

Figure 3.1. Ambiguity between the source and vocal tract models is illustrated with three examples. In case 0, a normal glottal source and vocal tract give rise to a normal voice; the normal LF glottal flow derivative waveform, normal vocal tract log spectrum transfer function magnitude, and normal voice time series and spectrum for /a/ are plotted. Case 1 shows a pathological glottal source missing the normal sharp return; it is convolved with a normal vocal tract model to give rise to a pathological /a/ with sinusoidal appearance and missing high frequency formants. This voice is perceived as "breathy." Case 2 combines the normal glottal source with a vocal tract model with greatly attenuated high frequency formants; the resulting voice time series and spectrum is exactly the same as case 1. Thus, working backwards from the resulting time series of cases 1 and 2 (as is attempted in inverse filtering and formant analysis), it is impossible to arrive at the correct vocal tract and source model without additional assumptions or information.

PC #1: STIMULUS GEN.

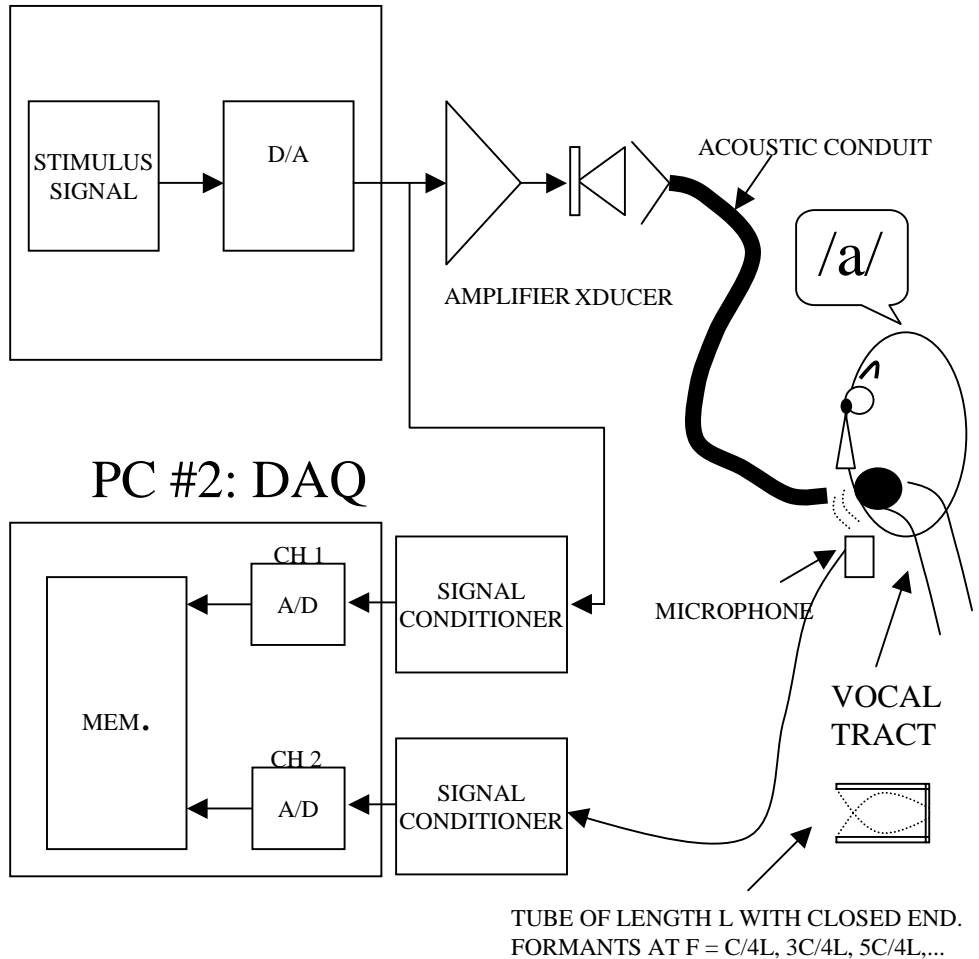


Figure 3.2. System setup for external stimulation of the vocal tract. Two PCs are used: PC #1 generates a stimulus signal to excite the vocal tract. An audio amplifier and transducer transform the stimulus to sound, which is then conducted via an acoustic conduit to either a quarter wave tube model or the subject's vocal tract. PC #2 is used to record the stimulus and response signals. A sample of the stimulus is acquired on channel 1 of an analog to digital converter and is stored to memory. The response is recorded with a small microphone, conditioned and acquired on channel 2 of the analog to digital converter. The tube model is used for proof of principle and system validation.

