

# Guidelines for using Mavis to analyze Hungarian data

## Opening the program

- 1) Open Matlab – there should be an icon for this on the desktop; if not, then find Matlab under All Programs.
- 2) Download MAVIS\_PC.zip to the desktop from the course Blackboard site and unzip to the desktop.
- 3) From within Matlab, go to the newly unzipped MAVIS\_PC folder – this can be done by clicking on the "Browse for folder" button in the current directory box in the upper right of the Matlab window or in the "Current Directory" pane to the left of the Command Window. You can also paste the newly unzipped folder name, including its Windows path which should look like ... C:\Documents and Settings\User\Desktop\MAVIS\_PC, into the "Current Directory" pane.
- 4) In the Matlab command window, type *mavis* and then the return key – this will lead to windows opening on your screen (called TRAJECTORY at the left of your screen, SPATIAL at the upper right corner, and MAVIS at the lower right corner). You can access a technically detailed “help” facility of Mavis by TRAJECTORY → Mavis → About Mavis.

## Selecting a data file to view in Mavis

- 1) Once in Mavis, from the Trajectory window menu, select Mavis, and then Open Dataset.
- 2) Data files can be found in the downloaded folder. A sample data file is *bib\_FL\_B1\_R2.mat*. Click on one of these files to open. You can listen to the entire data file by choosing Trajectory, Yack Selection from the drop-down menu. About the file names: In the file name, the B\_R mnemonics stand for Block and Repetition. Thus *bib\_FL\_B1\_R2.mat* is the second repetition (R2) in the first block (B1) of data acquisition of the word *bib*. The Hungarian data were acquired in two blocks. Within each block there are two, and for some subjects four, repetitions of an utterance containing the target word. So the other “bib” data files are named *bib\_B1\_R1*, *bib\_B2\_R1*, *bib\_B2\_R2*. Note that the word in question is uttered two times in the utterance recorded in each data file. So in total there are 8 (= 2 blocks x 2 repetitions of the utterance containing the word within each block x 2 instances of the word in each data file) renditions of each target word (*bib*, *szeditett*, *zefirben* and so on). For the subjects that did four instead of two repetitions of the word per block the total renditions would be 16. The data files for three speakers of Hungarian can be found under the Data Sets tab in the course Blackboard site. Click on Data Sets and then click on Hungarian and download and unzip the data for any given subject.

## Labeling files

Now that you can view data files, the next step of the analysis is to label them, i.e. place labels in the data file that identify the things you want to measure. There are two ways to label data files – manual labeling and automatic labeling. The majority of the steps are

## Guidelines for using Mavis to analyze Hungarian data

the same for each of these methods. Let us assume that our objective is to take measurements at the vowel contained within the target word “bib”. We are thus looking for the positional extrema (whether you want maxima or minima will depend on the vowel and the context) – how front the vowel is, how back the vowel is, how low the vowel is, etc. Such extrema can be used to punctuate the positional targets for vowels.

Select Labels, Configure labeling, Set User Proc (for manual labeling, leave this blank, click OK, and skip ahead to the section on manual labeling), select ml\_extrema (or feel free to experiment with other procedures), select the trajectory to label (x for horizontal, y for vertical), minimize this window and click OK in the Configure Labeling window.

### *Automatic labeling*

- a) You can listen to the entire data file by choosing Trajectory, Yack Selection from the drop-down menu. Now we will zoom in the region of the data file containing where the target word with the vowel of interest. You can zoom in a selection by clicking on the left edge of the top audio channel and (when a vertical line appears) moving it with the mouse to the left edge of the area you wish to enclose. Again using the mouse, click on the right edge of the audio channel and move the vertical line to the right edge of the area you wish to enclose. When a section of the audio trajectory is framed in this way, Yack Selection will play just that portion. You can use this to make sure that the portion of the signal you have zoomed in corresponds auditorily to what you intended to zoom in. It is not at all unusual to go through several repetitions of doing some zooming, then Yack Selection, then zooming (in or out), and so on until you home in on the portion of the signal you are interested in.
- b) For each trajectory/receiver that you wish to measure, right-click in the respective channel to find the positional ‘max’ and ‘min’ in that frame.
- c) To see the values for x and y in at the specific point in time, left click on the channel trajectory and the measurements appear in the separate small window called MAVIS. Mouse dragging while left clicking updates the values continuously.
- d) *Note that if you are labeling more than one trajectory (TB1, TB2, TD, LL ...), the location of extrema will not necessarily line up on the screen.*
- e) To check the labels, go to Labels, Edit labels. You can delete a label by selecting the label and clicking ‘clear’. The numbers which appear in the label box are the time points (in milliseconds) where the labels appear within the utterance.

