社団法人 人工知能学会
Japanese Society for
Artificial Intelligence

人工知能学会研究会資料
JSAI Technical Report
SIG-Challenge-B102-3

UI-ALT: 音の選択聴取を可能とする実世界アバタのためのユーザインタフェース

UI-ALT: User Interface for Avatar-based Listenable Telepresence

○植田 俊輔，今井 倫太，中臺 一博，中村 圭佑

Shunsuke Ueda, Michita Imai, Kazuhiro Nakadai and Keisuke Nakamura

慶應義塾大学

Keio University

(株) ホンダリサーチインスティチュートジャパン

Honda Research Institute Japan Co., Ltd.

ueda@ayu.ics.keio.ac.jp

Abstract

In a telepresence situation, a remote user has difficulties in catching and joining conversations because the user has to listen to the mixture of sound sources via a user interface. To relax this problem, this paper proposes User Interface for Avatar-based Listenable Telepresence (UI-ALT). A remote user can see scenes and listen to conversations via a real world avatar like a telepresence robot having a camera and microphone array. The user selects a conversation by marking persons of interests as a circle or a line on a UI-ALT display. The user can listen only to the selected conversation even when several conversations occur simultaneously because sound source separation with the microphone array eliminates non-target sound sources. Through offline evaluation, we showed the effectiveness of UI-ALT in a telepresence situation.

1 はじめに

人間は雑音環境においても音を聴き分けることができる．例えば，パーティのような多くの雑音が存在する環境の中でも人間は自分が興味のある会話を選択的に聴き取ることが出来る．この現象は「カクテルパーティ効果[1]」という名称で知られている．しかし，テレプレゼンスロボットがこのような雑音環境に置かれた場合，遠隔ユーザは遠隔地でどのような会話が行われているのかを理解することは困難である．

近年，テレプレゼンスアバタとしてのロボットが様々な方法で研究されており[2][3][4]，Anybots 社の QB[4]のように実用化されている例もある．これらのロボットは遠隔地で存在感を示し，人間の代わりにタスクをこなすことが期待されている．しかし，これらのロボットは人間とのインタラクションに必要な音声情報をうまく処理することが出来ないため，高雑音環境での人間とのインタラクションが難しいと考えられる．日常環境の中には大抵音声を含む複数個の音源が存在しており，人間とインタラクションを行うにはこうした複雑な音環境の理解が必要となる．

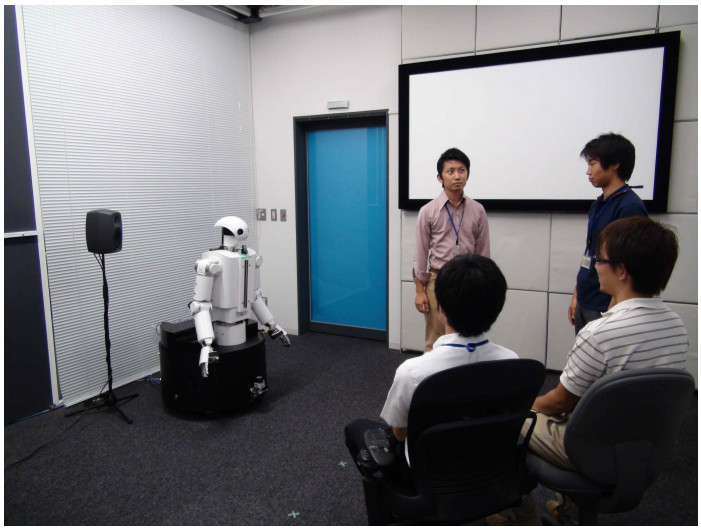

Figure 1: Avatar robot in a noisy room

本稿では，実世界アバタを対象として音の選択聴取を可能とするユーザインタフェース UI-ALT を提案する．UI-ALT はマイクロフォンアレイを搭載したアバタロボットを使用したインタフェースであり，マイクロフォンアレイ処理によって提供される音源定位および分離機能によりユーザは UI-ALT を通して望む方向の音を選択的に聴取することが出来る．つまり，UI-ALT を用いることで音の聴き分けを行うアバタロボットが実現可能である．