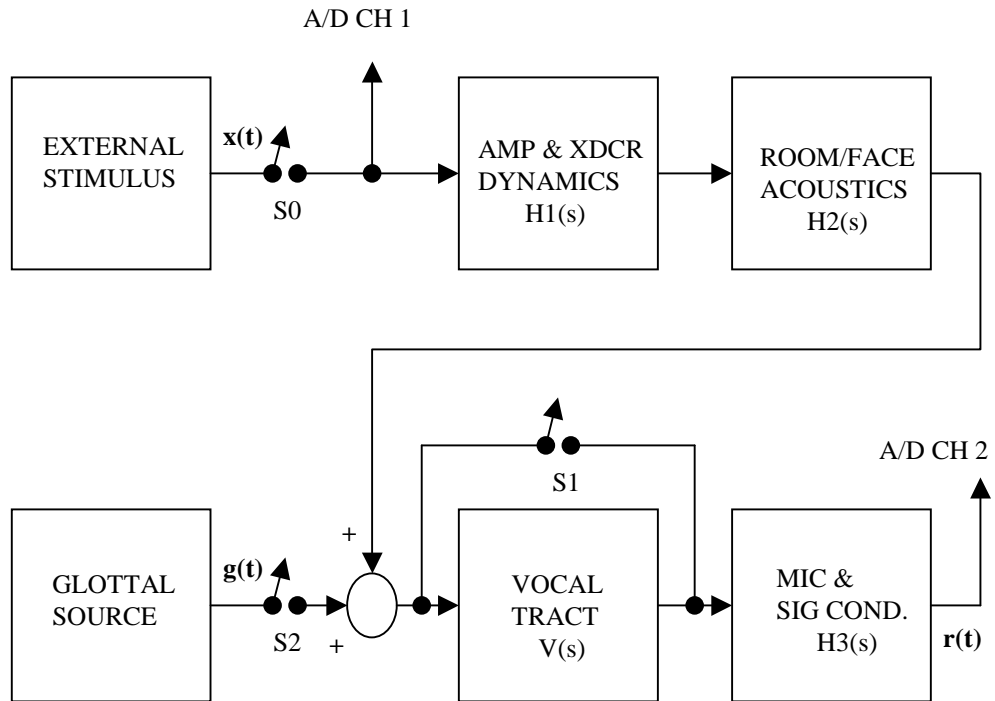


Figure 3.3. Model of external source formant analysis system. In addition to the normal glottal source $g(t)$ and vocal tract $V(s)$ pathway, an external source $x(t)$ is added to stimulate the vocal tract. Stimulus electronics, transducer, and environmental acoustics are modeled in $H1$ and $H2$; microphone and signal conditioning are modeled in $H3$. Switches $S0$ and $S2$ represent presence/absence of $x(t)$ and $g(t)$; $S1$ represents presence/absence of the vocal tract (mouth open or shut).

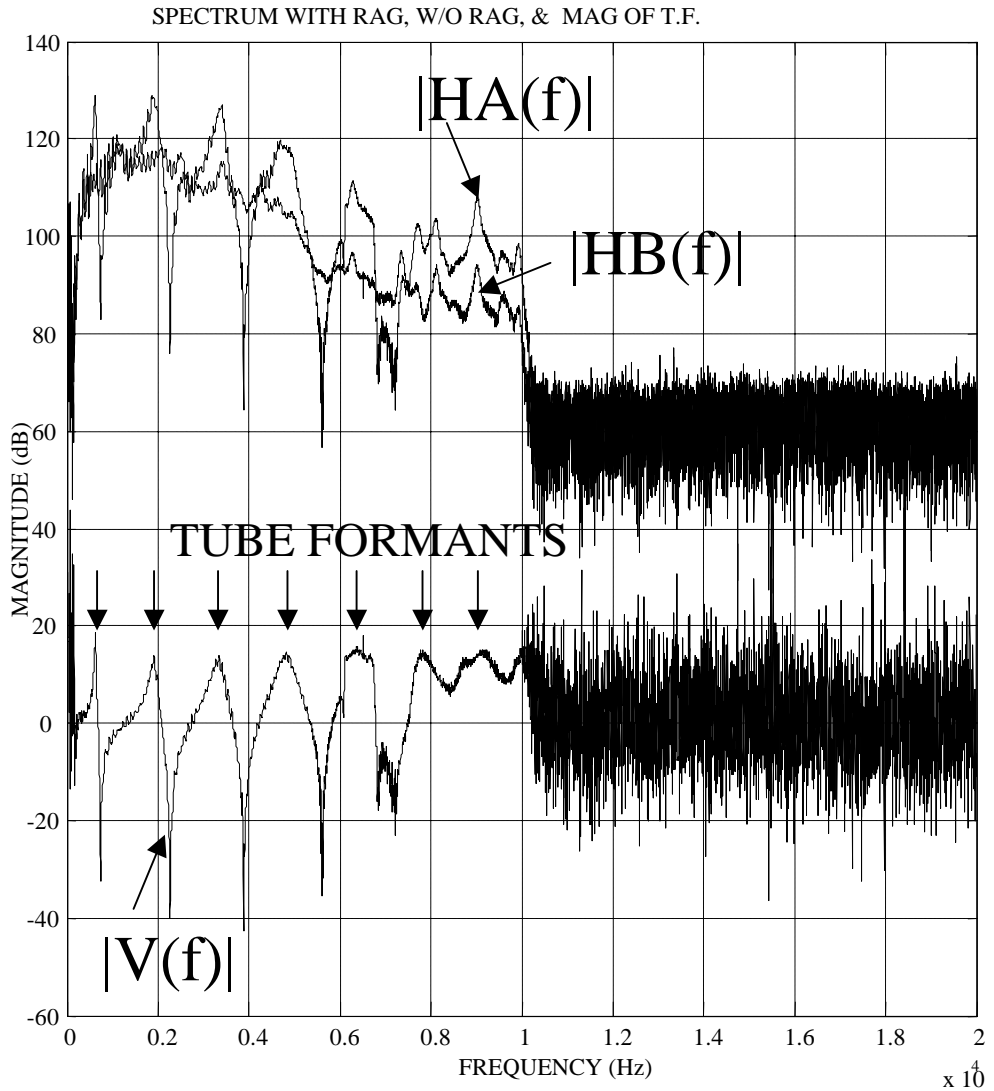


Figure 3.4. Validation of external source testing setup using a simple quarter wave tube closed at one end. HA is the spectrum recorded by the microphone near the open end of the tube when a 0 – 10KHz stimulus sinewave sweep is executed. HB is the “background” spectrum recorded when the tube is stuffed with a rag, effectively eliminating the tube formants. V is HA – HB, revealing the expected quarter wave tube formants. All spectra become noise above 10KHz at the end of the sweep.

