Modeling of Psychoacoustics and Auditory Perception: How Far Can We Go?

Matti Karjalainen

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

This presentation is available in:

www.acoustics.hut.fi/~mak/ICADkeynote.pdf

For refs, see also: www.acoustics.hut.fi/publications

CONTENTS

- Introduction and motivation
- Psychoacoustics and auditory functions
- Computational auditory modeling
- Modeling of basic psychoacoustic features
- Binaural modeling and spatial hearing
- Advanced modeling and CASA
- Applications of auditory modeling
- What cannot be done (yet or ever)?
- Future challenges
MOTIVATION

• Computational modeling of how we hear and perceive sounds is a target of active research because:
  – It adds to our knowledge on human auditory functions by enabling the testing of theories on complex auditory functions
  – It is a key enabling approach to new applications in audio, speech, and multimedia
  – Dream of advanced ”artificial listener”

FRAMEWORKS for MODELING Communication by Sound and Voice

• Sound source modeling
  – physical modeling

• Signal modeling
  – channel modeling and DSP

• Listener modeling
  – auditory modeling
HEARING RESEARCH

• Physiology of hearing
  – Physically/chemically measurable (objective) properties of hearing
• Psychoacoustics (auditory psychophysics)
  – Subjective responses to objective stimuli
• Cognitive properties of hearing
  – High-level functional properties of hearing

PERIPHERAL vs. CENTRAL HEARING

• Auditory periphery
  – Relatively well known
  – Basic properties easy to model
• Central auditory system
  – Relatively weakly known
  – Difficult to model
ICAD’01

Psychoacoustics Concepts: MASKING

- Frequency masking
  - Spreading of masking effect of a sound component in the frequency domain
  - More prominent upwards in frequency

- Temporal masking
  - Spreading of masking in the time domain
  - More prominent forward in time (lasts up to 200 ms)

ICAD’01

Psychoacoustics Concepts: PITCH

- Pitch = subjective sensation of sound on low/high scale
  - Relation to frequency not linear, rather logarithmic-like
  - ERB (Equivalent Rectangular Bandwidth) scale theoretically best motivated
  - Bark scale technically most frequently used
Psychoacoustics Concepts: LOUDNESS

- **Loudness**
  - Unit: 1 sone
  - 1 sone = 40 dB sine at 1 kHz
- **Loudness level**
  - Logarithmic measure
  - Unit: 1 phone
- **Wideband loudness**
  - Each critical band contributes equally to total loudness

**COMPUTATIONAL AUDITORY MODELING**

- **Physiological models**
  - Goal: accurate simulation of physiological details
- **Psychoacoustical (perceptual) models**
  - Modeling of results of psychoacoustical facts and theories
- **Functional (hypothetical) higher level models**
  - Any potentially useful algorithms simulating auditory functions
- **Cognitive models**
  - Modeling cognitive processes in auditory perception
- **Simplified models for technical applications**
  - Mixtures of auditory, signal, and source modeling
AUDITORY SPECTRUM THROUGH FFT

- A typical frequency-domain auditory spectrum computation principle
  - Frequency domain properties (particularly steady-state loudness) can fairly easily be modeled accurately
  - Zwicker’s model
  - Moore’s model
  - Problem: temporal properties not easily modeled accurately

EXAMPLES OF SIMPLE AUDITORY SPECTRA

- Sine wave (a)
  - Shows the spreading of excitation pattern for sine wave of 400 Hz
- White noise (b)
  - Shows the frequency sensitivity curve of the auditory system
FILTERBANK-BASED AUDITORY MODELS

- Filterbank models of peripheral hearing
  - Bandpass filters simulate the frequency selectivity (critical bands) of the inner ear (basilar membrane and hair cells)
  - Half-wave rectification in each channel by hair cell neural firings
  - Neural firings (statistically) synchronized up to about 1-3 kHz
  - Adaptation of firing rate after onset
  - Temporal integration (lowpass filtering) in loudness formation and temporal masking

GAMMATONE FILTERBANKS

Temporal and magnitude response of a filterbank channel

\[ g(t) = at^{n-1}e^{-2\pi f_c t} \cos(2\pi f_c t + \phi) \]
NEURAL ADAPTATION

Neural adaptation model by Dau et al
Automatic gain control feedbacks used

EXAMPLE OF TEMPORAL PROCESSOR

• Neural adaptation, temporal integration, and temporal masking model (Karjalainen 1996):
  – Neural feedback model
  – Adaptation (AGC) in firing rate simulation
  – Loudness (level) computation
  – Temporal masking effect
  – Temporal integration and firing rate adaptation shown to be complementary functions
PERIODICITY (PITCH) ANALYSIS

- Periodicity analysis model (Meddis)
  - Bandpass filterbank + half-wave rectification + lowpass
  - Periodicity analysis in each critical band by autocorrelation
  - Summary autocorrelation function (SACF) a good periodicity indicator in many cases
  - Valid only at low frequencies (< 1kHz)

