Repeat After Me

User's Guide

1. General overview

System Requirements: Mac OS X (version 10.2 or later).
Supported sound file formats: 16-bit 22 kHz, mono uncompressed AIFF.
Document format: package, document bundle with .ram extension.

Repeat After Me is a tool that is designed to improve the pronunciation of text generated by the Text-To-Speech (TTS) system, by means of editing the pitch and duration of phonemes.

Brief instructions:

1. Type in source text.
2. Convert the source text to phonemes and "tune" representation (with information of phoneme pitch/duration).
3. Record with a microphone or load from a sound file a voice representation of the entered text.
4. Extract pitch and duration information from the recorded voice (or imported file).
5. Apply the recorded pitch and duration to source phonemes.
6. Interactively edit pitch/duration representation.
7. Obtain pitch/duration phonemes representation in text form for further use with the Text-To-Speech component.

2. Document window layout

The document window consists of the following text fields and functional panes:

1) The source text field.
2) The Phonemes field.
3) The Phonemes pane (shows phoneme duration).
4) The Pitch pane (a pitch/duration diagram).
5) The Recording waveform pane (an amplitude/duration diagram).
6) The status bar.
7) The Comments pane (generated phonemes with pitch/duration information - "tune" format).

The Phonemes and Pitch panes display the same information in different forms, and hence are both named "Tune" (implying that it is the graphical representation of the tune format). The Phonemes, Pitch and Recording Waveform panes look like the one pane, and therefore in some cases they are referred to as the Combined pane. The panes inside the combined view are in some cases referred to as Special panes. Below the Combined pane is the status bar. The Comments pane is located at the bottom of the window divided from it by the splitter.

Status bar

The status bar, located between the Combined pane and the Comments pane, displays information about the current pointer position. When the pointer is over a Special pane the status bar displays the horizontal pointer position in milliseconds. For recorded sound, the location in milliseconds is counted from the displaced point, not from the beginning of the sound. If the pointer is over the Pitch pane, it also displays the vertical pointer position in Hz.

3. Special panes functionality

All panes are represented as separate objects, which means that you may select each pane and perform editing tasks within it. Certain operations, for example Speak, can be applied to the currently active pane. Click any pane to make it active.

Panes background

All panes in the combined view share horizontal coordinates. The background in panes is colored according to the type of phonemes used when the synthesized pitch graph is built. The background color indicates the type of phoneme, as follows:

red – vowel;
blue – voiced consonants (b, d, g, j, v, z, D, Z, m, n, and so on);
green – voiceless consonants (p, t, k, c, f, s, T, S, and so on);
gray – pause.

Scrolling and zooming

You can change the vertical scales of the Pitch and Waveform panes using the slider control located above the vertical scroll bar of each pane. In some cases, the content of the panes may exceed the window borders after zooming. If this happens, you can take advantage of the vertical scroll bars that allow you to scroll the content of the Pitch pane along the Frequency axis and the content of the Waveform pane along the Amplitude axis.

For the content of the Combined pane, horizontal zoom is also available, using the slider control located in the bottom left corner of the pane. You may also scroll the content of the Combined pane along the Milliseconds axis. Because all special panes are aligned along the milliseconds axis, horizontal scrolling and zooming affects content of all special panes.

Choosing the ***Zoom In*** and ***Zoom Out*** commands from the **View** menu enlarges or reduces the scaling of the panes' content in increments of 10% for both horizontal and vertical scaling. If the active pane is being scaled vertically, it also zooms it in or out vertically by 10%. You cannot zoom in or zoom out beyond the predefined minimum and maximum zoom factors.

Rulers and grids

The Pitch and Waveform panes can have rulers and grids. The horizontal ruler is placed at the bottom of the combined pane and is shared by all panes. You can turn rulers and grids on and off using the corresponding commands from the **View** menu.

Phonemes pane

The topmost pane in combined view displays phoneme durations on a colored background. This is the only pane which displays phoneme borders and the phoneme itself.