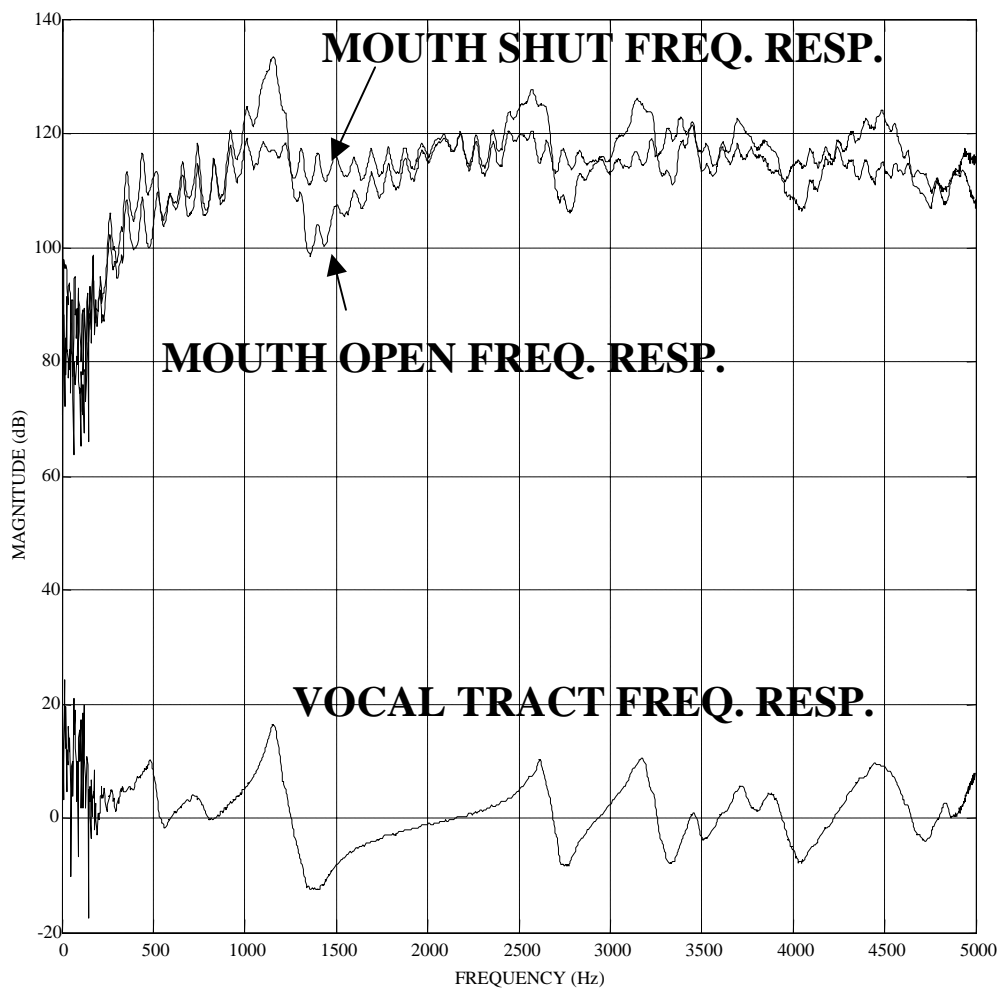


Figure 3.5. Vocal tract transfer function for /a/ with a normal male voice. Analogous to Fig 3.4, the chirp response spectra of mouth open/shut (top curves) are subtracted to yield the vocal tract frequency response (bottom). The expected formants for /a/ are present, plus additional peaks. Ripples at 94 Hz spacing in the top curves due to acoustic conduit resonances are effectively cancelled by the subtraction process.