また，UI-ALT のユーザは UI 上で簡単なコマンドを入力することで音の選択聴取が出来る．水本らの研究[5]で

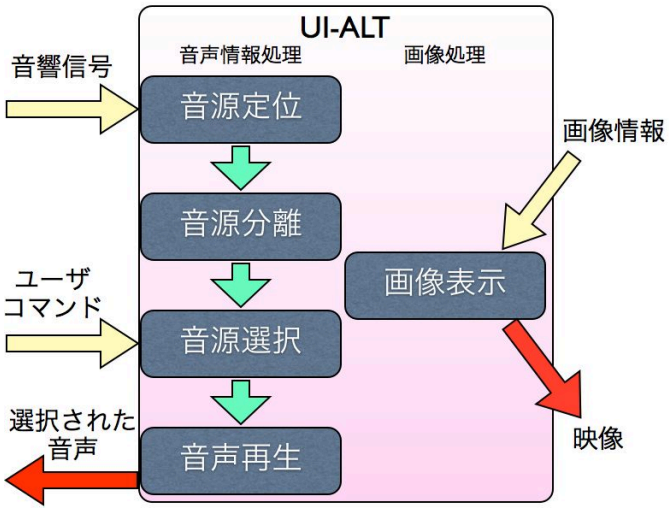

Figure 2: System architecture of UI-ALT

Figure 3: Snapshot of UI-ALT screen

は Willow Garage 社のテレプレゼンスロボット Texai に音の選択聴取が出来るユーザインタフェースを実装した．しかし水本らの研究では遠隔ユーザが分離された音を聴く際に，音の方向と幅の２つのパラメータを操作しなければならないため，実際にユーザが分離音を聴く際に煩雑な操作が要求される．このため，実際に遠隔ユーザはスムーズなインタラクションを行うことが出来ない．UI-ALT では，ユーザが UI の画面上で聴きたい方向を囲む，もしくは線を引くことで分離音が聴取出来るため，ユーザにとって簡単な操作で分離音を聴くことが出来る．

本稿は次の通りに展開する，第２節では UI-ALT のシステム構成について述べる．第３節では UI-ALT が応用可能なインタラクションの例について述べる．さらに第４節では，UI-ALT の有用性を示す為にオフラインで行ったディクテーション実験について述べ，最後に第５節でまとめと今後の課題を示す．

2　システムアーキテクチャ

UI-ALT のシステム構成図を図２に示す．

UI-ALT のユーザは，図３に示すように画面上で複数人が同時に喋っている中で会話を聴きたい方向の人に対してマウスを用いて線を引いたり丸で囲ったりすることでその方向の分離音を聴くことが出来る．この機能は，図２の中にあるオープンソースなロボット聴覚ソフトウェア HARK[6][7]を利用した音源定位・分離のモジュールによって実現される．

音源定位や分離，遠隔地のカメラ映像などはすべて ROS（Robot Operating System）[8][9]のメッセージで通信を行う．UI-ALT では音声データとカメラデータを同時に扱うため，処理が重くなってしまう可能性がある．そこで ROS が提供するメッセージを用いて通信を行うことにより，音声波形信号や音源 ID，カメラ画像情報など多種にわたるデータを小さい遅延で通信することが可能である．

以下の小節では UI-ALT の主要モジュールである音源定位，音源分離，音源選択の各モジュールについて詳しく述べる．

2.1　音源定位モジュール

入力である音響信号は最初に定位モジュールに送られる．定位モジュールではどの音がどの方向から来ているのかを推定することが出来る．音源の定位には HARK で提供されている雑音に頑健で，複数音源の定位が可能な MUSIC(MUltiple SIgnal Classification)[6]を用いる．MUSIC により，複数音源の水平方向の定位が可能となる．定位情報は入力音響信号とともに音源分離モジュールへ渡される．

2.2　音源分離モジュール

音源分離モジュールでは，選択的な会話の聴取を実現するために．定位情報と入力音響信号（混合音）から各音源信号を分離する．UI-ALT では HARK で提供されている GHDSS(Geometric-constrained Highorder Decorrelation-based Source Separation)[6]を用いて音源分離を行う．分離された音源情報は UI-ALT の音源選択モジュールへと送られる．

