TP 3.1 Software: A Tool for Designing Audio, Visual, and Audiovisual Perceptual Training Tasks and Perception Tests

Andréia Schurt Rauber[1], *Anabela Rato*[2], *Denise Cristina Kluge*[3], *Giane Rodrigues dos Santos*[1]

[1] Catholic University of Pelotas, Pelotas, Brazil
[2] University of Minho, Braga, Portugal
[3] Federal University of Paraná, Curitiba, Brazil

asrauber@gmail.com, asrato@gmail.com, deniseckluge@gmail.com, giane.rsantos@hotmail.com

Abstract

Many speech researchers (linguists, psycholinguists, speech therapists) who work with speech perception still struggle to prepare their data collection tests mostly due to limitations as regards programming skills and knowledge of scripting languages. Concerned with this scenario and methodological constraints, we developed an open source application software (http://www.worken.com.br/tp) that enables researchers to design not only speech perception tests, but also perceptual training tasks with immediate feedback. The objective of this paper is to show some of the main features of the software and to highlight its user-friendliness.

Index Terms: perceptual training tool, second language testing/training, immediate feedback

1. Introduction

Since the compilation of speech perception studies in the book organized by Winifred Strange [1] in 1995, studies on second/foreign language (L2) speech perception have increased considerably. Experiments that test/train L2 speech perception make use of software that either have restricted access because they are developed for specific projects or that *do* have open access, such as Praat [2], Alvin [3], Percy [4] and WebExp2 [5], but might not be considered accessible to researchers with little expertise in programming or scripting languages.

Moreover, researchers using open source software may have limitations to elaborate perceptual training tasks using video material. Concerned with these constraints, we designed TP 3.1, which stands for *Teste/Treino de Percepção* (Perception Testing/Training). TP 3.1 tests/trains the perception of sounds (speech segments and suprasegments) in an audio-only condition, in audiovisual and video-only conditions.

Thus, the aim of this paper is to describe the main features of TP 3.1 and to highlight its user-friendliness, which can be very helpful to many undergraduate/graduate researchers who want to design speech perception experiments and lack programming knowledge and technical support.

The paper is organized as follows: in Section 2 we will present general features of the software and give a brief explanation about types of perception tests/tasks; in Section 3 we will show how to set identification and discrimination tests/tasks in TP; in Section 4 we will present some TP testing/training screens with video/audiovisual material; and in Section 5 we will state our final remarks. We attached a zip folder to this paper containing audio and video files that can be used as samples to create identification and perception tests.

2. TP 3.1: General features

The main motivation to design TP was the need of a software that could not only test speech perception, but also provide users with perceptual training tasks with immediate feedback. Although the program runs in one of three languages, namely English, Portuguese, and Spanish, one general concern was to allow the testing of any language; thus, the experimenter can configure the necessary instructions and buttons according to the language being tested/trained. As shown in Figure 1, no programming language to do so is necessary, only predefined fields need to be filled in.

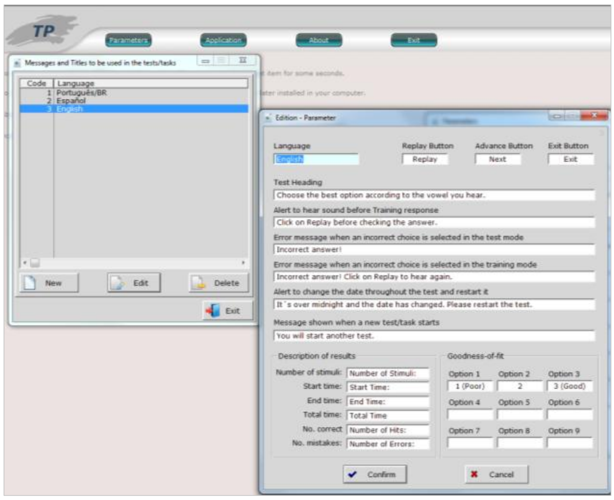

Figure 1: *Screens to set messages and buttons.*

TP runs in any Windows version, but it can also run in other operating systems, such as Linux and Macintosh, provided that adequate emulators are previously installed.

