

# Short manual for the software TractSyn

Peter Birkholz

September 28, 2004

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Oscillogram view</b>	<b>2</b>
<b>3</b>	<b>Vocal tract view</b>	<b>4</b>
<b>4</b>	<b>Time domain simulation (TDS) view</b>	<b>5</b>
<b>5</b>	<b>Segmental control model view</b>	<b>6</b>
<b>6</b>	<b>Fricative view</b>	<b>7</b>

## 1 Introduction

I developed the program TractSyn for my dissertation about articulatory speech synthesis at the institute for computer science at the university of Rostock (Germany). The parts of the software that I have made available on the project web site are intended to be used for educational and scientific purposes. Originally, the software was written for our internal use only and may thus not fulfill all the requirements that you would expect from "professional" software.

This manual will give you a quick overview of the possibilities of the program. The installation just requires to unpack the zip-file that you can download from the website. The software is written for the Microsoft Windows operating systems and should run from Windows Me upwards (2000, XP). It requires OpenGL and DirectX 5.0 or higher. You start the program by calling "speak.exe" (The file "speak.res" must be in the same folder).

The software is arranged in several "views", of which each has a focus on a particular topic. The individual views can be accessed by the buttons in the left-bottom corner of the program window or by the menu. You find

- an oscillogram/spectrogram view,
- a vocal tract view,
- a view for the visualization of the acoustic simulation in the time domain (TDS),
- a view for a segmental-kinematic control model,
- and a view for the generation of fricatives.

Each individual view is organized in frames: two narrow frames on the left side and one or two big frames on the right side. The bottom-left frame is the audio control frame and looks the same for all views. It contains the buttons to play, record or delete sounds. The upper left and right frames are specific for the current view and contain diverse buttons, controls and pictures. From now on, the upper-left and the upper-right frames will be called "side frame" and "main frame", respectively.

The following sections of this manual will explain the most important functions of the different views.

## 2 Oscillogram view

The main frame of the oscillogram view shows two oscillograms and one sonagram picture. The lower oscillogram shows the same signal(s) as the upper one but with a larger range (it is "zoomed out"). You can display at most three different signals in the oscillograms that are called "main track", "EGG-track" and "extra track". Each track has a length of 30 s and represents an audio- or EGG-signal with a sampling rate of 22050 Hz and 16 bit resolution. Whether a signal is displayed in the oscillogram or not is determined by the state of the corresponding check-box. Within the oscillogram pictures, you can set two marks: a start mark (green vertical line) and an end mark (red vertical line). Click the left mouse button (LMB) to set the position of the start mark and the right mouse button (RMB) to set the position of the end mark.

When you load a wav-file into one of the tracks (via the menu), it will be loaded to the position of the start mark. With the menu item "File->Load WAV+EGG" you can load a stereo wav-file, where one track is interpreted as audio (loaded to the main track) and the other one as EGG (loaded to the EGG track). You can exchange both signals with the button "Track 1 <-> 2" in the main frame.

When you select a region of the signal(s) by setting the start mark and the end mark in the oscillogram, you can perform different operations on this part of the signal with the buttons in the side frame or the menu.

**Save a wav-file.** Select "File->Save WAV-file" from the menu and the signal in the selected region will be saved.

**Calculate a sonagram.** Click the button "Sonagr.", and the sonagram of the selected signal region will appear in the sonagram picture. You can change the brightness of the sonagram with the corresponding "+" and "-" buttons.

**Determine the  $F_0$ .** Click the button "Determine F0" and select "Calculate F0" in the appearing dialog. The  $F_0$  is calculated for the selected region. The results can be saved to a txt-file. The txt-file contains one  $F_0$ -estimate for approximately every 8 ms. The first, second and third column in each row of the file are the time,  $F_0$ -estimate and voicing decision of each point in time. The course of the calculated  $F_0$  is displayed as a yellow curve in the sonagram picture on a log. frequency axis. The algorithm for the pitch detector is shortly described in [2].

If you want to determine the  $F_0$ -curve of more than one wav-file, you can press the button "F0 from file list". In the appearing file dialog you can select multiple wav-files. The results of the  $F_0$  determination are saved in txt-files again, that have the same name as the corresponding wav-files, but the extension ".txt" instead of ".wav".

