Spatial audio

Matti Gröhn

Matti.Grohn@ttl.fi

Outline

- Spatial hearing
- Reproduction
  - Headphones
  - Loudspeakers
- Applications
- Experiments in localization, navigation and orientation in virtual environment
- Subjective comments and discussion on design issues

Matti.Grohn@ttl.fi

Spatial hearing

- Sound source can be associated with a direction and distance.
- Most of the sound sources have a definite spatial location
- Hearing has no directional limitations

Matti.Grohn@ttl.fi

Spatial hearing

- Only information used in spatial hearing is the sound pressures in ear canals.
- Listeners use different type of cues: spectral content of ear canal signals and level, temporal or spectral differences in ear canal signals

Matti.Grohn@ttl.fi

Binaural cues

- There are two main binaural cues
  - Interaural time difference (ITD)
  - Interaural level difference (ILD)
- ITD is the primary cue in frequencies below 1.5 kHz and ILD is the primary cue above that threshold.

Matti.Grohn@ttl.fi

Monaural spectral cues

- Monaural spectral cues are important in elevation perception.
- The head, torso and shape of the pinna affects spectral cues.

Matti.Grohn@ttl.fi

Head-related transfer functions

- Monaural and binaural cues can be represented as free field transfer functions from a sound source to each ear canal.
- These functions are called head-related transfer functions (HRTFs).

Matti.Grohn@ttl.fi

Precedence effect

- The precedence effect is a mechanism, that helps in the reverberant room to localize sound sources.
- It ensures that localization is based on direct sound and reflections are not able to affect localization.

Matti.Grohn@ttl.fi

Cone of confusion

- Cone of confusion is defined as the set of all possible sound source locations, with the same time difference between the ear canal signals.
- Front-back confusions are common in static localization experiments

Matti.Grohn@ttl.fi

Head movements

- Tilting of the head changes the monaural spectral cues
- Rotating of the head changes the binaural cues
- Head movements decrease the effect of the cone of confusion and improve the localization accuracy.

Matti.Grohn@ttl.fi

Accuracy

- Localization (in front of the listener)
  - Azimuth: 1-4 degrees depending on signal
  - Elevation: 4-17 degrees depending on signal
- Distance
  - works relatively well only for sounds near the listener

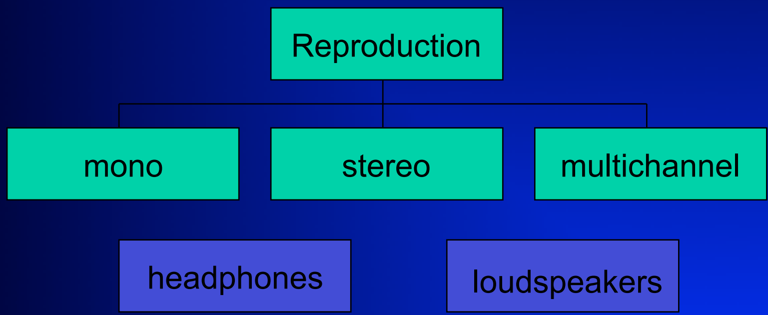
Matti.Grohn@ttl.fi

Spatial sound reproduction

- Sound source could be produced in any location around the listener
- Sound sources are divided in two types
  - Real source is a a physical sound source like a loudspeaker or a musical instrument
  - Virtual source denotes an auditory object that is perceived in location that does not correspond to any physical sound source.

Matti.Grohn@ttl.fi

Spatial sound reproduction



Matti.Grohn@ttl.fi

Headphone reproduction

- Sound signal is reproduced directly to ears
- Perceived azimuth angle can be changed manipulating the ITD and/or ILD values
- Effective spatialization is achieved by using HRTFs

Matti.Grohn@ttl.fi

HRTFs

- HRTFs are divided in individualized and non-individualized HRTFs
- Invidualized HRTFs are measured individually for each listener (time consuming task)
- Non-individualized HRTFs can be based on measurements of the other person or the dummy head.
- Typically individualized HRTFs are more accurate than non-individualized HRTFs

Matti.Grohn@ttl.fi

HRTFs

- In headphone reproduction a sound signal can be positioned in any direction if HRTFs for both ears are available
- If listener's head or the sound source moves, then this movement should be taken account in processing.
- In a dynamic environment, a head-tracking and real time processing of HRTFs are needed for accurate reproduction.

Matti.Grohn@ttl.fi

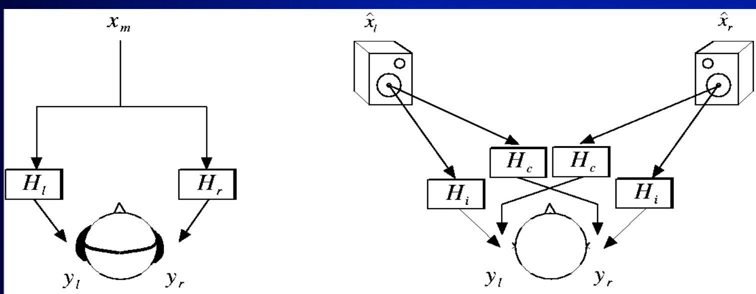
Loudspeaker reproduction

- A variable number of loudspeakers can be used for 3D spatial audio
- Theoretically 2 loudspeakers is enough
- Most common surround system is 5.1
- In addition, there are three commonly applied multi-loudspeaker methods: wavefield synthesis, Ambisonics and vector-based amplitude panning

Matti.Grohn@ttl.fi

2 loudspeakers system

- HRTF processing and cross talk cancellation are both needed in this solution.