Two types of tests/training tasks can be applied with this software: speech identification and speech discrimination. In the former, a single stimulus can be presented and the listener has to choose from several response buttons shown on a computer screen which one of them most resembles the sound heard. For example, a listener hears the word "pet" and decides if the first sound of the word most resembles [p], [t], [k], [b], [d], or [g].

As regards discrimination tests, their objective is to test if the listener can discriminate sounds in a sequence of two or three stimuli. These stimuli can be isolated segments or syllables or even words. Discrimination tests can have several formats: categorical AX tasks, AXB tasks, ABX tasks, and category change tasks, for instance. In a categorical AX task, participants hear a pair of stimuli and decide whether the two stimuli are the same or different. For example, they hear the stimuli "think"-"sink" and consider them as different, or hear the stimuli "think"-"think" and consider them as the same. In the AXB test participants hear three stimuli and decide if the second stimulus is similar to the first or the third stimulus. For example, participants hear "bit"-"beat"-"beat" and consider the third stimulus to be similar to the second. A variant of this test is the ABX format, an oddity task in which three stimuli are presented and participants decide whether the odd item is in the first, second or third position. The position of the odd item varies in each trial. For example, participants hear the trial "bit"-"bit"-"beat" and are expected to consider the third stimulus as the odd item. Finally, in a category change task several stimuli from one category are presented followed by stimuli from a different category and participants have to indicate when a change in category has happened. For example, listeners hear a sequence of "da"-"da"-"da" tokens and when a stimulus "ga" is played they signal it.

In TP, when setting the language instructions and the kind of test (testing mode) or task (training mode) the experimenter can choose to automatically randomize the sequence of stimuli each time a task begins to minimize ordering effects. If this option is not selected, the order of presentation is the one predefined by the researcher. If the test mode is selected, the researcher can include a category-goodness-of-fit scale with a maximum of 9 points, so that stimuli can, for instance, be judged from poor (1) to good (9) exemplars of a category. A 100-point sliding scale (Figure 2) can also be added, to rate, for example, overall degree of perceived foreign accent in second/nonnative language speech. In addition, experiments can be configured so that the listener can (1) replay the stimulus more than once before pressing one of the response buttons, and (2) go back one trial and change a response once given. These two options are set by the experimenter and the number of possible replays is also pre-defined. If the training mode is chosen, the software application automatically activates the trial-by-trial feedback by showing an error message when the wrong button is chosen. The user has to listen to the stimulus/trial again and choose the correct answer in order to proceed.

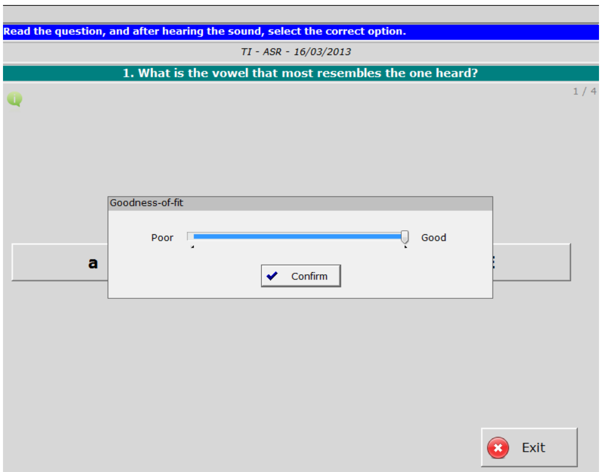

Figure 2: *Sliding scale to rate goodness of fit.*

At the end of each task a summary of the results is displayed on the computer screen, namely the total number of correct and incorrect answers, time spent, and total number of stimuli heard. This cumulative feedback can also be provided at the end of the training session.