### *Manual labeling*

- a) Zoom in – enclose just the vowel/segment being looked at (in actuality, you may want to start with a larger range and continue to minimize the window until the vowel of interest is located). When a section of the audio trajectory is framed, Yack Selection will play just that portion.
- b) For each trajectory that you wish to measure, you need to find the extrema – the ‘max’ or ‘min’ values during a vowel’s production on either the x or the y dimension. To see the positional x and y values at any point in time, left click

## Guidelines for using Mavis to analyze Hungarian data

on the channel trajectory and the positional values appear in the 'X val' and 'Y val' fields appear in the MAVIS window at the lower right corner.

Assuming you are looking to label the occurrence of a maximum in the y dimension of the TB1 receiver, scroll along some area where height appears to be greatest holding down the left mouse button until you find this point, and place a label at that time point with a right-click. When you click at the desired point, a *Create Label* box will appear on the screen. In the *Name* box, you can edit the name of the label.

- c) *Note that if you are labeling more than one channel trajectory, the location of labels will not line up perfectly on the screen.*
- d) To check, edit or delete labels, go to Labels, Edit labels (you can delete one or more labels by selecting the label and clicking 'clear').

Once you have gotten used to labeling, you can make your job easier by changing the **display options**, thereby removing items from the screen that are not necessary for your analysis. Using the Trajectory window, select Trajectory, Configure Trajectory Display to remove items not to display (for example, jaw movement may not be important for your purposes; but remember that it might be useful to include the information on lips when there is a [b] in the word). From here you can also select to only look at a particular axis – if working on the x-axis, you can remove the check from the y-axis box (and vice versa). This step is recommended to ensure labeling the axis you want. This setting sticks throughout your labeling, but the other items (jaw, lips, tongue body, etc.) have to be re-selected for each data file opened.

### *Saving files*

After labeling both instances of the word in each sentence, you must save the labeled file. Otherwise you will lose the label information. On the menu bar, select Mavis, Save dataset. Save the labeled files in your desktop Mavis directory. Save each labeled file as its exact name in the source folder (this is the default option). You will later set up a list file (explained below) based on the names of labeled files you have saved.

### *Data Extraction*

In order to do a quantitative analysis on the files that have been labeled, it is necessary to extract your labels and their information into a file. Therefore, once labeling is complete, the next step is to make a file containing a list of the data files from which labels are to be extracted and then to use a procedure called `tabulate2.m` to do the extracting of label information. [Note: you may want to test extracting labels after completing a small batch of labeling to ensure accuracy of methods.] You will need Matlab open for this but you can close any Mavis windows.

- 1) Open a window where you can see all of the names of the labeled data files you have saved to your project folder. You will need their names in the rest of this step. From the Start button, select Accessories, Notepad (do not use WordPad).

## Guidelines for using Mavis to analyze Hungarian data

Make a list inside the Notepad file with the names of all the data files for which you will label extract (but do not include the extension .mat). Specifically, each line in your Notepad window should start with the name of the data file, followed by a space (or a tab), followed by the name of the trajectory where your labels are in the data file, followed by a space (or a tab), followed by the axis (x or y) which is labeled. For example, if you have a file `zefirben_B1_R3` and you want to look at the TD trajectory on the y-axis, write `zefirben_B1_R3 TD y`. Make sure each such line has the file name, the trajectory, and the axis to extract. Once your list is complete, save the file (into your project directory) and give it a name. A sample file name for your list file might be `test_TB1_list` or `list_TB1Y`.

Example of list file:

```
zefirben_B1_R2 TB1 y
zefirben_B1_R3 TB1 y
zefirben_B1_R4 TB1 y
zefirben_B2_R1 TB1 y
zefirben_B2_R2 TB1 y
zefirben_B2_R3 TB1 y
zefirben_B2_R4 TB1 y
kretenben_B1_R2 TB1 y
kretenben_B1_R3 TB1 y
kretenben_B1_R4 TB1 y
kretenben_B2_R1 TB1 y
kretenben_B2_R2 TB1 y
kretenben_B2_R3 TB1 y
kretenben_B2_R4 TB1 y
...
```

The example file above will only look for labels in the y-axis. For the other dimension, substitute in the dimension name (e.g. above, substitute y for x). You can have one such Notepad file for each trajectory/receiver you want to extract from: one for TB1, another for TB2 and so on. This is one way of organizing your measurements. However, you can also construct your list file above so that it contains, say, `bacsiban_B1_R1 TB1 Y` in one line and `bacsiban_B1_R2 TD Y` in the other line. Such a list file will allow you to extract labels from multiple trajectories/receivers.