Matti.Grohn@ttl.fi

2 loudspeakers system

- Two main limitations
  - The best listening area (sweet spot) is very limited
  - It is critical to listening room conditions and the full spatial information can be retained only in anechoic listening rooms

Matti.Grohn@ttl.fi

5.1 system

- Most common home-theater system
- Standardized method to reproduce surround sounds provided in movie soundtracks
- It does not reproduce 3D spatial audio

Matti.Grohn@ttl.fi

Wavefield synthesis (WFS)

- Reconstructs a whole sound field.
- The WFS produces the sound field accurately only if the loudspeakers are at a maximum distance of a half wavelength from each other.
- In the horizontal plane using a 100 loudspeaker wave field can produced spatially accurately for frequencies up to 1000 Hz.
- Accurate 3D wave field would need hundreds of loudspeakers -> not practical for 3D spatial audio

Matti.Grohn@ttl.fi

Ambisonics

- In Ambisonics a sound signal is applied to all loudspeakers with different gains.
- For optimal result, the loudspeakers should be in symmetric layout.
- The directional accuracy degrades radically outside the sweet spot

Matti.Grohn@ttl.fi

Vector-based amplitude panning (VBAP)

- In 3D sound reproduction the virtual source is reproduced using the three closest loudspeakers at a time.
- The directional accuracy degrade much less outside the sweet spot than with Ambisonics
- VBAP enables arbitrary positioning of loudspeakers.
- VBAP is a computationally efficient method

Matti.Grohn@ttl.fi

Headphones vs. loudspeakers

- Headphones
  - HRTFs needed for 3D spatial audio
  - Decrease the effect of background noise
  - Suitable for mobile systems and Head-Mounted-Displays
- Loudspeakers
  - Suitable for multi-user environments
  - Tracking of the user is not necessary
  - Work well with projection displays

Matti.Grohn@ttl.fi

Applications

- Flight simulators
- Computer aided design process
- Architecture
- VR-, console and computer games
- Scientific data representation
- Artistic installations
- etc.

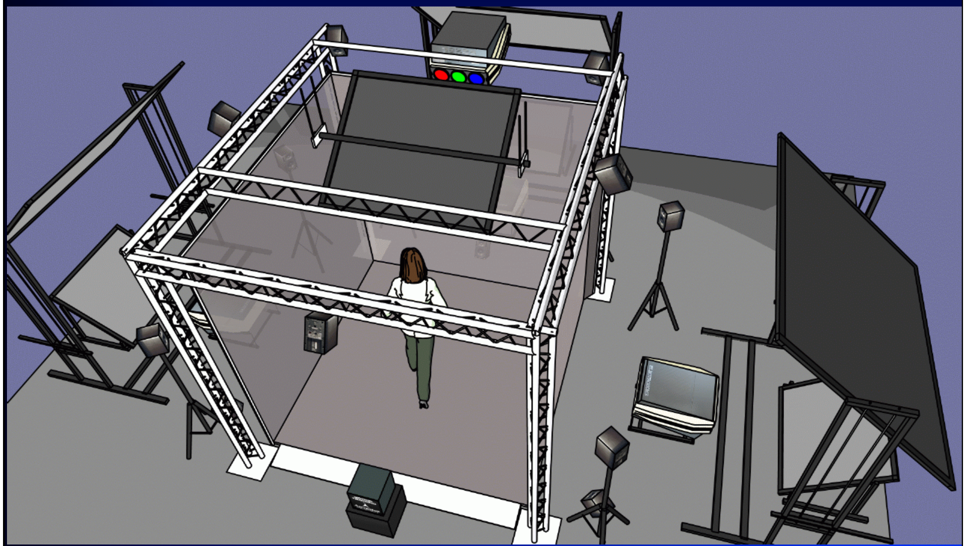
Matti.Grohn@ttl.fi

Experiments in virtual environment

- Localization
  - Static sound source
  - Moving sound source
- Navigation
- Auditory artificial horizon

Matti.Grohn@ttl.fi

Test environment (= EVE)



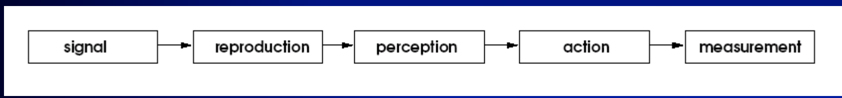
Matti.Grohn@ttl.fi

Method

- Goal: determine how well and accurately the user can perform in localization, navigation and orientation tasks
- Research method: Task based user tests.

Matti.Grohn@ttl.fi

Experiment design



- Scheme for task based auditory experiments

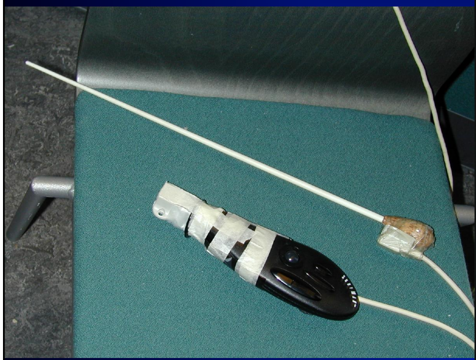
Matti.Grohn@ttl.fi

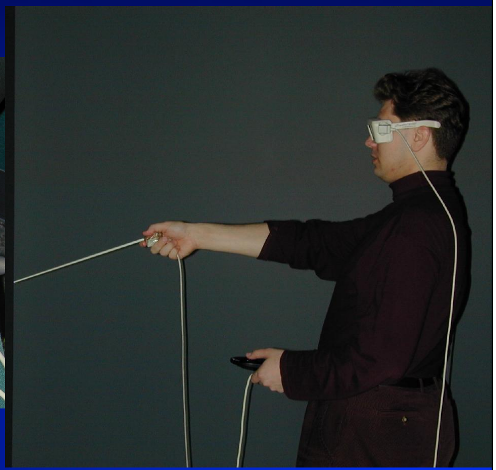
Localization experiment

- Eight non-paid test subjects
- The subject's task was to point to the direction of the perceived location of the sound source.
- Static sound source
- Moving sound source

Matti.Grohn@ttl.fi

Test devices





Matti.Grohn@ttl.fi

Static sound source

- 14 real source (loudspeakers
- 14 virtual source
- Each position presented twice
- 56 source positions in randomized order

