M. Cler & J. Intoy
Semester Projects
SH901 – FA 2015
Neural Control of Speech

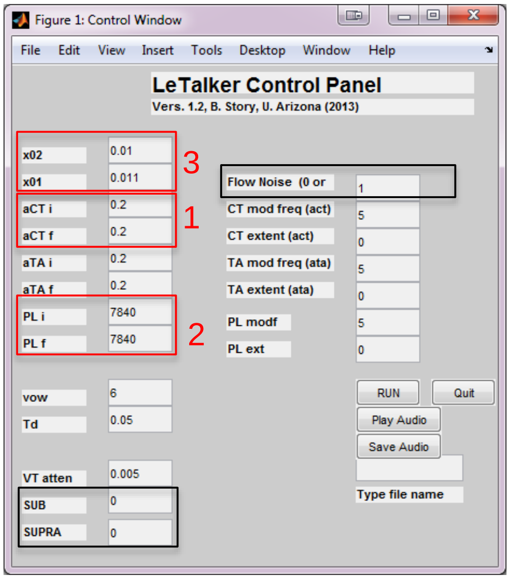
Goal:

The goal of this semester project was to improve the DIVA model's control of voice. Prior to this project, the DIVA model used three voicing parameters: (1) Fundamental frequency (F0), (2) Pressure, (3) Voicing, which was either on or off, depending on whether there should be phonation. Although these parameters were used by the model to generate sound, the model was not incorporating them into its learning, and the control of these parameters was not based on physiology.

Meredith's main contribution:

It was decided prior to the project's start that the Story-Titze model of vocal folds should be used to control DIVA's voicing. This model (henceforth "LeTalker") is a biomechanical, three-mass model of vocal fold vibration. LeTalker[1] allows users to set a wide variety of physical parameters, including cricothyroid (CT) muscle tension, thyroarytenoid (TA) muscle tension, lung pressure, prephonatory distance from the midline of the lower cover mass, and several options for vocal tract shape (see Fig 1). A main component of this project, then, was to determine first whether the two models were compatible, and if so, which parameters should be under DIVA's control.

It was determined that, in order to stay roughly compatible with DIVA's current state, the three extant voicing parameters would now represent (1) CT tension[2], (2) lung pressure, and (3) vocal folds' prephonatory distance from the midline (see Fig 1). This last is essentially a marker for how approximated the vocal folds are from -1 (quiet breathing) to 1 (tightly closed). LeTalker now takes in all three parameters from -1 to 1 and rescales them to the parameter ranges it uses (i.e., CT tension should go from 0 to 1, lung pressure from 2000 to 20000, and voicing from 0 to 0.03). A series of LeTalker simulations was run with many permutations of inputs generating three respective outputs: (A) F0, (B) Intensity (root mean square of glottal pressure), and (C) Harmonic to Noise ratio – that is, how much the vocal folds are vibrating (see Table 1). Figure 2 shows the output parameter of F0 (y-axes) plotted against each of the three input parameters; note that these parameters interact with each other nonlinearly. Figure 3 shows a more in-depth description of these interactions; here, vocal fold distance was held constant while CT tension and lung pressure were varied. The output of F0 is shown on the y-axis, with lung pressure indicated with color. Figure 4 shows the interactions between the output parameters, which are not completely dependent. The aforementioned permutations of LeTalker were used to generate DIVA's internal model (see *Janis's main contribution*).



x02 0.01
x01 0.011
aCT i 0.2
aCT f 0.2
aTA i 0.2
aTA f 0.2
PL i 7840
PL f 7840
vow 6
Td 0.05
VT atten 0.005
SUB 0
SUPRA 0
Flow Noise (0 or 1
CT mod freq (act) 5
CT extent (act) 0
TA mod freq (ata) 5
TA extent (ata) 0
PL modf 5
PL ext 0

Figure 1. LeTalker control Panel showing parameters that are (easily) modified. Parameters outlined in red are actively controlled by DIVA. Parameters outlined in black are set programmatically to values as shown (all others set to LeTalker default).

