Current Directions in Audio Signal Processing Research at TKK

Matti Karjalainen

TKK - Helsinki University of Technology
Department of Signal Processing and Acoustics
(Espoo, Finland)

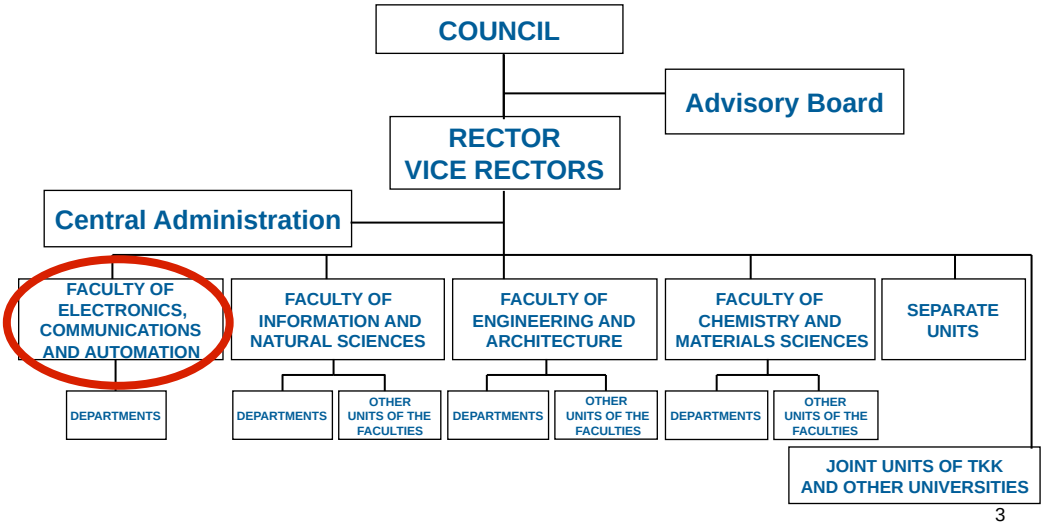
Helsinki University of Technology - TKK

- Located in the city of Espoo
  - About 10 km from the center of Helsinki

www.tkk.fi

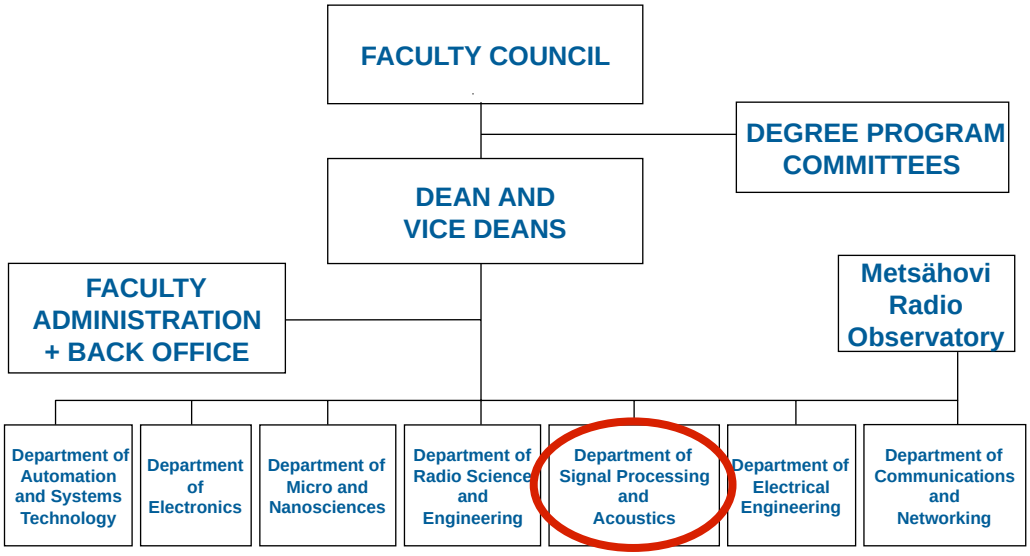
- To become the **Aalto University** in 2009!