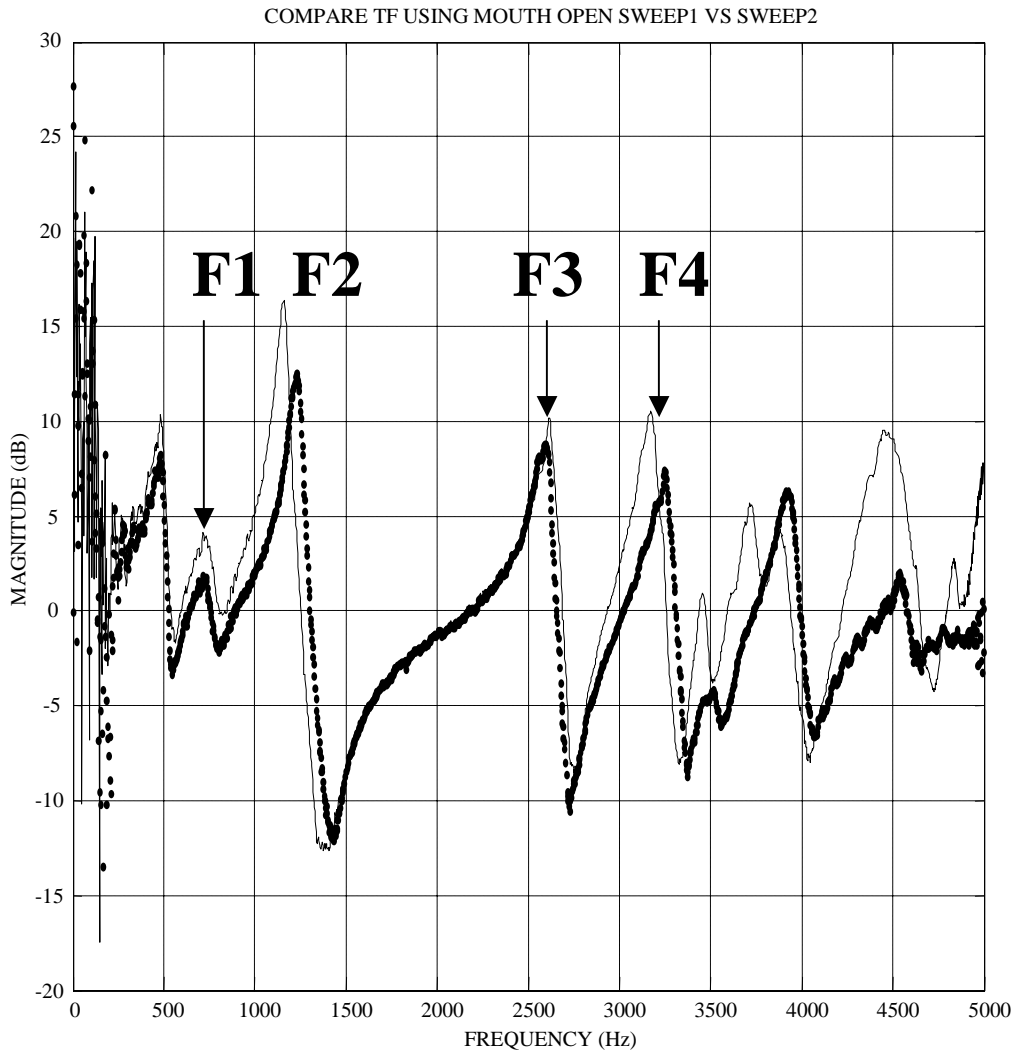


Figure 3.6. Drift in formant peaks during testing. In order to estimate effects of articulator movements, the vocal tract calculation is repeated using different chirps from the same time series. In this case, the peaks remained fairly constant with about 100 Hz change at 3200 Hz (3%).

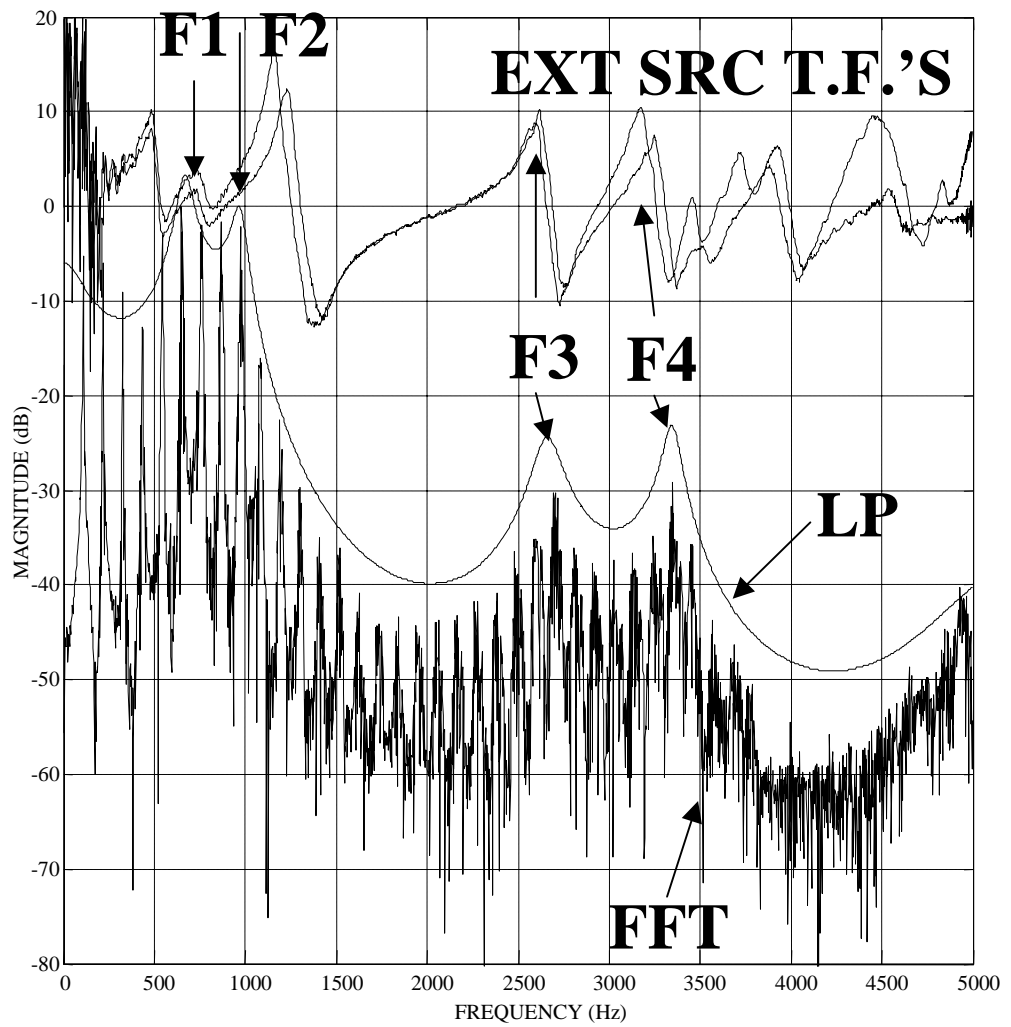


Figure 3.7. Comparison of traditional FFT/LP analysis with the ES analysis of Fig. 3.5. Formant peak shifts between FFT/LP and ES resonances are probably due to involuntary articulator movement from the vocalizing to non-vocalizing transition. Many resonances are shown in the ES data that are not revealed by FFT/LP.

