HUTear – a Free Matlab Toolbox for Modeling of Human Auditory System

Aki Härmä and Kalle Palomäki

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
Espoo, Finland

Abstract

This paper introduces a Matlab toolbox for modeling of human hearing. The toolbox consists of a generic frame and a set of optional building blocks which can be used to reconstruct several different auditory models published at fields of psychoacoustics and hearing research. The toolbox is freely available at `www.acoustics.hut.fi/software/HUTear`.

1 Introduction

In principle, modeling of the peripheral part of the auditory system is straightforward. Main functions of outer ear, middle ear with its ossicles, and cochlear spiral of the inner ear can be studied in terms of conventional mechanical physical models which usually can be approximated using digital filters. Computational models of the periphery of the hearing mechanism are rather powerful. Many aspects of human hearing which can be found in psychoacoustic listening tests, for instance, the frequency selectivity of hearing, can already be seen at the level of the mechanical periphery.

In the hearing mechanism, the mechanical part is followed by neural processing. Usually individual neural cells higher at the auditory pathway are not explicitly modeled but some functional computational model is used. Usually a such model have been developed and adjusted so that some particular effect observed, e.g., in a listening test, can be predicted by the model. The reasons for this are obvious – the number of neural cells is huge and their interaction at higher stages of the system is poorly understood. This difficulty of modeling neural processes has led to the situation where depending on a research group or goal in modeling, several different functional models are used.

Models of human hearing are becoming increasingly important at the fields of speech and audio signal processing. Modern audio and speech coding algorithms rely on modeling of frequency masking characteristics of the hearing mechanism. Objective sound quality evaluation is one of the most important technical applications for computational auditory models. Sound quality evaluation of audio signals [1] was also the application which drove the authors of this paper to reproduce some of the models presented in psychoacoustics literature and collect them to a modular toolkit.

The main design goals are listed below:

- The toolbox should be useful for both research and education.
- The toolbox should simple and easy to learn to use.
- The toolbox should be easily customizable, easy to update, and straightforward to include into users own application.
- All elements should be computationally well optimized.

2 Auditory models and HUTear

Depending on an application, quite different models can be used. A model for human auditory processing may consist of the following elements.

2.1 Outer and middle ear

Outer ear is usually modeled as a simple linear filter. In the case of a *binaural* model this may also contains a model for directivity of the ear, i.e., so called Head-Related Transfer Functions, HRTFs. Model for inner ear is usually integrated into a model of outer ear because it can be approximated by a linear time-invariant system. In HUTear, there are several alternative modules for this part.

2.2 Inner ear and cochlea

Frequency selectivity of the hearing mechanism is usually modeled as a filterbank. Currently, the most common choice is so called *gammatone* filterbank [2]. In HUTear, there are three different implementations for the gammatone filterbank. It would be also possible to use a transmission line implementation for a cochlear model [3]. Currently, the toolbox does not contain alternatives for filterbank models. However, it is possible to integrate such models easily into the toolbox because the filterbank module can be replaced with some other function. Typically the number of channels in the filterbank ranges from 32 to 256. The number of channels in a filterbank also determines the number of channels at higher stages of the model. HUTear has an interactive function **Make_GTbank**, which may be used to design coefficients for various types of gammatone filterbanks. The toolbox also contains an implementation of Irino's [4] level-dependent asymmetric compensation filterbank, which effectively makes the filterbank stage time-varying.

2.3 Models for mechanical-to-neural transduction

In the ear, the conversion of mechanical motion on the cochlea to neural activity in the auditory nerve is done by inner haircells (IHC), which are placed on the organ of *Corti* in the inner ear cochlea. This may be modeled using some functional model or by using a realistic electrochemical model for the transduction, e.g., Meddis's IHC synapse model [5; 6]. HUTear contains an implementation of Meddis's model and various tools that can be used to implement typical functional models for transduction.