TKK ORGANIZATION



3

HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Signal Processing and Acoustics
DAFx-08 Tutorial 2, Sept. 1, 2008 / *MK*



4

TKK Department of Signal Processing and Acoustics

- The only university unit in Finland that offers teaching and conducts research primarily in
  - Acoustics, audio, speech processing
  - Also communication systems, measurement technology, and optical technology
- The best facilities in Finland
  - 3 anechoic chambers
  - Listening test room
- Personnel about 90
  - 8 professors
  - Prof. Jorma Skyttä, Head of Dept.

www.acoustics.hut.fi

5

HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Signal Processing and Acoustics
DAFx-08 Tutorial 2, Sept. 1, 2008 / *MK*

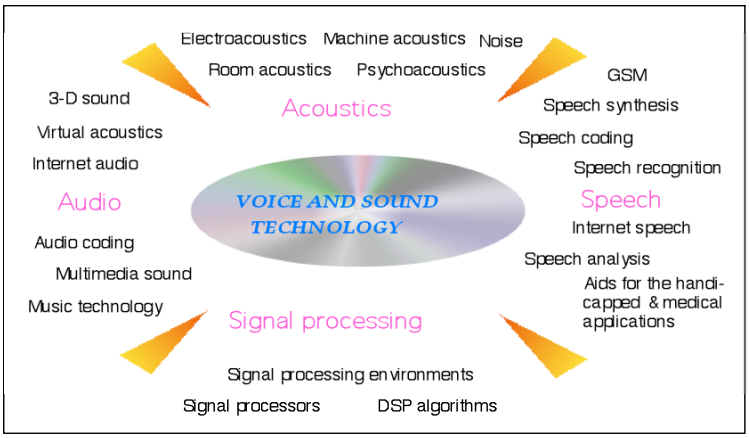
Research Teams in Acoustics at TKK

1. **Communication Acoustics** (Prof. Matti Karjalainen)
   Including Dr. Ville Pulkki and his group
2. **Speech Analysis** (Prof. Paavo Alku)
3. **Speech Technology** (Prof. Unto K. Laine)
4. **Audio Signal Processing** (Prof. Vesa Välimäki)
   Including Dr. Cumhur Erkut and his group

Strong collaboration with Department of Media Technology:
Lauri Savioja, Tapio Lokki, Tapio Takala & others

6

Research Domains in Acoustics



Research topics in Communication AcousticsTeam:

- Spatial sound (Ville Pulkki's group)
  - perception, reproduction and, coding
- Evaluation of sound quality
  - perceptual studies and auditory modeling
- Augmented reality audio (ARA)
  - presented later in this talk
- DSP for loudspeakers and headphones
  - response equalization by DSP
- Physical modeling techniques
  - together with Vesa Välimäki's group

Prof. Matti Karjalainen

Dr. Ville Pulkki

Spatial sound (Ville Pulkki's group)

- VBAP (Vector Base Amplitude Panning), SIRR (Spatial Impulse Response Rendering), DiRAC (Directional Audio Coding), perception of spatial sound, auditory modeling of spatial sound perception, etc.
- See Ville's tutorial and demo today!

9

Physical modeling paradigms (Matti K)

- **Hybrid modeling using**:
  - Mixed, wave-based (W-modeling) and Kirchhoff-based (K-modeling)
  - *Digital Waveguides* (DWG), W-modeling
  - *Wave Digital Filters* (WDF), W-modeling
  - *Finite Difference Time-Domain* (FDTD), K-modeling
  - *Modal decomposition* techniques (K or W-modeling)
  - *Source-filter modeling*
  - *K/W-converters*
- **BlockCompiler** http://www.acoustics.hut.fi/software/BlockCompiler/
  - Software platform for multi-paradigm physics-based modeling
  - Model builder, scheduler, code generator, compiler, realtime synthesis