A phoneme's duration can be changed by dragging the phoneme borders. Using the **Shift-click** technique on a phoneme border locks or unlocks it. When a phoneme border is dragged, all phonemes to the left of the dragged border up to the locked border or the first phoneme are resized proportionally. If there is a locked border to the right of the dragged border, the phonemes to the right are also resized proportionally, otherwise the phonemes to the right are merely shifted. To change the duration of a single phoneme, hold down the **Option** key while resizing its right border.

Pitch pane

The Pitch pane displays a graph of phoneme pitch and duration (sometimes referred to as the synthesized pitch graph) and a graph of pitch extracted from recorded sound (sometimes referred to as the recorded pitch graph).

Synthesized speech graph is represented by lines colored according to the corresponding phoneme types, with dotted lines used for voiceless consonants, solid lines for other phonemes. Points where pitch can be edited (pitch targets) are represented by dots. You can change a pitch target by dragging it vertically. By default, the neighboring pitch targets of this phoneme are lined up. To change a single pitch target within a phoneme, hold down the **Option** key while

modifying the pitch value.

To add a pitch target, hold down the **Shift** key and click in the place where you want to add a pitch target on the pitch curve. If it is possible to add a pitch target there, it will be added. You can delete pitch targets or move them horizontally by altering their numerical values in the Comments pane and then applying its content to the Pitch pane (see Build Graph section).

Recorded pitch graph is displayed if there is a sound attached to the document and pitch has been extracted from it. The voiced sections of the sound are represented by thick cyan lines behind the synthesized graph, and voiceless sections are represented by gray lines at the bottom of the pane (at the minimum pitch level).

To align the beginning of speech in the recorded sound with the time axis zero, the recorded pitch graph can be moved horizontally using the **Left** and **Right** arrow keys. Default scrolling is 1 pixel per key press. To speed up scrolling, hold down the **Option** key. The recorded waveform is aligned with the recorded pitch graph automatically.

To increase or decrease the pitch range of the recorded sound, use the **Up** and **Down** arrow keys. This can be useful if you are mapping speech of a male person to a female TTS voice or vice versa. Holding down the **Option** key while scrolling also speeds up this process.

Waveform Pane

The Recording waveform pane displays the amplitude/duration graph of the recorded sound. This pane is not editable, but its content can be spoken. For the recorded waveform, a portion of sound can be selected, zoomed to, and played.

As in the Pitch pane, you can align the beginning of the sound to match the beginning of speech with the left and right arrow keys. If pitch is extracted, the recorded pitch graph is aligned with the waveform automatically. If horizontal alignment is changed, remember that the speaking of the whole sound will start from the point at which you have aligned the beginning of the sound, not from the beginning of the recording.

The Waveform pane is optional and can be turned off either by choosing ***Hide*** from its contextual menu or by choosing ***Hide Recorded Waveform*** from the **View** menu. You can turn this pane on by choosing ***Show Recorded Waveform*** from the **View** menu.

4. Main functionality

Most of the tasks you generally use RAM for are gathered in the **Tools** menu, all of its items are duplicated in the window's toolbar.

Convert to Phonemes converts the content of the source text field to phonemes and displays them in the phonemes field. If you press Enter in the source field, in addition to converting text, a synthesized pitch graph will also be built from the source text. The corresponding toolbar item is named "To Phonemes".

Build Graph builds a synthesized pitch graph out of the content of the currently active pane. The graph is displayed in the Pitch pane. This command is applicable to the Text field, Phonemes field and Comments pane. For added convenience, the corresponding items "Text to Graph", "Phonemes to Graph" and "Tune to Graph" are also available on the toolbar. (Not all of the available items are included in the default toolbar set).

Note that when the graph is built from the Comments pane, it could be built from a selection or from the whole content of the Comments pane. Due to the fact that this pane may contain not only the tune representation of the working phrase, but also actual comments, the string requested to build the graph from is considered to be in tune format. However, if this string contains tune representation enclosed by [[inpt TUNE]] ... [[inpt TEXT]] embedded commands, everything but the text enclosed by these commands will be skipped.