2.4 Models for neural processing

The neural processing stage in hearing is usually modeled using some functional model, which has been designed to predict some specific effects occurring in hearing. HUTear contains three different temporal adaptation models which have been mainly developed to explain certain time-varying aspects of hearing.

2.5 Model for binaural interaction

The sound in different direction produces direction dependent interaural time and level differences (respectively ITD and ILD). The brain is capable to extract cues for sound direction from those. Models of binaural auditory system include typically some version of interaural cross correlation function which typically describes the interaural time difference (ITD). Input for the to the binaural model need pairs monaural channels, thus pair of monaural models of HUTear can be used as input for binaural model. Typical assumption is that similar monaural model is used for both ears. Difference between ears is then only in directivity which is modeled using HRTF-functions. HUTear contains an implementation of Lindemann's cross correlation function model [7; 8; 9]. Lindemann extended the pure cross correlation with contralateral inhibition and monaural detectors. With those also the influence of interaural level differences (ILD) can be included to the correlation model. The model includes also the dynamic inhibition component which models the precedence effect.

3 Structure of the toolbox

The HUTear toolkit is a modular package which contains a generic frame for an auditory model and a set of modules which can be hooked into it to produce a model for some particular application. The basic building blocks of such a model are shown in Fig. 1.

The name of the frame function is **AudMod**, which is used in the following way:

```
>> A=AudMod(signal,model);
```

where `signal` is a single row or column vector of signal samples and `model` is a *struct* variable which defines a model or a *string* which contains a file name of a text file where the fields of the struct variable are given. This text file is called *profile description file*. Currently, there are four ready-made profile description files in the toolbox each implementing some particular model from the literature. A user may design an own model profile or use some of the ready-made profiles. The current version contains ready-made profiles which are in line with the following models:

1. Sliding window model by [10]
2. A model based on a Püschels nonlinear adaptation network [11]
3. A model based on Karjalainen's nonlinear adaptation network [12]
4. A peripherical part of a pitch model based on Meddis's inner haircell synapse model [6]

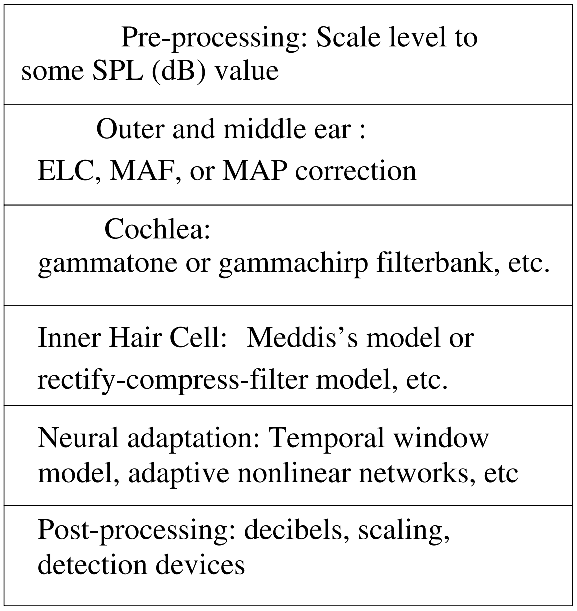

Figure 1: A block diagram of a generic auditory model

The output `A` is a matrix of output column vectors each of which represent an individual channel of the model. The number of output channels is specified in profile description file, in fact, it is the same as the number of channels in the filterbank. By using only one filter, the model is reduced to a single channel model.