Finally, TP automatically registers detailed information about the performance of each user in an Excel spreadsheet. These data include the total number of correct and incorrect answers, the response chosen to each stimulus/sequence of stimuli, reaction time, and a confusion matrix (Figure 3). In the training mode, the answer saved is the first option chosen by the participant. Reaction time is measured from the moment a stimulus starts playing to the mouse clicking on the response button. The confusion matrix provides overall percentages of misidentifications, i.e., the number of times a certain phonetic segment was perceptually misidentified as (an)other segment(s), which helps to understand the patterns of perceptual errors.

	A	B	C	D	E	F	G	H	
1	Stimulus	Response	Result	Time(seco	Stimulus q	Category			
2	a	a	C	1.14	Good	71			
3	E	E	C	0.72	Good	100			
4	a	a	Err	1.19	Good	100			
5	a	a	C	2.45	Regular	45			
6									
7	Total:	Test	Date	Time Used	# of Stimul	Hits	Errors	% Hits	
8		Ident.	16/02/2013	5 s	4	3	1	75.00	%
9									
10									
11	Summary:	Response	Total	Hits	Errors	% Hits			
12		a	2	2	0	100.00	%		
13		E	2	1	1	50.00	%		
14									
15	Confusion:	Test	Heard	Identified	Errors	%			
16		E	/E/	a	1	25.000			

Figure 3: *Test results automatically saved in a spreadsheet.*

It is also possible to organize several experiments in a single testing/training session by choosing the sequence of presentation. A screen with a message previously set by the researcher will inform participants when a new experiment begins.

In the next section we will show how to set identification and discrimination tests in TP.

3. Identification and Discrimination Tests in TP

Setting identification and discrimination tests/tasks in TP is simple and intuitive. First, it is necessary to give the test/task a name and to check the "activate" box. After activating the test/task, the labels that will be shown on the response buttons must be specified. In the audiovisual and visual modes, five response buttons can be added. In the audio mode, 28 orthographic labels or six response buttons with pictures can be included in an experiment (Figure 4).

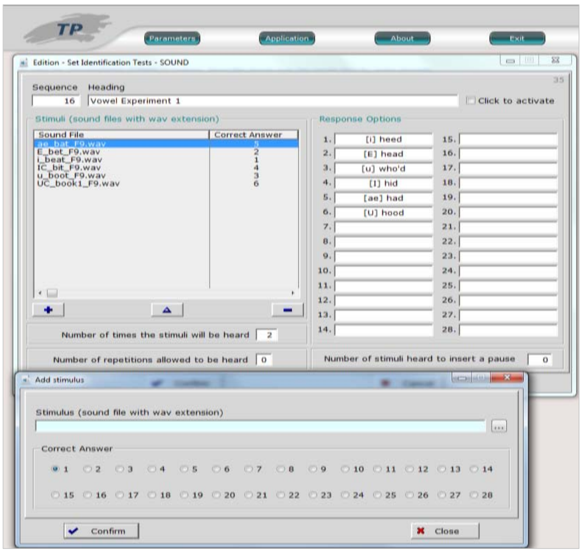

Figure 4: *Loading stimuli and setting response buttons.*

In this screen, besides determining the correct response button for each stimulus, the researcher can also set (1) the number of times each token will be presented during the task, (2) the number of repetitions the listener will be allowed to hear each sequence of stimuli, and (3) add a pause between a certain number of stimuli. An example of a test screen participants will see is shown in Figure 5.

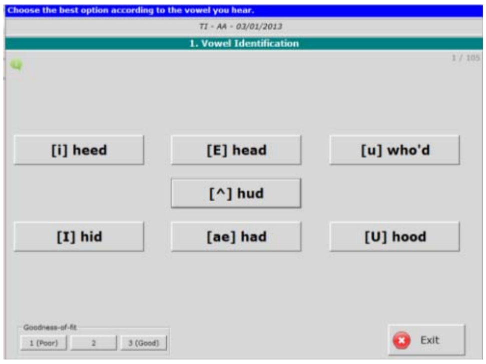

Figure 5: *Example of an identification test screen.*

Figure 6 shows an example of an identification test screen with images as response buttons.