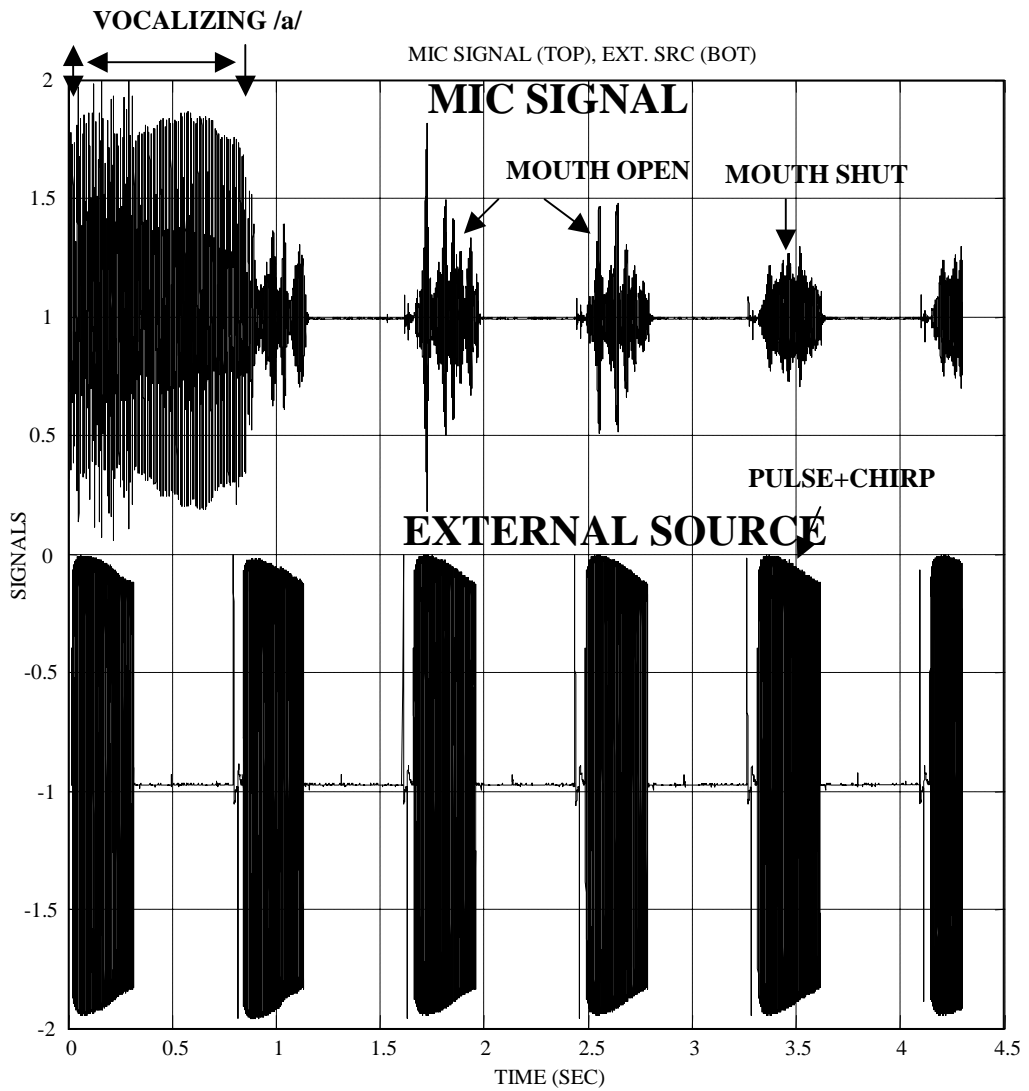


Figure 3.8. Time history of vocal tract audio output (top) and the ES (external source) stimulus (bottom). Periodic pulse + sine chirps are interspersed with periods of silence, during which vocalization (alone) can be recorded (0.3 s to 0.8 s). The resonances of the vocal tract are clearly visible in the mouth open responses at 1.7 s and 2.6 s. Note that the first chirp occurs during vocalizing; this was done to allow collection of data for future analysis using correlation techniques to separate the responses to chirp and glottis.