**Find the glottal closure instants in a laryngogram.** When you have loaded a signal to the EGG-track, press the button "Edit EGG-signal" in the side frame. In the appearing dialog, press "Filter laryngogram". After that, call the dialog "Edit EGG-signal" again and press "Determine glottal closure instants". The  $F_0$ -curve corresponding to the GCI's of the laryngogram is displayed as big yellow points in the sonagram picture.

**Calculate a long-time average spectrum.** Press the button "RMS average spectrum" and adjust the parameters in the appearing dialog. The starting point for the calculations is the start mark in the signal. The resulting spectrum is shown as red curve in the spectrogram in the vocal tract view. The red curve can be moved up and down with the slider right of the spectrum picture.

**Add noise to the audio signal.** Press the button "Add noise" in the side frame and select a SNR. You can add either white noise or noise from a wav-file.

**Create a test signal.** Press the button "Create test signal" in the side frame and select the signal you want.

**Play the audio track.** Press "Play all" in the audio control frame in order to play the whole 30 s signal in the main track (in a loop). A shortcut for this button is STRG+P on the keyboard. Press "Play part" (or STRG+SPACE) when you want to play the selected signal part (in a loop).

**Clear all tracks.** In order to clear the signals of all three tracks, press the button "Clear all tracks" in the audio control frame.

### 3 Vocal tract view

The main frame of the vocal tract view contains an image of the three dimensional vocal tract model [3, 1], an image with one particular cross-sectional area in the vocal tract, an image with a glottal flow pulse and an image of the vocal tract area function. The shape of the vocal tract can be changed by dragging the yellow control points in the vocal tract picture with the LMB. Please read [3] for a detailed description of the vocal tract geometry and its parameters. The perspective and the distance of the model can be changed, when you press the left or right mouse button in the picture and move the mouse around. Display options for the vocal tract can be found in the side frame. One of the control points indicates the position of the current cut-area along the vocal tract center line (check "Final center line" and "Cut area" in the side frame). Dragging this control point will change the displayed cross-sectional area.

The underlying model for the glottal flow pulse is the LF-model [5]. This flow pulse model is used for the acoustic synthesis of voiced static sounds. You can change the parameters of the glottal flow pulse when you click on the corresponding picture with the LMB or display its derivative with the RMB.

The vocal tract area function displayed in the picture below is derived from the three dimensional vocal tract model and changes immediately when you change the vocal tract parameters by means of the control points. The gray area functions are side branches, namely the nose cavity and the piriform fossa. You can change the display options for the area function when you click on the picture with the LMB.

The picture on the bottom of the application window shows the vocal tract transfer function from the glottis to the lips that is calculated by means of the area function (The check box "Primary" must be checked). You can change the type of the displayed spectrum when you click on the picture with the LMB. In the dialog that appears you can for example choose to display the vocal tract input impedance (seen from the glottis) or the transfer function from volume velocity at the glottis to radiated sound pressure ("Transfer function P/U"). The options that influence the calculation of the transfer function can be changed in the dialog that appears, when you click on "Acoustics" in the side frame. You can choose one type of radiation impedance, change the losses included in the vocal tract simulation, and so on.

The button "Pole-zero plan" in the side frame opens an editor, where you can edit a volume velocity transfer function by placing poles (resonances) and zeros (anti-resonances) in the top-left quadrant of the Laplace-plane. Both poles and zeros can be created or deleted with a context menu within the picture (right mouse button). In order to change the position of the poles and zeros, you can drag them with the LMB. The resulting transfer function is displayed as a red curve in the spectrum picture, when the check box "P-Z-Spec." is checked (right of the spectrum picture).

In reality, a transfer function has got an infinite number of poles. When you want to approximate the effect of the higher poles (the ones above the highest pole that you have placed in the plane) then check the box "Higher pole correction" in the Pole-zero dialog.

Do not use the buttons "Test" or "Spectrum" in the side frame.

When you want to play the vowel corresponding to the current vocal tract geometry, press the button "Play part" in the audio control frame (Clear the tracks before this, if necessary). The way the audio signal is generated is described in [4] and [1]. The button "Play all" simply plays the audio signal of the main track in the oscillogram view.

