6.542J, 24.966J, HST.712J LABORATORY ON THE PHYSIOLOGY, ACOUSTICS, AND PERCEPTION OF SPEECH
Fall 2005

Kenneth N. Stevens Joseph S. Perkell
Stefanie Shattuck-Hufnagel

Lab 16 *11/08/05*

Higher–level synthesis with a formant synthesizer, using quasi-articulatory parameters

References

Hanson, H.M. and K.N. Stevens (2002), *Control of acoustic source parameters in speech synthesis: A quasi-articulatory approach*. **J. Acoust. Soc. Am.**, 112, 1158-1182.

Introduction

In this lab, you will do some speech synthesis exercises using a quasi-articulatory synthesizer called HLsyn (HL for "higher-level"). This synthesizer is controlled by a set of 13 parameters, rather than the 40-odd parameters used to control the acoustic sources and transfer functions in KLsyn that was used in the earlier synthesis lab. The HL parameters are converted, through a set of mapping relations, into the KL parameters, as schematized in Fig. 1 (taken from Hanson and Stevens). The figure also lists the 13 HL parameters, and shows how these parameters are related to the vocal tract shape, the laryngeal configuration, and the respiratory pressure. The parameters are also listed in Table I.

Rather than controlling directly the acoustic source parameters and the filtering, HLsyn controls the articulatory parameters that give rise to the sources and filters. A central component of the mapping relations is the calculation of airflows and pressures in the vocal tract, and the estimation of acoustic sources (periodic or turbulence noise) that result from these airflows and pressures at constrictions in the airway. The methods for calculating the sources are described in Hanson and Stevens (e.g. see Fig. 4 of that paper).

Procedure

In this lab, you will do exercises that illustrate how the HLsyn synthesizer works. The synthesizer is installed on three computers running Windows, and it will be necessary to do this lab at workstations different from the ones used in other labs. Two of the computers you need to use are

in the Speech Lab. You can log on to the computer using the course account and password. In Windows, you will find on the desktop icons for two programs that you are going to use for the lab: SpeechStation and HLsyn. You will use HLsyn to do the synthesis and SpeechStation (speech-analysis software developed by Sensimetrics Corp.) to examine the spectrogram and F0 contour of the synthesized utterances.

Task: Modifying voicing and manner of articulation for labial consonants.

When you double click on the HLsyn icon, the HLsyn window will appear. Click on the "Open" button and navigate to *'Killick1\NT_Users'(W:)* and then to the file **abit_grpx.hld** (where "x" represents your group number, for example, group 1 should use the file **abit_grp1.hld**); these files should be in the *labc* User folder. The file contains the HL parameters used to synthesize (by hand) the brief utterance "a bit". A spectrogram of this synthesized utterance is attached (Fig. 2). A number of HL parameters were manipulated to synthesize this utterance. You can look at these parameters either in numerical form or graphically (click on "view" and select HL graphic or HL parameters). If you click on "view", you can also select "Spectrogram" to look at the spectrogram of the synthesized utterances. If you want to use SpeechStation, you need to save the synthesized file as a .wav file and open it with SpeechStation.

Notice that the parameter *al* (cross-sectional area of lip opening in mm^2) goes to zero to make the /b/, and is then released at time 250 ms. The subglottal pressure *ps* is about 8 cm H_20 through much of the utterance, but is lower toward the beginning and end, reflecting the onset and termination of the utterance. The area of the glottal opening *ag* remains at about 4 mm^2 through much of the utterance, particularly during the /b/ closure. The cross-sectional area of the velopharyngeal opening *an* remains at zero since there are no nasal sounds in this utterance. Notice that the parameters *dc* and *ue* are used to help to maintain vocal-fold vibration during the /b/ closure. (See paper by Hanson and Stevens.)

You can also view airflow and pressure parameters (choose PF graphic or PF parameters under "view"). Observe how intraoral pressure (*Pm*) builds up in the /b/ closure, and note the peak in airflow at the mouth (*Uacx*) immediately following the /b/ release. Explain these changes in *Pm* and *Uacx*.

You can also look at the KL parameters that emerge from the mapping relations. For example note how AV (amplitude of voicing) changes with time through the /b/ closure and just before and after this closure. Also observe AF near the release, and look at the HL and KL parameters used to synthesize the final /t/.

(i) As a first exercise, change the utterance from "a bit" to "a mitt". To do this, introduce the parameter *an* (area of the velopharyngeal opening). Start *an* increasing about 40 ms before the labial closure, to a maximum of about 20 cm^2 during the closure, and then decreasing to zero about 40 ms into the vowel /ɪ/. You may want to undo the parameters *ue*, *dc*, and *ap* that are introduced in "a bit". After you have made changes, click on "Synthesize". You can also click on "Play" to hear the utterance. Experiment with the timing and maximum value of *an*. Observe the airflow and intraoral pressure during /m/.

