音声言語情報処理 32−12
(2000. 7. 15)

文脈情報と韻律情報とを用いたあいづち語彙の使い分け — ユーザ印象の検討

塚原 渉[1]　　Nigel Ward[2]
[1]日立製作所　　[2]東京大学大学院工学系研究科

概要

音声認識の実用化に伴い，応用分野もビジネス用途から娯楽・教育用途にまで広がってきた．娯楽・教育用途ではインタラクション自体の魅力が大切だが，その際，従来システムのような不自然な応答は致命的となる．そこで，人間同士の友好的な会話では相手の状態を推測しながら応答を微妙に変えていく点(レスポンシブネス)に着目し，機械との対話におけるレスポンシブネスの有効性を検討した．学習ゲーム形式の会話(山手線駅名当てクイズゲーム)において，システムの確認発話応答あいづち(はい，うんなど)の使い分けルールを実装し，被験者13人に対して音声認識をWizard of Oz方式で行う会話実験を行った．その結果，コーパス中の出現比率であいづちを使い分けるよりも印象が良くなることが分かった．

Evaluating the Effectiveness of Subtle Choices in Acknowledgements in Japanese based on Prosodic Cues and Context

Wataru Tsukahara　　Nigel Ward
Hitachi, Ltd.　　School of Engineering, University of Tokyo

Abstract

As advances in speech recognition enable applications in entertainment, education and so on, users will demand that the interactions themselves be pleasant. Human-human interaction is pleasant in part because of the feeling that the other person is really listening and caring. That is, the other person picks up cues regarding the speaker's internal state at each moment and responds appropriately. To emulate this ability, a system must be able to infer the user's internal state and to use this information when choosing responses. We implemented this "responsive" ability for a memory game, using prosody and context to determine choice of acknowledgements. Most users did indeed prefer interacting with the responsive system, when preferences were measured using suitably sensitive techniques.

1. 背景

計算機能力および音声認識技術の向上により，音声認識が実用化されつつある．それに伴い，応用分野もビジネス用から娯楽・教育用へと広がってきた．これらの用途ではユーザのタスク遂行への動機は弱いため，システムとのインタラクション自体が自然で魅力的でなければ使用されない．しかし，この要素は従来研究が対象としてきた時刻表案内・観光案内・航空券予約などのビジネスタスクドメインにおいては重要視されていない．このような場面では，タスクを達成できればフラストレーションが生じない限りはインタラクションの質自体が問題になることはないと考えられていた．

[1]本研究は著者の東京大学大学院在学中に行われたものである．

[1]w-tsuka@crl.hitachi.co.jp
[2]nigel@sanpo.t.u-tokyo.ac.jp, http://www.sanpo.t.u-tokyo.ac.jp/~nigel/

一方，人間同士の友好的な会話では，不安や機嫌などの内部状態の推測に応じた応答変化 (レスポンシブネス) が観察される．従って発話の内容 (何を言ったか) や韻律 (どう言ったか — 図 1A) から推測される，ユーザの内部状態を考慮した応答は，自然で優しい印象のシステムを演出する (図 1B) のに有用であると予想されるが，その検証は行われていなかった．

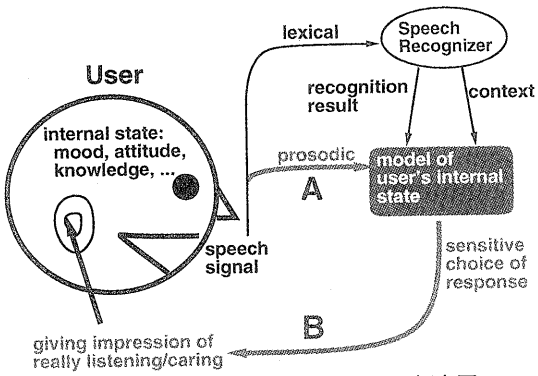

図 1　レスポンシブな対話システム概念図
A vision for responsive dialog system

また，韻律情報と感情の相関に関する報告は多い [4] が，怒りや悲しみなど，会話における表出頻度の低い感情が多く，感情認識には不向きである．ユーザの自信の度合いの推測 [1] のように，より実際的で微妙な内部状態の推測が必要である．また，研究の多くは分析に留まり，音声対話システムにおける韻律情報利用の有効性の検証は，ユーザの問い返しの意思の推測 [3] などに限定されている．