2.3　音源選択モジュール

音源選択モジュールはユーザのコマンドによって分離された音源を選択して音声再生モジュールに渡すモジュールである．ユーザがどの音源も選択していない場合は入力音響信号がそのまま再生モジュールに渡される．UI-ALT では図４に示すように選択したいグループを丸で囲う，選択したいグループの上に線を引くといった２種類の方法で音源選択をすることが出来る．

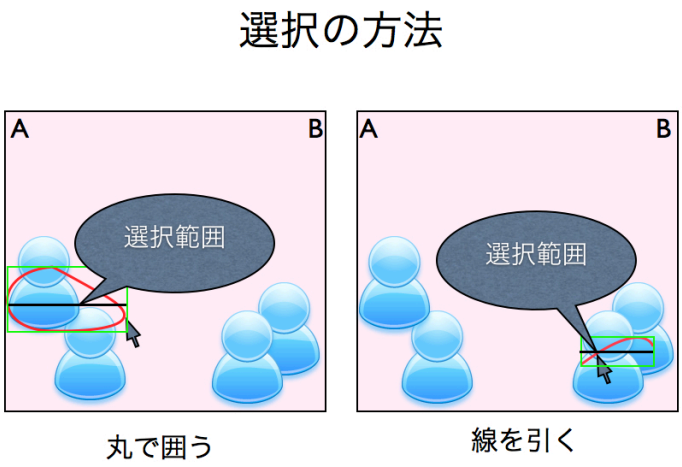

Figure 4: How to select sound source

ユーザがマウスモーションにより UI-ALT 上に円もしくは線を描き終わると，UI-ALT は以下の処理で音源選択を行う．

1. 描かれた円もしくは線の画像内における x, y 座標の最大値および最小値を取得する．
2. 選択範囲の x 座標の最大値および最小値をあらかじめ決められている USB カメラの画角から以下の式で角度に変換する．画像サイズは 640 × 480 であり，画像の中心が 0° である．

$$\theta = \pm \arctan\left(\frac{|x - 320| \times \tan\left(\frac{\text{カメラ画角}}{2}[deg]\right)}{320}\right) \quad (1)$$

3. 算出された角度範囲と音源の角度を比較して範囲に含まれていれば音源が選択されたと判断する．
4. 音声再生モジュールへ選択された分離音情報を送る．

UI-ALT では複数の音源を選択することも可能であり，複数選択された場合には選択された音源の数分の混合音が再生される．また音源選択を解除することも可能である．ユーザがマウスを右クリックすることで，音源選択状態をリセットして何も選択していない状態に戻すことが出来る．

3　応用可能なインタラクション場面

本節では，UI-ALT が実世界において応用可能であると考えられる場面について考察していく．具体例として以下に述べる 3 つの例を挙げる．

3.1　パーティ参加

ここでは，アバタロボットがパーティ会場にいて遠隔でユーザがパーティに参加する場面を考える．パーティ会場内では様々な場所で会話が行われていたり音楽が流れており，多様な音源が存在する．このため，遠隔ユーザはど

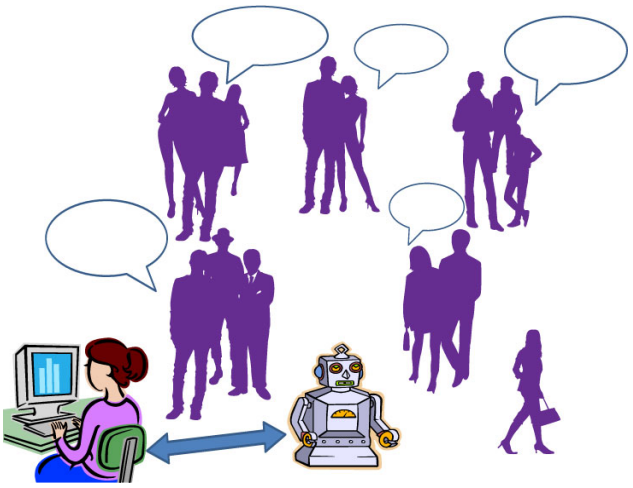

Figure 5: Avatar robot with UI-ALT at a party

のような会話が行われているのかを理解するのは難しい．仮に友人をパーティ会場で発見した場合でも遠隔ユーザは彼らが何を話しているのか理解することは難しい．そこでユーザは UI-ALT を用いて友人らを画面上で囲むことで友人らの会話の内容を聴くことができ，ユーザが実際にアバタロボットを操作して会話に参加することも可能になる．つまりユーザはまるで自分がそのパーティに参加しているかのような感覚を得ることができる．本例により，UI-ALT が可能とする音の選択聴取の有効性，またその結果として会話参加の容易性を表している．