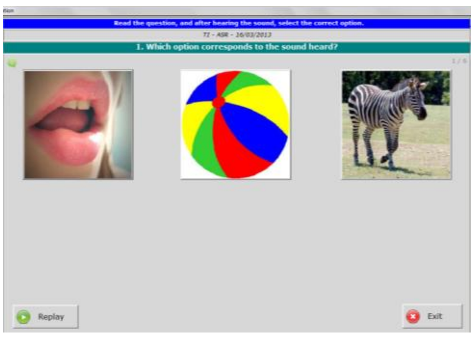

Figure 6: *Identification test screen with images as response buttons.*

Text with instructions for each test/task can be added. These instructions will be available to read before a test/task begins and during the experiment (Figure 7).

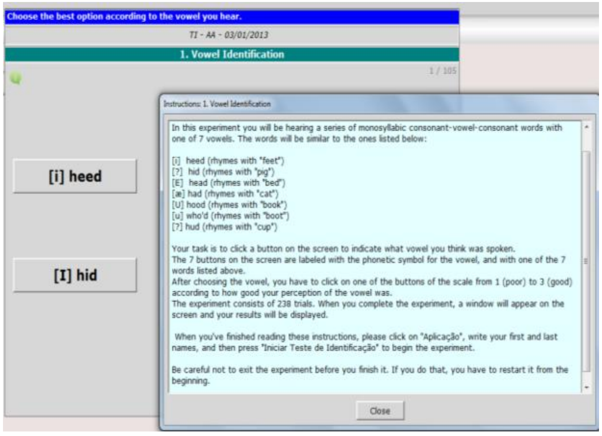

Figure 7: *Instructions for an identification test.*

An example of an AX training task screen is shown in Figure 8. In this example, the participant chose the incorrect response button. TP shows the correct answer and in order to proceed the listener must hear the sequence of words again and choose the right answer.

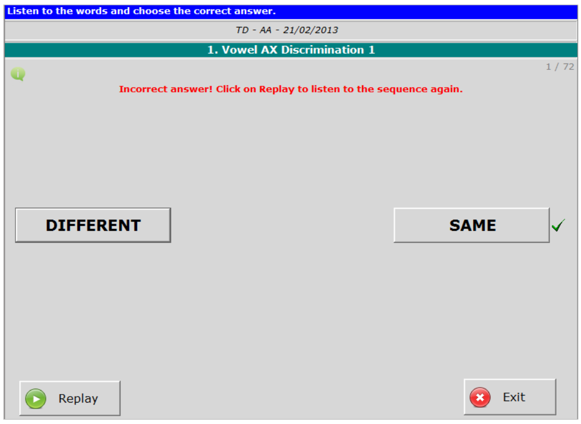

Figure 8: *Example of an AX discrimination task screen.*

4. Video-only and audiovisual tests/tasks

According to Rosenblum [6], "it is becoming increasingly clear that human speech is a multimodal function, usually apprehended by visual (lipreading) as well as auditory (hearing) means" (p. 51). TP 3.1 allows the use of video in order to test/train the perception of sounds using visual cues. The same procedures used to load and label sound files must be followed with video files used for visual and audiovisual experiments.

It is worth noting that TP 3.1 is not a video or sound editor, hence the files to be used in tests or training tasks must be ready to be loaded in the software.

Figure 9 shows an example of an audiovisual test/task whose aim was to test the perception of English nasals in word-final position [7]. In this study, the visual articulation cue of [m] and [n] was shown to give relevant information to participants, in this example, Brazilians who nasalize the vowel preceding an English word-final nasal. Thus, given the nature of the target stimuli, the focus in the video was only on the mouth of a native speaker of American English.

