音声対話システムにおける音環境への反応表出による アフォーダンスの評価

Evaluation of Affordances by Representing Reaction to Sound Environment in Spoken Dialogue Systems

夏目龍司[1] 李晃伸[1]
Ryuji Natsume[1] Akinobu Lee[1]

[1]名古屋工業大学
[1]Nagoya Institute of Technology

Abstract: This study focuses on affordances in spoken dialogue systems, and proposes a spoken dialogue system which reacts to the sound environment around the system using agent. In order to allow people to talk to spoken dialogue systems easily, we hypothesize that the systems need to provide affordances that cannot only be used through voice interaction, but can also interact intellectually. Finally, we shall report the results which compare the proposed methods and traditional main affordances of voice input, such as sound level meter and icons.

1 はじめに

音声技術の発展により，音声を入力インタフェースとして取り入れた音声対話システムが普及してきている．音声対話システムはユーザと機械がコミュニケーションを行うシステムである．近年では，カーナビゲーションシステムやスマートフォンなどに搭載され，キーボードやマウスなどを用いず直感的に操作できるため，今後さらに広く利用されることが期待される．しかし，機能的に充実した音声対話システムが実用化される一方，現状音声対話システムや音声インタフェースを日常的に利用する光景を見る機会は少ない．その要因の一つとして話しかけにくさが挙げられ，ユーザにとって話しかけやすいインタフェースを有した音声対話システムの構築のためには，ユーザの機械との対話に対するバリア低下に向けた研究が必要である．

人間はモノを利用するとき，そのデザインや特徴から使い方を直感的に捉え判断することができる．このようにモノの使い方を決定する基礎的な特徴のことをアフォーダンス[1]と呼ぶ．音声対話システムにおいても，システムを前にして自然に使えるようなインタフェースをデザインすることが重要である．そこで本研究では，話しかけやすさの改善を目指して音声対話システムにおけるアフォーダンスを考え，システム周囲の音環境に対して対話エージェントが反応を表出する音声対話システムを提案する．

2 音声対話システムにおけるアフォーダンス

音声対話システムは，2 つのアフォーダンスを表出する必要があると筆者らは考えている．1 つ目は入力手段は音声であること（本研究では音声入力のアフォーダンスと呼ぶ），2 つ目は知的なインタラクションが可能であること（本研究では理解のアフォーダンスと呼ぶ）である．以下，本研究で仮定する音声入力のアフォーダンスと理解のアフォーダンスについて述べる．

2.1 音声入力のアフォーダンス

音声入力のアフォーダンスとは，システムへの入力手段が音声であることを表出し，システムの使用方法を提供する特徴と定義する．主要な表出方法例として物理マイクやレベルメータ，アイコン，文字表示などが挙げられる．システムへの入力しやすさは優れており，音声認識や音声検索において効果的である．

しかし，人間同士のコミュニケーションにおける現象や概念が基幹となる音声対話システムでは，ただモノを扱う場合と異なり，知的なインタラクションができることも表出する必要があると考えられる．そこで，音声入力のアフォーダンスだけでなく，新たに理解のアフォーダンスを導入する．

2.2　理解のアフォーダンス

理解のアフォーダンスとは，システムが物事や環境などを理解しているよう表出し，システムとのインタラクションが可能であることを提供する特徴と定義する．雑談やインタラクション自体を目的としたシステムも存在するため，ユーザがシステムを前にして，発話をシステムが理解してくれるということをユーザに感じさせる必要がある．音声対話システムの設計に音声入力および理解のアフォーダンスを活かすことで，ユーザは人間に話しかける場合と同様の感覚で話しかけることが可能となり，ユーザのシステムに対する話しかけやすさの改善が期待できる．理解のアフォーダンス表出方法例は，語彙リストやタスク一覧表示などが挙げられる．しかし，これらは対話タスクの依存性が高い表出方法であるため，発話が限定的になってしまい，本来，自由な発話で操作することができる音声対話の良さを活かせない可能性がある．対話タスク非依存で理解のアフォーダンスを表出することが重要である．

本研究では，音声入力のアフォーダンスおよび理解のアフォーダンスを同時かつ対話タスク非依存で表出する方法として音環境への反応表出を提案する．

3　音環境への反応表出

音環境への反応表出とは，ユーザがシステムと対話を開始する前から，その環境で生じた音に対してシステムが反応を表出することによって，音が入力されているという音声入力のアフォーダンスとユーザが聞こえている音と同じ音を理解しているという理解のアフォーダンスを表出する．これによりシステムへのスムーズな話しかけの実現が期待できる．図 1 に提案する音声対話システムのイメージを示す．