Table 1. LeTalker Inputs and Outputs

DIVA Art #	Input (to LeTalker)	Range [3]	Output (calculated from LeTalker glottal pressure output)	Range[4]
(11)	CT tension	0 - 1	F0	0; 68 - 613
(12)	Lung pressure	2000 - 20000	Intensity	1.7 - 2384
(13)	VF prephonatory distance	0 - 0.03	Harmonic-to-noise ratio	-.0087 - 52.4

[1] LeTalker 1.2 downloaded from http://sal.arizona.edu/node/26

[2] TA tension is kept constant (0.2) for now; if future simulations require both CT and TA control, an additional control parameter will need to be added to DIVA, run_all_permutations_LeTalker.m will need to be rerun to include dynamic TA activity, and the outputs of this script will need to be fed into a new forward model / RBF.

[3] These are the ranges in LeTalker; they are linearly rescaled to -1 to 1 in DIVA; VF prephonatory distance is rescaled somewhat non-intuitively to 1 to -1 (that is, VFdist 0.03 = DIVA -1 and VFdist 0 = DIVA 1)

[4] These ranges are rescaled to 0 to 1 in DIVA for now; see details in *Janis's main contribution*

Finally, both LeTalker and DIVA were modified to be compatible. The details of incorporating these models were the bulk of the work undertaken, but are not particularly interesting or germane for future reference. In brief: a variety of modifications to LeTalker were necessary in order for it to be called iteratively by DIVA. The LeTalker code was modified to make all current variables persistent, as future vocal fold dynamics depend on the current vocal fold state. A variety of internal parameters were hard-coded as well, including those that bypass LeTalker's vocal tract DIVA's synthesizing module was modified so that instead of generating glottal pulses using glotlf.m (Liljencrants–Fant model) it uses LeTalker. LeTalker parameters are set so that it does not use its own vocal tract, but instead passes DIVA a train of glottal pulses.

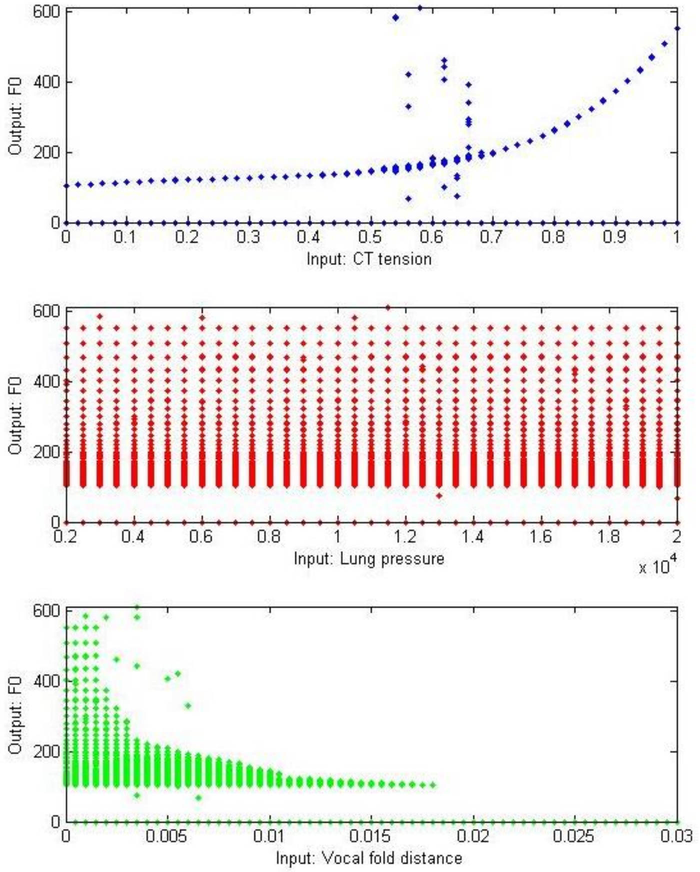


Figure 2. F0 output (y-axis) plotted against the three different inputs (x-axes; TOP: CT tension; MIDDLE: Lung pressure; BOTTOM: VF distance).

