

Using the Klatt Synthesizer

On the PC, it is a DOS program: SenSyn.
This is the Klatt & Klatt (1990) synthesizer.

On the Mac (OSX), the program is NXKlatt.
This is the Klatt (1980) synthesizer.

These are resonance synthesizers. The user specifies the source characteristics and the formant characteristics at each point in time. The synthesizer will compute the corresponding sound.

To make a set of tokens that form a series, one or more of the parameters that control the source and formant characteristics are systematically changed from values appropriate for one token or phoneme to values appropriate for another token or phoneme.

I. An Example

To make a series that varies in voicing from /ba/ to /pa/, the acoustic correlates of VOT (voice onset time) are manipulated.

A series of VOTs with values of 0, 10, 20, 30,... 60 msec could be generated. For any particular VOT, the initial part is voiceless. In this segment, set F0 to 0, AV to 0, B1 to 400 and AH to 60. This turns on aspiration and turns off voicing.

At the point where the source is to shift from voiceless to voiced, set AH to 0, B1 to the bandwidth for the rest of the vowel (e.g. 60), F0 to the target value (e.g. 119 Hz), and AV to 60.

This description neglects any formant transitions, release burst, etc.

II. An Overview

A) Cascade versus Parallel mode

The synthesizer converts a description of the formants (frequencies, bandwidths and amplitudes) and source characteristics (F0, amplitude, frication, other source qualities) into the equivalent sound.

The synthesizer has two modes of operation: serial (cascade) and parallel.

In cascade mode, with voiced or aspirated sources, the user specifies the frequency and bandwidth for each formant.

Amplitudes are determined automatically (based on vocal tract constraints).

In parallel mode and in either mode with frication as the source, the user specifies the frequency, bandwidth and amplitude for each formant.

B) Source characteristics

The source characteristics include F0 (fundamental frequency), AV (amplitude of voicing), AH (amplitude of hiss), AF (amplitude of frication), AN (nasal amplitude), OQ (open quotient, SenSyn only) and an assortment of parameters to tune the source to mimic specific voices.

The user specifies the parameters for each point in time. By default, this is every 5 msec. However, in SenSyn, *updating of the fundamental is only done at the beginning of a vocal pulse*. So, at the onset of a vocal pulse, the control parameters from the most recent point in time are used. This can make precise manipulation of timing (such as VOT) tricky.

C) Formants

There is a complete set of parameters for each of F1 through F6. In SenSyn, these are:

F_n	Frequency of formant
B_n	Bandwidth of formant
A_nF	Amplitude of fricative formant
A_nV	Amplitude of voiced formant (parallel mode, 1-4 only)
AB	Amplitude in fricative bypass path

There is also a set of values for a nasal formant (pole and zero).

In NXKlatt, A_nF and A_nV are not separate. There is one set of formant amplitudes: A_n .

(Note – n is 1, 2, 3, 4, 5, or 6)

III. A Sample Utterance

A) /bae/

The starting point for this syllable is to set the overall duration. The syllable will be 290 msec.

The beginning of the syllable /bae/ has initial formant transitions, and these are followed by a “steady-state” vowel. We choose a series of points in time to specify the formant frequencies. The synthesizer will interpolate the values between these points in time for us.

For a male syllable, we might specify F1 as follows:

time	value
0	400
30	632
150	652
285	695

The specification of F2 and F3 would be similar (rising initial transition). F4 and F5 would be fixed (not time varying).

AV would be 60 in the first frame (time 0) and 64 in the next frame (time 5). From 5 to 100 it stays at 64 then tapers to 61 at time 220. From time frame 220 to 270 it would taper down from 61 to 50 and then to 34 for frame 285.

F0 is 105 at onset (specified as 1050 in SenSyn), rising to 124 (1240) at time 35 and then falling to 120 (1200) at time 100. It then falls to 101 (1010) at time 230 and then to 98 (980) at time 285.

This is the “standard” stressed syllable profile for a male talker.

Finally, a release burst can be added by turning on AF to 63 for time 0 only (0 for the rest, time 5 to the end) and setting AB to 63 for time 0 through 20, then to 0 for the rest. This creates a brief, wide-band release burst.

B) The VOT series

For each VOT (in 5 or 10 msec steps from the 0 msec /bae/ above), you would:

Turn off AV (0), set F0 to 0, turn on AH (60), and widen B1 (to 400).

Your Assignment:

Make a VOT series that ranges from /d/ to /t/ or /g/ to /k/. In doing this, you can choose any vowel that you wish.

There must be a minimum of 7 stimuli in the series. The step size in VOT between stimuli should be 5 or 10 msec.

You need to hand in the parameter files and the audio files. If the synthesizer is not the SenSyn (Klatt & Klatt, 1990) or the NXKlatt (Klatt, 1980) synthesizer, then I will also need a copy of the synthesizer. The files will be returned to you after they have been examined.

In doing this project, you may work with a second person. If you do so, hand in only one copy of the materials with both names on it.

Your Assignment (Optional, Part 2):

The synthesizers, particularly SenSyn, are “universal”. They can be used to do any phonetic contrast in any language. They can also mimic any voice.

Talkers differ in vocal track length, vocal fold size, and idiosyncrasies of how they speak (produce the gestures of speech). Record a talker saying an isolated vowel. Now, make a synthetic replica of the vowel. It should sound like the correct vowel and the original talker.

In doing this, you will need to use the synthesizer in parallel mode so that you can control the formant amplitudes to match them to the target talker. If you want to match a female voice, use SenSyn as it has parameters such as OQ that are needed.

Use Praat to evaluate the result of your synthesis and help you focus on the parameters that need to be adjusted.

The parameters that you will need to work with are:

F1, F2, F3, F4
B1, B2, B3, B4
A1, A2, A3, A4

F0, AV, OQ, possibly AH (breathy voice)

Again, turn in both the parameter listing/file and the audio files.