# Multi-Context Voice Communication
## In A SIP/SIMPLE-Based Shared Virtual Sound Room With Early Reflections

Yasusi Kanada
Hitachi Ltd., Central Research Laboratory
Japan

## Background

- **Voice is the original and a most important communication medium among people.**
- **Various voice communication media (VCM)**
  - Telephone
    - "Inconvenient" user interface kept unchanged for 130 years.
  - Teleconference systems
    - Solved some inconvenience of telephone.
    - Introduced other inconvenience.
  - Others: transceivers, amateur radio, ...

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

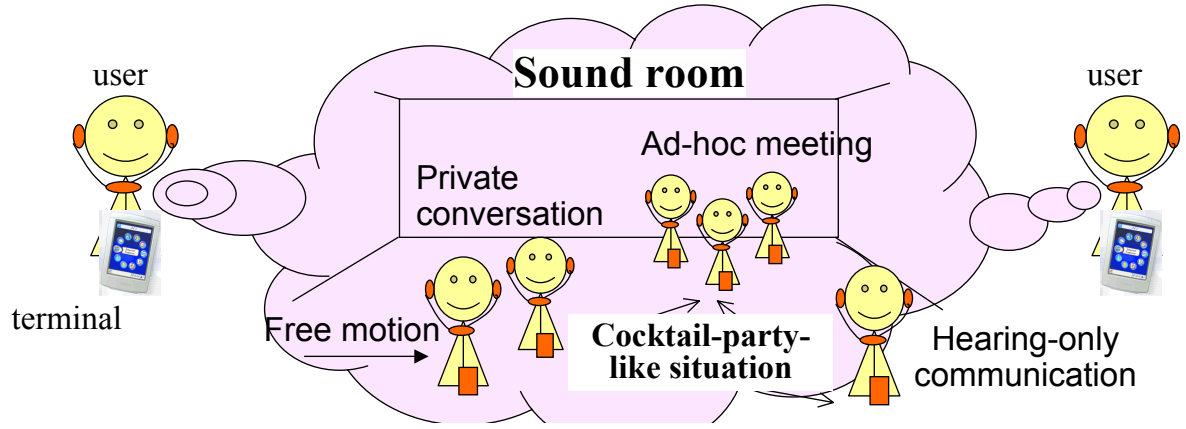
# Background (cont'd)

- **VCM should be innovated.**
  - In face-to-face communication, various communication patterns are available.
    - E.g., free conversation with two or more talkers.
  - Communication patterns through VCM are limited.
- **Specific problems in VCM**
  - Speaker identification problem
    - Difficult to *identify* and to *remember* the speaker especially in conventional audio only environments.
  - Multiple talker problem
    - In face-to-face communication, parallel conversations often occur.
    - They are difficult through VCM.



# Introduction to voiscape

- **"Sound room"**
  - A virtual space is expressed by sound directions and distances (i.e., by spatially located sounds).
  - People in the room can move freely.
- **voiscape is a type of VCM that uses sound rooms.**
  - "Places of communication" are created in a sound room.




**Multiple places of communication in a room**

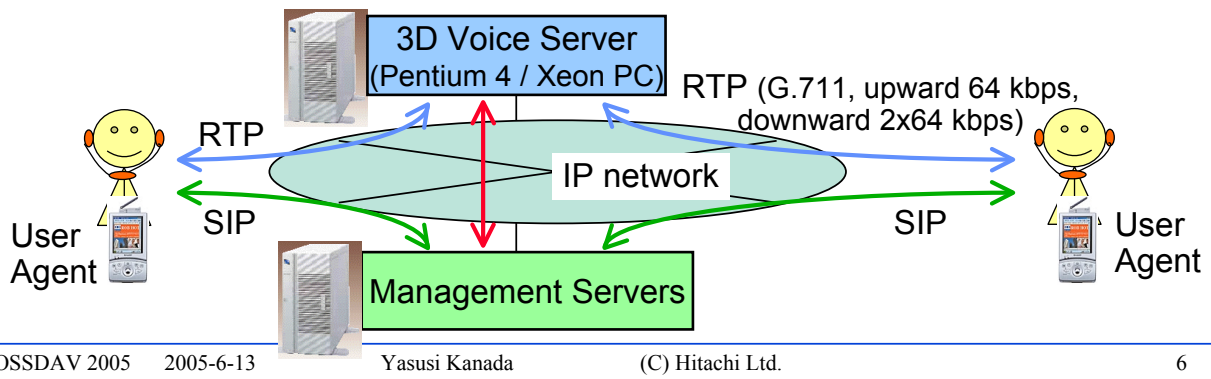
# Prototypes of voiscape

- **Jasper:**
  **The first prototype was presenced in CCN 2004.**
  - Java-based (JMF, Java3D, and LWJGL (light-weight Java Game Library))
  - Built-in VoIP and 3-D audio
    — sound quality was not good
- **VPII (Voiscape Prototype II):**
  **This presentation focuses on the second prototype.**
  - C++ and C based — to get better performance
  - VoIP (RTP) and 3-D audio are developed from scratch.