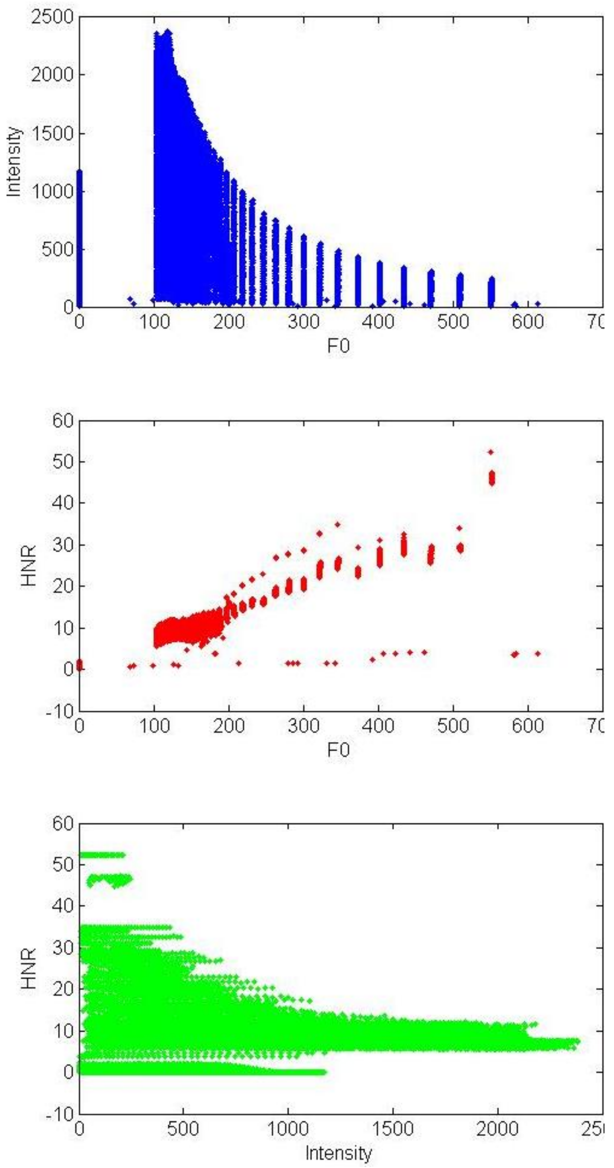


Figure 4. Relationships between all outputs. TOP: F0 vs Intensity; MIDDLE: F0 vs HNR; BOTTOM: Intensity vs HNR.

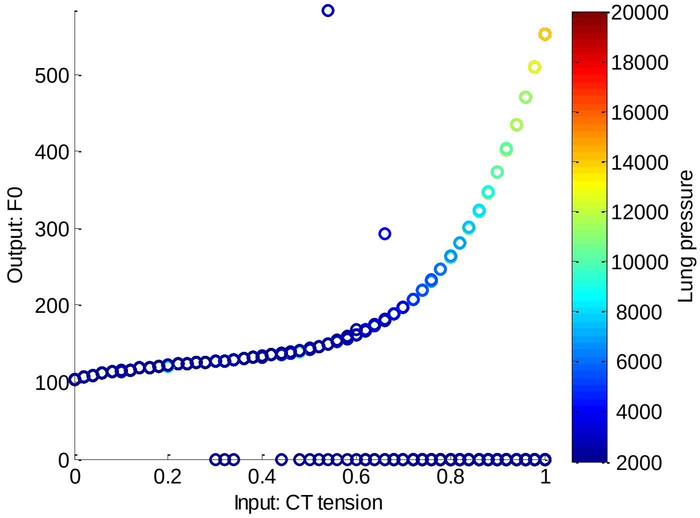


Figure 3. CT tension input plotted against F0 output with lung pressure plotted in color. For this plot, vocal fold distance was held constant at 0.001.

Janis's main contribution:

Several steps were required to give DIVA real-time control of these voicing parameters in addition to the extant articulatory parameters. First, it is not feasible for DIVA to call LeTalker every time a parameter needs adjustment (900000 calls); instead, DIVA needs to have an internal model of all combinations of input parameters and what the outputs are from those combinations. Radial basis functions were used to approximate LeTalker in a fashion similar to the forward model that transforms the articulator positions into formant frequencies and somatosensory outputs. The internal model was trained on the aforementioned LeTalker permutations. Since LeTalker outputs are monotonic in certain input regions, the forward model was trained on slightly smoothed outputs to reduce the size of the region with a nonzero gradient, particularly in the F0 output.

In order for DIVA to learn how to generate the desired voicing outputs F0, intensity, and HNR, an inverse model was added to DIVA specifically for the voicing parameters. The voice inverse map acts exclusively on the voicing parameters and outputs to determine a corrective feedback motor command (see Fig 5). This motor command is also combined with cerebellar and speech sound map inputs to train the feed forward weights for learned commands ("Learned Motor Command"). The learning delay caused by cerebellum is set to match that of the auditory learning.