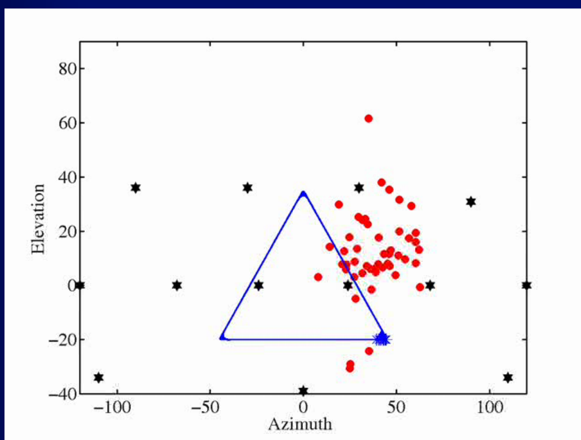
Matti.Grohn@ttl.fi

Moving sound source

- Three different stimuli
  - Pink noise
  - Music
  - Frog croak
- Two different paths
- Each path-stimulus combination was presented four times -> 24 tasks/subject

Matti.Grohn@ttl.fi

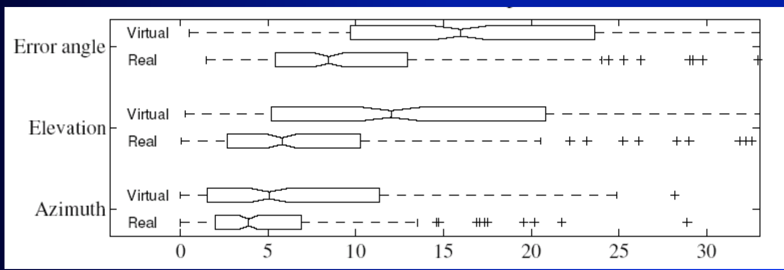
Animation



Matti.Grohn@ttl.fi

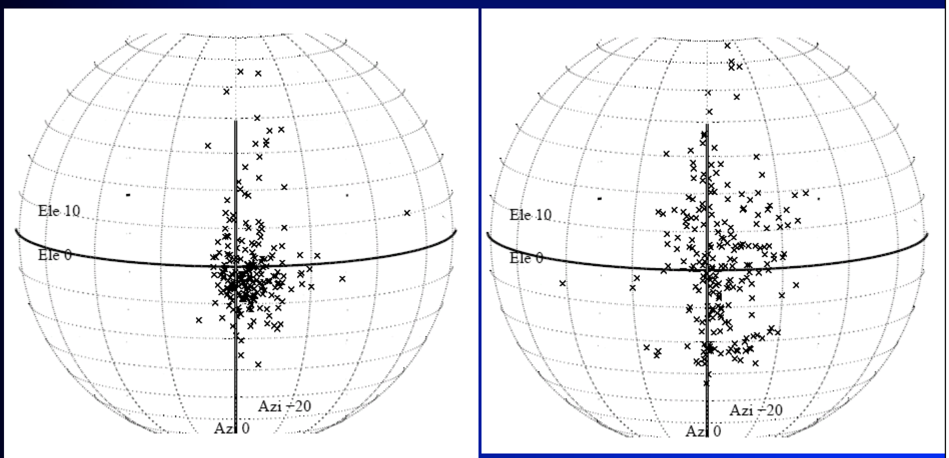
Static sound source

	Azimuth	Elevation	Error angle
with screen compensation			
Real source	4.5^o	5.7^o	8.5^o
Virtual source	4.9^o	11.6^o	15.5^o



