ITU-T Workshop on
"From Speech to Audio: bandwidth extension, binaural perception"
Lannion, France, 10-12 September 2008

Spatial audio conferencing.

Peter Hughes.

Broadband Applications Research Centre, BT Group.

Bringing it all together

Introduction.

- Personal background:
  - BT Research labs since 1975.
  - Work areas: data communications, speech technologies, audio and accessibility.
- Current and recent activities:
  - Spatial audio.
  - Bandwidth expansion.
  - Accessibility – telephony for hearing impaired.
  - TA2 (EU 7th Framework project on social networking).

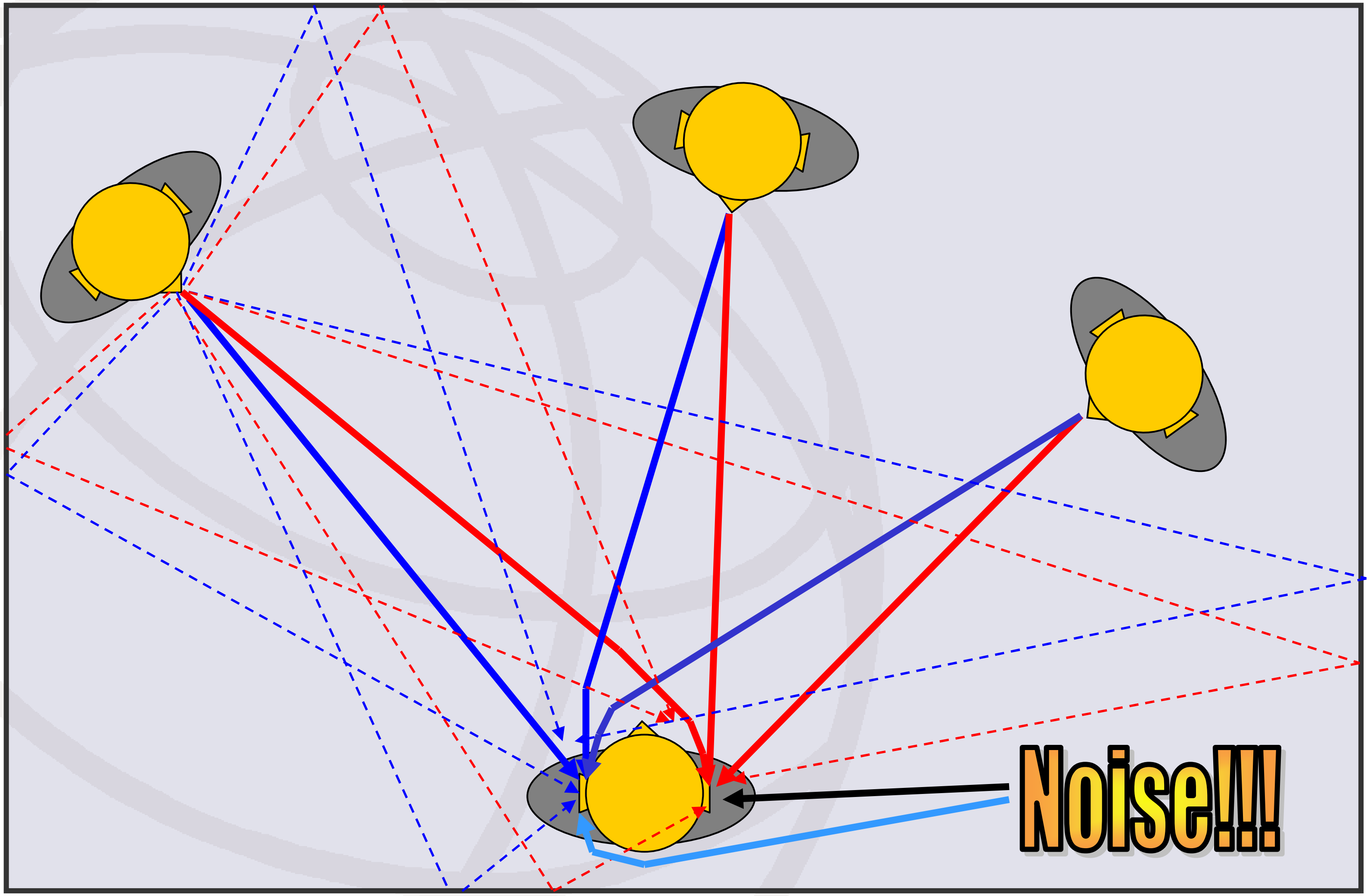
Natural audio for conferencing

What do users want from a conference?