3.2　レストランでの注文取り

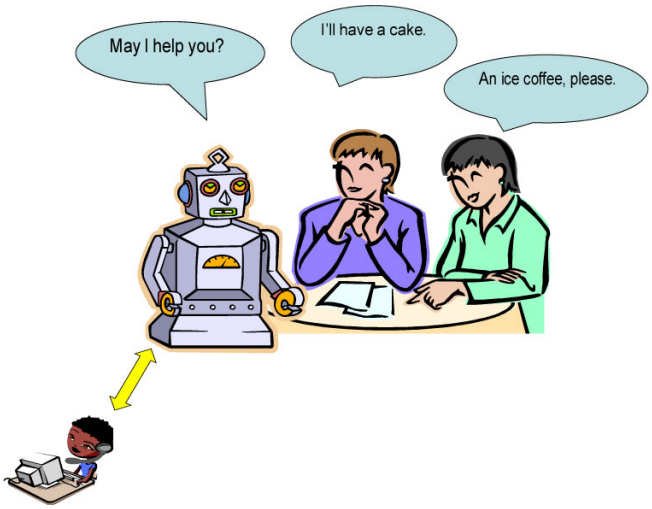

Figure 6: Avatar robot takes orders at a restaurant

ここでは，ファミリーレストランにおいてアバタロボットが従業員に変わって注文を取るという場面を考える．ファミリーレストランは家族連れなど様々な客層で賑わいをみせる場所であり，会話の音以外にも食事中に発生する音（フォークが皿に当たる音，グラスがぶつかる音など）が

ある高雑音環境である．遠隔ユーザはこのような雑音環境においても正しく注文を取るために，UI-ALT を用いて注文を取る人を画面上で選択することにより，遠隔ユーザは注文を正しく取るというタスクを遂行することができる．本例は，UI-ALT はファミリーレストランのような雑音環境において遠隔ユーザが対話タスクを遂行するために有用なシステムであることを表している．

3.3 会議

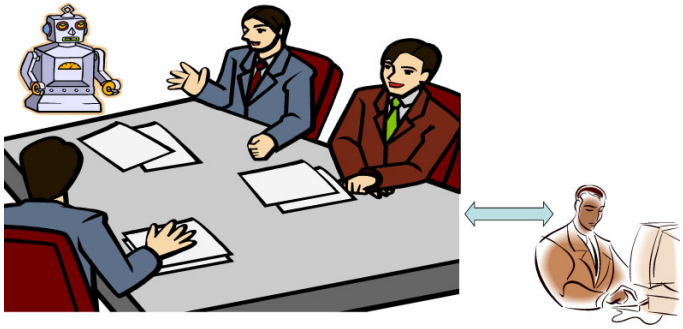

Figure 7: Avatar robot attends a meeting

ここでは，アバタロボットを通して遠隔でユーザが会議に参加する場面を考える．会議を行う際，時に活発な議論が行き過ぎて他の人の発言を聴かずに好き勝手に話し出してしまい，会議自体が収拾がつかないことがある．アバタロボットを通じて遠隔で会議の様子を見ているユーザにとっては会議室で発生しているすべての発言を聴き取ることは困難である．しかし，こうした発言の中に重要なキーワードが含まれている可能性もあるため，ユーザは出来るだけすべての発言を拾いたいと考える．UI-ALT を利用することで画面を見ながら気になる発言をしているユーザの発言を選択的に拾うことができる．UI-ALT は遠隔で会議のログを取る際にも有用であると考えられる．