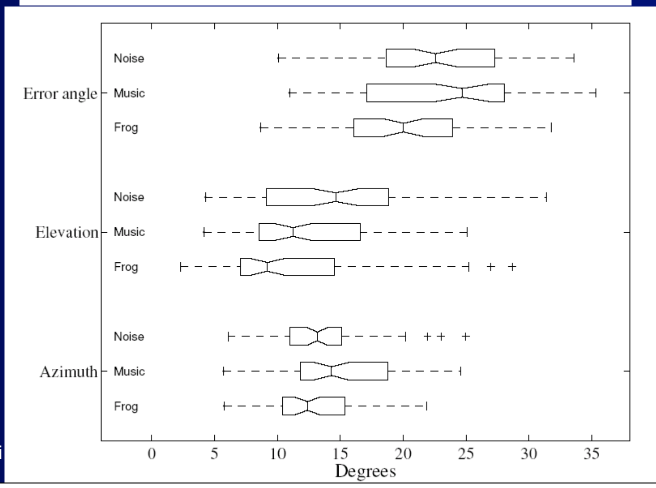
Matti.Grohn@ttl.fi

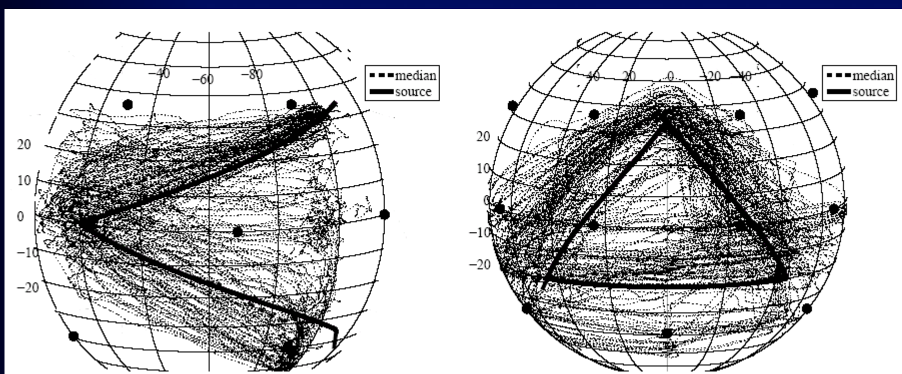
Real source and virtual source

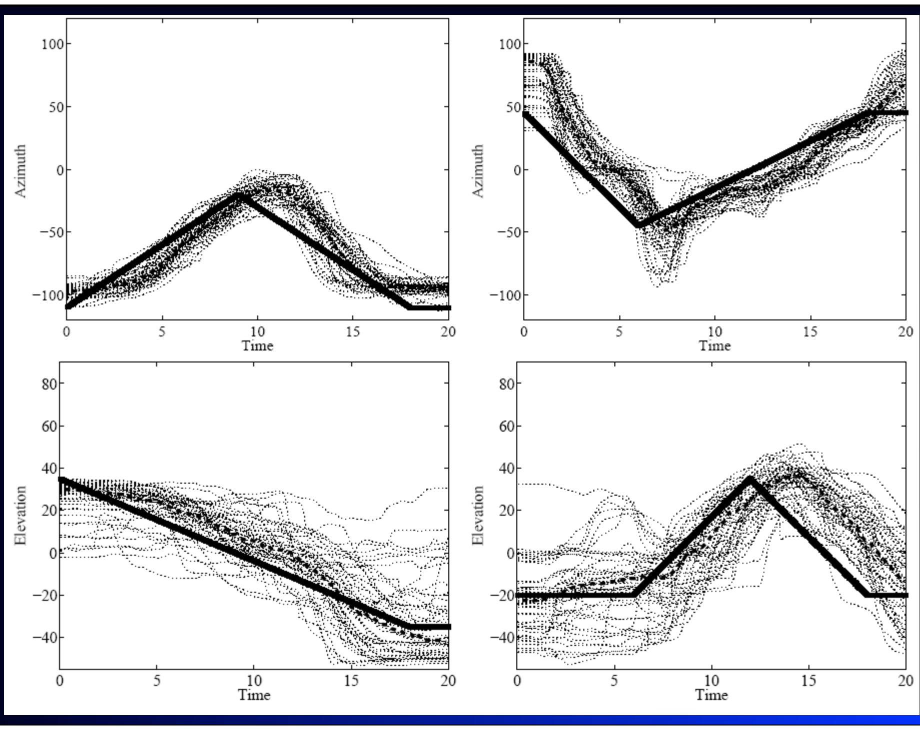


Matti.Grohn@ttl.fi

	Azimuth	Elevation	Error angle
Noise	12.8^o	13.2^o	21.8^o
Music	14.6^o	11.7^o	21.7^o
Frog	12.6^o	10.1^o	19.3^o
All	13.4^o	11.5^o	20.8^o







	Azimuth	Elevation	Error angle
Static experiments			
Sandvad 1996			
Real source	7^o	9^o	
HRTFs	7^o	10^o	
Djelani et al. 2000			
HRTFs			12.4^o - 21.1^o
Martin et al. 2001			
Real source			8.0^o - 11.0^o
HRTFs			9.6^o - 9.7^o
Our experiment			
Real source	4.5^o	5.7^o	8.5^o
Virtual source	4.9^o	11.6^o	15.5^o
Moving sound source experiment			
Our experiment	13.4^o	11.5^o	20.8^o

Matti.Grohn@ttl.fi

Localization

- Moving sound source is harder to localize than static sound source.
- Achieved localization accuracy as good as other research groups have achieved.

Matti.Grohn@ttl.fi

Navigation

- Smooth navigation in virtual reality is a desired feature.

Matti.Grohn@ttl.fi

Goals

- Experiment: Find out how well people can navigate in virtual environments with different kind of cues
- Long term: Effective utilization of spatial audio in navigation

Matti.Grohn@ttl.fi

Navigation experiment

- Eight test subjects
- Task: Find as many gates as possible in three minutes time

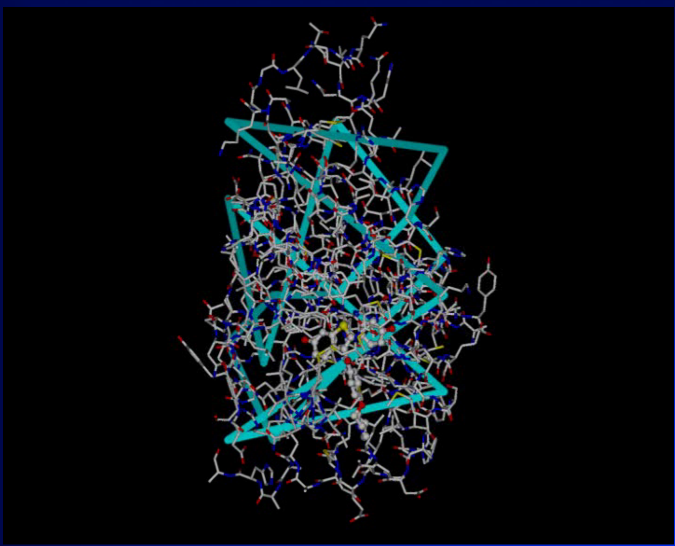
Matti.Grohn@ttl.fi

Procedure

- Three types of conditions:
  - auditory
  - visual
  - audio-visual
- Each task was started in the middle of the molecule
- The index of the first gate was randomized

Matti.Grohn@ttl.fi

Track configuration



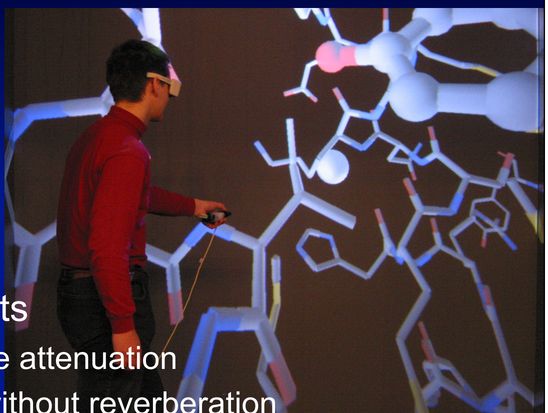
Matti.Grohn@ttl.fi

Procedure

- Subject followed the predefined track gate by gate
- Both directions were used
- Finding of the gate was indicated with auditory signal
- Only one gate at a time was visible/ audible