- Everybody has a natural voice and can be heard clearly.
- I can see who is present and who is talking
- I can share information easily and be confident I’m understood.
- I don’t have to hold a piece of bent plastic to my head
- I can actually enjoy taking part!

Because life is not monophonic….

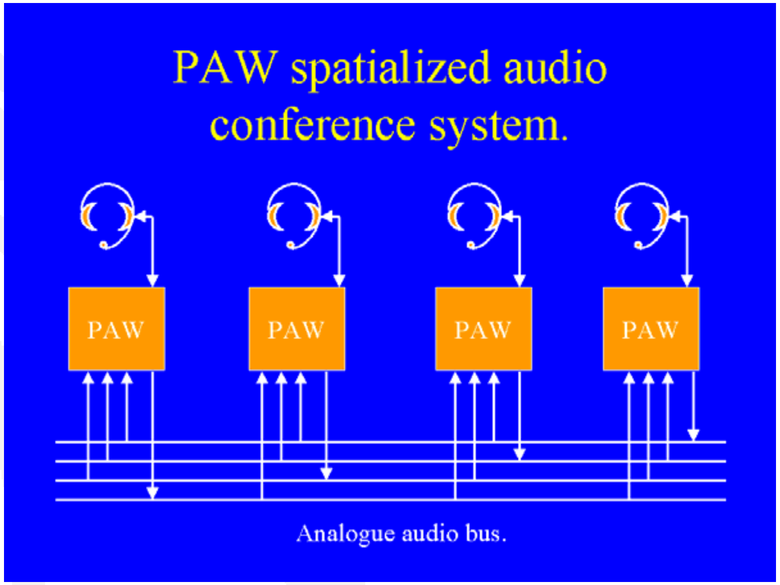
Why use spatial audio?



Earlier work at BT Labs 1995-99.

- **NSAS: Shared network audio server.**
  - Virtual meeting space featuring avatars meeting in several different sized virtual rooms with .
  - Spatial audio – various renderings – ambisonic, binaural, stereo etc - available using Lake Huron platform.
  - Current status: Online virtual worlds (e.g. Second Life) gaining popularity.
- **SmartSpace:**
  - Concept demonstrator of alternative office desk/chair.
  - Immersive sound field using 3 loudspeakers to support video display.
- **Virtual conferencing.**
  - Large display screens (typically back projection) with spatial sound to merge boundary between local and distant realities.
  - Current status: techniques commonly used in top-end teleconferencing.
- **Spatial telephony conferencing**.
  - Simple spatial audio applied to audio conferencing – the Personal Audio World (SAW) system.

PAW conference system.



- Hardware based 4-way system (Motorola DSP56302 processor).
- 46 coefficient HRTFs, 10 degrees.
- 16 kHz sample rate - 7kHz BW.
- 3 artificial rooms with spatial reverberation.

