Multi-Context Voice Communication Controlled By Using An Auditory Virtual Space

Yasusi Kanada
Hitachi Ltd., Central Research Laboratory
Japan

Background

- **History of voice communication media (VCM)**
  - Telephone has been used for about 130 years.
    - A. G. Bell invented the telephone in 1876 and the first telephone network was built in 1878.

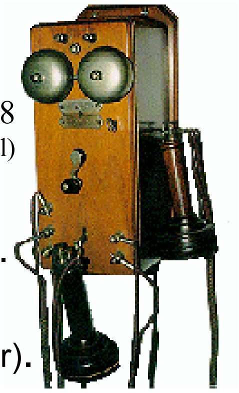
http://sln.fi.edu/franklin-/inventor/bell.html

The first telephone exchange (in Connecticut in 1878) (http://www.att.com/history/history1.html)

  - Audio (and video) conferencing systems has been developed since 1970's (or earlier).
  - Telephone is still the most popular VCM.

Background (cont'd)

- **The telephone user interface has not been changed since its invention!**
  - The interface:
    - To connect to (to call) the remote site.
    - To talk/listen using one microphone and *one* speaker.
    - To disconnect (to hung up).
  - This interface has serious problems (explained later).
  - The reasons why the interface has not been changed.
    - People has been supported this interface.
    - *The network was stiff so that it could not be changed.*
- **It is time to change the interface!**
  - IP networks are going to replace telephone networks.
  - IP telephony is much more flexible than the telephone.
    - Especially, there is no need to disconnect explicitly, because IP networks are always connected.



A telephone set in 1878
(http://www.atcaonline.com/phone/coffin.html)

User-interface Problems of VCM

- **Current VCM do not support natural n-to-n communication, and do not use human communication ability fully.**
  - Face-to-face communication is basically n-to-n.
  - Conversations by VCM are one-to-one (in telephone), or n-to-n but strictly constrained (in conferencing systems).
  - When talking with two or more persons by VCM,
    - Sometimes it is difficult to recognize and to remember who is talking.
    - The "cocktail party effect" is not supported -- people cannot speak concurrently.

User-interface Problems of VCM (cont'd)

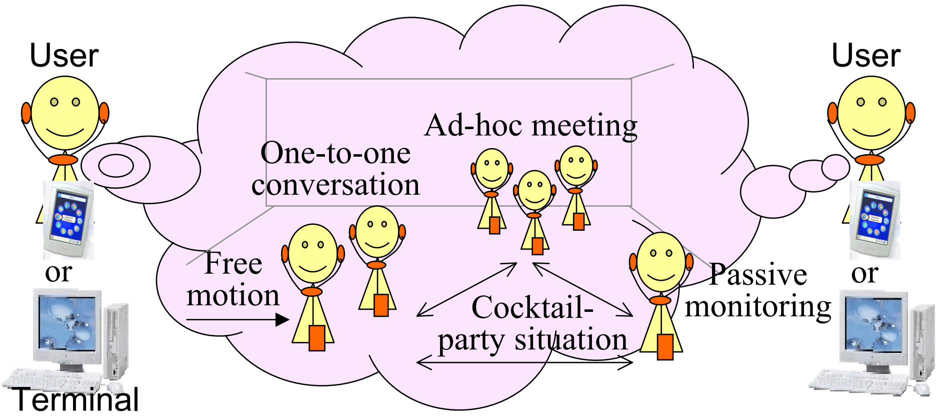
- **People can talk/hear only when they are (intentionally) connected.**
  - No information comes when they are not connected.
    - Eg. She cannot see if he is ready to answer a telephone call.
  - In face-to-face communication, there are unintentional but important communications.

A Solution called "voiscape"

- **A new medium that solves the above problems will arise in several years -- I call this "voiscape".**
- **Expected features of voiscape**
  - *Spatial hearing*: Both ears should be used.
  - *"Sound room"*: People can move within an auditory virtual space called sound room.
  - *Full-time connection*: People can always recognize (hear) the presence of others.

More on “Sound Room”

- **Virtual-location-based conversation is enabled.**
  - Voice communication in the real world is simulated.



  - Each voice in the sound room is spatialized.
    - The direction and the distance are expressed by spatial audio technologies.
  - Each user can freely move in the room.
    - The user can walk up to or run away from another user.

More on “Sound Room” (cont’d)

- **Personalized policy-based communication-control is required.**
  - I.e., each user (and the manager) should be able to specify policies to control sessions and resources.
  - Because it is necessary
    - to control limited communication resources such as network bandwidth, and
    - to avoid privacy problems – a user’s voice may be heard by an unrecognizable user in the room.
- **Comparison to conventional virtual environments:**
  - Most virtual environments express the space by graphics but not by spatial sounds.
  - Real-time bi-directional communication was not the main focus of conventional auditory virtual environments such as DIVA of Helsinki University.