10

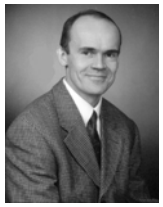
Example: Guitar modeling

- **Accurate modeling of guitar bridge admittance**:
  - Measurement of guitar bridge and string behavior
  - DWG and WDF modeling for realtime synthesis
  - Parallel (mode-based) admittance modeling
    - by DWG or consolidated WDF block
  - Six DWG string models
  - For details, see:
    - *Balazs Bank* and *Matti Karjalainen*: "Passive admittance synthesis for sound synthesis applications," Acoustics'08, Paris, 2008
    - *Matti Karjalainen*: "Efficient Realization of Wave Digital Components for Physical Modeling and Sound Synthesis," IEEE Trans. ASLP, 2008 vol. 16 (5) pp. 947 – 956
  - **SOUND SYNTHESIS DEMO IN BLOCKCOMPILER**

HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Signal Processing and Acoustics
DAFx-08 Tutorial 2, Sept. 1, 2008 / *MK*

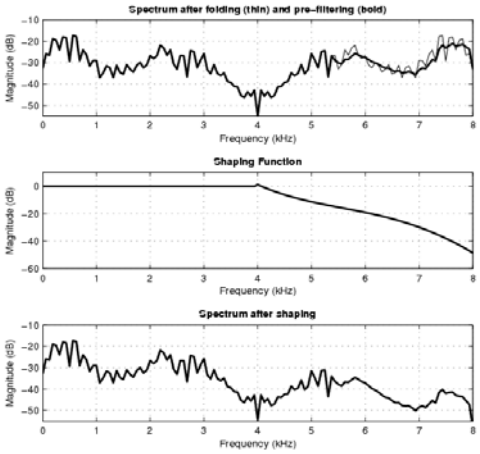
Research topics in Speech Analysis Team:

- Speech enhancement
- HMM-based speech synthesis
- Robust spectral modelling of speech
- Brain functions of speech and 3D sounds
- Voice production modelling
  - analysis of vocal emotions
  - occupational voice care



Prof. Paavo Alku

Artificial Bandwidth Extension (ABE) of speech signals



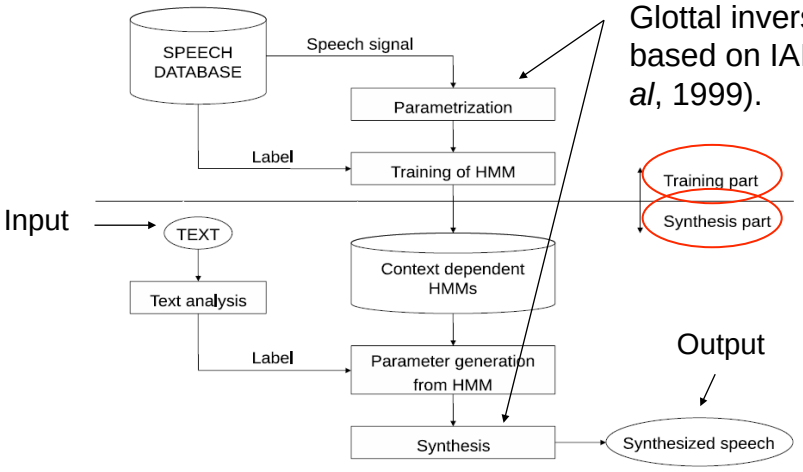
Improve the quality of narrowband speech in speech transmission

Research is conducted together with Nokia: the ABE method developed at TKK was implemented in spring 2007 in Nokia's 3G phones (e.g., model 7390)

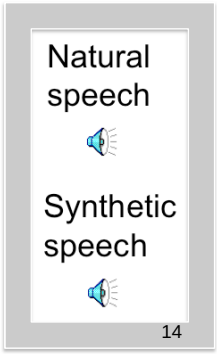
13

HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Signal Processing and Acoustics
DAFx-08 Tutorial 2, Sept. 1, 2008 / *MK*

Hidden Markov Model (HMM) based speech synthesis utilizing glottal inverse filtering



Glottal inverse filtering based on IAIF (Alku *et. al*, 1999).