When you want to load or save a vocal tract configuration, select "File->Load vocal tract parameters" or "File->Save vocal tract parameters" from the menu. The vocal tract parameters are saved in ASCII-files with the extension ".vtp". Three of these files for the vowels /a/, /e/ and /u/ are included in the download as examples.

## 4 Time domain simulation (TDS) view

This view is meant to visualize the distribution of sound pressure and volume velocity in the vocal tract during phonation. The picture at the bottom in the main frame shows the vocal tract area function of the entire speech apparatus, including the subglottal system, the glottis the vocal tract and the side branches piriform fossa and nose cavity. The four circles connected to the nose cavity area function represent the sinus cavities of the nose. The picture right above the area function shows the distribution of the pressure or volume velocity along the vocal tract. In order to start the simulation, select the radio button "Vowel (Entire system)" in the side frame and then press "Start synthesis" (You should clear the sound tracks before this). The speed of the animation can be changed with the corresponding sliders in the side frame. The controls in the top-right corner of the application window allow you to select the physical quantity (pressure or volume velocity) and the sound paths you want to display during the simulation.

In the picture with the area function you see a vertical dashed line. This line marks a particular position in the vocal tract and can be moved with the LMB. During the simulation, the selected physical quantity at this particular place is displayed as a time signal in the top picture of the main frame.

The button "Glottis dialog" in the side frame opens a dialog, where you can adjust the glottal settings for the synthesis. The glottis is based on a model by Titze [7] and has been extended for the simulation. You can for example change the abduction at the lower and upper edge of the arytenoids with the sliders "Top displacement" and "Bottom displacement". The glottis dialog can remain open when you start the simulation with "Start synthesis". To set the area function that you modeled in the vocal tract view, press the button "Set area function" in the side frame. For the "Kind of synthesis", you should only select either "Vowel (LF-excitation)" or "Vowel (Entire system)" in the current version of the program. Do not press the button "Test" in the side frame.

## 5 Segmental control model view

In this view, you find an editor for the temporal control of the vocal tract parameters. This allows you to produce short utterances of speech. The utterances may contain vowels, nasals and voiced plosives. Fricatives and unvoiced plosives can not be generated by the vocal tract model at the moment.

The editor is based on the segmental-kinematic control model proposed by Kroeger [6]. Here, each phone of the utterance is represented by one segment or production interval. Within each production interval exist three time marks. At each of the time marks you can either fix the value of a particular parameter or not. The temporal change of each parameter is determined by means of interpolation between the points in time, were it has been fixed.

In the main frame, the horizontal axis represents time and the vocal tract parameters are arranged along the vertical axis. Altogether, there are 18 parameters for the vocal tract shape and six parameters for the excitation at the glottis. With the LMB, you can select a phone and a time mark (vertical dashed line) within the phone. With the sliders in the side frame on the left side you can change the duration of the selected phone and the relative position of the time mark. A click with the RMB within the editor frame calls a context menu. The items in this menu allow you to make a connection (fix a parameter value at one of the time marks), delete a connection, insert a new phone (production interval), delete a phone or load the vocal tract parameters from a ".vtp"-file at one of the time marks. Furthermore, you can transfer the parameter values of the selected time mark to the vocal tract model in the vocal tract view ("Push parameters") and vice versa ("Pull parameters"). Fixed parameter values are indicated by small yellow rectangles with a black border. You can change these parameter values by dragging the yellow points with the LMB in the editor frame or with the corresponding slider in the side frame.

To see how a complete word looks like in the editor you should load the phone chain file "mandy.phc" (which is part of the download) into the editor via the menu "File->Load phone chain".

When the check box "Show vocal tract" at the bottom of the side frame is checked you will see a vocal tract model in a separate non-modal dialog box. The shape of this vocal tract always corresponds to the values of the parameters at the selected time mark. When you want to see the vocal tract shape at an arbitrary point in time, you can hold down STRG on the keyboard as well as the RMB and move the mouse through the editor frame.

Another way to change the selected phone or parameter, you can hold down the STRG-key and press one of the arrow-keys.