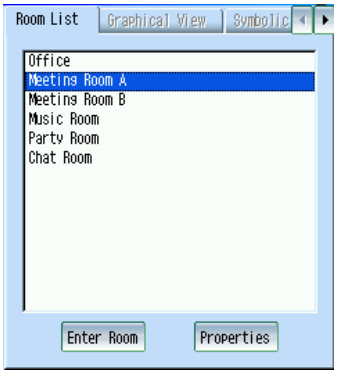
# Architecture of VPII

- **Three major elements of VPII**
  - User Agent (UA)
    - Terminal software on Linux PDA (Zaurus) or Windows PC
    - Ethernet or wireless LAN
  - Management Server Collection (RMS, RLS, SIP registrar)
    - Room, user locations, and room list management by using SIP and SIMPLE (SIP for Instant Messaging and Presence Leveraging Extensions).
  - 3D Voice Server (or media server)
    - Spatialization and mixing
    - No DSP now

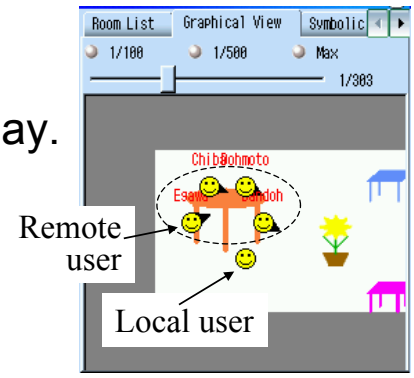


# User Interface of VPII

- **User select a sound room from a list.**
  - RLS sends the room list to UA.



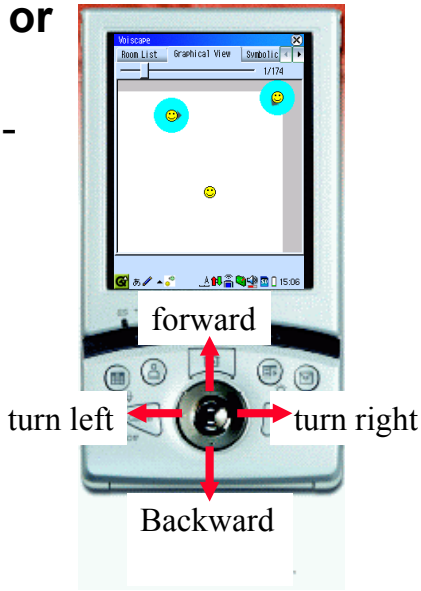

- **UA displays the sound room.**
  - Auditory display — the main display.
  - Visual map — a supplementary display.
  - Combination
    - User can map a voice and an icon.





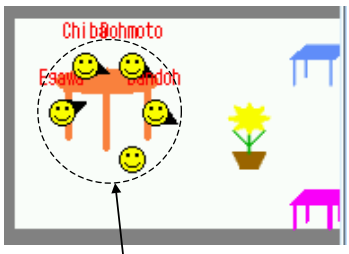

# User Interface of VPII (cont'd)

- **User can move by using cursor keys or other pointing devices.**
  - This motion is independent from real-world motion.



# Features of VPII

- **Low-delay motion-tracking spatial audio**
  - The sampling rate is 8 kHz.
  - HRIR (HRTF) and early reflections are computed.
  - Spatialization delay is minimized to enable bi-directional communication.
  - User motion is reflected in the sound in real time.
- **Virtual-place-based selective communication**
  - User can select a “place of communication” by using a map and icons.
  - Icons can be used as “landmarks”.
- **SIMPLE-based sound room management**
  - User’s location and orientation are treated as part of room presence.
  - SIMPLE is used for presence event (motion) notification.