How to Make N-to-N Communication Natural?

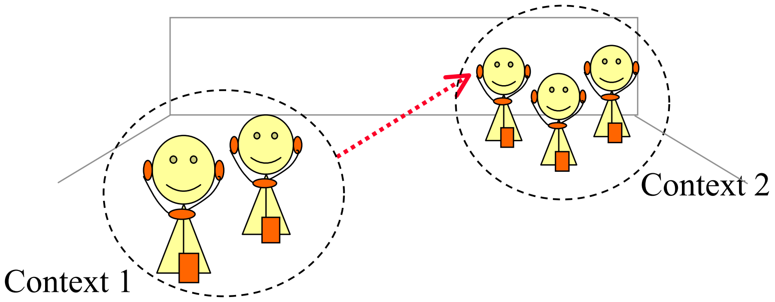
- **Two points**
  - Explicit conference control is avoided by sound room.
  - Multi-context communication is enabled by sound room.
- **Explicit conference control**
  - In conventional conferencing systems,
    - *Explicit session control*: Connection and disconnection are explicitly controlled by the users.
    - *Explicit floor control*: One or several concurrent speakers are selected in a centralized method, if floor control functions are supported.

How to Make N-to-N Communication Natural? (cont'd)

- **Explicit conference control (cont'd)**
  - *Solution*: In voiscape,
    - Connection and disconnection should be implicitly controlled by distance-based policies
      - Example policy:
        if another person comes within 5 m, connect to that person, and
        if another person goes over 6 m away, disconnect from that person.
    - Policies should be arbitrated between each pair of persons.
      - The weaker policy, i.e., the policy that more strongly protects privacy, should win.
    - Speakers should be selected by each person
      - *Attention-based selection*: If the user pays attention to one of concurrent speakers, the user can listen to the voice (by the cocktail party effect).
      - *Motion-based selection*: If the user comes close to one of concurrent speakers, the user can hear the voice better.

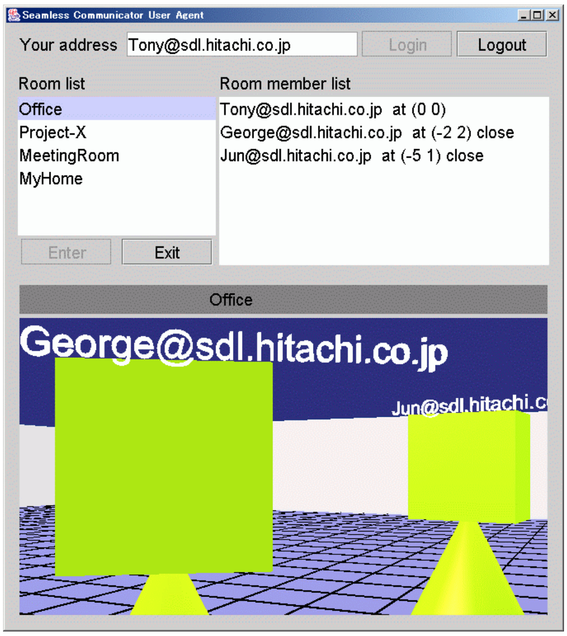
How to Make N-to-N Communication Natural? (cont'd)

- **Multi-context communication**
  - What is multi-context communication?
    - People can talk concurrently -- multiple contexts can coexist.
    - Contexts can crossover -- people in different contexts can talk each other.
  - *Solution*: In a sound room,
    - Multiple contexts can coexist (if they are distant).
    - People can hear other contexts in the same sound room.
      - They may be able to find a more interesting context.



Outline of voiscape-based communication*

- **The server sends a room list.**
- **The user selects and enters a room.**
- **The user agent (UA) shows inside the room.**
  - It shows the users and objects in the room.
- **The user selects another user for conversation by moving and turning in the room.**
  - Pointing devices (such as a mouse or cursor keys) are used.



