A computer-based tutorial on Models of Speech Perception

Francesco Cutugno, Cécile Fougeron

A computer-based tutorial on Models of Speech Perception

Francesco Cutugno and Cécile Fougeron^*

*CIRASS - Università degli Studi di Napoli Federico II
^Lab. de Psycholinguistique Expérimentale, Univ. de Genève & ILPGA,Paris III
frix@unina.it, cecile.fougeron@pse.unige.ch

Abstract

Learning material on speech perception has appeared in a recent survey to be less documented than other topic in phonetics. This tutorial was built in order to fill, at least in part, this lack. We aim to provide a comprehensive overview of some models of speech perception process to both basic and advanced students. The topics covered in the tutorial include a broad description of the mechanisms involved in the speech perception process; difficulties that listeners have to overcome and that models have to account for; a comparative description of 5 of the most famous models: Fant's auditory model, Motor Theory, Analysis by Synthesis, Quantal Nature of Speech, Hyper- and hypo-speech models. The description and comparison of the models are based on their main assumptions, historical background, and supporting experiments and results. This computer based tutorial rely on an interactive learning strategy based on multimedia presentations, hypertext, pop-ups, and animated diagrams. This work is freely available on World Wide Web, at the Elsnet site (www.elsnet.org).

1. Introduction

In a current survey of phonetic education in Europe conducted by the Phonetic Working Group of the Socrates Thematic Network in Speech Communication Sciences [1,2], we observed that speech perception is one of the less thought topics in phonetic courses and a topic in which teachers wished to have more teaching material. Therefore we developed, as one of the task of the Working Group, a tutorial on models of speech perception in order to enrich the available didactic set of materials in this field.

Some teaching support about speech perception are currently available in the form of computer assisted tutorial or more general interactive self study courses. These materials mainly concern aspects of psychoacoustics [e.g. 3], examples of well known perception facts [e.g. 4] and even simulations of perceptual experiments [e.g. 5]. On the contrary, description of more theoretical models of the speech perception process seems to be rarely treated in these kind of materials. References about this topic are more frequent in textbooks [e.g. 6, 7], but very often the models are briefly summarised and rarely systematically compared. Alternatively, teachers have to refer students to original published papers (most often written by the

model's author himself) or chapters of highly specialised books not directly conceived as approachable teaching materials.

It was thus a challenge for us to try to built an interactive and interesting tutorial on these theoretical aspects that are quite hard to explain, illustrate, synthesise, and compare. As a result, a large part of our tutorial is a text-based course covering these theoretical aspects, but we have tried to provide the students with simple illustrations made as links to non-text materials such as simulations, sounds samples, references and citation from original papers, and animated figures.

The four models we review in this tutorial have been regarded as fundamental in the development of Phonetic Sciences.

These models, listed in chronological order, are:

- the Auditory model by G. Fant [8, 9];

- the Motor Theory by A.M. Liberman and colleagues [10, 11];

- the Analysis by Synthesis by K.H. Stevens [12];

- the Quantal Nature of Speech (QNS) by K.H. Stevens [13];

- the Hyper and Hypospeech by B. Lindblom [14].

This tutorial is aimed at basic and advanced students in phonetics. Even though most of the specialised terminology and difficult concepts are defined and illustrated, students are supposed to have acquired a basic knowledge of phonetics.

This tutorial is supposed to be integrated into a basic course on Speech Perception. It is a self-study material, although its modularity will allow teachers to build on its contents and adapt it to students needs. It can easily be used as a complement to a course covering other aspects of speech perception.

2. Contents

The course is divided into three main parts:

1. Generalities about speech perception process:

This section includes an overview of speech perception processes, including an introduction to questions such as: "why is the perception of speech a complex task?"; "which properties of the speech signal are important for perception?"; "what kind of information the listener has to compute?". Furthermore this section introduces the

student to terms such as: features, invariance, linearity, segmentation, normalisation, eccetera.

This part also cover a series of facts that each model should describe and interpret. The process of speech perception is described in terms of signal properties, extraction of auditory units, central treatment. Particular attention is given to the problem of how listener integrate informations coming from different sources.

2. Description of a variety of speech perception models:

In this section, students find accurate descriptions of the various models listed in the tutorial. Models are first classified according to some main characteristics such as: active or passive, where invariance lies in the models, what kind of minimal unit is the object of recognition. Along with this classification, a series of ID cards illustrate main aspects of each model.