(ii) As a second example, start with the original "a bit" and change it into "a pit". This should be done by increasing the glottal opening *ag*, beginning about 20 ms prior to the labial closure. Make it increase linearly from this point to a maximum of about 30 mm^2 at the time of the labial release. Then decrease *ag* back to its nominal value of 4 mm^2 about 50 ms after the release. This should lead to an interval of aspiration. You should undo the parameter *ue* used for /b/, and you should change the sign of *dc* to simulate increased stiffness of the vocal folds and of the vocal-tract walls. Observe and record various aspects of pressures, flows, and KL source amplitudes (AV, AF, AH).

(iii) Make a few changes to the parameters in the original "a bit". Some suggested changes are:

- make *ue* = 0 in the /b/.
- set *dc* = 0 throughout.
- set *f0* = a constant (say 1200) throughout. Note that the KL parameter F0 is not constant due to other effects. Explain.
- set *ps* = a constant throughout.

In each case, observe the result, and comment.

(iv) If time permits, try synthesis of other variations of "a bit", such as "a bid" or "a bat".

Table I. Description of HLsyn parameters

Parameter	Range	Description
f1	150-1300	First natural frequency of vocal tract, assuming no local constrictions (Hz)
f2	550-3000	Second natural frequency of vocal tract, assuming no local constrictions (Hz)
f3	1200-4800	Third natural frequency of vocal tract, assuming no local constrictions (Hz)
f4	2400-4990	Fourth natural frequency of vocal tract, assuming no local constrictions (Hz)
f0	0-5000	Fundamental frequency due to active adjustments of vocal folds (Hz x 10)
ag	0-40	Average area of glottal opening between the membranous portion of the vocal folds (mm^2)
ap	0-40	Area of the posterior glottal opening (mm^2)
ps	0-20	Subglottal pressure (cm H_2O)
al	0-200	Cross-sectional area of constriction at the lips (mm^2)
ab	0-200	Cross-sectional area of tongue blade constriction (mm^2)
an	0-100	Cross-sectional area of velopharyngeal port (mm^2)
ue	-200 - 200	Rate of increase of vocal-tract volume (cm^3/s)
dc	-100 - 100	Change in vocal-fold or wall compliances (%)

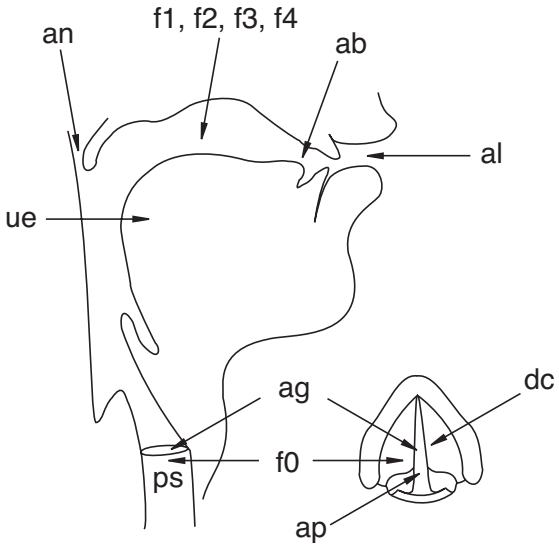

Figure 1(a)

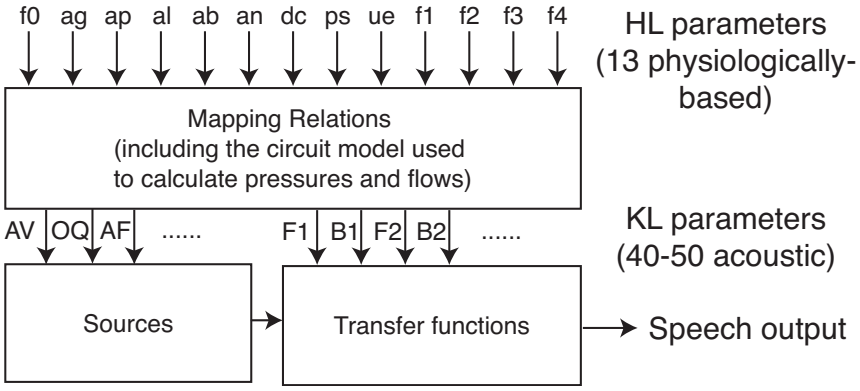

Figure 1(b)

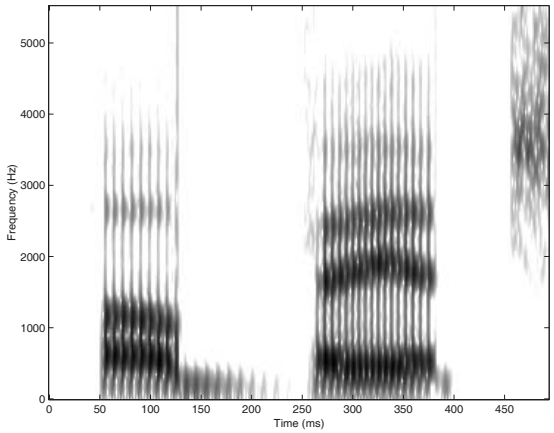

Figure 2