- 2) Return to Matlab (and the Mavis directory, as described in step 3 of ‘Opening the Program’). From the Command Window, execute `tabulate2.m` by typing `tabulate2('infile', 'outfile')`, where `infile` is the list file you made in step 1 above, and `outfile` is the file you want to save your output to. Make sure that `infile` matches the name of the file exactly – so if you created a Notepad file called `infile.txt` then write `tabulate2('infile.txt', 'outfile')`. In some operating systems, if you are just looking at the icon of the `infile` you may not be able to see whether it has an extension (in Windows, for example, you need to right-click at the icon and then choose Properties to see the whole file name and its possible extension, e.g. `infile.txt`). The output file name can be any name you wish, but it is best if it is similar to the input filename. For example, assuming you want `infile` and `outfile`

## Guidelines for using Mavis to analyze Hungarian data

to be called `infile_TB1Y` and `outfile_data_TB1Y`, and you want them to be in your Mavis directory, type `tabulate2('infile_TB1Y', 'outfile_data_TB1Y')`. Your infile and your labeled datasets should all be in the same directory or folder as Mavis. Your outfile will be created in that same directory.

- 3) Once `tabulate2.m` is executed, the outfile created by the function is saved to your designated folder and can be opened from within Excel. The output file specifies the times and values of each label in the desired trajectory, along with the filename, the trajectory name itself, and the axis the label represents. For example:

```
zefirben_B1_R1 TB1 y 1230 5.58 1360 9.82
```

tells you that in the y dimension of the TB1 trajectory of the data file `zefirben_B1_R1` there are two labels. The first appears at 1230 ms and has the y value 5.58 mm and the other appears at 1360 ms and has the y value 9.82 mm.

**To sum up**, extract your labels by typing `tabulate2('infile', 'outfile')`, where `infile` is a file with a list of the information you want to extract and `outfile` is the file you want to save your output to. You type `tabulate2('infile', 'outfile')` from within the Matlab Command Window after you have changed the directory to the MAVIS\_PC directory on the Desktop.

### Viewing your measurements in Excel

At this stage, the data has been extracted to outfile. Now the data must be converted to an Excel format so as to facilitate statistical analysis.

- 1) Open Excel. From Excel, go to the drop-down menu for File → File Open and select the output file you designated. Be sure to set the *Files of type* option to *All Files* so that your file appears in the list of options.
- 2) The data must be converted to a format that is Excel-like. Inside the *Original data type* box, select the delimited option. Select *Next*. Inside the *Delimiters* box, select *Tab* and *Space*. Select *Finish*.
- 3) Line up the words with their values by inserting cells inside Excel accordingly. Add a row with the names of the labeled values. Set-up the spreadsheet in a systematic way to analyze the data (use the example below – which combines three trajectories – for reference purposes). `Tabulate2.m` outputs the time stamps in the data file where each labeled value is extracted from, but these numbers can be removed (as we have done in the example shown below).
- 4) Save the Excel file (using the Microsoft Excel Workbook file type).

**zefirben, with**

<b>long i</b>	<b>TB1Y_1</b>	<b>TB1Y_2</b>	<b>TB2Y_1</b>	<b>TB2Y_2</b>	<b>TDY_1</b>	<b>TDY_2</b>
<code>zefirben_B1_R1</code>	5.5872	7.0354	9.82	10.8574	-1.6472	-0.9722
<code>zefirben_B1_R2</code>	6.8669	7.3151	12.0201	12.1119	-0.6786	-1.1088

That's it ! All that is left to do now is to think.

The Hungarian vowel inventory is summarized below, using Hungarian orthography and IPA symbols (the latter in the usual brackets).

Vowel inventory of Hungarian (Ringen & Vago 1998)<sup>1</sup>

	[-back]		[+back]	
	[-round]	[+round]	[-round]	[+round]
<b>high</b>	i [i]	í [i:]	ü [y]	ű [y:]
<b>mid</b>	e [ɛ̃]	é [e:]	ö [ø]	ő [ø:]
<b>low</b>	e [ɛ]		á [a:]	a [ɔ]

## Hungarian stimuli (in orthographic form)

### Non-sense monosyllables

bib	bőb
bíb	bab
beb	báb
béb	bob
büb	bób
búb	bub
böb	búb

### Real-word tokens – monosyllabic

ív	vív
íz	híd
hír	ír
míg	víg
cím	síp
hisz	nyit
szél	cél
éj	héj

<sup>1</sup> The mid short [ɛ̃] is a dialectal variant of low [ɛ] in several non-Budapest dialects.

Real-word tokens – multi-syllabic

Second vowel is [i:]

zefírben	zafírban
zefírból	zafírból
szédített	zúdított
tömítő	tompító
primitívnel	normatívnál
esszívhez	passzívhoz
bővítéd	szolítod

Second vowel is [i]

bécsiben	bácsiban
bilivel	bulival
öcsitől	kocsitól
Filipem	polipom
rövidnek	szolidnak
érintett	bólintott
imihez	Tomihoz
csitrihez	lutrihoz

Second vowel is [e:]

bidétől	kadéttól
kreténben	szaténban
tenyérnél	tányérnál
filéhez	máléhoz
esszével	sasszéval
békésen	málésan
intézett	ganézott