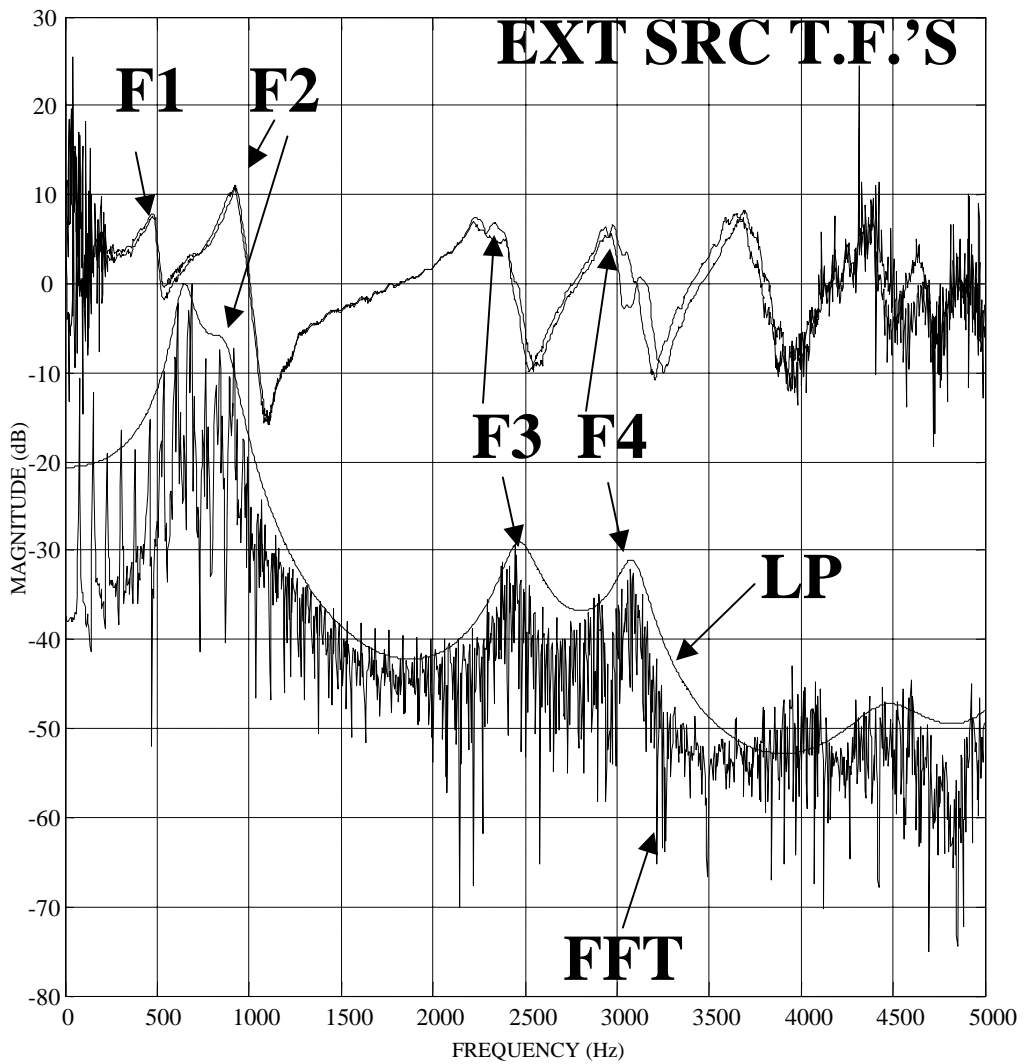


Figure 3.9. FFT/LP/ES analysis of another instance of normal male /a/. Compare with Fig 3.9b which is a breathy /a/ from the same time series.

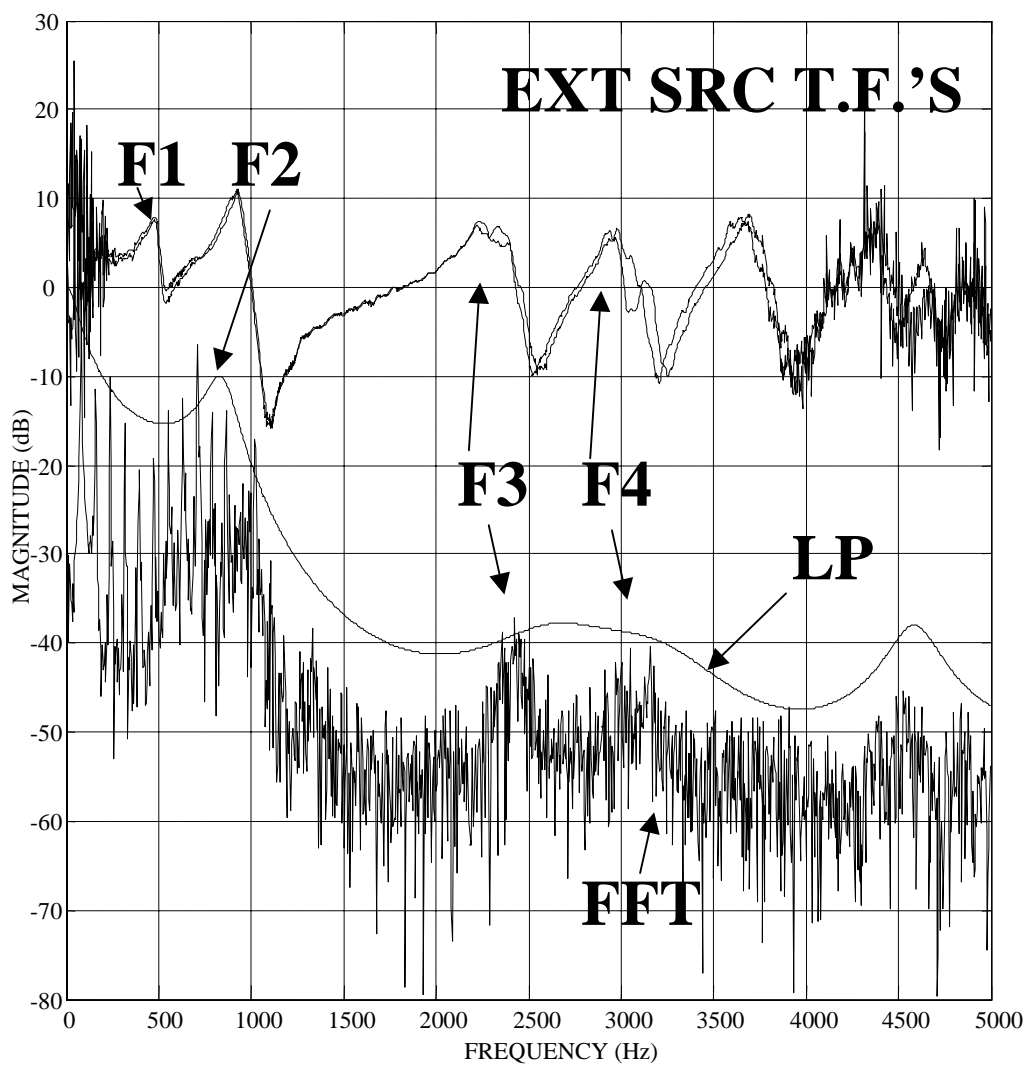


Figure 3.9b. FFT/LP/ES analysis of a breathy /a/ from the same time series as Fig. 3.9.