Matti.Grohn@ttl.fi

Stimuli



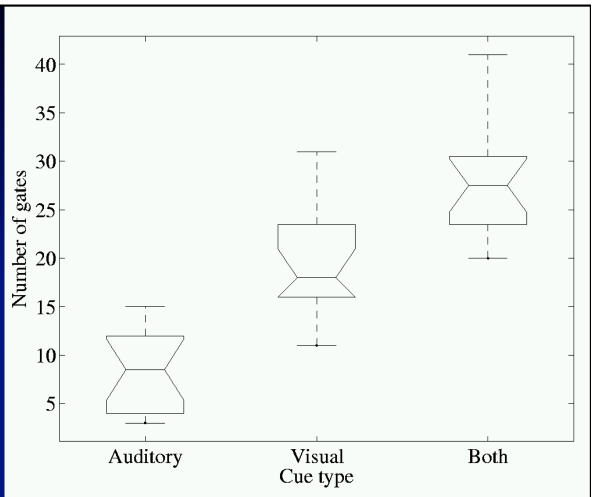
- Visual
  - white ball
- Auditory
  - Pink noise bursts
    - 1/r-law distance attenuation
    - in a free field without reverberation
- Audio-Visual
  - Both auditory and visual stimuli

Matti.Grohn@ttl.fi

Audio-Visual example

Matti.Grohn@ttl.fi

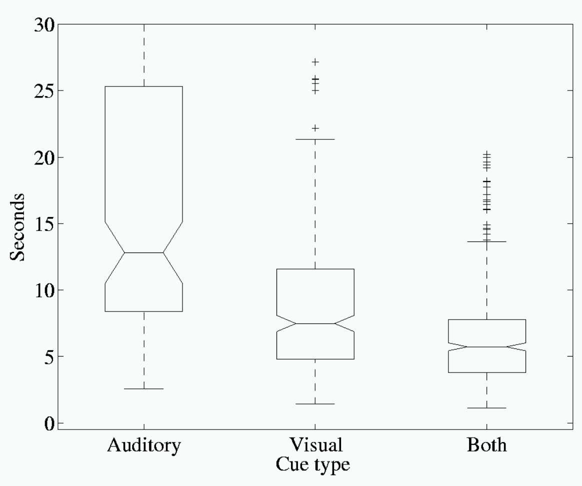
Number of gates



Cues	Median	Min	Max
Auditory	8.5	3	15
Visual	18	11	31
Audio-visual	27.5	20	41

Matti.Grohn@ttl.fi

Time between gates

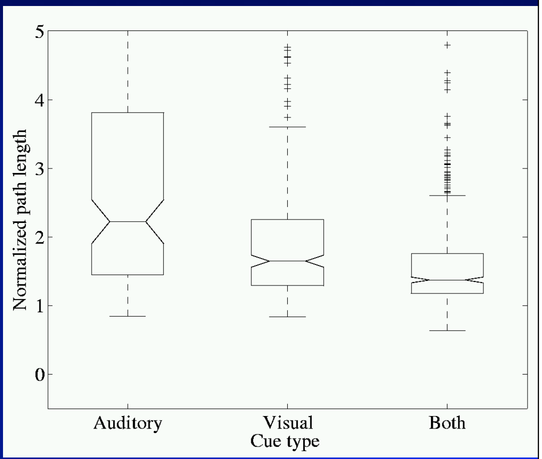


Cues	Median	Min	Max
Auditory	12.8	2.6	99.2
Visual	7.5	1.4	52.7
Audio-visual	5.7	1.2	20.2

Matti.Grohn@ttl.fi

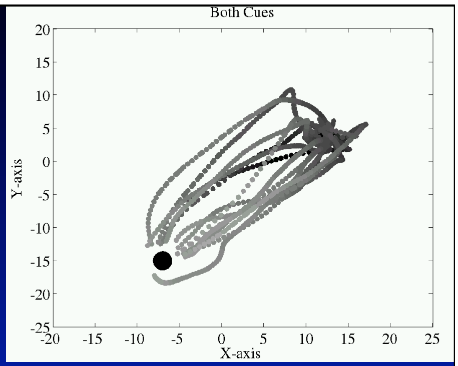
Normalized path lengths (NPL)

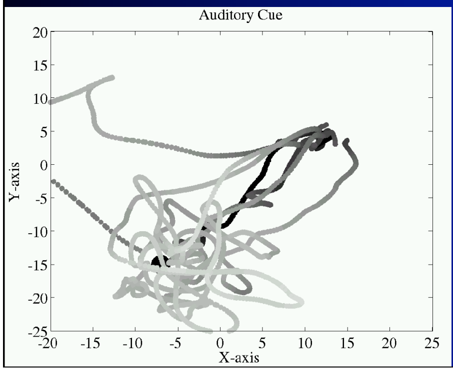
- NLP = Path length / Distance between gates

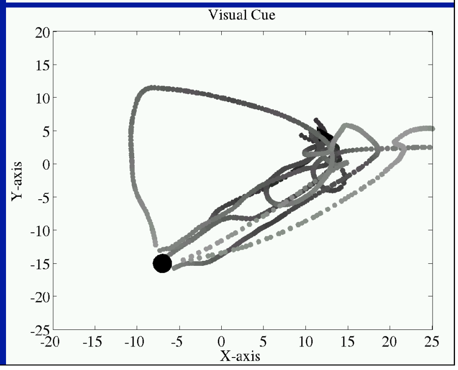


Matti.Grohn@ttl.fi

Paths for one gate pair







Summary

- Audio-visual cues were remarkably better than other cues
- Visual cue more efficient than auditory
- Auditory navigation possible even in 3D