レベルメータなどは音声入力のアフォーダンスを表出するが，理解のアフォーダンスは表出しない．タスク一覧などは理解のアフォーダンスを表出するが，対話タスクの依存性が高いうえ，音声入力のアフォーダンスは表出しない．音環境への反応表出は音声入力および理解のアフォーダンスを同時かつ対話タスク非依存に表出することができる．音環境への反応表出の位置づけとして，表 1 に各表出方法とアフォーダンスの関連を示す．

4　実験条件

4.1　実験システム

図 1：提案する音声対話システム

表 1：各表出方法とアフォーダンスの関連

	レベルメータなど	タスク一覧など	音環境への反応表出
音声入力	○		○
理解		○ (タスク依存)	○ (タスク非依存)

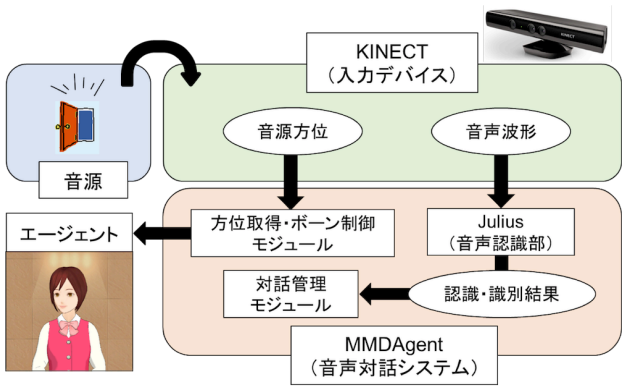

図 2：システムの概要

本システムは，エージェントを用いた音声対話システムを対象とし，MMDAgent[2]を用いる．

音環境への反応表出を実現するために音源方位推定と特定音の識別を実装する．音源方位推定により音の方向，また特定音の識別により音の種類を理解しているよう表出可能となり，アフォーダンス表出により有効であると考えたため 2 種類の実装を施した．図 2 にシステムの構成を示す．

音源方位推定では，マイクアレイを利用したビームフォーミングを利用し，各マイクへの音の到達時間差とマイク間の距離から音源の位置を推定している．外部から入力された音の音源方位を取得し，方位取得・ボーン制御モジュールでエージェントの首，目のボーンを音源の方向へ向けることで任意音に対して常時反応を表出する．

特定音の識別では，ガウス混合モデル(Gaussian Mixture Model; GMM)に基づいた環境雑音識別[3]を行うことで，特定音として識別可能にしている．外

部から入力された音の音声波形を Julius[4]によって処理し，特定音と識別された場合，コマンドメッセージを対話管理モジュールに流すことで音の種類に応じた反応を表出する．

今回の実験においては影響を一つずつ検証するため，音源方位推定のみを用いた反応表出とした．評価実験はモニタとして画面比率 32V 型の液晶ディスプレイを設置し，ディスプレイにエージェントとしてメイちゃん[1]を表示した．メイちゃんは，MMDAgent をベースに開発された双方向音声案内デジタルサイネージに使用されているキャラクタである．実験の様子を図 3 に示す．

音環境への反応表出方法として音源に顔を向ける動作を行う提案システム RS と表記する．音声入力のアフォーダンスを表出するシステムとしてレベルメータとアイコン表示（LI），対話タスク非依存で理解のアフォーダンスを表出するシステムとして吹き出し表示（BA）を採用し，比較検証する．図 4 にレベルメータとアイコン表示したシステム（LI），図 5 に吹き出し表示したシステム（BA）を示す．

4.2 タスク設定

実験は大学生, 大学院生の男性被験者 18 名で行った．被験者は 3 種類の音声対話システムを順不同で利用している．実験環境は研究室内環境を想定し，被験者と観察者の 2 名以外のいない静かな屋内で実験を行った．システム側から見た実験を行う部屋の様子を図 6 に示す．

被験者への事前説明として A4 の紙媒体で以下を提示し，口頭で説明した．

- これから 3 種類のシステムを利用してもらいます．
- 「エージェントのプロフィール」を尋ねてみてください．
- 時間の目安として 1 分 1 ベル，2 分 2 ベル鳴らします．
- 観察者が室内を行動しますが，気にしないでください．
- 実験中，観察者は何も答えることができません．
- システムごとに気持ちをリフレッシュして，気楽に臨んでください．