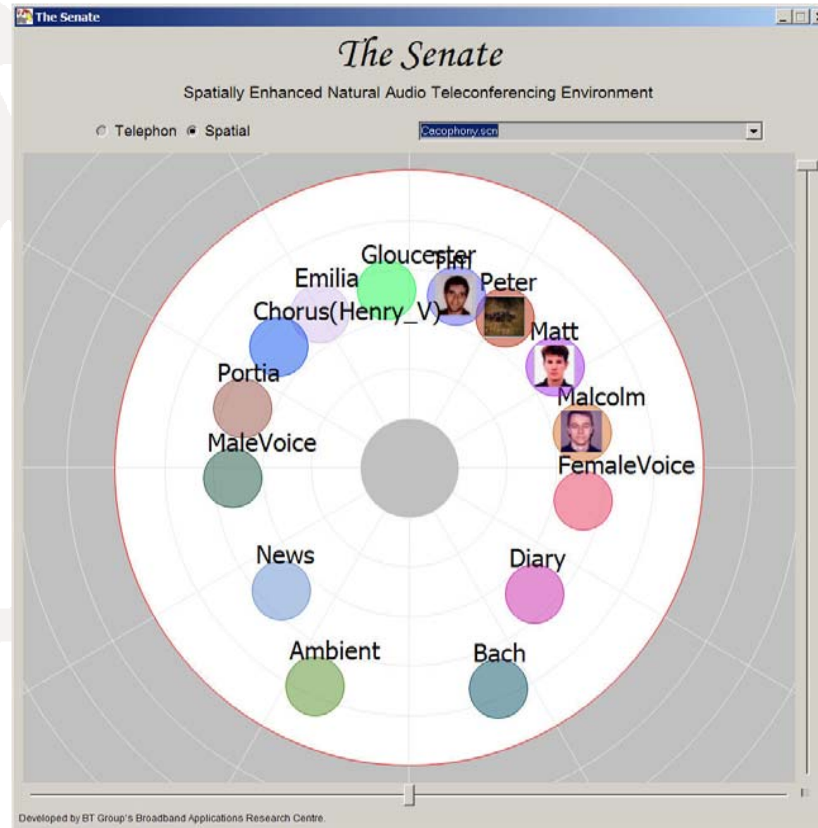
Results from PAW

- Results of informal tests:
  - Users were universally very impressed by both increased clarity and naturalness.
  - Artificial rooms were very popular.
  - Comments: ‘Like being in the same room’, ‘Hearing in colour’ and ’When can we have it?
- Results of formal tests:
  - Positive, but largely inconclusive.
  - Demonstrated need for more rigorous design.
- Difficult to convince people that the system could be delivered!

The Senate

Spatially Enhanced Natural Audio Teleconferencing Environment.

- Fully interconnected SIP VoIP conference
- 7kHz bandwidth using G.722.2.
- HRTF (5 degree spacing CIPIC Kemar model) or 5 channel audio.
- Simple graphical interface to control volume and position.
- Visual talker indiction.
- Text to speech for text data
- Support for video streaming.
- Audio smileys, background music.



Run demo.

Senate extensions.

- Artificial room acoustics
- Voice caricaturing and enhancement.
- Selectable GUI ‘skins’ for domestic, business, teenage markets.
- Loud speaking systems.
  - Stereo,
  - cinematic surround systems.
- Groups of people.
  - Microphone positioning
  - Wearable microphones (e.g. ‘tie-clip’).
  - Need spatial audio over wide listening area.
- Echo control.
  - Difficult for multi-channel due to cross correlations between channels.
  - Easier for HRTF or intensity panning.

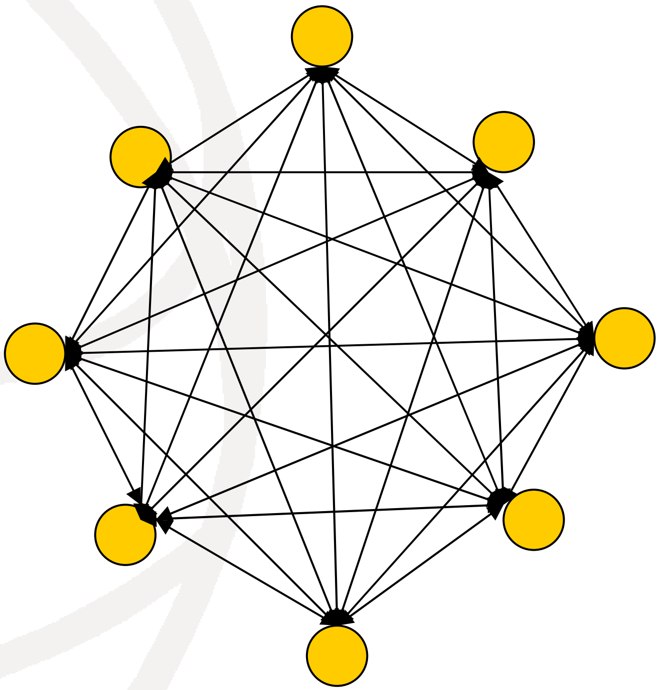
Network design.

- Fundamental to designing spatial audio conference system.
- 4 basic options are:
  1. Fully interconnected Peer to Peer with all processing performed at the client terminals.
  2. Centralised processing – all processing performed in server.
  3. Distributed processing – processing performed at several points in the network to optimise processing and network resources.
  4. Server concentration schemes using efficient multi-channel audio compression – e.g. Spatial Audio Object Coding (SAOC) or channel concentration.

Fully interconnected mesh.

- Each conferee gets a direct stream from all other conferees.
- Audio rendering performed independently at each client.
- Network loads assume G.722.2 at 24kbits/sec.