応答の生成 (図中 **B**) に関しては，応答のイントネーション変化による印象評価は行われているが [5]，応答語彙の変更の影響に関する知見は十分でなく [2]，検討の必要がある．

2. 目的

本研究では，ユーザの内部状態に応じた応答 (レスポンシブネス) が，ユーザに自然で優しい印象を与えると仮定し，これを検証することを目的とする．また，レスポンシブな音声対話システムの構築・評価手法に関する知見を得ることも目的とする．

そのための方法として，日本語会話において意味情報としては等しい応答語彙「はい」「うん」「そう」などのあいづちが使い分けられていることに着目し，これらを使い分ける効果を聴取実験，対話実験で評価した．

3. レスポンシブネスの有効性の評価実験

対象とするタスクドメインとしては家庭教師の授業を単純化した「山手線駅名クイズゲーム」を選択した．これは，家庭教師が「山手線の駅名を順番にあげてごらん，つまっちゃったらヒントあげるからね」のように始め，生徒に駅名を次々に答えさせていくものであり，英単語や年号，九九などの暗記を支援する学習支援ソフトを単純化したものである [6]．

ユーザが正解した場合は，表 1のルールに基づいて応答が選択される (導出は 4.章を参照)．ルールは上から順に適用される．

このルールを評価するため，音声対話システムとして実装した．その際，音声認識誤りがシステム評価に与える影響を排除するため，ヒント発話，あいづち応答，発話区間検出などはシステムが行い，ユーザ回答の正解・不正解判定をオペレータがキーボードから入力する Wizard of Oz 法で実装した．システム応答は編集合成音声を用いた．

コントロール条件には出現確率がコーパスと同比率になるようにランダムにあいづち語彙を選択するルールを用いた．これは，この方式を内部状態を考慮しない準最適なルールと考えたからである．

被験者はルール条件，ランダム条件 (コントロール条件) それぞれのシステムと山手線クイズ対話を約 1 分 30 秒行い，その後自分とシステムとの会話を聞き直して使いたいシステムを選択した．どちらのシステムと最初に話すかはランダムに決まるので，学習効果や実験条件を知ることによるオペレータの非言語的な態度変化は排除される．実験中はオペレータと被験者はついたてで仕切って見えないようにし，またオペレータは一切発話しなかった．

その結果，被験者 13 名中 10 名が「より使いたいコンピュータ」としてルール条件のシステムを選択した ($p < 0.05$)．従って，ルールに基づいた応答語彙変化はユーザに好印象を与えることが分かった．

表 1　あいづち使い分けルール
Rules for choosing acknowledgement responses

ルール略称	条件	応答内容
「沈黙」ルール	応答発話時にユーザが発話中	あいづち省略
「順調」ルール	ヒントも誤答もなし，回答時間 1 秒以下,「はい」が 3 回連続	「うん」
	ヒントも誤答もなし，回答時間 1 秒以下，(「そう」または「そうそう」) が 3 回連続	「うん」
	ヒントも誤答もなし，回答時間 1 秒以下，(「うん」または「うんうん」) が 3 回連続	「はい」
	ヒントも誤答もなし，回答時間 1 秒以下	前と同じ応答
「長考」ルール 1	回答時間 12 秒以上	＜駅名復唱＞
「疑問」ルール 1	ヒント一個以下で，発話末尾正規化ピッチ勾配が $0.1[1/s]$ 以上	「うん」
「褒める」ルール	誤答またはヒント後，前 2 回は「そう」か「＜駅名復唱＞」のみ	「そう，＜駅名復唱＞」
	誤答またはヒント後，回答時間 2 秒以下	「そうそう」
	誤答またはヒント後	「そう」
「長考」ルール 2	回答時間 1.5 秒以上で，前回・前々回のどちらかは＜駅名復唱＞以外	＜駅名復唱＞
「疑問」ルール 2	発話末尾正規化ピッチ勾配が $-0.02[1/s]$ 以上	「うん」
「元気」ルール	元気さ ： $\overline{f_{0norm}} + 1.5\overline{E_{norm}} > 3.5$	「そうそう」 「はいはい」 「うんうん」
「デフォルト」ルール	該当ルールなし	「はい」