実験は，被験者が室内を自由に行動できるようにエージェントのプロフィールを尋ねるというタスクを設定し，音声対話システムを実稼働させた．

観察者はシステムに音環境に対する反応を表出を

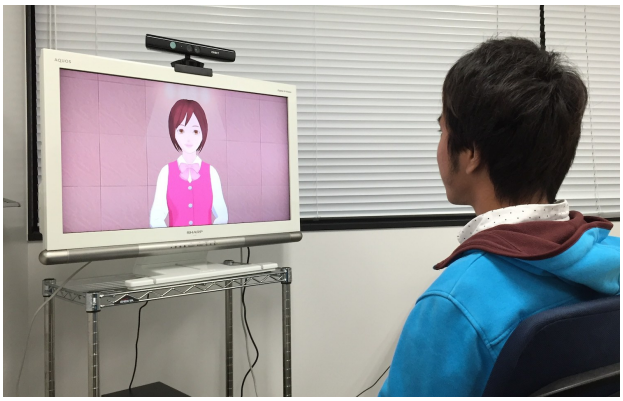

図 3：実験の様子

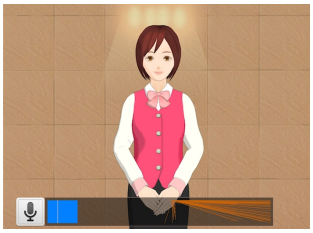

図 4：レベルメータとアイコン表示（LI）

図 5:吹き出し表示(BA)

図 6：システム側から見た実験環境の様子

させるため，タブレットからアラームを鳴らす，冷蔵庫を開閉する，棚から書籍を取り出す，椅子を動かす, 咳をする, ホワイトボードにペンを置くなど，音を発生させるための特定の行動をランダムにとる．それにより被験者の注意が散漫してシステムとの対話の妨げにならないよう事前説明を必要とするが，観察者の行動によってシステムに起きる変化は説明しないようにした．また，被験者は音声対話システムとの対話に関する一対一の評価をするため，被験者とシステムとの対話に観察者が介入しないようにした．

4.3 評価方法

被験者は，1 システム終了ごとに SD 法による 5 段階評価アンケートに回答した．アンケートは Bartneck らが提案した，ロボットの印象を評価するためのアンケート[5]を基に作成した．このアンケートには，人間らしさ，生き物らしさ，好ましさ，知

[1]

嫌い	1	2	3	4	5	好き
機械的な	1	2	3	4	5	有機的な
人工的	1	2	3	4	5	生物的
論理的に使える	1	2	3	4	5	直感的に使える
偽者のような	1	2	3	4	5	自然な
話しかけにくい	1	2	3	4	5	話しかけやすい
活気のない	1	2	3	4	5	生き生きとした
不親切な	1	2	3	4	5	親切な
無関心な	1	2	3	4	5	反応のある
ぎこちない動き	1	2	3	4	5	洗練された動き
意思を持たない	1	2	3	4	5	意思を持っている
親しみにくい	1	2	3	4	5	親しみやすい
話しにくい	1	2	3	4	5	話しやすい
人工的な	1	2	3	4	5	生物的な
ひどい	1	2	3	4	5	良い
不活発な	1	2	3	4	5	活発な
使いにくい	1	2	3	4	5	使いやすい
機械的	1	2	3	4	5	人間的
死んでいる	1	2	3	4	5	生きている
不愉快な	1	2	3	4	5	愉快な

総合評価

悪い	1	2	3	4	5	良い

図 7：被験者に提示したアンケート

性，安全性の指標を評価するための対立した形容詞項目が示されている．本研究では，人間らしさ，生き物らしさ，好ましさの指標を採用し，アフォーダンスに関連する項目を作成，追加した．なお，各指標を評価するための項目は被験者に意図を読まれないよう，ランダムに並び替え配置した．図 7 に被験者に提示したアンケートを示す．

5　実験結果・考察

図 8 にシステムの印象評価実験の結果を示す．まず，「直感的に使える」の項目で提案手法である RS が LI，BA を上回っているため，RS がアフォーダンスの表出方法として成立していたことが確認できる．

話しかけやすさの改善については，「話しかけやすい」の項目で RS，LI，BA の順で評価されており，1 発話目のバリア低減に RS が有効であることが示された．しかし，「話しやすい」の評価は LI，RS，BA の順という結果となった．これは一定以上の対話が続いた場合，すでに被験者はシステムがインタラクション可能であり，発話を理解してくれると分かったためであると推察される． よって，この差は LI と RS の音声入力のアフォーダンスとしての好ましさの影響であると考えられる．

第 2 節で述べた 2 つの音声対話システムにおけるアフォーダンスに関して述べる．音声入力のアフォーダンスについては，すべての追加項目で LI および RS が BA よりも良い結果となった．音声対話システムにおいて，音声入力のアフォーダンスの有効性を改めて確認できた．理解のアフォーダンスについては，対話タスク非依存での表出（BA）を単体で用いた場合の有効性は低かったが，音声入力のアフォーダンスに加えることで，話しかけやすさを改善することが分かった．

