocal tract. The pre-vocal tract slider should be set as high as possible without the system clipping.

The "Fear" and "Aggression" sliders (figure 13) attempt to map more emotive qualities to control changes consistent with Morton's Motivation/Structure rules. Increasing "Fear" simply increases the base frequency of the sound, while more "Aggression" increases both the FM amount (ratio of dry to FM signal) and strength of the Distortion (modulation depth) envelope, which effectively increases the "harshness" of the signal in most cases.
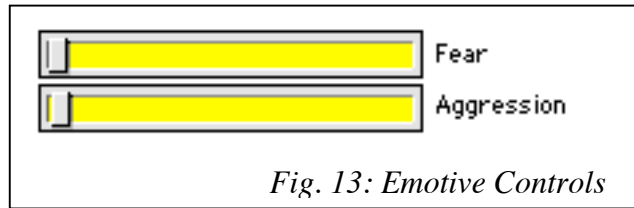
Fig. 13: Emotive Controls

An overall duration (figure 14) slider provides control for the length of the vocalization. This duration can also be randomized to a limited degree for each playback. Play controls (figure 15) are simple and include a repeat function so that sounds can be heard continuously while editing. The "Variations" toggle activates the random feature of the pulse width, vocal movement, and frequency sections so that each playback of a sound is a bit different. Presets are saved by shift-clicking in the preset box, and recalled by double-clicking.
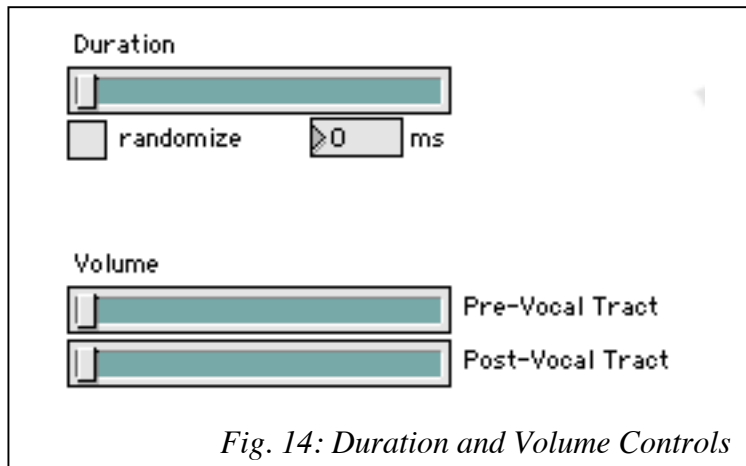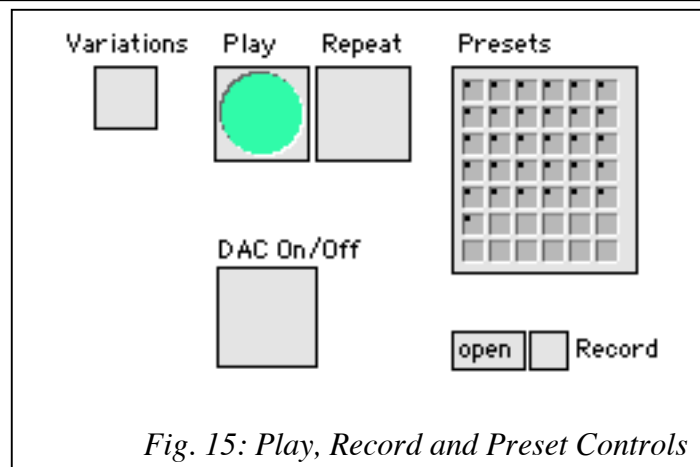

Fig. 14: Duration and Volume Controls

Fig. 15: Play, Record and Preset Controls

Considerations for Future Development

This incarnation of Synthasaurus attempts to make a useful step in the synthesis of emotive, easily controlled animal sounds. There are many ways in which this design could be further developed.

This model focuses on the creation of relatively short, one-oscillator timbres. A useful method of working with these sounds would be a "compositing" environment that enables both sequential and simultaneous mixing of voices to create more complex vocalizations.

Currently the user can only specify a pre-defined set of envelopes for vocal tract movement. Enabling the user to draw custom vocal tract envelopes on the user interface screen would be a useful feature, so that studies of actual animal mouth movements could be incorporated into sound design. Custom envelopes can be drawn if the user owns the development version of MAX (rather than just the stand-alone MAXPlay application), since the envelopes reside a couple patch layers underneath the user interface screen.

More realism could be incorporated by adding a feedback feature in the vocal tract model, which simulates the effect of reflected air influencing the nature of vocal cord oscillation, especially at higher air pressures. This may provide a more realistic distortion or harshness to the signal.

The approach in this project has been to simulate a general-purpose animal tract that could create a wide variety of textures. Further nuances related to specific kinds of animals could be modeled with a more configurable oscillator/vocal tract system (allowing one to model the dual bronchial passages in birds, with independently controlled oscillators, for example), or an altogether different kind of mechanism that

coupled the animal's vocal cords to the surrounding air (such as air sacs in frogs). Species like arthropods and employ quite different kinds of sound production mechanisms that may be interesting to model.

This synthesis engine might be a useful resource within a larger artificial intelligence environment, whereby a created "organism" interacting in an environment (either an abstract being in a virtual world, or a physical robot) may make life-like, emotive vocalizations in response to various stimuli. The pure synthesis approach used in this model (as opposed to using modifications of sampled audio) lends itself to more flexible control of sound parameters.

Conclusion

Hopefully this project encourages further research in synthetic animal-like vocalizations. There are many applications for such a synthesis engine in the game and film industries, as well as robotics and other kinds of artificially created "lifeforms". The unique nature of these kinds of sounds offers a new domain of timbrally rich and expressive qualities that could be used to interesting musical effect as well.

Acknowledgements

Northwestern University professors Gary Kendall (Music Technology), Ian Horswill (Computer Science) and Charles Larson (Communication Sciences) were extremely helpful in the design and implementation of this project. Collaboration with these three professors brought much insight into the relationships between the fields of

sound synthesis, robotics, artificial intelligence, and animal vocal tract physiology and anatomy.

References

Bradbury, J. W. and Vehrencamp, S. L. *Principles of Animal Communication*, Sunderland, Massachusetts: Sinauer Assocaites, Inc., 1998

Cook, P. "SPASM, a Real-Time Vocal Tract Physical Controller*" Computer Music Journal*, 17(1), 1993

Darwin, C., *The Expression of Emotions in Man and Animals*, Chicago: University of Chicago Press,1965

Morton, E. S. and Page, J. *Animal Talk*, New York: Random House, 1992