被験者には会話終了後，自分とコンピュータとの会話に現れた全てのあいづちについて自然さを 7 段階で評価させた (図 2)．ランダム条件会話

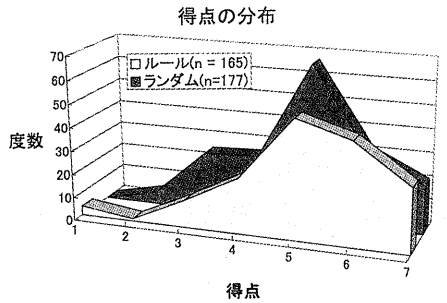

図 2　二つの条件でのあいづちの自然さ得点のヒストグラム (ルール条件あいづち: 165 個，ランダム条件あいづち: 177 個，被験者 13 名)

の評価では分散が大きいのに比べ，ルール条件会話の評価は高い点数側に偏っており，ルール条件で出されたあいづちの方が全体として好印象を与えられることが分かった ($p < 0.05$)．

ルール条件/ランダム条件を選んだユーザの比率はヒントをオペレータが出す予備実験 (12 : 3) や，聴取実験 (3 : 1) でも同様だったので，ランダム条件を好む少数のユーザも存在する可能性がある．

また，被験者が各ルール毎のあいづちの自然さを評価した平均点を図 3 に示す．ここで,「ランダム好き」はランダムあいづちを選択した被験者,「ルール好き」はルールあいづちを選択した被験者である．ルール好きの被験者は褒めるルールを

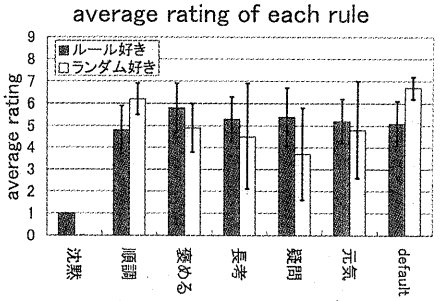

図 3　ルール好きユーザとランダム好きユーザのルール毎の評価の平均：7 段階評価，誤差棒は 95%信頼区間．

順調ルールより高く評価し，ランダム好きのユーザはルール好きのユーザよりデフォルトルールや順調ルールをより高く評価していた ($p < 0.05$)．また，この図より韻律情報のみから計算される「元気さ」ルールも他のルールと同様に機能していることが分かる．

次に，被験者の評価基準を調べるために，聞き直し後に様々な形容詞についてその当てはまり度

合いを3段階で評価させた(図4). すると,「優し

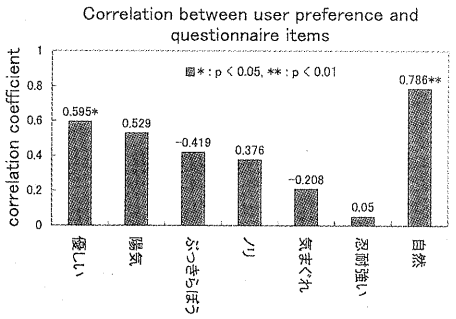

図4 ユーザの選択と印象を表す形容詞との相関関係(正負の符号は除いた)

かった」と「自然だった」が相関した. 被験者は第一に会話の自然さを重要視しており, これは会話システムにおいてほんの一要素の誤動作でも全体評価が落ちるという懸念を裏付けるものだろう.

「優しさ」の重要性は本研究の仮定に合致するが,「忍耐強さ」などは重要ではない. 被験者は「優しい」↔「ぶっきらぼう」($r = -0.76$, $p < 0.01$),「忍耐強い」↔「気まぐれ」($r = -0.63$, $p < 0.01$), と連想しているが, 忍耐強いか気まぐれかは気にしているとは言えない.

4. 応答選択部の構成手法

この節では, 前節で有効性を示した応答選択ルールの構成法について述べる.

応答選択ルールをコーパスから学習させるために39人の家庭教師役人物で41対話(146分, 重複2名)のコーパスを収集した(図5). しかし, 家庭教師(横軸)によって応答語彙が大きく異なることから分かるように, 必ずしも全ての家庭教師がレスポンシブな応答をする訳ではなかった. これは家庭教師の個性が影響が大きいためであると考え, 39人から優しい家庭教師のモデルとして, (1)あいづちを打ち, (2)ヒントを出し, (3)楽しそうに進行する一人を選び, 新たに6対話30分のコーパスを再構築した(図6)[6].