以上で挙げた例から，日常環境におけるインタラクションにおいて，音声情報が必要不可欠であることがわかる．UI-ALT は雑音環境における人間とアバタロボットとのインタラクションに有用なユーザインタフェースとなりうる．

4 オフライン実験による評価

雑音環境における UI-ALT の有用性を示すために，本稿では UI-ALT を用いてユーザにディクテーションを行ってもらうオフライン実験を行った．本節では実験設定，実験結果，結果に対する考察を述べる．

4.1 実験設定

ユーザ実験を行う前に，別室でパーティ会場を想定した環境で音声と画像の録画を行った．実験室では雑音としてバックグラウンド音楽を流し，パーティに近い設定とした．4 人の大学生を実験室に集めて 2 人 1 組のグループを作ってもらい，図 8 のようにアバタロボットの正面から ±30° の方向に立ってもらった．音声の録音は頭部に 8 チャネルのマイクロフフォンアレイを搭載したアバタロボットを使用し，映像の録画には USB カメラを使用した．

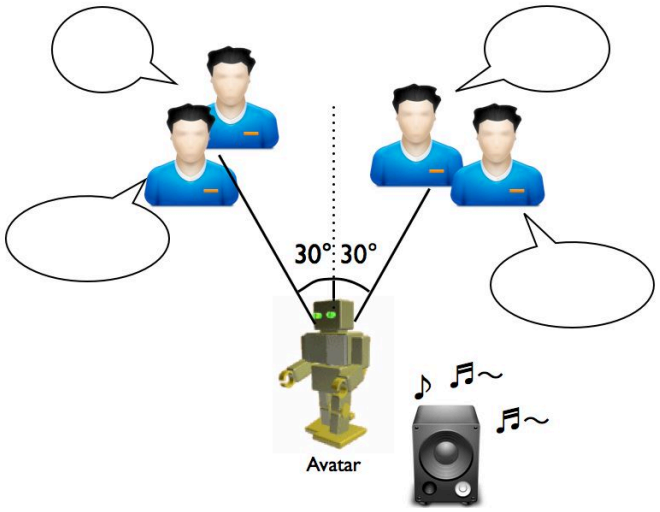

Figure 8: The location of the avatar robot and people during experiment. A loudspeaker plays background music.

ディクテーションのトピックとして両方のグループでお互いの自己紹介を行ってもらった．具体的な話題として，会話中にお互いの名前，出身，所属，趣味の 4 つの話題についてかならず触れてもらった．UI-ALT を使う場合と使わない場合を比較するために，同じようなシーンをグループ構成を変えて 2 種類録画を行った．各グループの発話の様子の例を図 9 に示す．

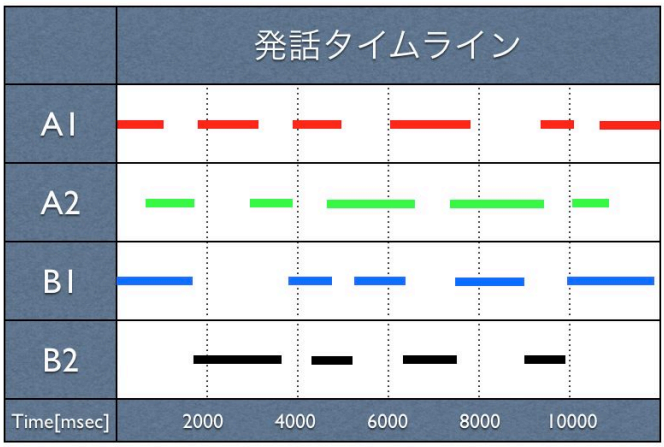

Figure 9: Timeline of each person's remark

本実験では UI-ALT のユーザとして 8 人の大学生に実

際に UI-ALT を用いてディクテーションタスクを行ってもらった．８人のうち４人は事前に UI-ALT の使い方を学ばずに使用してもらい，残りの４人は事前に１回だけ使い方を学んだ上で使用してもらった．実験では，各被験者は事前に撮影した２種類のビデオをランダムな順番で観てもらった．１度目は UI-ALT を使わずに映像と音声をそのまま流し，２度目は UI-ALT を用いて聴きたい会話を選択しながら実験を進めてもらった．被験者には映像内の２つのグループによる自己紹介の話題としてあげられていた内容を解答用紙に書き出してもらった．