14

Speech Technology Team:

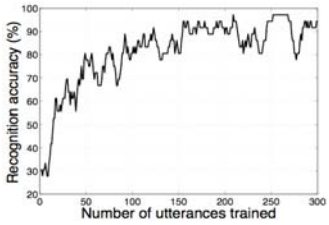
- Speech analysis and recognition
- EU project:

Acquisition of Communication and Recognition Skills (ACORNS)

➢**Prof. Unto K. Laine**

- Development, implementation and testing of computational models capable of acquiring human-like verbal communication behavior, i.e., ***simulation of infant language acquisition***

- ***Learning of emergent dynamic patterns*** of speech and non-speech stimuli from multimodal input through interaction between the learning agent and a caregiver.



15

Research topics in Audio Signal Processing Team:

- Model-based sound synthesis and acoustics of musical instruments
  - Piano, guitar, clavichord, harpsichord, guqin, kantele, tanbur, ud, …
  - Virtual analog synthesis and audio effects electronic instruments, amplifiers …
- Signal processing
  - Digital filters e.g. for sound synthesis
- User interaction and sound design

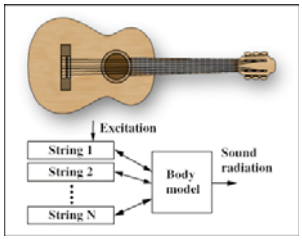
➢**Prof. Vesa Välimäki**

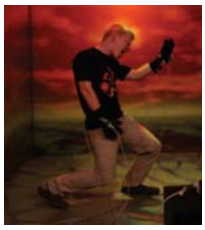
➢**Dr. Cumhur Erkut**

16

Sound Synthesis and Physical Modeling

- Physical and mathematical modeling of musical instruments
- Simplification of models and signal models
- Real-time simulation
- Computationally efficient synthesis algorithms
- User interfaces





- Synthesis, Control, and Hierarchical Modeling Algorithms for Sonic Interaction Design (SID)
- Tools for interaction-centered sound design
  - Scenes with a multitude of sounding objects
  - Physical sound synthesis blocks
  - Block and user interaction managed similarly
  - Parametric control models in different time scales

➢Don't miss the SID presentation on Wednesday !

Augmented Reality for Audio Applications

- Goals / Background
- Real vs. virtual vs. pseudo-acoustic vs. augmented reality audio
- Headsets, terminals, and systems
- User positioning and environment scene analysis
- Applications: Case examples

Future of your personal audio communications ?

Concepts / goals / background

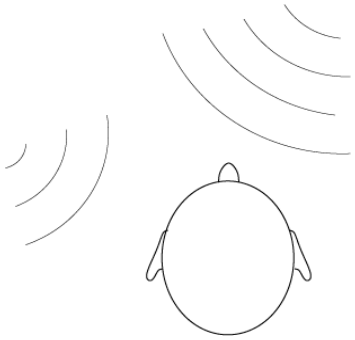
Mobile Augmented Reality Audio = MARA

Wearable Augmented Reality Audio = WARA

- *Wearable systems with wireless communication*
- *Maximal flexibility, to be used any time, anywhere*
- *Highly personalized, tiny earplugs for user interface*
- *Support for all audio and voice communication plus support for hearing protection and aiding*
- *User positioning and tracking is often needed*
- *Environment aware techniques (scene analysis)*
- *Application scenarios are rich*