3. How these models do or do not work.

In this resuming section, arguments in favour and against each models are developed. Data coming from various experiments reported in literature are used in the discussion together with multimedia examples illustrated by pictures and speech sounds.

3. Tutorial Structure and Organisation

This tutorial has been developed with Authorware, an object authoring software distributed by Macromedia, Inc. The tutorial is compiled as an executable application that can be run both on local Macintosh or PC computers and on the World Wide Web with a Shockwave plug-in. The Authorware software is not needed to run the compiled application in any case.

The tutorial is organised as a sequence of different kinds of cards (i.e. windows):

   - The "course cards" include the theoretical content of the tutorial. Navigation between these cards is made by buttons directing to particular section of the course (e.g. Figure 1) and by basic "back", "next" and "menu" buttons. In this way, students can access to content in a linear order or can go directly to particular chapters. The presentation of the concepts is made either as plain text or as animated objects (text and/or sound) appearing successively on the screen as the student click on a button to continue (e.g. Figure 2).
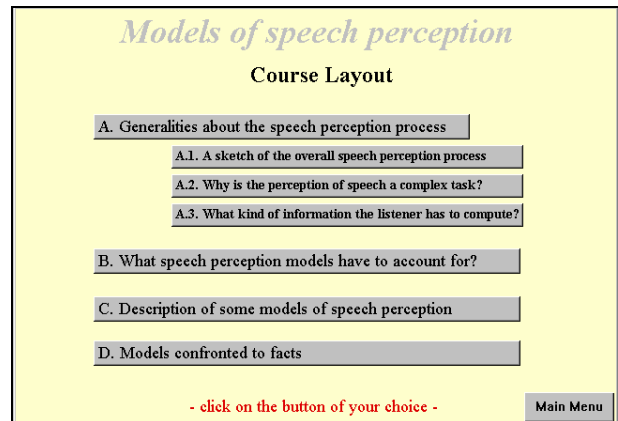

Figure 1: Clickable buttons (always in grey) allow the student to navigate between sections of the tutorial. Instructions, written in red, help the student on how to proceed.
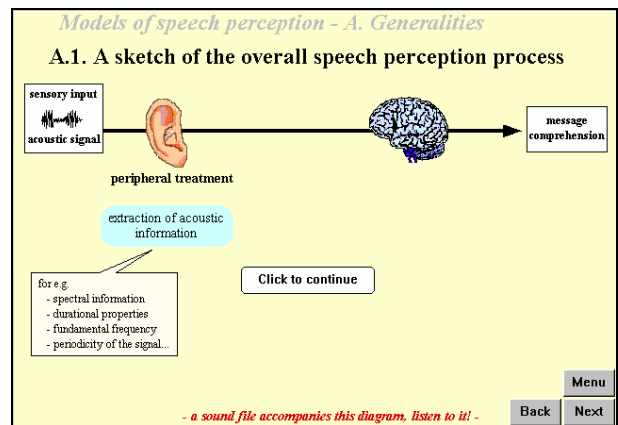

Figure 2: Example of a "course card" where the content appears successively in the form of pictures and block of text complemented by a more comprehensive explanation recorded in a sound file. A "click to continue" button allow the student to proceed step by step. Instructions, written in red, tell the student that he has to listen to the audio explanation.

   - The "example cards" are accessed within the "course cards". They contain examples or explanation complementing the notions given in the course card. These example cards are accessed by a white «click here» button and are exited by an OK button returning to the course card (e.g. Figure 3).