Seamless Communicator User Agent

Your address Tony@sdl.hitachi.co.jp

Room list

Office
Project-X
MeetingRoom
MyHome

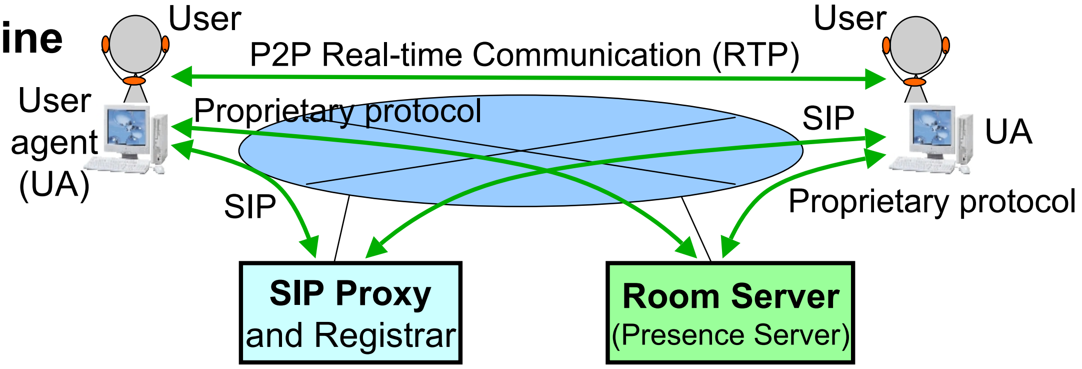
Room member list

Tony@sdl.hitachi.co.jp at (0 0)
George@sdl.hitachi.co.jp at (-2 2) close
Jun@sdl.hitachi.co.jp at (-5 1) close

Office

Voiscape Prototype Implementation

■ Outline

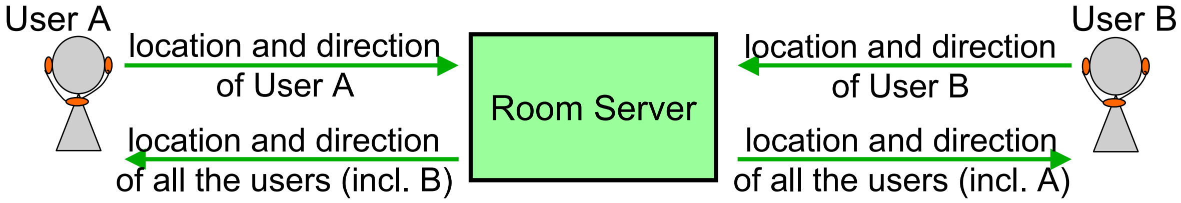


- ◆ PCs with Java-based UAs were used for terminals.
  - Each PC had 3D sound functions (a sound card with a CMI8738).
- ◆ Two servers were used: a Room Server and a SIP Proxy.
- ◆ The protocols and codecs:
  - The voice communication is P2P and unicast, and RTP is used.
  - UAs spatialize received voice streams and mixes them.
  - The sampling rate is 8000 Hz (G.711 u-law 64 kbps format).
  - Sessions are controlled by SIP (Session Initiation Protocol).
  - The room server uses a proprietary protocol.

Voiscape Prototype Implementation (cont’d)

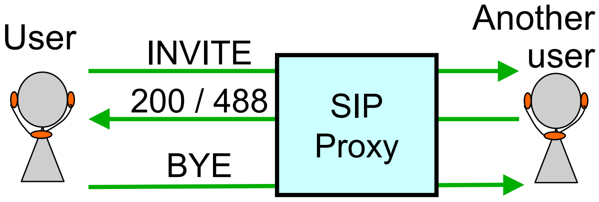
■ Locations / presence management

- ◆ The room server manages
  - rooms (creation, deletion, etc.),
  - room properties (such as room sizes), and
  - room users (i.e., presence and locations in the room).
- ◆ UAs and the the room server exchage users’ location information while the users are in the room.
  - Each UA sends the user’s location and direction to the server.
  - The server distributes gathered users’ locations and directions to all the UAs.


User A
location and direction of User A
location and direction of all the users (incl. B)
Room Server
User B
location and direction of User B
location and direction of all the users (incl. A)

Voiscape Prototype Implementation (cont'd)

- **Policy-based session control**
  - A UA sends a SIP INVITE/BYE message to another UA according to the policies.
    - The UA sends an INVITE when it comes within the connection distance.
    - The other UA responds with "200 OK" to the INVITE when it is within its own connection distance, but responds with "488 not acceptable here" when out of it (when it has a weaker policy).
    - The UA sends a BYE when it goes over the disconnection distance.
  - This mechanism implements the policy arbitration: a weaker policy wins.



Evaluation*

- **Virtual-location-based conversation**
  - Most people felt the direction and distance in the sound room in the intended way.
    - They recognized a virtual speaker walking on a circle trace.
    - They roughly distinguished the direction and distance of speakers.
- **Policy-based session-control**
  - It worked, but the response was rather slow in Java-based implementation (needed 7 sec. to connect).
- **Extensive evaluation on multi-context communication has not yet been conducted**
  - because it was very difficult to improve the voice quality of the Java-based implementation.

Conclusions and Future Work

- **Conclusions**
  - Virtual-location-based conversation and policy-based control enable an n-to-n communication environment.
  - Such an environment can be built by low-cost hardware and software.
    - However, required quality was not achieved by the Java-based implementation.
- **Future work**
  - An improved prototype with better voice quality, with wearable terminals, and with SIP event-notification mechanisms will be developed.