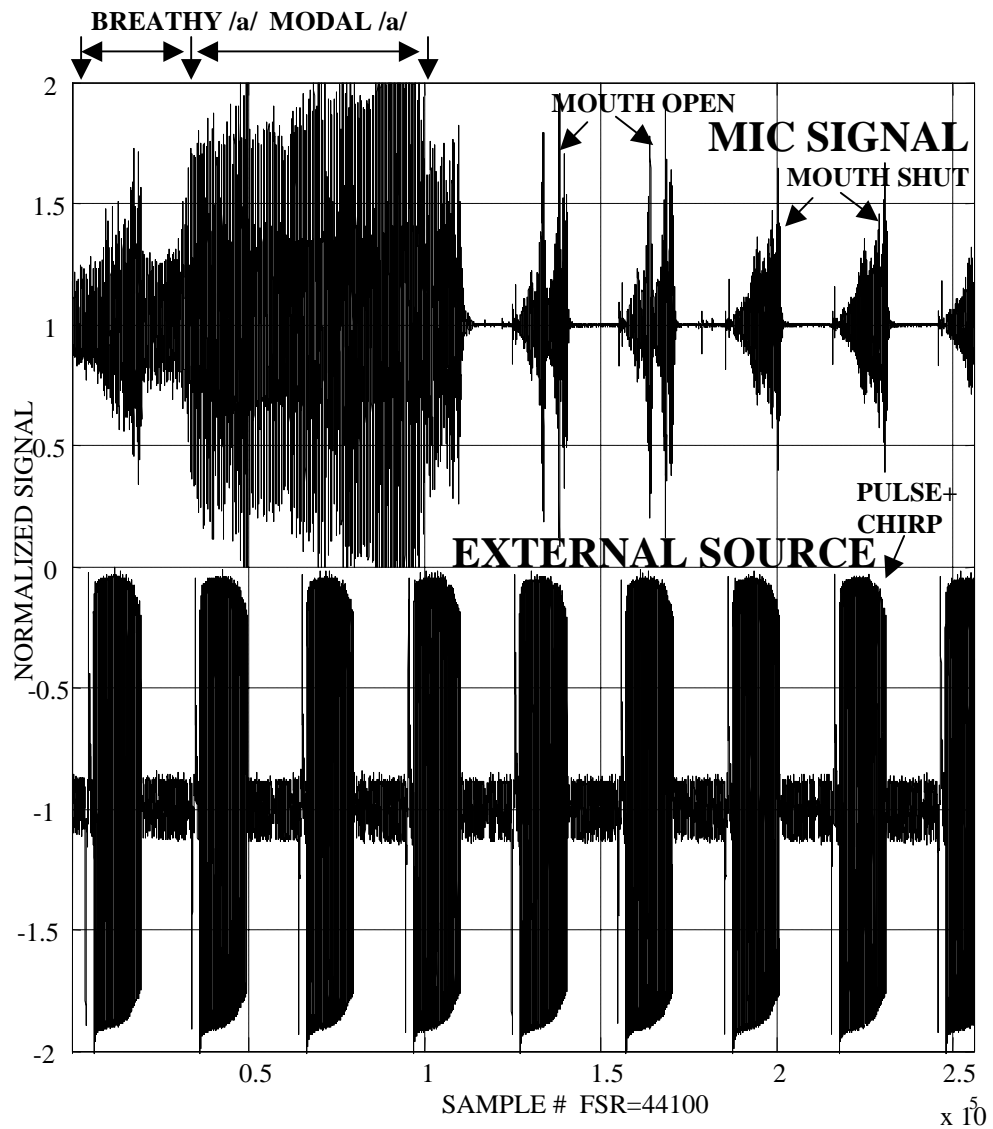


Figure 3.10. Time history of vocal tract output and ES (analogous to Fig. 3.8) with an added breathy segment at the beginning. The voice is the author's.