   - The "reference or definition cards" are pop-up balloons giving either references of articles and related web sites, or short definitions of some specialised words. These pop-ups are activated by clicking on particular words signalised in blue (e.g. Figure 4). The references and definition cited in the tutorial are also available in alphabetic order in one section of the tutorial.
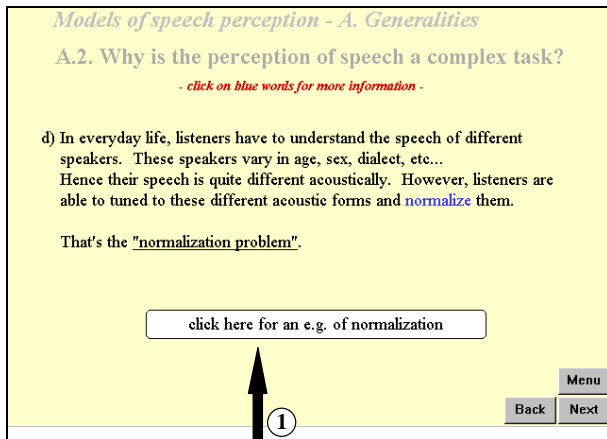
Figure 3: Example of a course card explaining the concept of normalisation. The white button (1) leads to an example card where the process of normalisation is illustrated by a comparison of the same sentence produced by a male and a female speakers (with sound files & spectrograms).
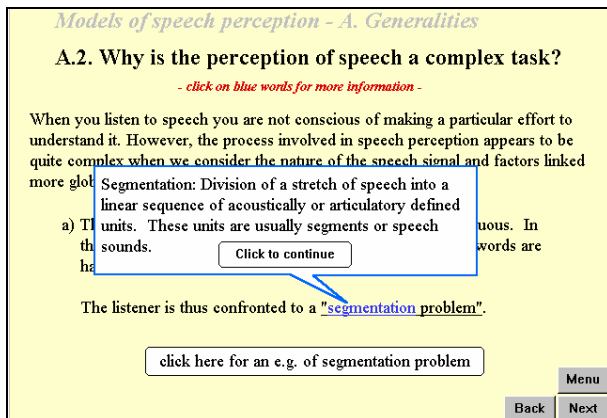

Figure 4: Example of a definition pop-up balloon explaining the word "segmentation". Students have instruction telling them that words written in blue are clickable.

The organisation adopted in this tutorial allows us to develop an interactive learning strategy. Students are able to navigate alone through the tutorial with the help of explicit buttons. When needed, they are helped by short instruction written in red (e.g. «click here to see...,to listen to..., to have an example of...»). This structure, comparable to an hypertext frame, give the students relative freedom in organising their learning session. They can go directly to particular independent section, but within a section, access to subsections (examples, more information) always leads them back to where they were before. In this way the students never get lost as sometimes can occur when they are consulting the Web.

Moreover, this kind of presentation allow us to offer the student with different ways to approach the description of the models. One choice is to learn about the models using the sequence of ID cards structured the same way

for each models (basic ID (e.g. Figure 5), assumptions, explanatory diagram...). The other approach is more comparative: specific points of comparison (units of perception, active or passive theories (e.g. Figure 6) are selected and the students learn how each model is positioned relative to each point.
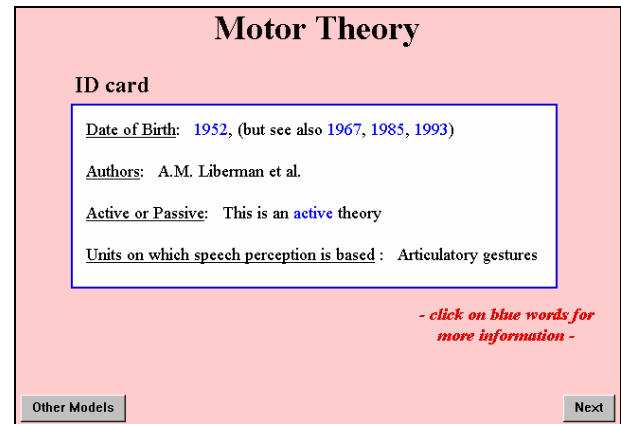

Figure 5: Example of basic ID card made for the "Motor Theory" model.
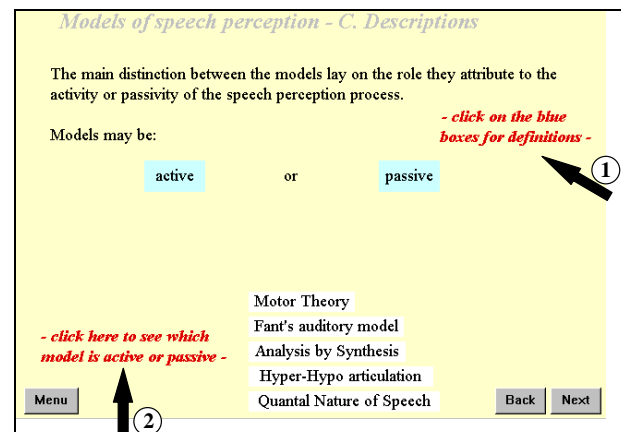

Figure 6: Example of comparative card based on the distinction between "active and passive" models. In(1), students are invited to click on the words "active" or "passive" to see their definition. In (2), students have access to an animation showing which of the models are active or passive.