このコーパスは小規模であるため, コーパスからのルール抽出には分類木生成ツール(C4.5など)は用いず, コーパス中で目立つ特徴のみから予備ルールを構築し, ユーザにこのルールの不備を指摘してもらい改良する方法を取った. 予備

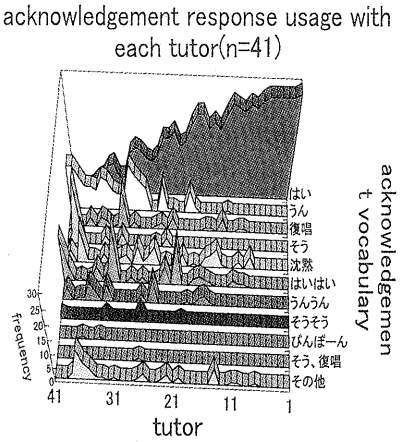

図5 41対話毎のあいづち頻度(横軸:家庭教師)

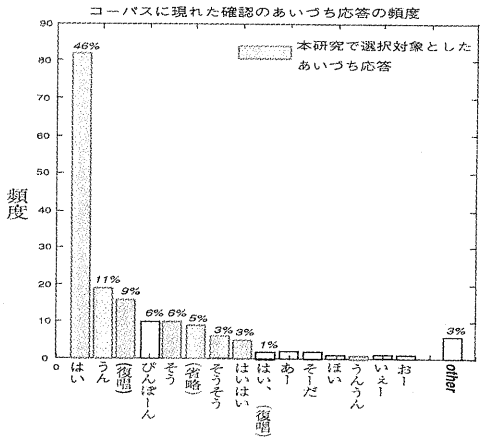

図6 6対話コーパスのあいづち

ルールは「はい」「うん」<駅名復唱>を使い分けるものであった.

次に予備ルールの予測に従ってコーパス中の一対話(4分)のあいづち(30個所)を編集合成で入れ替え, 8人の被験者に聴かせた. その際の指摘をもとに再構成したのが本ルール(表1)である.

本ルールを構築する際の指摘には「元気に答えた時には」「順調に進んでいる時には」など, ユーザの内部状態への明示的な言及があった. 従って, これらのルールは指摘者の内部状態推測を直感的に反映していると言える(表2).

また, 本ルールは大まかには「自信」と「元気

表 2　ルール構成の起源，応答印象，ルール条件の内部状態としての解釈
origin, subjective impression, and reflection of user's internal state, for each rule

ルール	ルールの起源	応答の主観的印象	ユーザの内部状態の解釈 (その起源)
「沈黙」ルール	指摘	特になし	(発話権)(なし)
「順調」ルール	指摘，コーパス	一貫性がある	順調だ (指摘)
「長考」ルール	コーパス	ペースを落とす	難しかった (文献)
「疑問」ルール	指摘	優しく	自信がない (文献)
「褒める」ルール	指摘	褒める	ヒントからうまく答えられた (指摘)
「元気」ルール	指摘→コーパス	元気に	元気に答えた (指摘)
「デフォルト」ルール	なし	なし	なし

さ」の二つの内部状態に関係していると解釈できる．これらは事前に想定していたものではないが，コーパスと指摘から作られたルールが結果として内部状態の推測として解釈可能であったと言えるだろう．

4.1　予備ルールとの比較

本ルールがコーパス特徴のみから構成した予備ルールよりも自然で優しい印象かを調べるため，コーパス中のあいづちを予測した (図 7)．すると，本ルールでは全体の予測精度が低下した．しかし，本ルール/予備ルールそれぞれについてあいづち (29 個) を編集合成した会話 (1 対話 4 分) を 9 名の被験者に聞かせ，自然な方を選択させたところ，8 名が本ルールを選んだ ($p < 0.05$)．