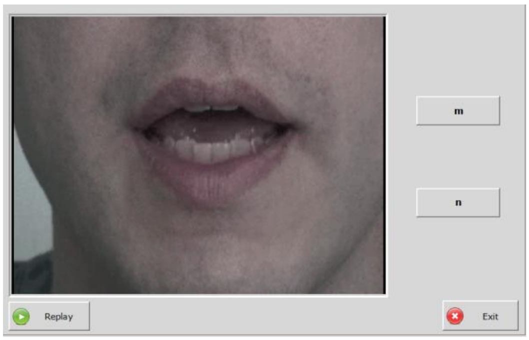

Figure 9: *Example of an identification test/task set in TP 3.1: focus on the speaker's mouth.*

Differently, Figure 10 shows the screen used in a study [8] whose aim was to test the perception of the English voiceless interdental fricative by native speakers of European French and Brazilian Portuguese. The participants' native languages do not have interdental fricatives, and it was expected that the target sound would be identified either as [t], [f], [s] or the interdental fricative itself.

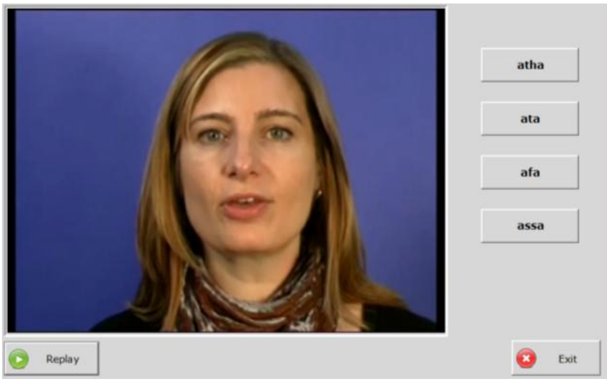

Figure 10: *Example of an identification test set in TP 3.1: focus on the speaker's face.*

It is possible to export perception tests/tasks set up in TP to a single folder so that experiments can be sent to other researchers. To run these tests/tasks in computers in which the program is installed, the option "third-party tests" must be selected in the Test Environment Selection menu. Another option to export experiments is the TP Direct menu that can be used to copy activated tests/tasks to an external drive in order to install them in other computers. The difference between these two menus is that in the first option tests/tasks can be edited and changed after being exported, and with TP Direct they are not editable. This is explained by the fact that the former is aimed at sharing experiments among researchers, and the latter is more suitable to administer the experiments to a group of informants.

5. Final remarks

In this paper we briefly described the main characteristics of the TP 3.1 software, which was created to offer researchers an intuitive tool to prepare their perception tests/tasks. TP can easily run perceptual training tasks and give immediate feedback to participants. This free software is available at http://www.worken.com.br/tp. We are now working on adapting the software for visually-impaired participants.

6. Acknowledgments

We would like to thank Marcos Figueiredo (Worken Systems) who developed our software and Mara S. Reis for the video (Figure 10) that illustrates an audiovisual experiment in this paper. Our special thanks to the donors who have helped us pay expenses to implement changes in TP (this software is not funded by any institution or research grant).

7. References

[1] Strange, W. [Ed], "Speech perception and linguistic experience: issues in cross language research", Timonium, MD: York Press, 1995.

[2] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer", v. 5.3.04, availabe at: www.praat.org, accessed on 13 March 2013.

[3] Hillenbrand, J. M., "Alvin", available at: http://homepages.wmich.edu/~hillenbr/, accessed on 13 March 2013.

[4] Draxler, C. "Percy- An HTML5 Framework for Media Rich Web Experiments on Mobile Devices". Interspeech 2011 Proc.: 3339-3340, 2011.

[5] Keller, F *et al.* "WebEXP2", available at:http://www.inf.ed.ac.uk/research/isdd/admin/package?download=143, accessed on 12 March 2013.

[6] Rosenblum, L. D., "Primacy of multimodal speech perception", in D. B. Pisoni and E. R Remez [Ed], The handbook of speech perception, 51-78, Blackwell Publishing, 2005.

[7] Kluge, D. C., "Brazilian EFL learners' identification of word-final m-n: Native/nonnative realizations and effect of visual cues". Saarbrücken: Lambert Academic Publishing, 2010.

[8] Reis, M. S., "The assimilation and discrimination of the English /th/ by European French and Brazilian Portuguese speakers", in A. S. Rauber, M. A. Watkins, R. Silveira and R. D. Koerich [Ed], The acquisition of second language speech, 169-192, Florianópolis: UFSC, 2010.