Press the button "Play all" in order to calculate and play the the output signal of the entire phone sequence (You should clear all tracks before you start the synthesis). The speed of the calculation can be changed with the corresponding slider in the side frame but is limited by the power of your computer. Press the button "Play part" in order to play the sound corresponding to the vocal tract configuration at the selected time mark.

## 6 Fricative view

In this view, you can examine the production of fricative consonants. Therefore, you can create an arbitrary supraglottal area function consisting of 32 tube sections and place monopole and dipole noise sources at different positions along the tract. You can select a tube section and change its cross-sectional area when you click at it with the LMB in the corresponding picture. The length of the selected tube section (painted in yellow) and its axis ratio (all tube sections are actually assumed to be elliptic in shape) can be changed with the sliders in the top-left corner of the main frame. Here, you can also change the overall length of the vocal tract.

The small icons on top of the area function represent the potential positions of the noise sources. Dipole pressure sources are symbolized by a voltage source (one circle) and monopole flow sources are symbolized by a current source (two intersecting circles). A click on one of the sources opens a dialog where you can adjust the source parameters like spectral shape and amplitude. In order to shape the source spectrum you can choose a low-pass filter of either first or second order. The cutoff frequency and the damping factor (for second order filters) of the filters can be changed with the sliders. When the check box "Is active" is checked for a particular source then it contributes to the output of the vocal tract according to its amplitude. The total output sound pressure spectrum is shown in the spectrum picture below the main frame. In this spectrum you can immediately observe the acoustic consequences of changing area values or source parameters. As examples, I have provided the area functions and sources for the fricatives /s/ (alveolar) and /sh/ (postalveolar) as the files files "s.af" and "sh.af". They can be loaded into the program when you press the button "Load areas+sources" in the main frame. With the button above, you can save your own fricatives. When you want to listen to the fricative corresponding to the calculated spectrum, select "FDS with sources" as the kind of synthesis in the side frame and press "Start synthesis" (You should clear the tracks before starting the synthesis!).

At the bottom of the side frame you find two additional sliders to change the pulmonary pressure and the glottal area. The pulmonary pressure is only used to determine the mean flow through the vocal tract. Its value of 800 Pa should not be changed a lot. The glottal area can be changed in order to observe the damping of the formants and antiformants in the fricative spectrum due to subglottal coupling. A greater glottal area causes a greater damping. The button "FDS acoustics" opens a dialog where you can adjust the parameters for the acoustic synthesis of the fricative. It is the same as the corresponding dialog in the vocal tract view.

I have used this program module to estimate the area functions and noise source characteristics of different fricative consonants by matching the synthetic fricative spectra with those of natural fricatives. The results of this study were presented at the German-French Summerschool on "Cognitive and physical models of speech production, perception and perception-production interaction" in Lubmin, Germany, 2004. A paper was also submitted to the ICASSP-2004 in Philadelphia.

## References

- [1] P. Birkholz. Entwicklung eines dreidimensionalen Artikulatormodells für die Sprachsynthese. Master's thesis, Institute for Computer Science, University of Rostock, 2002.
- [2] P. Birkholz. Grundfrequenzbestimmung unter Berücksichtigung linearer Frequenzänderungen. In *Fortschritte der Akustik, DAGA '03*, pages 768–769, 2003.
- [3] P. Birkholz and D. Jackèl. A three-dimensional model of the vocal tract for speech synthesis. In *Proceedings of the 15th International Congress of Phonetic Sciences*, pages 2597–2600, Barcelona, Spain, 2003.
- [4] P. Birkholz and D. Jackèl. Influence of temporal discretization schemes on formant frequencies and bandwidths in time domain simulations of the vocal tract system. In *Interspeech 2004-ICSLP*, Jeju, Korea, 2004.
- [5] G. Fant, J. Liljencrants, and Q. guang Lin. A four-parameter model of glottal flow. *STL-QPSR*, 4:1–13, 1985.
- [6] B. J. Kröger. *Ein phonetisches Modell der Sprachproduktion*. Niemeyer, Tübingen, 1998.
- [7] I. R. Titze. Parameterization of the glottal area, glottal flow, and vocal fold contact area. *Journal of the Acoustical Society of America*, 75(2):570–580, 1984.