本実験では我々は以下に挙げる項目について観察を行った．

- ディクテーションの正答率
- ユーザによる音源選択の仕方
- ユーザによる音源選択のスピード

4.2 結果

図 10 はディクテーションタスクにおける各被験者の正答率，事前練習を行わなかったグループの平均正答率，事前練習を行ったグループの平均正答率，および全体の平均正答率を UI-ALT を使った場合と使わなかった場合で比較した結果である．グラフの縦軸は正答率，横軸は各ユーザの ID を表す．UI-ALT を使った場合の全体の平均正答率は 76%であったのに対し，UI-ALT を使わなかった場合の平均正答率は 35%にとどまった．また，UI-ALT を事前に練習しなかったグループが UI-ALT を使った場合の正答率が 67%であったのに対し，UI-ALT を事前に練習したグループが UI-ALT を使った場合の正答率は 85%となった．平均正答率の結果を見ると，ユーザが UI-ALT を使った場合は使わなかった場合より２つのグループの会話の内容が理解出来ているということが言える．

ユーザによる音源選択の仕方については，一つのグループを長い時間選択しているユーザもいれば，頻繁に選択するグループを変えるユーザも見受けられた．選択のスピードについても，素早く選択しているユーザもいれば，ゆっくり選択しているユーザも見受けられた．

4.3 考察

実験結果より，UI-ALT を使った場合，ユーザのディクテーションの正答率にかなりの向上が見受けられる．このことから，UI-ALT は高雑音環境であっても会話内容の理解を支援するツールであると言える．

しかし，UI-ALT を事前に練習しなかったグループの中に，どちらの音声を選択してよいかわからずにビデオの再生が終わってしまい，ディクテーションタスクに回答出来なかったユーザも存在した．この現象の原因の一つとして考えられるのは UI-ALT の映像のフレームレートの低さである．今回の実験では遅延をなるべく小さく抑えるためにビデオのフレームレートを落として実験を行った．しかし，実験後に行ったアンケートからユーザは話者を選択する際に話者の口元や表情を見てある程度決めているという知見が得られた．ディクテーションタスクに回答出来なかったユーザはどちらのグループが何の話題について話していたのかが音声情報だけでは理解出来ず，映像のフレームレートも悪かったためにどちらのグループを選択してよいか混乱してしまったと考えられる．このことから，音源選択の際には視覚情報が聴覚情報と同じぐらい重要な役割を果たしているということが言える．

また，実験後のアンケート結果から，被験者のうちの半数が UI のマウス操作が複雑なため音源選択に苦労したという回答を得た．実験から，ユーザによって選択の仕方やスピードの違いが様々異なることが見受けられたが，ディクテーションタスクの正答率と比較してみると，素早く選択しているユーザほどより良い正答率を出しているという傾向が見られた．このことから，UI-ALT は音源選択の際に有効ではあるが，必ずしもすべてのユーザに対して直感的なインタフェースではないことがわかる．今後はユーザが望む音源を素早く選択出来るように最適な選択方法を調べていく必要がある．

5 まとめ

本稿では，実世界アバタを対象として，音の選択聴取機能を有するユーザインタフェース UI-ALT を提案した．UI-ALT は人間とアバタロボットとのインタラクションにおいて欠かすことの出来ない音声情報を扱えるインタフェースであるため，実世界の様々な環境に適用可能であると考えられる．本稿では実際に UI-ALT の応用が可能であると考えられる３つのインタラクションシナリオを示し，UI-ALT を用いることによって遠隔ユーザが雑音環境の音をアバタを通して聴く際に聴きやすくなったことをディクテーション実験により示した．

今後の課題として，まずインタフェースの改善が挙げられる．オフライン実験から，ユーザは話者を選択する際にある程度画面を見ながら選択しているという傾向が見られたので，UI の画像を見やすくする必要がある．また，選択の仕方も人それぞれであるということから，どのような選択の仕方が一番ユーザに取って使いやすいのかを調査する必要がある．