Matti.Grohn@ttl.fi

Discussion

- Auditory cue was utilized to define the approximate location of the gate
- Visual cue was utilized in the final approach

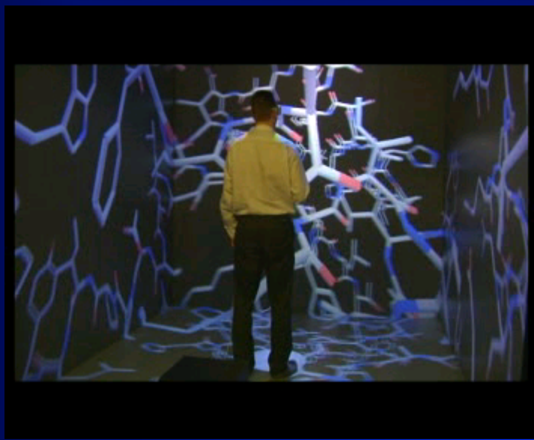
Matti.Grohn@ttl.fi

Audio-only navigation experiment

- Same task with four different auditory stimuli and no visual cue
- The first was the same auditory cue as in previous experiment
- In the second distance to the gate was indicated with the a density of the bursts
- In the third signal narrow-band noise was added to original cue to provide information about the height of the sound source
- Fourth signal combined all cues