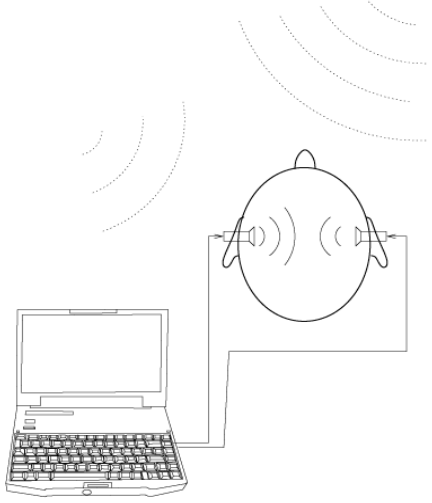
Concepts 1: Real acoustic environment

- **Listener in a *real* acoustic world**:
- *Maximally natural perception*
- *Limited distant communication*
- *No virtual effects*



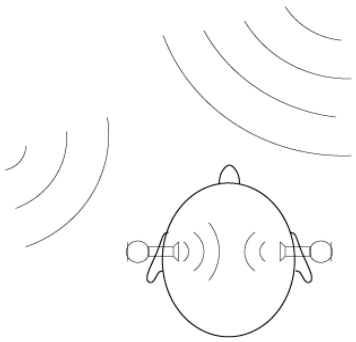
Concepts 2: Virtual acoustic environment

- **Listener in a *virtual* audio world**:
- *Virtual auditory objects perceived*
- *Real acoustic environment may be attenuated or absent*
- *Only virtual effects*



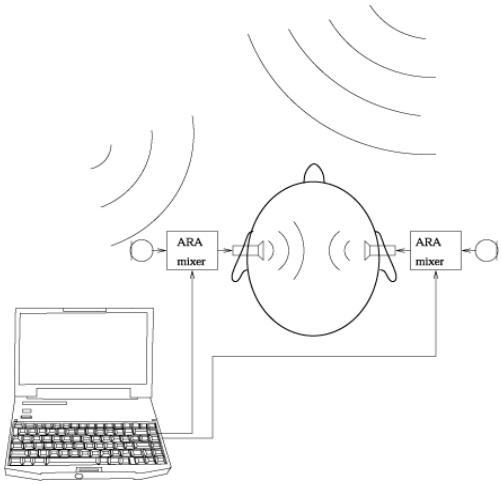
Concepts 3: Pseudo-acoustic environment

- ***Pseudo-acoustic*** **representation of the real acoustic environment**
- *Binaural microphone-earphone*
- *Like binaural hearing aids*
- *The real acoustic environment around the user is heard and even amplified/enhanced*



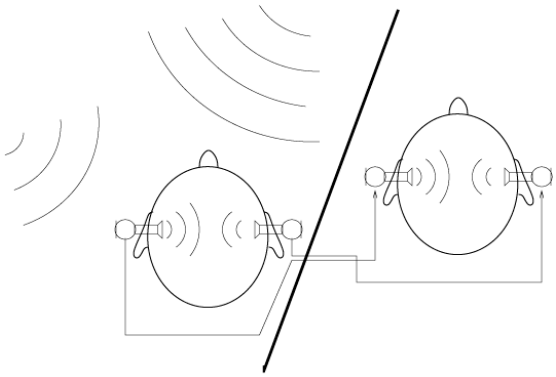
Concepts 4: Augmented acoustic environment

- ***Augmented*** **reality audio (ARA) system**
- *Binaural microphone-earphone*
- *User can hear the pseudo-acoustic environment*
- *Virtual sounds superimposed onto the pseudo-acoustic environment*



Example: Binaural telepresence

- ***Telepresence*** **audio system**
- *Two-way binaural communication between two (ore more) persons*
- *Binaural telephony: merging of speech transmission and binaural technology*