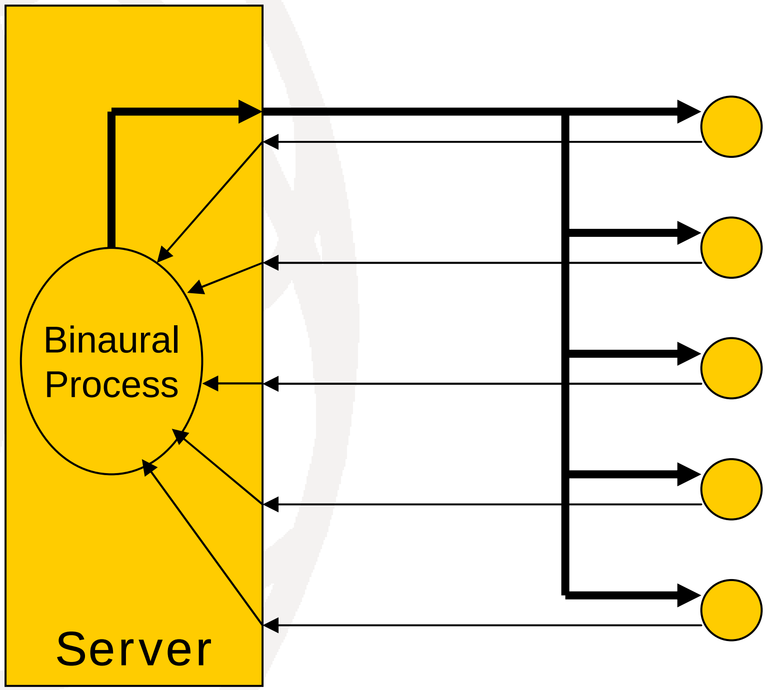
N	Total bi-directional network hops	Total Network load (kbit/sec)
2	1	48
3	3	144
4	6	288
5	10	480
6	15	720
7	21	1008
8	28	1344
9	36	1728
10	45	2160



Central processing.

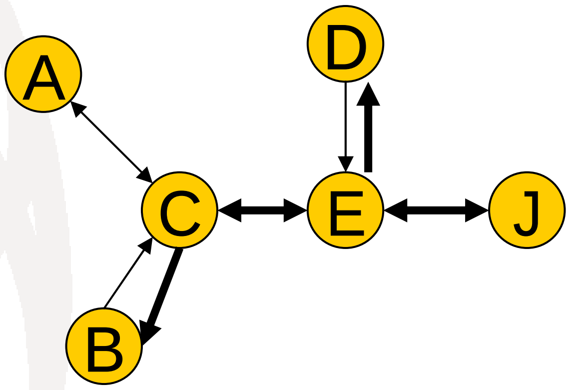
- All spatial processing done in central bridge and 2-channel spatial mix broadcast to all users.
- Note higher bit rate is required to preserve spatial cues.
- 24kbits/s upstream; 128kbits 2-channel downstream.
- No processing required at local terminals.

N	US and DS. (kbit/s)	Total worst case network load (kbits/sec).
2	24 / 128	304
3	24 / 128	456
4	24 / 128	608
5	24 / 128	760
6	24 / 128	912
7	24 / 128	1064
8	24 / 128	1216
9	24 / 128	1368
10	24 / 128	1520



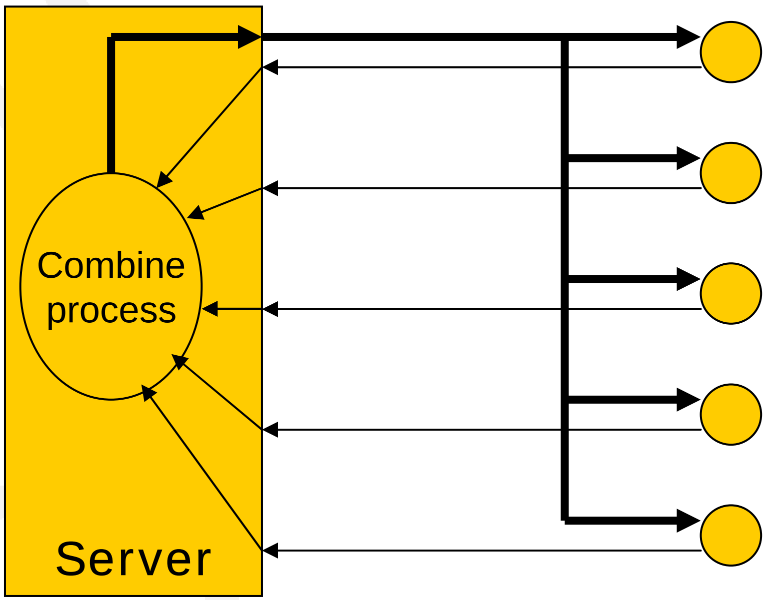
Distributed processing.

- Terminals transmit and receive mono or multi-channel spatial audio, allocated to achieve optimum usage of processing and network resources
- Heavy lines indicate multi-channel spatial signals.
- Application Layer Routing (ALR) can have a major impact on efficiency.
- Area of research at BT*.



Server concentration.

