

# SYNTHESIS OF THE SINGING VOICE USING A PHYSICALLY PARAMETERIZED MODEL OF THE HUMAN VOCAL TRACT

**Perry R. Cook**

**Center for Computer Research in Music and Acoustics (CCRMA),  
Stanford Univ.**

This is SPASM, Singing Physical Articulatory Synthesis Model. It is the result of my research in voice synthesis at CCRMA, Stanford University. The research was performed while working toward a PhD. in Electrical Engineering, and continues since I completed the degree in January 1991. The idea of using this type of model is to model the production mechanism rather than the resulting speech sound. This approach results in naturalness of sound and control. Synthesizing transitions in the shape space generates acoustically plausible time variations, unlike many other voice synthesis techniques.

Running the program and selecting **Help** from the menu will pop up a window with a short paper explaining the SPASM project.

Some things to know about running SPASM:

After running, you might need to select **Set Defaults** from the main menu, and set the complete file path with final slash. This sets the directory where the directories shapes, glots, etc. will be found. The program will pop up an alert panel if it needs you do to this. Example file path:

/user/prc/Library/SPASM/

In order to synthesize any new soundfiles, you must be the owner of the SPASM directories.

Most stuff works in the DSP, including now the noise injection. The ShapeInterpolator and TractView windows have gotten lots of attention, allowing much smoother and faster real-time DSP control. Look for MIDI control soon.

Only a few shapes are stored in the shapes library in the demo version. The shapes available are: ahh, eee, ooo, rrr, mmm, nnn, shh, and nothing. The shape called nothing is really a non-shape, which is a straight tube with no reflections at either end. This allows you to hear the glottis or noise generator alone.

The glottis files which come with the demo version are:

pulse10.glt, pulse20.glt, pulse30.glt (band-limited pulses)  
test.glt, test20.glt (standard parameterized glottis waveforms)  
loud.glt, soft.glt. (used in the crescendo example)

The soundfiles in the DEMO window are:

sounds/aheooo	three steady state vowel sounds
sounds/machines	same vowel sounds, but without vibrato
sounds/arpeg	same vowel sounds, but on different pitches
sounds/nasals	three steady nasal (liquid consonant) sounds
sounds/fricatives	some steady state fricative consonants
diphones/diphs	transitions between vowels, called diphthongs
diphones/nasals	transitions between nasals and oral vowels (listen in stereo!)
diphones/plosives	voiced plosive consonants (importance of throat radiation)
diphones/cresc	a musical crescendo (amplitude AND spectral evolution)
nphones/shiela	her first word, and as a result, her name
nphones/vocaliz	a vocal exercise commonly used to train singers

The files nphones/shiela and nphones/vocaliz were generated using the singer program, which is the ANSI C companion to SPASM. The singer program takes text input and generates connected singing. A new companion program, LECTOR, reads Ecclesiastical latin text and generates singer code for synthesis. Look for singer and LECTOR on the ftp site soon.

Have fun!!