4 Profile description

As mentioned above, an auditory model which is implemented using **AudMod** is defined by a **model** struct variable which is typically written in a profile description file, for examples, see [13]. In the following example we run one of the ready-made profiles to get a struct **model**.

```
>> Profile_Dau96
>> model

model =

          id: 'An auditory model by
               (Dau et al., 1996)'
          fs: 44100
          ds: 1
      outmid: [1x1 struct]
     cochlea: [1x1 struct]
    haircell: [1x1 struct]
      neural: [1x1 struct]
```

In its simplest form a **model** is an empty struct giving a model which relays the original signal unchanged. Above, the structure contains a hierarchical set of parameters. The variables model.fs, and model.ds set the sampling rate of the input signal and a number by which the sequence is sampled down in the model. The four following fields define the models for outer and middle ear, inner ear, neural transduction, and neural processing. It is also possible to include a fifth field into to the model which makes it possible to define various post-processing functions such as normalization or mapping to a decibel scale. Each field includes a more detailed description of that particular function. For example, in the previous model the cochlea is modeled using a 64 channel gammatone filterbank where channels range from 200 to 22050 Hz as can be seen in the following:

```
>> model.cochlea

ans =

    id: 'Gammatone filterbank'
    gt: [1x1 struct]

>> model.cochlea.gt

ans =

    design: 'gammatone'
       nch: 64
    frange: [200 22050]

>>
```

It is not a purpose of this paper to give a detailed description of the models that are contained in the toolbox and can be used in **model** structure because the toolbox is supplied with an accompanying manual. Basically, there are several alternatives for each of the fields. In addition, all the five main fields can also define an external function, e.g., by setting `model.cochlea.function="CochleaB"`, the model for inner ear is replaced by some new function called "CochleaB" and which should have the following form: `y=CochleaB(x)`. In fact, since this type of additional function can be defined at all the stages of the model, a user may also design a model which has nothing to do with auditory modeling.

5 Additional functions

The toolbox has a set of supporting functions. Those are mainly tools that can be used to design or customize filters that are used at various stages of a model. For example, **Make_GTbank** is an interactive function which can be used to design a gammatone filterbank which models cochlear processing and **Make_OEM** may be used to design a filter which approximates the transfer function from outer ear to inner ear.

Since typical models for mechanical-to-neural transduction and neural adaptation are level-dependent HUTear follows a widely used convention that a Root-Mean-Square value of 1 in a digital signal corresponds to the sound pressure level of 0 dB. In order to make it easier for users of the toolbox to work with various signals, the toolbox has a function, **Pascalize**, which scales an input signal so that its RMS corresponds to a given SPL (dB).

The psychoacoustic loudness measures give better approximation for perceived loudness than common desibel scale [14; 15]. As a tool for loudness estimation, HUTear provides Moore's revised version of Zwicker's loudness model [16]. The model can be used for calculation of loudness in *sones* and loudness level in *phones* from any input signal. It is the only truly separate function in the toolbox, but it was considered to be important. Source code of the model has been kindly provided by Prof. Brian C. Moore.

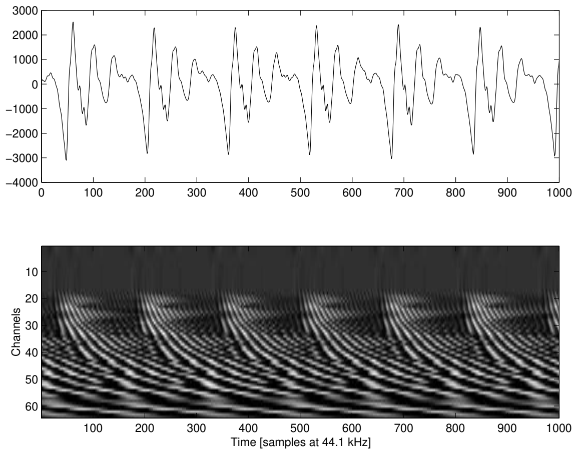

Figure 2: Top: A brief speech segment, finnish vowel /a/. Bottom: outputs of a model profile `Profile_Meddis91`. The rows of the image show simulated neural activity patterns in 64 channels of an auditory model.