Place of communication

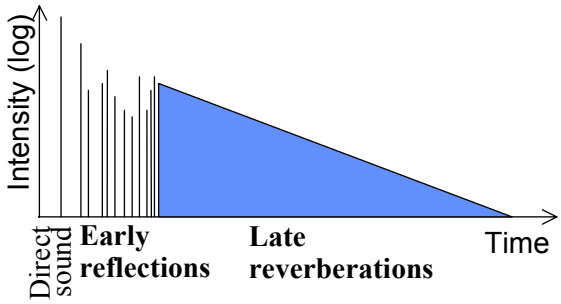


# More on Low-delay Motion-tracking Spatial Audio

- **HRIR (head-related impulse response)**
  - To minimize the delay, HRIR is applied to direct sounds in time domain.

# More on Low-delay Motion-tracking Spatial Audio (cont'd)

## ■ Reverberations



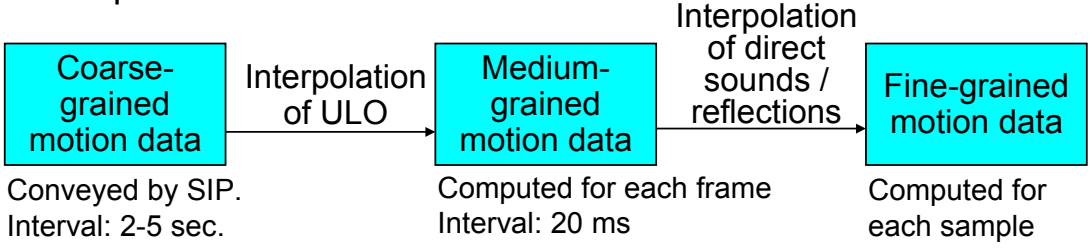

- ◆ Reverberations consist of
  - ▮ Early reflections
  - ▮ Late reverberations
- ◆ Only early reflections by the sound room walls are computed in VPII by a 2-D image source method.
- ◆ Early reflections are added because they cause
  - ▮ Out-of-head localization.
  - ▮ Feel of distance.
- ◆ No late reverberations because they have
  - ▮ No explicit advantage
    - – They are unnecessary for out-of-head localization or feel of distance.
  - ▮ Harms
    - – They tend to make the voices unclear.
    - – They are computationally expensive.



# More on Low-delay Motion-tracking Spatial Audio (cont'd)

## ■ Motion tracking

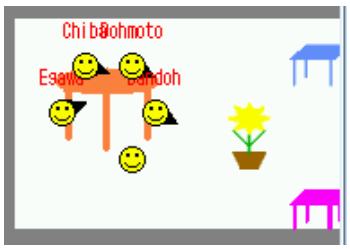
- ◆ Problems caused by a quick user motion
  - ▮ Click noises
  - ▮ Users' identity misses: fail to identify a user before and after a motion.
- ◆ Three interpolation methods for solving the problems
  - ▮ Interpolation of user locations and orientations (ULO)
  - ▮ Interpolation of direct sounds
  - ▮ Interpolation of reflections




- ◆ Interpolation of reflections is omitted in VPII because it is expensive and noises caused are small.

# More on Virtual-place-based Selective Communication

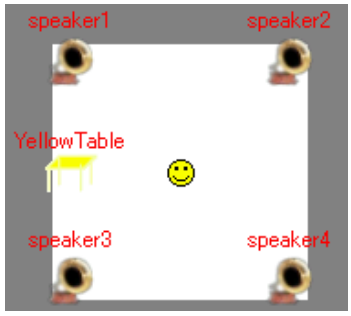
- **A 2-D view is used because**
  - Easier to map sound sources in auditory and visual displays than 3-D views.
    - Mapping the direction
    - Mapping the distance






# More on Virtual-place-based Selective Communication (cont'd)

- **Icons and landmarks**
  - Three types of objects in VPII
    - Persons (users)
    - Speakers (streaming sources)
    - Stationary objects — tables, plants, etc.
  - Objects are represented by icons.
    - Visual icons are shown on the map.
    - Auditory icons are heard in some situations.
    - Each user can use a default icon or his/her own icon.
  - Stationary objects can be used as landmarks.
    - You can specify a place by a landmark: "Let's meet at the pink table." — a new place of communication will be created.