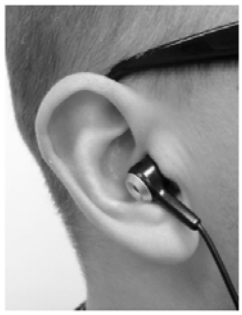


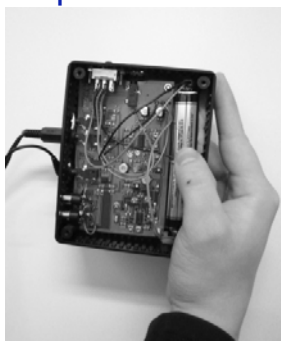
Headset design for ARA systems

- Use of fixed position loudspeakers
  - *Not mobile or wearable*
- Shoulder-top loudspeakers
  - *Discomfort ? Privacy ? Feedback problem*
- Open headphone models:
  - *Feedback problem*, *Privacy?*
- Bone conduction etc. special headphones
  - *Ok for special applications*
- Earplug type of headphones (closed ear canal):
  - *Discomfort (by some users) ?*

Experimental hear-through headset system

Microphone + earphone
Mixer + equalizer





Problems: 1) Leakage, 2) Equalization

27

Headset properties and problems

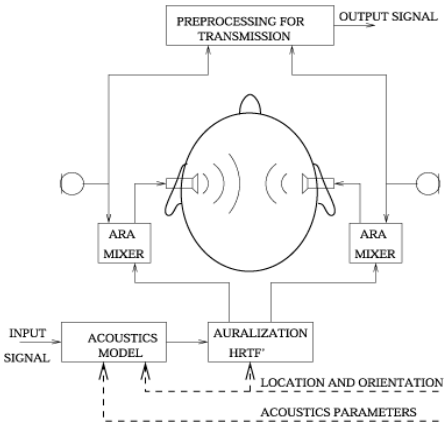
- Discomfort of wearing (ventilation, wire hiss)
- Occluded ear canal problems (own voice)
- Microphone noise & directivity
- Wiring / wireless connection to terminal device
- Earplug leakage (problem in equalization)
- Pseudoacoustics equalization (latency << 1 ms)
- Hearing aid functions possible
- Hearing protection functions possible
- Acoustic positioning can be integrated

28

MARA system diagram

Augmented Reality Audio (ARA) mixer:

- The MARA key component
- Mix between virtual and pseudo-acoustic environments

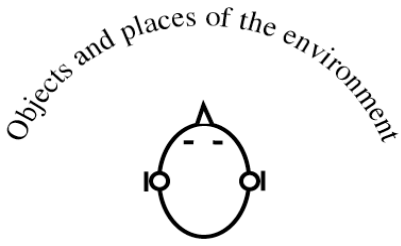


(*+ positioning and tracking & environment scene analysis*)

29

Tracking and positioning of a user

- Need of position-aware applications and services
- User tracking is so far a bottleneck in MARA applications
- GPS doesn't work inside buildings; head position and orientation needed
- Inexpensive positioning techniques to be developed
- Acoustic tracking ?



1) What is my position in a room or outside ?
2) Which is the direction I am looking to ?

30

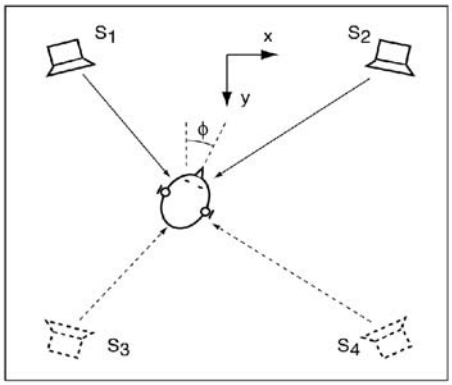
Acoustic tracking and positioning

➢ We have studied acoustic positioning techniques:

- User position and orientation is tracked by anchor sounds (e.g. non-audible 20 kHz pulses)

- Position estimated based on delays of anchor sounds to headset microphones

- Estimation of orientation (face direction) by interaural time differences (difficult)



31

About MARA applications

- ➢ Hands-free, eyes-free, anytime, anywhere !
- ➢ Human-to-human vs. human-to-machine
  - speech communications or information services
- ➢ Recorded vs. synthesized vs. real-time
  - messages and real-time conversation
- ➢ Localized or freely-floating virtual sources
  - sources rendered to objects in the environment
  - sources rendered relative to user's head
- ➢ Context- and location-aware services