Currently, the outputs of the internal model of LeTalker are scaled to be from 0-1. Since low values of CT tension, lung pressure, and distance between vocal folds are relaxed at low values, the null space projection was not updated for these values.

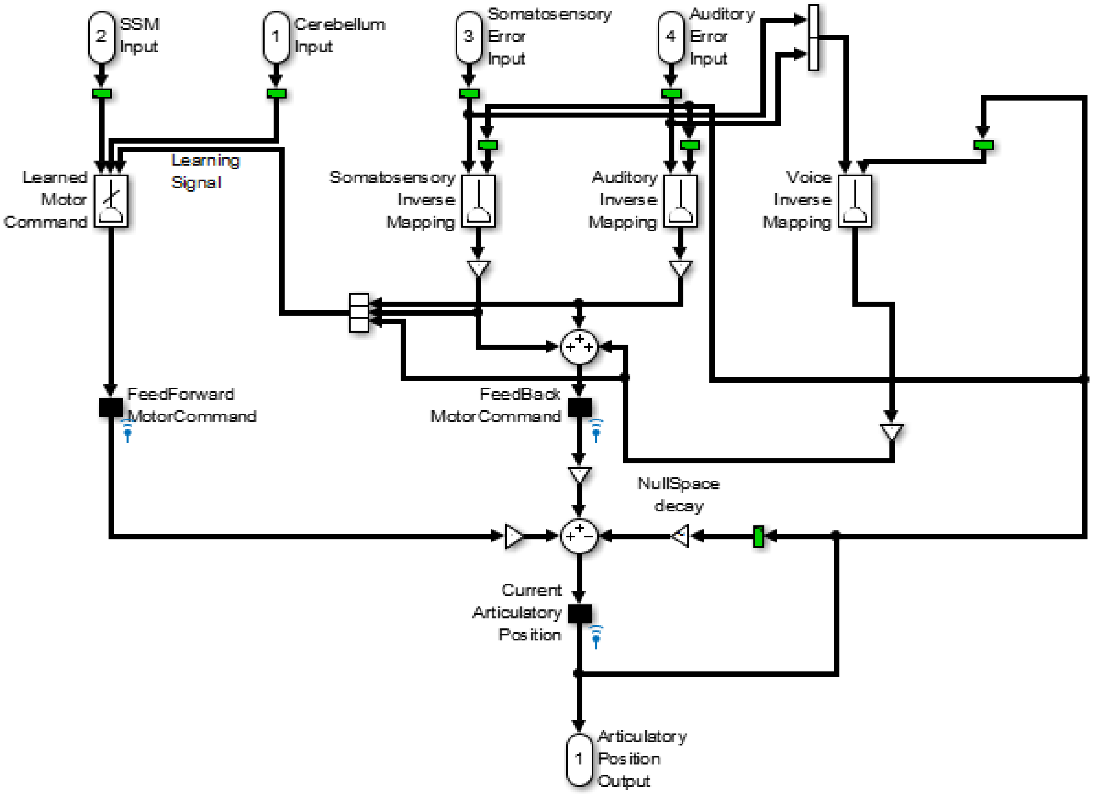


Figure 5. Simulink schematic of Articulatory Velocity and Position Maps with added Voice Inverse Mapping.

Outcome:

The DIVA model is successfully using the RBF-estimation of LeTalker and LeTalker to learn and synthesize the voicing aspects of speech for simple targets. Targets were determined based on values recommended for LeTalker phonations. There are some problems with the learning and synthesis that are discussed in the next section. See Fig 6 for example of an /a/ with increasing F0 produced by: (1) the original DIVA, (2) DIVA with static inputs to LeTalker (i.e., glottal source is generated by LeTalker and filtered through DIVA, but inputs are generated manually and not under DIVA's control), and (3) new DIVA/LeTalker controlling all voicing parameters.