Graph to Phonemes dumps a selected portion of the graph, or the whole graph if no selection has been made, in its textual representation ('TUNE' format) to the Comments pane. It replaces the selection, or if there is no selection when the pane is inactive replaces the whole content of the pane, and encloses this representation with [[inpt TUNE]] ... [[inpt TEXT]] embedded commands. This textual representation can then be modified, selected and read back into the Phonemes/Pitch panes. The corresponding toolbar item is named "Tune".

Impose Recorded Durations attempts to align phoneme durations with the recorded sound. It calls the external tool, PTAligner. PTAligner will fail if your phrase has more than 100 phonemes.

Extract Pitch extracts pitch information from the recorded sound and displays the recorded pitch graph in a Pitch pane. When the pitch is extracted, you can align the synthesized pitch graph with it, either manually or automatically.

Impose Recorded Pitch aligns the synthesized pitch graph to the recorded pitch graph if pitch is extracted from the recording. For this alignment to be more precise, you can add pitch targets (see the paragraph on adding pitch targets in the Pitch pane functionality description).

Attaching sound to the document can be achieved in two ways. You can choose ***Attach...*** from the **Sound** menu (or click the Attach Sound button on the toolbar) and then choose a file with already recorded sound (16bit mono uncompressed AIFF only). Alternatively, if you have microphone, you can choose ***Record...*** from the **Sound** menu (or click the corresponding button on the toolbar) and record the sound in the application. When you have a sound attached to the document, its waveform is displayed in the Recorded Waveform pane.

Speaking pane content

Choosing ***Speak*** from the **Sound** menu plays the content of the active pane. The ***Stop*** command from the **Sound** menu stops the speech. Speaking behavior is slightly different for different panes.

For the source text field and Phonemes field, choosing ***Speak*** from the **Sound** menu plays the whole content of the active field (in the 'TEXT' and 'PHON' speaking mode, respectively).

Speaking phonemes and the Pitch pane content will produce the same result, as they display the same information and have common selection. When several phonemes are selected, choosing ***Speak*** from the **Sound** menu plays the selected portion. When nothing is selected, it plays the whole phrase.

For the Waveform pane, either the selected portion of the recorded sound is played, or, if there is no selection, the sound is played from the very beginning. The point at which you align the sound is considered to be the starting point, rather than the actual beginning of the sound (see the paragraph on sound alignment in the Pitch/Waveform panes functionality description).

For the Comments pane, the selection (or the whole content if there is no selection) is spoken in the 'TEXT' speaking mode. So, if you want to play phonemes or tune you have in this Comments pane, you should specify the desired speaking mode yourself using embedded commands (when a graph is dumped to this pane using Graph to Phonemes, this is done automatically).

5. Additional features

Selection

Portions of phonemes and sound can be selected separately. To select a sound portion, activate the Waveform pane, and then drag the pointer over the selection until you have selected the portion you want. If the portion of sound is selected, it will be spoken instead of speaking the whole sound.

To select a range of phonemes, activate the Phonemes pane. Click inside the phoneme you want to be the start (or the

end) of the selection, and drag over the selection to the required phoneme. The selection is displayed in both the Phonemes and Pitch panes, but it can be altered only in the Phonemes pane.

Zooming to fit selection

During certain actions, such as attaching a new sound, building a synthesized graph, or opening a document, the combined pane will be scaled so that most of its content is visible (all if possible).

If one of the special panes is active and contains selection, choosing ***Fit to Window*** from the **View** menu will attempt to scale and scroll its contents so that the selected range fits the whole combined pane and nothing else is visible. However, if the scale factor necessary for this exceeds the possible scaling range, the closest appropriate scale factor will be chosen. If there is no selection, the above actions will be applied to the whole pane content.

Zoom to fit selected words

If you want to concentrate on working with just several words rather than with a whole phrase, you can select those words in the *source text field* (or *phonemes field*) and choose **View** - ***Fit to Window***. The effect will be the same as selecting the respective phonemes in the Phonemes pane. This feature relies on the fact that the word number is the same in the source text field, Phonemes field, and tune representation, and may fail if your graph does not show the entire combined pane.

6. Shortcomings and Limitations

Working with phrases containing more than 100 phonemes is not possible.