UI-ALT を用いたオンライン実験も計画している．今回のオフライン実験で得られた知見を基にアバタロボットを操作出来るようインタフェースを改良し，実際にパーティにアバタロボットを参加させて遠隔でユーザに参加してもらうといった実験を行っていく予定である．

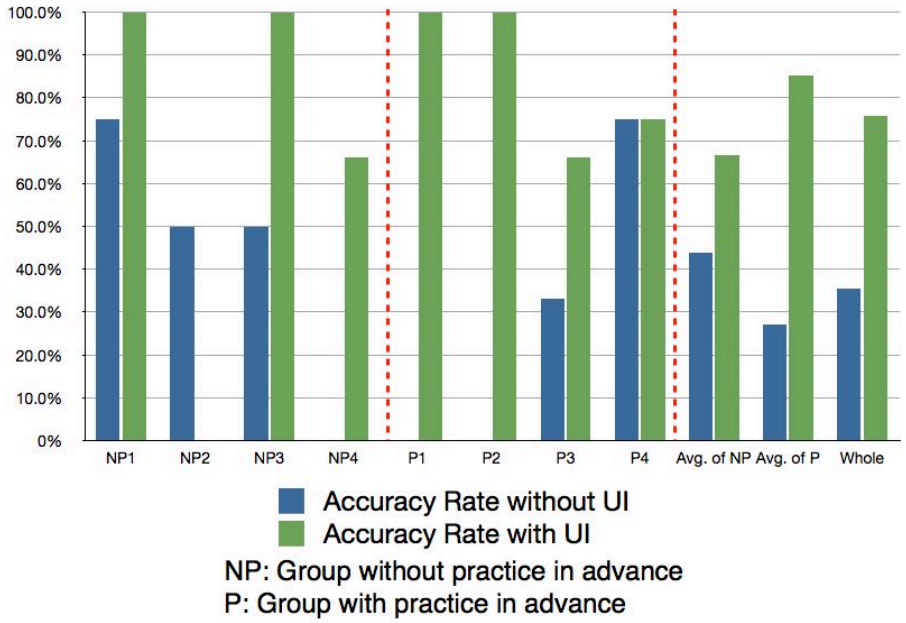

Figure 10: Result of accuracy rate in dictation task

参考文献

[1] Cherry E. Colin: Some Experiments on the Recognition of Speech, with One and with Two Ears, in *The Journal of the Acoustical Society of America*, *vol.25*, pp.975–979, 1953.

[2] Sigurdur Orn Adalegeirsson, Cynthia Brezeal: Mebot a robotic platform for socially embodied telepresence. in *Proc. of ACM/IEEE International Conference on Human-Robot Interaction(HRI)*, pp.15-22, 2010.

[3] Nishio, S, Ishiguro, H., Anderson, M., Hagita, N.: Repesentating personal presence with a teleoperated android: A case study with family. in *Proc. of AAAI 2008 Spring Symposium on Emotion, Personality, and Social Behavior*, pp.96-103, 2008.

[4] Anybots -Your Personal Avatar- : `http://www.anybots.com` .

[5] Takeshi Mizumoto, Takami Yoshida, Kazuhiro Nakadai, Ryu Takeda, Takuma Ohtsuka, Toru Takahashi, Hiroshi G. Okuno: Design and Implementation of Selectable Sound Separation on a Texai Telepresence System Using HARK in *Proc. of IEEE-RAS International Conference on Robotics and Automation(ICRA)*, pp.2130-2137, 2011.

[6] Kazuhiro Nakadai, Toru Takahashi, Hiroshi G. Okuno, Hirofumi Nakajima, Yuji Hasegawa, Huroshi Tsujino: Design and Implementation of Robot Audition System "HARK" in *Advanced Robotics*, *vol.24* pp.739-761, 2010.

[7] HARK Main Page: `http://winnie.kuis.kyoto-u.ac.jp/HARK/` .

[8] Morgan Quigley, Brian Gerkey, Ken Conley, Josh Faust, Tully Foote, Jeremy Leibs, Eric Berger, Rob Wheeler, Andrew Ng: ROS: an open-source Robot Operating System in *IEEE-RAS International Conference on Robotics and Automation (ICRA) Workshop on Open Source Software in Robotics*, 2009.

[9] ROS: `http://www.ros.org`