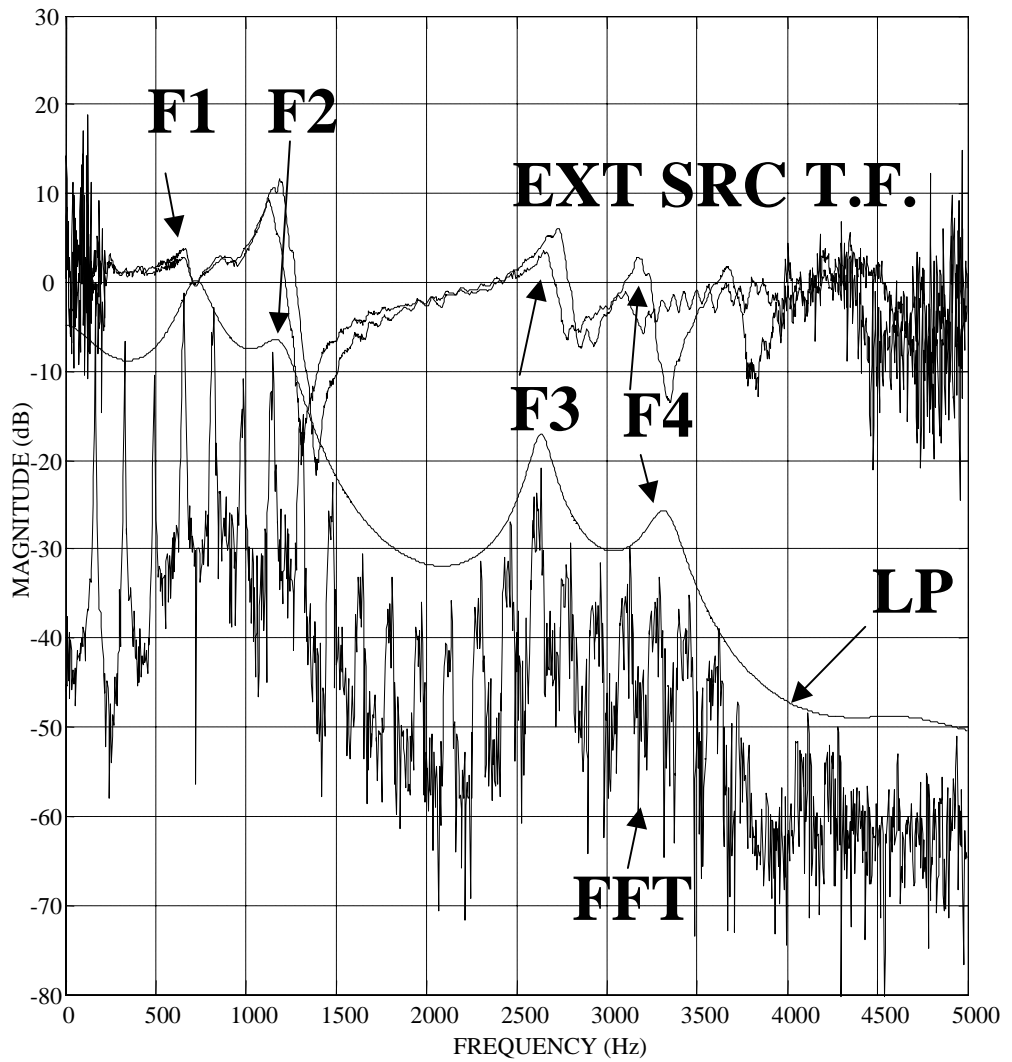


Figure 3.11. FFT/LP/ES analysis of a normal female /a/. Compare to Fig 3.11b which contains a breathy /a/ from the same time series.

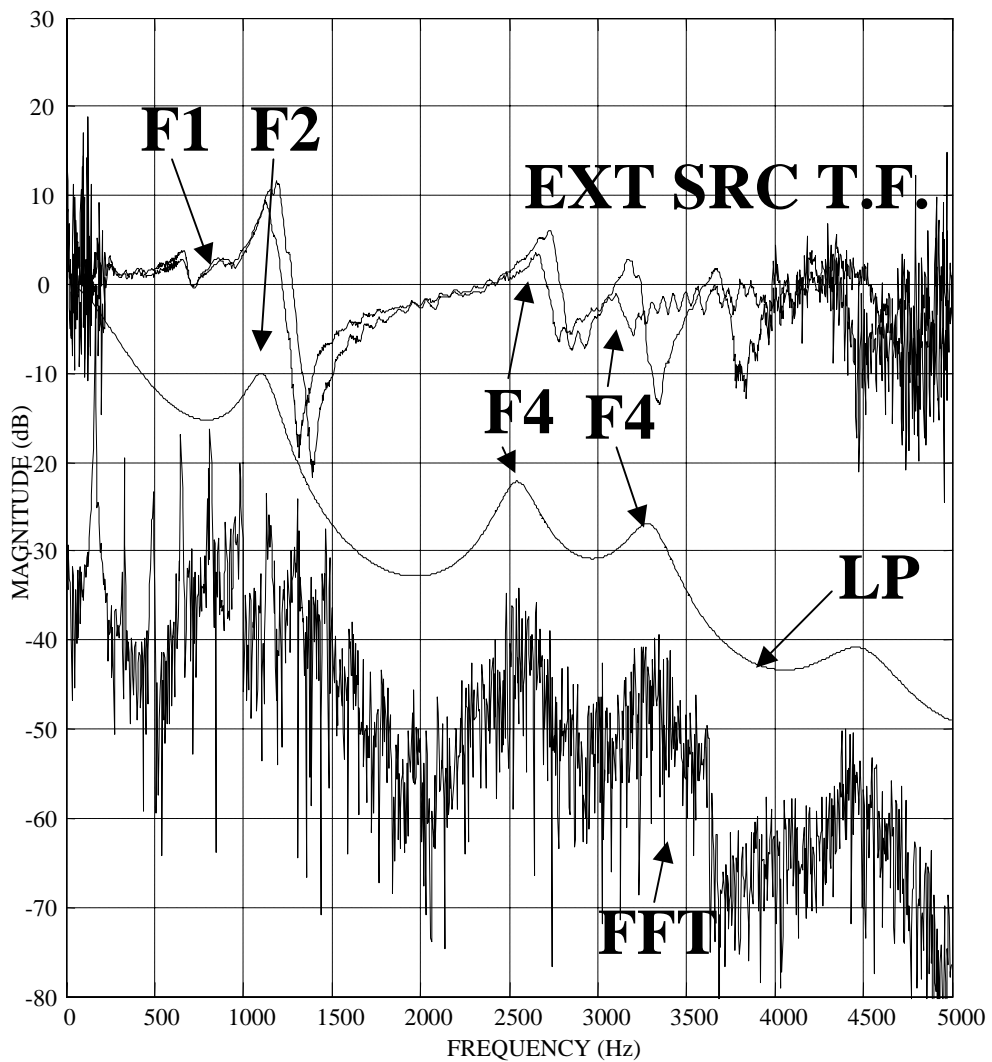


Figure 3.11b. FFT/LP/ES analysis of a female breathy /a/. Compare to Fig 3.11 which contains a normal /a/ from the same time series.