32

Acoustic Post-It

- *Localized audio message service*
  - User can leave or listen to audio messages like post-it stickers
- *Two approaches:*
  - Message rendered to a physical object (localized)
  - Message activates when user within a *range* of a message (freely-floating)
- *Personalized or public messages*
  - Warnings / general info / acoustic graffities
  - Selection profiles for receiving subjects

33

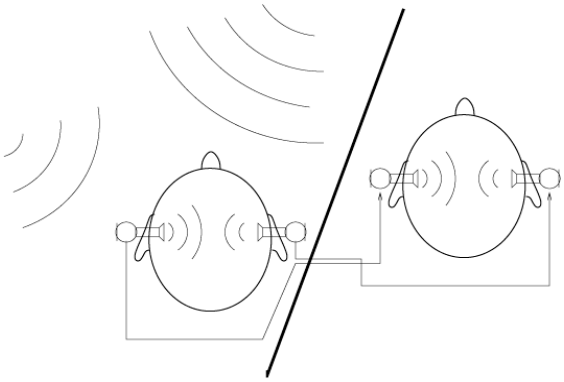
Audio memo & binaural recording

- *Recording of memo messages*
  - instead of making written notes & diaries
- *Contact management, e.g. memorizing names and events:*
  - Recalling persons name after being introduced briefly (Many people have difficulties in this!)
  - Storing important spoken messages or sound events
- *Recording interesting sound environments*
  - Like photographing but by storing sound scenes (audiography?)

34

Binaural telephony

- Augmentation of a pseudo-acoustic environment with another pseudo-acoustic environment
- Normally the remote speaker(s) need to be rendered (externalized) around the listener

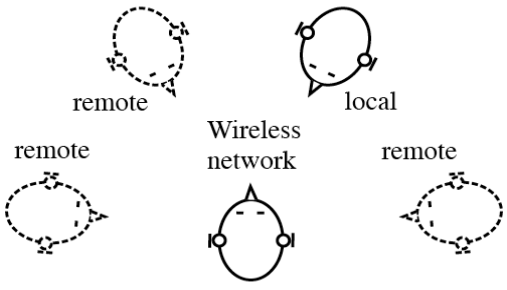


- Speech recognition and synthesis for man-machine communication

35

Audio meetings

- Easy creation of discussion groups consisting of remote and local participants
- Attending remote meetings
- Other speakers should be positioned around (different directions) and fused to similar acoustics (reverberation)



Telepresense audio meeting
(multiuser binaural telephony)

36

MARA Virtual band playing

- Possibly over Internet
  Problem: network latency

Hand position
sensors an MARA
headsets needed



MARA concerts, disco, theater & games

- Telepresence concerts & events
  - Legal issues involved (e.g., bootlegging)
- Virtual crowds
- Individual sound reproduction in a concert
  - Sound level, timbre and acoustics control
- Increased interactivity possible
- What about audio-only games ?
  - Orientation and control games (car driving by sounds)
  - Fight in the dark, Audio-Tetris, Audio-Puzzles

Virtual audio tourist guide

- Automatic spoken / audio information of touristic objects, monuments, inside museums, etc.
- Looking at an object and command/gesture to start a story
- Destination and path finding guidance (e.g., public transport)
- ”X-City today” info (cultural events of the city)
- GPS orientation by ”auditory display”

39

Audio-guided shopping

- Getting product information and pricing when approaching an interesting product or looking at it (in windows, stores)
- Advertisements from nearby shops, if enabled
- Audio instructions of use for products

40

Audio scenery and environments

- Nature sound experiences
  - Forrest & birds, seeshore & water sounds, wind & rain
- Virtual audio crowds & urban sounds
- Creative artificial (synthetic) sound scenes
- Noise canceling of real environment helps in experiencing virtual environments

Thank you for your attention!

Thanks are due to TKK Acoustics and Multimedia people