6 Example

Figure 2 illustrates the usage of an auditory model. The input signal in the upper part of the figure is a brief speech segment which has been scaled so that its RMS value corresponds sound pressure level of 60 dB. The lower image is the output of **A=AudMod(signal,'Profile_Meddis91')**, where the name a profile description file implementing the peripheral part of a model presented in [6] is given as a string. Individual rows of the image represent neural activity in the auditory nerve at different center frequencies.

7 Efficient implementation

Most of the elements of the models have been implemented as MEX-functions (C-language). All the stages following the model for outer and middle ear are multichannel models and hence the MEX-functions have been implemented as such that they take in and output a matrix of sequences representing individual cochlear channels. This leads to a significantly more efficient implementation than, e.g., implementation of a filterbank as a loop in a script-function.

8 Conclusions

A Matlab toolbox for auditory modeling was introduced in this paper. The toolbox is modular and flexible for various types of tasks.

9 Acknowledgments

The authors are grateful to Brian C. Moore, Matti Karjalainen, Toshio Irino, and Malcolm Slaney for permission to use some of their code fragments and functions in the toolbox.

References

[1] A. Härmä, "On the utilization of overshoot effects in low-delay audio coding," in *Proc of ICASSP'99*, vol. II, (Phoenix, Arizona), pp. 893–896, IEEE, March 1999.

[2] R. Patterson, K. Robinson, J. Holdsworth, D. M. andC Zhang, and M. H. Allerhand, "Complex sounds and auditory images," in *Auditory Physiology and Perception* (Y. Cazals, L. Demany, and K. Horner, eds.), pp. 429–446, Oxford: Pergamon, 1992.

[3] C. Giguere and P. C. Woodland, "A computational model of the auditory periphery for speech and hearing research. i. ascending path," *J. Acoust. Soc. Am.*, vol. 95, no. 1, pp. 331–349, 1994.

[4] T. Irino, "A time-varying, analysis/synthesis auditory filterbank using the gammachirp," in *Proc. ICASSP'98*, (Seattle), pp. 3653–3656, IEEE, 1998.

[5] R. Meddis, "Simulation of auditory-neural transduction: Further studies," vol. 83, pp. 1056–1064, 1988.

[6] R. Meddis and M. J. Hewitt, "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I:pitch identification," *J. Acoust. Soc. Am.*, vol. 89, pp. 2866–2881, June 1990.

[7] W. Lindemann, "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," *J. Acoust. Soc. Am.*, vol. 80, no. 6, pp. 1608–1622, 1986.

[8] W. Lindemann, "Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front," *J. Acoust. Soc. Am.*, vol. 80, no. 6, pp. 1623–1630, 1986.

[9] W. Gaik, "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 98–110, 1993.

[10] C. J. Plack and A. J. Oxenham, "Basilar-membrane nonlinearity and the growth of forward masking," *J. Acoust. Soc. Am.*, vol. 103, pp. 1598–1608, March 1998.

[11] T. Dau and D. Püschel, "A quantitative model of the effective signal processing in the auditory system. I. model structure," *J. Acoust. Soc. Am.*, vol. 99, pp. 3615–3622, July 1996.

[12] M. Karjalainen, "A binaural auditory model for sound quality measurements and spatial hearing studies," in *Pro. ICASSP'96*, vol. 2, (Atlanta), pp. 985–988, IEEE, May 1996.

[13] A. Härmä and K. Palomäki, "Hutear – matlab toolbox for auditory modeling." http://www.acoustics.hut.fi/software/HUTear, October 1999.

[14] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Berlin: Springer-Verlag, 1990.

[15] B. C. J. Moore, *Introduction to the psychology of hearing*. Academic Press, 4th edition ed., 1997.

[16] B. C. J. Moore and B. R. Glasberg, "A revision of Zwickers loudness model," *Acustica-Acta Acustica*, vol. 82, pp. 335–345, 1994.

Address for correspondence

Aki Härmä
Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, P.O. Box 3000, FIN-02015, Espoo, Finland,
Aki.Harma@hut.fi