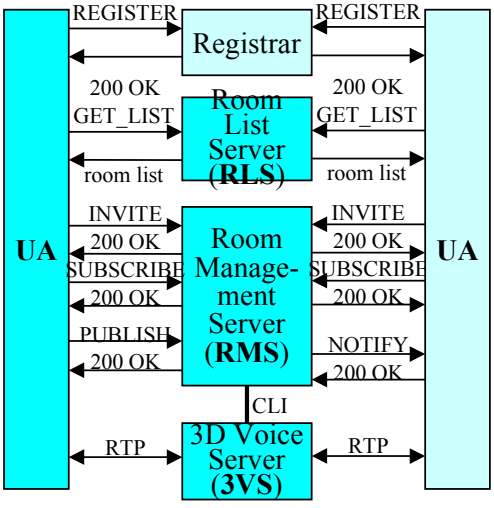
# More on Virtual-place-based Selective Communication (cont'd)

- **Other features**
  - Distance-based communication and awareness control
    - Each user is surrounded by a circular area called an *aura*.
    - If a remote user comes into the aura,
      - Both the local and remote user is made aware of this.
      - The local user hears the remote user's auditory icon.
      - The remote user hears a warning sound.
  - Privacy protection
    - Distance-based communication *policies* can be specified.
      - Connection and disconnection policies, etc. [Kanada 2004]
  - User-motion control
    - Long motion (long push of a cursor key)
    - No warping should not be allowed based on Benedikt's cyber space principles.



# More on SIMPLE-based Sound Room Management

- **Three types of messaging**
  - Room entrance and exit
    - To enter a room, UA sends INVITE to RMS.
    - To exit from a room, UA sends BYE to RMS.
  - Room presence management
    - UA sends PUBLISH that contains the ULO to RMS.
    - UA sends SUBSCRIBE that requests other users' ULO to RMS.
    - RMS "replies" with NOTIFY that contains other users' ULO.
  - Room list management
    - UA sends SUBSCRIBE that requests a room list to RLS.
    - RLS "replies" with NOTIFY that contains the room list.



# Informal Evaluation

- **VPII was informally evaluated with more than 200 people (who tried VPII mostly for only 5 to 10 minutes).**
- **Speaker identification and multiple talker problems**
  - People understand VPII can be used for cocktail-party-like conversations.
  - People could distinguish parallel conversations
    - by paying attention to, or
    - by moving toward one of them.



# Informal Evaluation (cont'd)

- **Three features of VPII**
  - Low-delay motion-tracking spatial audio
    - Most people were satisfied with 8-kHz sampling sound.
    - Vertical localization was not good (no vertical cue in 8-kHz sampling sound).
  - Virtual-place-based selective communication
    - Not yet evaluated.
  - SIMPLE-based sound room management
    - Presence propagation was delayed several seconds.
    - This delay should not be a major problem in conversation because no quick motion is required for conversation.

# Conclusion and Future Work

## ■ Conclusion

- ◆ VPII enabled parallel conversations in a sound room.
- ◆ SIMPLE-based management generally works well in VPII.

## ■ Future work

- ◆ The user interface requires much more evaluation and improvements.
  - ▮ A more detailed evaluation is ongoing.

No real demo is available here, but prerecorded sound samples are available.



# User Interface of VPII (cont'd)

## ■ Position and direction of local user on the screen is fixed.

- ◆ The room rotates when the user turns.

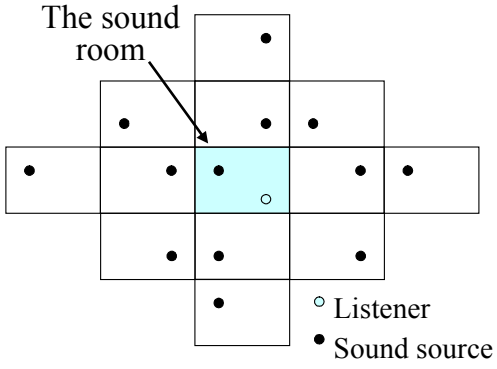
# More on Low-delay Motion-tracking Spatial Audio (cont'd)*

- **Sampling rate is 8 kHz.**
  - The reasons why 8 kHz is used are
    - Reasonable communication bandwidth and delay.
    - Real-time signal processing.
    - Narrow bandwidth of voice.



# More on Low-delay Motion-tracking Spatial Audio (cont'd)*

- **Method of calculating early reflections**
  - 2-D image source method with 12 reflections is used.
  - To reduce the amount of computation,
    - Early reflections are spatialized by controlling ITD and IID.
      - ITD = interaural time difference
      - IID = interaural intensity difference
    - Same HRTF is used regardless of the direction of early reflections.



# More on SIMPLE-based Sound Room Management (cont'd)*

- **The reasons why SIP and SIMPLE are used.**
  - Standard protocol
    - SIP and SIMPLE are IETF standards.
    - SIP can be used for interconnection with IP telephony systems.
  - Flexibility
    - VPII functions can easily be implemented by SIP/SIMPLE.
  - Economy
    - Other protocols are not necessary; SIP/SIMPLE can be used throughout VPII.