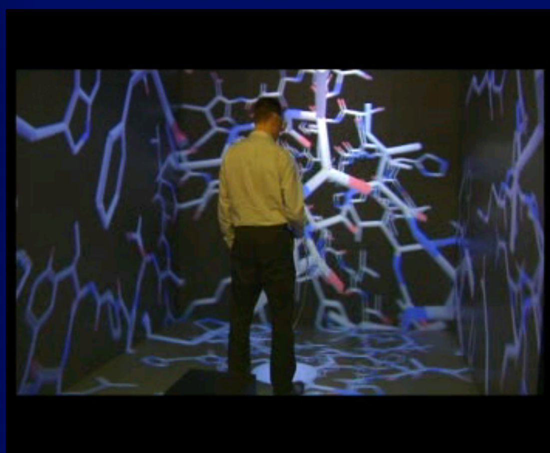
Matti.Grohn@ttl.fi

Gain + rate



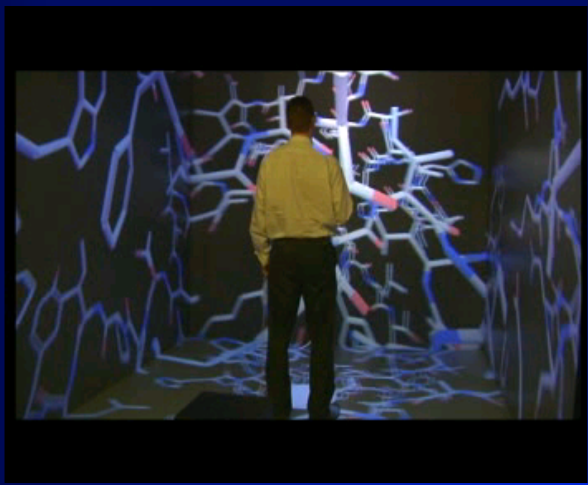
Matti.Grohn@ttl.fi

Gain + pitch

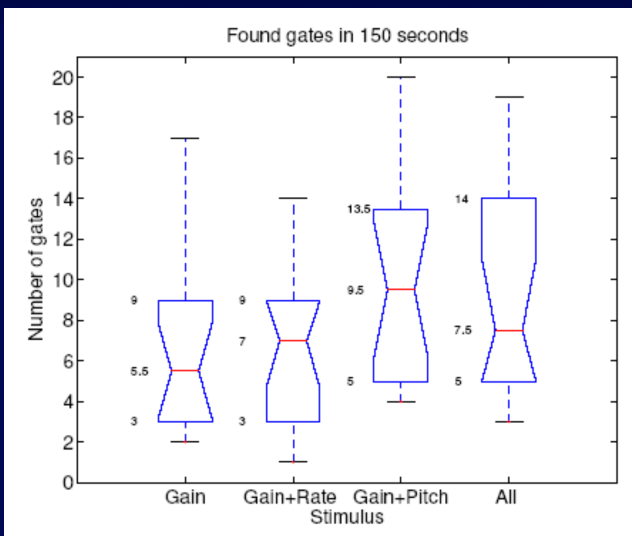


Matti.Grohn@ttl.fi

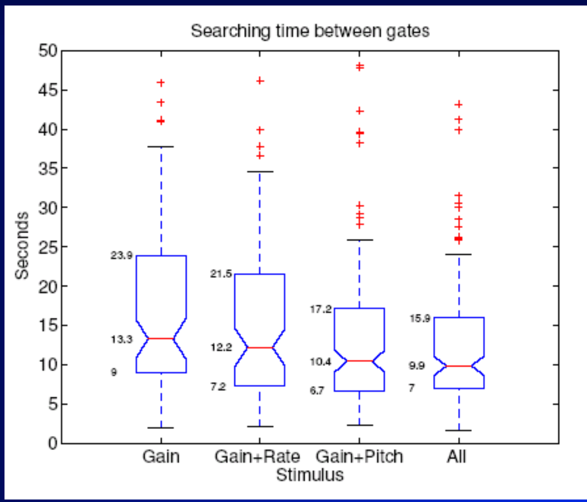
All



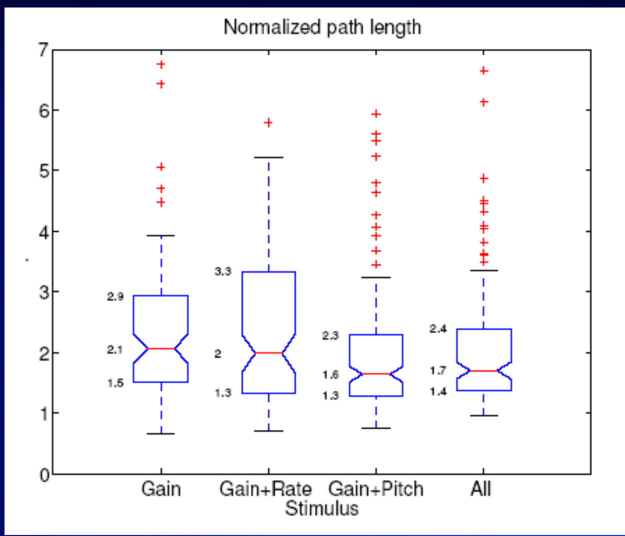
Matti.Grohn@ttl.fi



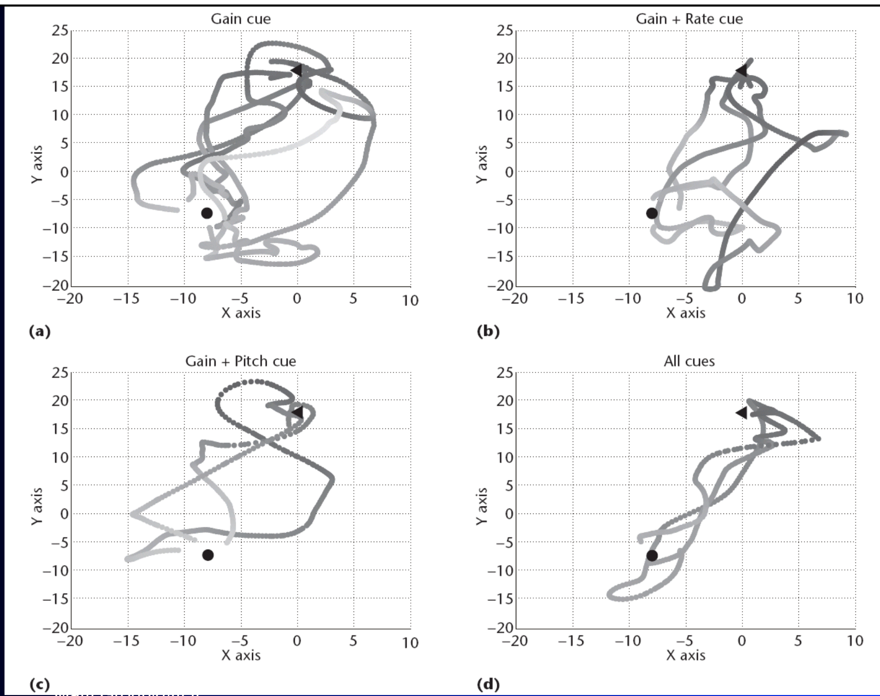
Matti.Grohn@ttl.fi



Matti.Grohn@ttl.fi



Matti.Grohn@ttl.fi



Result

- When the elevation information was encoded to the navigation cue signal, subjects found gates faster and the travel distance was shorter.

Matti.Grohn@ttl.fi

Orientation experiment

- Visualizations and architectural walkthroughs are common demonstrations in virtual rooms.

Matti.Grohn@ttl.fi

Motivation

- Typically in visualization demos user is disoriented in an unwanted way
- Visual artificial horizon might disturb the user

Matti.Grohn@ttl.fi

Goal

- Experiment: Find out the usefulness of the auditory artificial horizon
- Long term: Effective utilization of spatial audio in immersive visualization

Matti.Grohn@ttl.fi

Auditory artificial horizon

- "Ball on plate" metaphor: The ball starts to roll to the direction pointing downward
- Sound is heard from direction pointing downward
- Elevation information is mapped to azimuth angle

Matti.Grohn@ttl.fi

Stimuli

- Base stimulus: Pink noise burst.
- Gain stimulus:
  – Gain factor = amount of tilt
  – model oriented - stimulus inaudible
- Rate stimulus:
  – pulse rate = amount of tilt
  – model oriented = rate 0.7 Hz
  – max rate = 8 Hz

Matti.Grohn@ttl.fi

Stimuli

- Pitch stimulus:
  – narrow band-pass noise was added to base stimulus
  – model oriented = center frequency 50 Hz
  – max frequency 2 kHz
- Gain and pitch stimulus:
  – pulse rate 2.4 Hz

Matti.Grohn@ttl.fi

Experiment

- Eight test subjects
- Task: Navigate through a predefined route and keep the model as oriented as possible
- Four conditions: visual, gain, rate and pitch

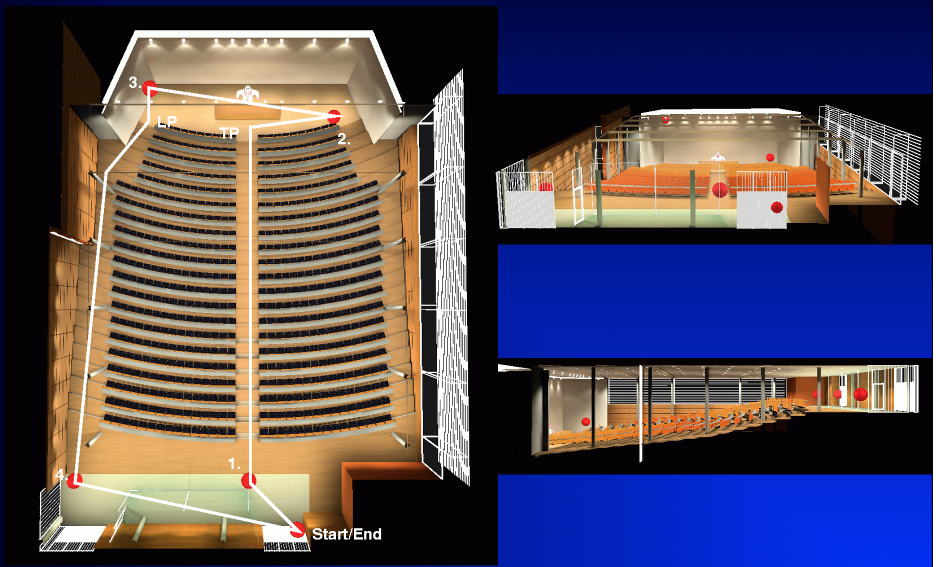
Matti.Grohn@ttl.fi

Experiment

- Training before test sets
- Two test sets:
  - Each condition once for each subject in test set
  - Order of the conditions randomized
- Location and orientation were recorded with 10 Hz sampling rate