In all the sections of the tutorial, multimedia presentation (visual and audio documents), animated figures, hypertext for definitions or references are used to make the learning of this theoretical subject entertaining and interesting.

4. Discussion

This tutorial is not a comprehensive course on speech perception. Our major purpose was to present some of the major models that have been elaborated in the last

40 years to account for the speech perception mechanism. The models described in the tutorial have been selected on the basis of the importance they have had in the field.

We decided to provide students with only a small review of well known perception facts (i.e. categorical perception, speech sounds recognition, ...). These facts are merely used as examples or arguments in favour of one or the other model. Always for this reason we decided to avoid any reference to psychoacoustic or basic audition facts.

Our work is freely available through World Wide Web: it is hosted on the elsnet web site (www.elsnet.org).

The work presented here is a first version of this tutorial. It will, in the future, undergo an evaluation on the basis of the recommandations of the "Computer Aided Learning" Working Group of the Socrates Thematic Network for Speech Communication Sciences. This evaluation should be based both on the basis of students reactions and of judgement offered by a wide selection of experts.

Improvement of this tutorial is planned. In this first version, no student self-assessment (i.e. "the ability to judge one's mastery of the material being presented" [1:128]) is available. This could be done in a future version by creating small test questions and exercises related to the models discussed and the problems occuring in explaining the process of speech perception. Furthermore, knowing the difficulties teacher may have to find teaching materials in languages other than English, a version of this tutorial in Italian and French (as a starting point) is planned.

Acknowledgements

References

[1] Bloothooft, G, van Dommelen, W, Espain, C, Green, P, Hazan, V, Huckvale, M and Wigfross, E, (1997). The landscape of future education in speech communication sciences – 1 analysis, OTS publications, Utrecht.

[2] Bloothooft, G, van Dommelen, W, Espain, C, Hazan, V, Huckvale, M and Wigfross, E, (1998). The landscape of future education in speech communication sciences – proposals, OTS publications, Utrecht.

[3] Acoustical Society of America, (1989). Auditory demonstrations on compact disk. asa.aip.org/discs.html

[4] Sensimetrics, (1998). Speech Production and Perception I, an interactive tutorial. www.sens.com/SPP1.htm

[5] Ericsson, A, web page for Categorical Perception, www.ling.umu.se/~anderse/education/Katper.html

[6] Borden, G, Harris K S and Raphael L J, (1994 3rd ed.), *Speech Science Primer. Physiology, acoustics and perception of speech,* William & Wilkins, Baltimore.

[7] Lieberman, P, Blumstein, S E, (1988), *Speech physiology, speech perception, and acoustic phonetics,* Cambridge University Press, Cambridge MA.

[8] Fant, G, (1960). *Acoustic theory of speech production*, Mouton, The Hague.

[9] Fant, G, (1967). Auditory patterns of speech. In Whaten-Dunn, W (ed.), (1967). *Models for the perception of speech and visual form*. MIT Press, Cambridge MA, 111-125.

[10] Liberman, A M, Cooper, F S, Shankweiler, D P and Studdert-Kennedy, M, (1967). Perception of the speech code, *Psychological Review,* 74:431-461.

[11] Liberman, A M and Mattingly, I G, (1985). The motor theory of speech perception revised, *Cognition,* 21:1-36.

[12] Stevens, K N and House A S, (1972). Speech perception, in Tobias J (ed.) (1972). *Foundation of modern auditory theory*, Academic Press, New York, 2:1-62.

[13] Stevens, K H, (1989). On the quantal nature of speech, *Journal of Phonetics,* 17:3-45.

[14] Lindblom, B, (1990). Explaining phonetic variation: A sketch of the H&H theory, in Hardcastle, W J and Marchal, A (eds.), (1990), *Speech production and speech modelling,* Kluwer Academic Press, Dordrecht, 403-439.