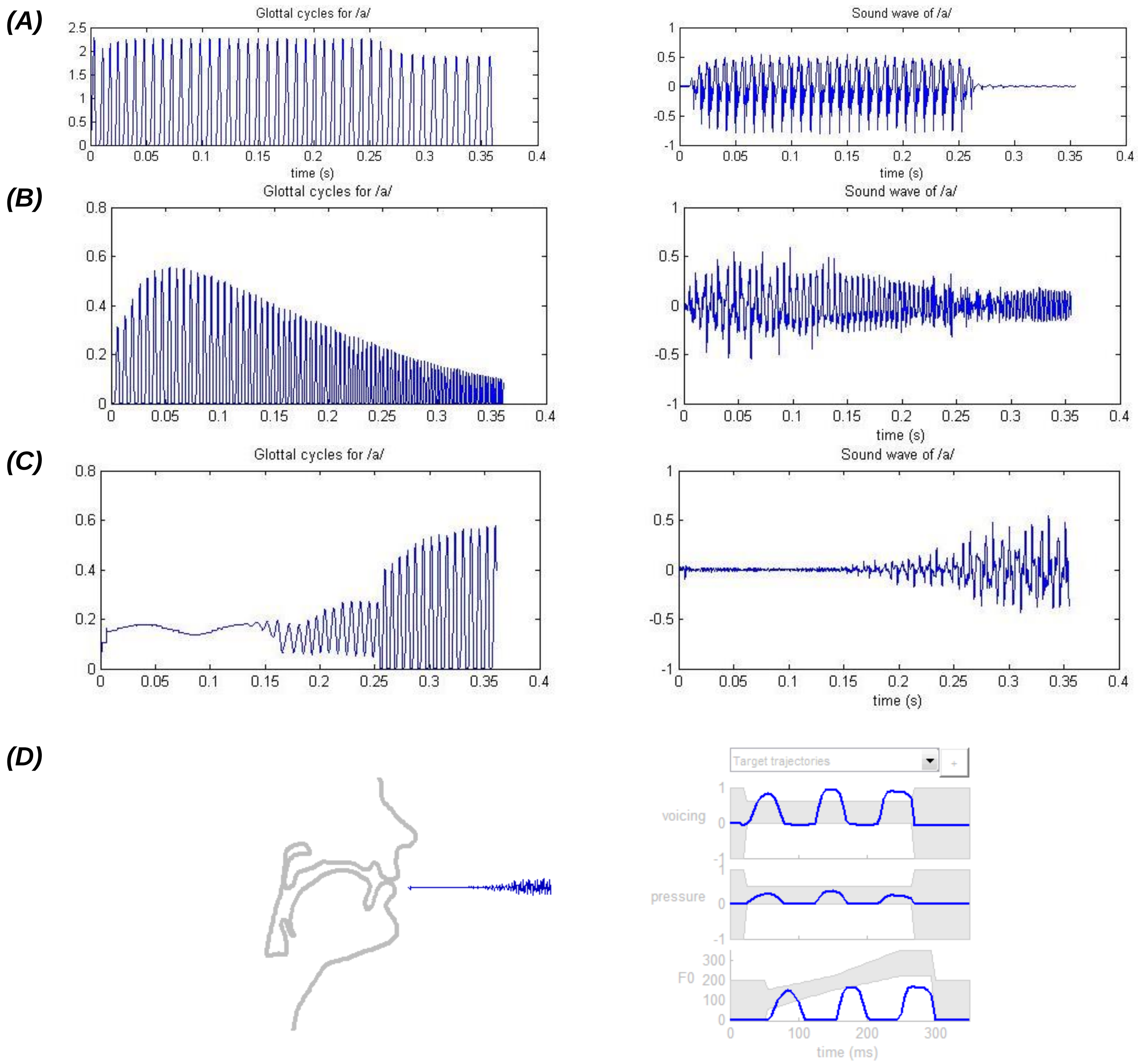


Figure 6. Three /a/ signals with increasing F0 targets generated by: ***(A)****: the original DIVA,* ***(B)****: Static inputs to LeTalker (i.e., glottal source is generated by LeTalker and filtered through DIVA, but inputs are generated manually and not under DIVA's control),* ***(C)****: DIVA controlling all voicing parameters. The* ***left*** *column shows glottal source; note the obvious increasing frequency in B and C.* ***Right*** *column shows filtered sound signal. This is not yet perfected-- note that the filter should have dampened all signal after 0.25s as in A. This is likely due to the different expected ranges in the output variables of pressure and voicing (see Future directions/recommendations → 5b).* ***(D)*** *shows the DIVA targets used to generate (C); note the oscillations, suggesting that some adjustments are needed (see Future directions/recommendations → 6).*

Future directions/recommendations:

Future directions for this project may wish to include the following objectives:

1. Update “stock” utterances in diva_gui with more appropriate targets
2. Update beginning/ending values for simulations so they don’t start at 0; update Null Space Projection to accommodate nonzero biases (See Fig 7);

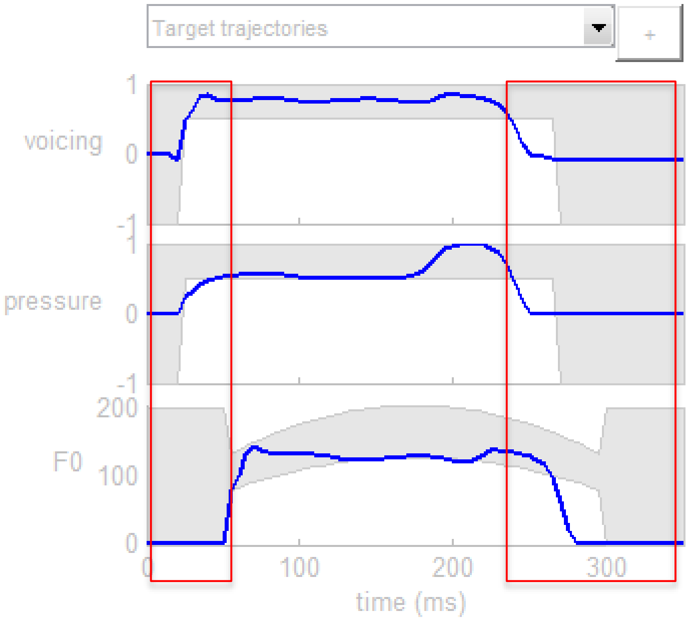


Figure 7. Example showing where different “base” values of voicing/pressure would be preferable.

3. Consider updating the words “voicing” and “pressure” to “HNR” and “intensity”? (See Fig 7).
4. Consider attempting to detangle F0 and HNR outputs (see Fig 4)
5. A variety of tweaks to the diva_synth_sound() function:
    a. The output of LeTalker is at quite a different scale than the Liljencrants–Fant model. At the moment, this is being rescaled at the end of the LeTalker function to more closely match what the original DIVA code was expecting. A better choice would be to rewrite the diva_synth_sound() function to expect this new range of glottal source.
    b. The code in this function is fairly compact and somewhat unclear and may need to be re-evaluated based on an important difference between the old DIVA and this new version in the expected values for Art(12)[pressure] and Art(13)[voicing]: these used to each only be very near 1 or very near 0 (see Fig 8, red and green). Now they can vary anywhere between 0 and 1, more like tension always was (Fig 8, blue); this may require an in-depth rewrite of the synthesizing code.

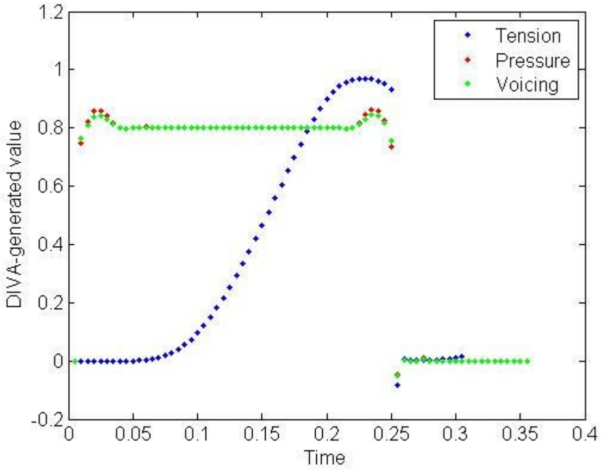


Figure 8. Original inputs from DIVA. Note that while tension varies from 0 to 1, pressure and voicing are either very near 1 or very near 0. The synthesizer code relied on this; because DIVA/LeTalker has new outputs that can vary evenly between 0 and 1, this code may need to be reevaluated

6. DIVA is not quite behaving as expected with the targets (see Fig 6D); adjustments to Simulink, particularly cerebellum, may help.
7. This version of DIVA may need a name? Suggestions include:
    - leDIVA (lumped element + DIVA)
    - mcDIVA (modulated cricothyroid DIVA)
    - DIVA/MVP (DIVA + modulated voicing parameters)
    - St. DIVA or DIVA ST or DIVAst? (Story-Titze + DIVA)