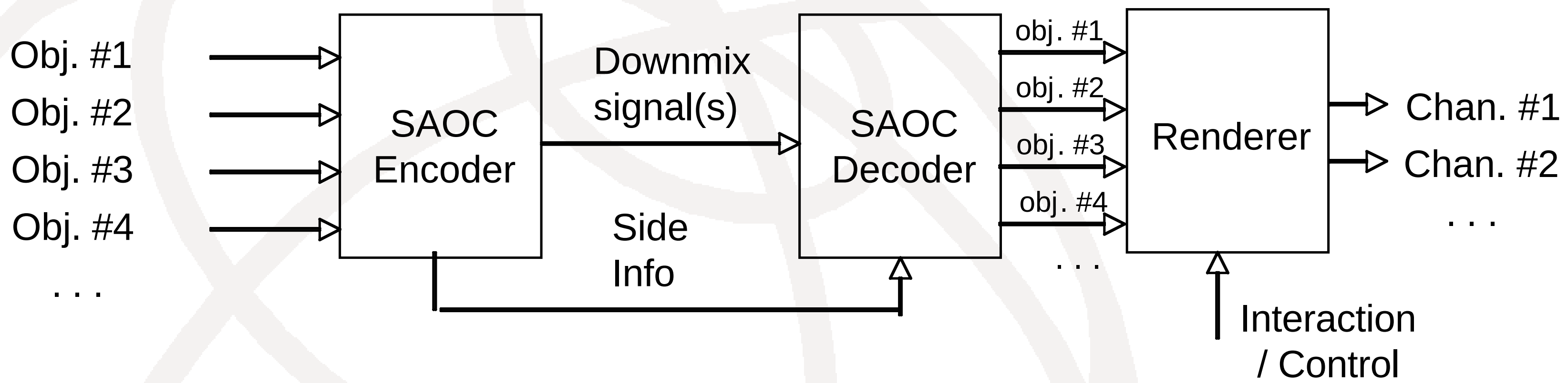
- Audio from M clients is compressed at the server into an efficient ‘N channel plus supplementary data’ form.
- This data is streamed to all clients and spatialized locally as required.
- This method is also suitable for Spatial Audio Object Coding.



24kbits/s mono upstream

128kbits 2-channel + supplementary data. downstream

Spatial Audio Object Coding.



- Parametric multiple object coding method. Based on MEG surround technology
- Very efficient transmission of multi channel audio data.
- Transmits N audio objects in a K channel audio stream. K < N, and K is typically 1 or 2 channels.
- Undergoing MPEG standardization process.

Project TA2.

Together anywhere, together anytime

- BT led EU 7th Framework collaborative project.
- New media experiences for homes and families.
- Enjoyable and fun - supporting family to family relationships that are served poorly by current ICT products and services.
- Promoting activities such as:
  - social interaction
  - building relationships
  - Entertainment
  - Relaxation
- Started January 2008.

What will TA2 address?

“New media experiences for households and families”

5 Prototype Applications:
Family game, My Videos, Child’s play, Sixth Age, Social Communication

- *Technology Capabilities*
- *System Architectures*

- *New patterns of consumption and production of digital media*
- *New converged business opportunities*
- *Improved social and emotional well-being*

TA2: The Technical Challenges

- How can we improve the **experience of audiovisual communications** between dislocated groups of people to a level where they are happy to use it for enjoyable purpose-driven social experiences?
  - Techniques include low-delay audio codecs, echo cancellation and spatial audio object coding.
- How can we support the **end-to-end delivery** of complex, interactive audiovisual services between homes?
- How can we bring support to the next generation of social networking applications?
  - We need to define the APIs, rules, protocols and network services which will enable applications developers to leverage network resources within these new applications.

The TA2 consortium

Industry partners		
BT	*Management, systems architecture*	
Alcatel Lucent	*Systems architecture*	
Philips	*Application design*	
Ravensburger	*Games*	
Limbic Entertainment	*Games*	
Eurescom	*Management*	
Research institutes		
Fraunhofer IIS	*Audio*	
TNO	*Users, markets, economic issues*	
IDIAP	*Audio/video scene interpretation*	
Joanneum Research	*Semantics and ontologies*	
CWI	*Media annotation*	
Interactive Institute	*Application design*	
Academic institutes		
Goldsmiths, University of London	*Computational modelling*	

Thanks for listening.

Peter Hughes.
Broadband applications research centre.
BT Group
Adastral Park
Ipswich
UK.

peter.j.hughes@bt.com
http://barc.visual.bt.co.uk/
www.ta2-project.eu

Bringing it all together