MULTIPITCH ANALYSIS

- Enhanced SACF in two channels (Tolonen & Karjalainen 2000)
  - Only two channels used in SACF computation (below and above 1 kHz)
  - Enhanced SACF (= ESACF) for better resolving multiple pitches by removing periodic and negative peaks in autocorrelation function
  - Can resolve up to 3-5 simultaneous harmonic sounds
  - ESACF example mixture of 3 harmonic sounds (a musical chord):
ICAD’01

SOURCE SEPARATION (example)

• Separation of simultaneous vowels (Karjalainen & Tolonen 2000)
  – Multipitch analysis applied to find pitch values of two or more vowels
  – Iterative technique can be used to improve multipitch analysis
  • Pitches are removed one by one from the mixture
  – Separated vowel spectra estimated (for 2 vowels)

ICAD’01

REVERBERATION VS. AUDITORY MODELING

Quality of late reverberation vs. modal density (Karjalainen & Järveläinen 2001)

• How many modes per critical band needed for perfect reverberation (or random noise)?

• How the auditory system resolves or analyzes a mixture of modes?

• Any perceivable periodicity higher in level than about -30 dB in a critical band envelope signal may degrade or color reverberation
Applications: AUDIO CODING

- Audio coding (such as MPEG-nn) is
  - probably the most important application so far utilizing auditory (perceptual) models
  - based on coding only auditorily relevant and non-redundant information of audio signals: data rate reduction about 10:1

Applications: SOUND QUALITY ESTIMATION

- Objective sound quality measures to correspond subjective perception must be based on auditory (perceptual) models
- Perceptual sound quality models can be applied to many problems, (although quite differently):
  - Quality of audio and speech systems
  - Sound quality of performing spaces, musical instruments, etc.
  - Noise quality (annoyance, disturbance), product sound quality
- Example: auditory principle for audio sound quality (Karjalainen 1981-85)
BINAURAL AUDITORY MODELING

- In binaural (two-ear) listening, modeling of:
  - Sound source direction
  - Sound source distance
  - Binaural loudness, timbre, etc.
- Perceived direction is based primarily on:
  - Interaural time difference (ITD)
  - Interaural level difference (ILD)
  - Spectral cues
- Binaural auditory modeling less developed than modeling of monaural properties, although basic ideas have existed since 1940’s (Jeffress)

BINAURAL AUDITORY MODEL (Example)

- Computational modeling of binaural hearing has turned out to be successful, e.g., in amplitude-panned sound reproduction (Pulkki et al)
- ITD estimated from auditory interaural cross-correlations
- ILD estimated left-right ear signal levels in critical bands
- ITD and ILD features can be combined to perceived direction by various ways (table lookup, neural nets)
- These simple models are successful as far as the precedence effect does not have to be taken into account
PRECEDENCE EFFECT MODELING

- Precedence effect
  - Repetitions of a sound within 1 – 30 ms after initial version don’t affect perceived source direction (dominance of first wavefront)
  - No essential precedence effect found for example for timbre
  - Precedence effect is probably a mixture of high- and low-level processes

Model proposed by Zurek

AUDITORY ANALYSIS OF SOUND FIELD

- Auditory modeling applied to room and concert hall responses (Lokki & Karjalainen 2000)
  - Auditory spectrogram (time-frequency plot) of room impulse response
  - Directional information, such as lateral or front-back 'spectrogram'
FREQUENCY-WARPED DSP vs. AUDITORY MODELING

- We have shown that frequency-warped DSP techniques are often an efficient and straightforward way of including basic auditory features in traditional signal processing algorithms (Härmä et al 2000).
- Frequency warped techniques are based on replacing unit delays by first-order allpass filters.
- By proper parameters, warping makes a very good match to the Bark scale (linear Hz scale mapped to Bark scale, Smith & Abel).

AUDITORY SCENE ANALYSIS AND CASA

- ASA attempts to explain the ability of the auditory system to organize the incoming sound into separate sound objects (streams, events) and their interrelationships.
- CASA (Computational ASA) has been developed for about 10 years based on ASA findings (Cooke, Ellis, etc.).
- Many challenging technical problems are more or less CASA type of problems:
  - Automatic transcription of music
  - Analysis of ambient and environmental sounds
  - Speech recognition in complex noisy environments
  - Audio content analysis
WHAT NEXT?

- Development of useful models for difficult phenomena
  - Precedence effect, separation of more complex sounds, etc.
- More accurate time-frequency modeling
  - Dynamic loudness, pre- and postmasking, level-dependent masking
- Modeling of timbre perception
  - Categorization of timbres
- Modeling specific phenomena in music perception
  - Consonance, dissonance, rhythm, sound textures
- Much improved binaural auditory modeling needed
  - Perceptual and cognitive modeling of sound environments (rooms, concert halls, etc.) including reverberant phenomena
- Integration of existing models into large-scale models
  - Needed for complex applications and for better overall understanding of auditory functions
- Much work needed for high-level and cognitive modeling
- Improved computational auditory scene analysis (CASA)
- Etc, etc ....

HOW FAR CAN WE GO?

**Personal opinion:**
In principle everything can be modeled computationally –
(How about in practice?)

**Time schedule of progress:**
The most difficult (sub)problems take at least tens of years
to solve even tentatively (cf., speech recognition)

**Big challenge:**
How to achieve much improved automatic self-learning and
organization principles?