Matti.Grohn@ttl.fi

Test route



Matti.Grohn@ttl.fi

Gain

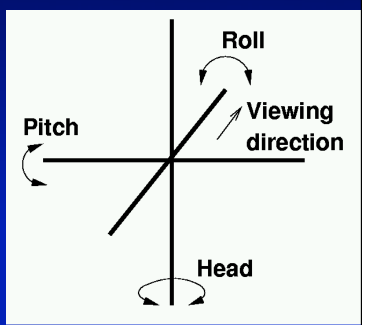
Matti.Grohn@ttl.fi

Pitch

Matti.Grohn@ttl.fi

Definitions

- Pitch angle = median(abs(recorded pitch angle))
- Roll angle = median(abs(recorded roll angle))

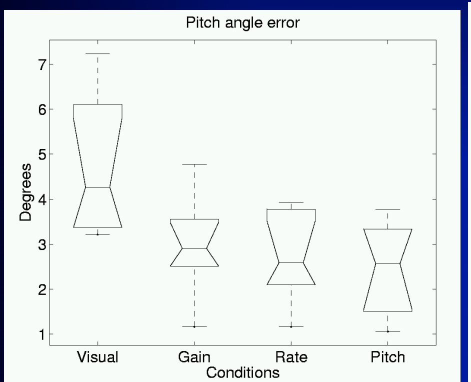


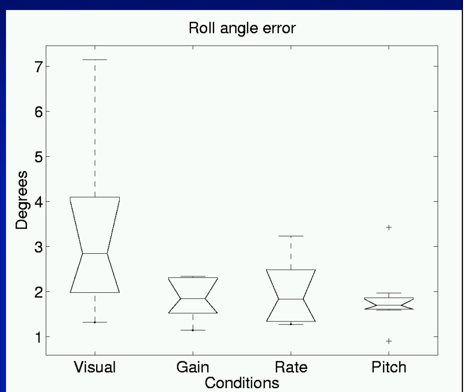
Matti.Grohn@ttl.fi

Results

		Visual	Gain	Rate	Pitch
First test set	Pitch angle	4.30	2.92	3.01	3.15
	Pitch std	2.08	1.88	1.47	1.62
	Roll angle	2.50	2.59	1.98	2.52
	Roll std	1.58	1.04	1.51	1.61
Second test set	Pitch angle	4.27	2.90	2.58	2.56
	Pitch std	1.60	1.05	1.02	1.04
	Roll angle	2.84	1.84	1.84	1.69
	Roll std	1.92	0.46	0.75	0.71

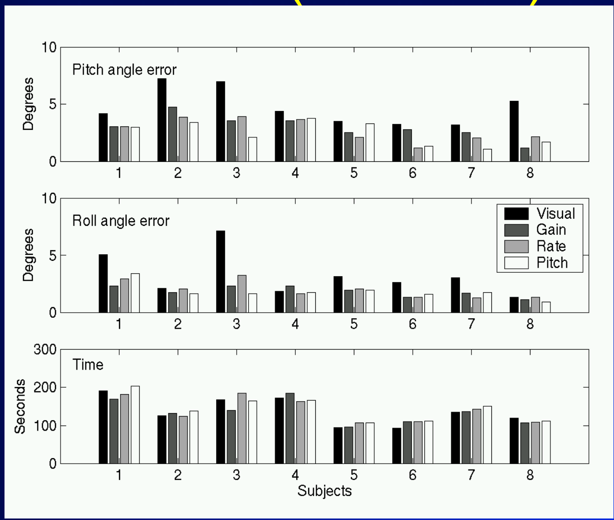
Matti.Grohn@ttl.fi



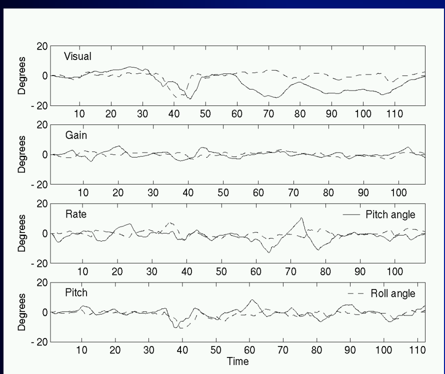


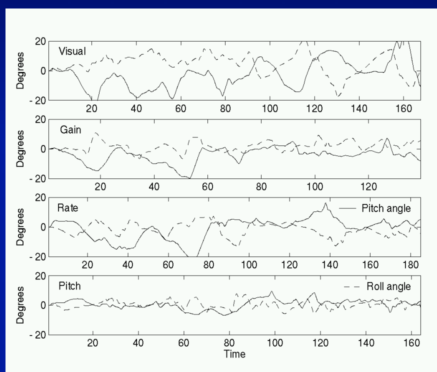
Matti.Grohn@ttl.fi

Results (continue)



Matti.Grohn@ttl.fi





Matti.Grohn@ttl.fi

Results

- Subjective order was the same for all the subjects:
  - Gain, pitch and rate

Matti.Grohn@ttl.fi

Discussion

- Subjects understood auditory artificial horizon and found it intuitive to use
- Amount of disorientation was larger in pitch than in roll
- Reference value is important in this kind of application

Matti.Grohn@ttl.fi

Spatial Audio

Matti Gröhn
CSC – IT Center for Science

Matti.Grohn@ttl.fi