副次的結果として，人間らしさと生き物らしさの指標について述べる．どちらの指標もすべての項目において RS が LI と BA の値を上回った．また，人間らしさは LI よりも BA に優位性があることが確認できる．しかし，生き物らしさは BA よりも LI に優位性が見られた．特に「生き生きとした」や「活発な」，「反応のある」の項目で上回っていることが確認できる．レベルメータのような機械的な表出は人間らしさにはネガティブな印象を与えるが，生き物らしさにはポジティブな印象を与えることが分かった．生き物らしさはたとえ機械的であっても何かしらの反応や動作の有無に大きく左右されることが考えられ，アフォーダンスとの強い関連も見込まれることから，今後ランダムにエージェントを動作させた場合と比較することで，新たな知見が得られると予測される．

6　おわりに

本研究では，音声入力のアフォーダンスおよび理解のアフォーダンスを同時かつ対話タスク非依存で表出する方法として音環境への反応表出を提案し，評価実験を行った．結果，音声入力のアフォーダンスの有効性を改めて確認でき，理解のアフォーダンスにおいてもその存在の可能性や関連がみられた．また，提案システムが初見の話しかけに対するバリア低下に有効であることが示唆された．今後の課題として，より効果的な反応表出方法を調査し，音声入力のアフォーダンスおよび理解のアフォーダンスを同時に表出する新たな方法，実験条件で検証することでさらなる発見が期待できる．

参考文献

[1] Donald A. Norman : The Design of Everyday Things, Basic Books, pp.10‐13, (1988)

[2] 大浦圭一郎，山本大介，内匠逸，李晃伸，徳田恵一：キャンパスの公共空間におけるユーザ参加型双方向音声案内デジタルサイネージシステム，人工知能学会誌，Vol.28, No.1, pp.60–67, (2013)

[3] 中村敬介，西村竜一，李晃伸，猿渡洋，鹿野清宏：実環境音声情報案内システムにおける環境雑音及び不要発話の識別，電子情報通信学会技術研究報告，Vol.103, No.632, pp.13–18, (2004)

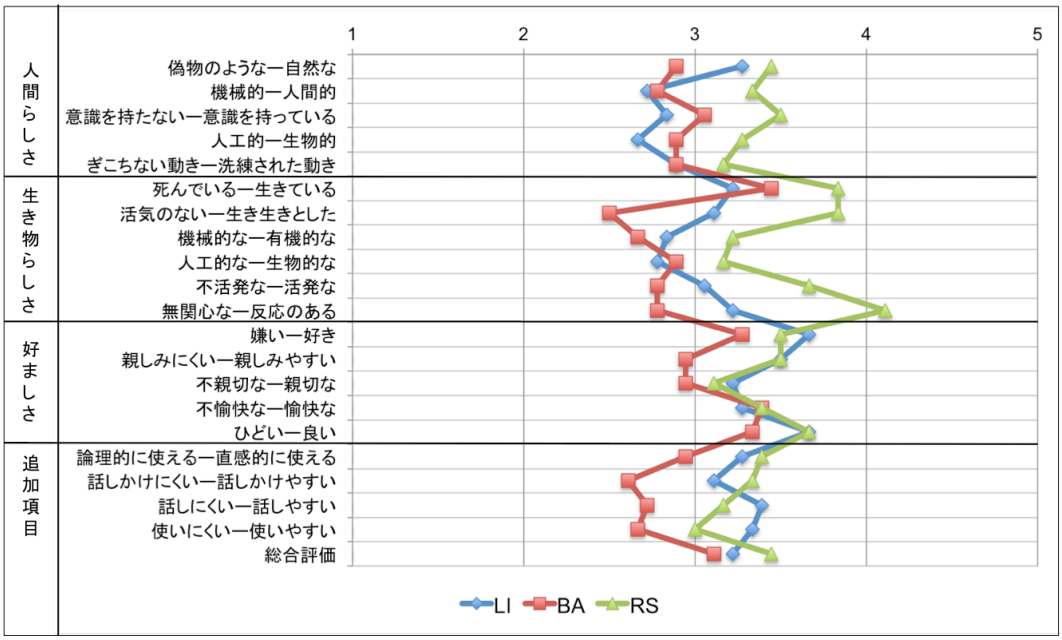

図 8：システムの印象評価実験結果

[4] 河原達也, 李晃伸：連続音声認識ソフトウェア Julius, 人工知能学会誌, Vol.20, No.1, pp.41-49, (2005)

[5] C. Bartneck, D. Kulic, E. Croft, and S. Zoghbi : Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots, International Journal of Social Robotics, Vol.1, No.1, pp.71-81, (2009)