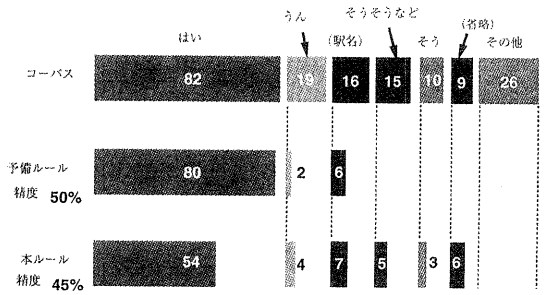

図 7　コーパス中のあいづちの予測結果
Prediction of acknowledgement responses in the corpus

従って，対象とする現象 (あいづち語彙) の多様性を知るためにはコーパス分析が有効だが，応答選択ルールの構築にはコーパス分析と聴取による指摘の双方が有効であると言える．また，聴取実験の結果から，コーパスデータへの最適化のみではなく複数の評価方法が必要なことが分かった．

5.　レスポンシブな音声対話システムの評価手法

本来，自然な応答は対話において意識されにくい．そのため，音声認識・理解を主目的とするシステムの評価とは異なり，自然さを重視したシステムのアンケート評価は精度が低くなる．

このため本研究では対話後にユーザが自分の対話を聞き直してから評価する方法を取った．聞き直しによってユーザは冷静に評価でき，また自分の会話なので会話中の状態を正確に想起できる．

この方法と会話直後での評価の比較を表 3，図 8に示す．図 8 は被験者が評価した「優しかった度合い」の，ルール条件とランダム条件の差についてのヒストグラムである．図 8 上は会話直後の，下は聞き直し後である．会話直後では「ルール好き」と,「ランダム好き」の分布に有意な差は見られず，被験者の選択と優しかった度合いは関連しているとは言えなかったが，聞き直し後では図 8下のように被験者の選択との間に相関が生じていた ($p < 0.05$)．

表 3　会話実験：好ましいシステムの選択
Preference of system by the subjects

順番	ルール条件	ランダム条件	
会話直後	6 名	7 名	—
聞き直し後	10 名	3 名	$p < 0.05$

また，あいづちの評価点数は聞き直し後の方が全体的に低くなり，聞き直し後のコメントにはあいづちに関するものが増えた ($p < 0.05$)．これらの結果から，聞き直しによってより細かな評価が可能になると言えるだろう．

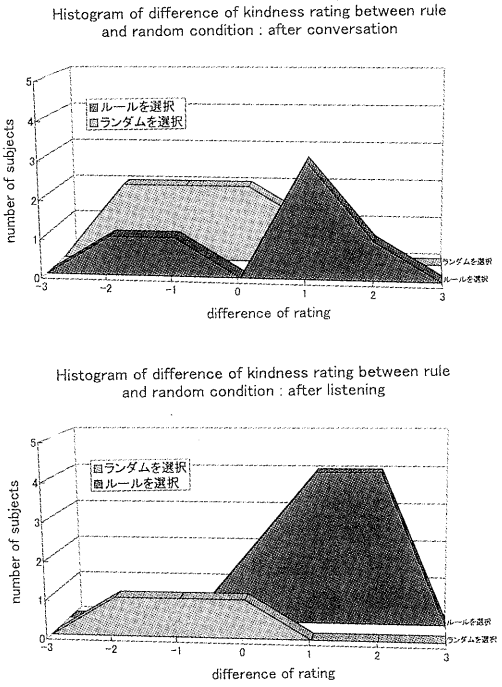

図 8　二条件での優しさ度合いの差のヒストグラム（上：会話直後，下:聞き直し後)

6. 結論

音声対話システムの構成に関する以下の知見を得た．

1. あいづち語彙の変化による印象制御の有効性を示した．
2. ユーザの言語・非言語情報に基づく繊細な応答選択戦略が印象向上に有効であることが分かった．これは，内部状態推測が有効である可能性を示すものである．
3. 韻律情報のみから構成した応答選択ルールも有効であることが分かった．

対話システムの構成・評価手法に関する以下の知見を得た．

1. コーパスの最適化とユーザ印象は一致しないため，双方の評価が必要であることが分かった．
2. 会話実験において聞き直し評価が有効であることが分かった．

謝辞

本研究は国際コミュニケーション基金，中山隼夫科学技術文化財団，稲盛財団，文部省の援助を受けて行われた．ここに記して謝意を表する．

参考文献

[1] S. E. Brennan and M. Williams. The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive status of speakers. *Journal of Memory and Language*, Vol. 34, pp. 383–398, 1995.

[2] Yasuhiro Katagiri, Miyoko Sugito, and Yasuko Nagano-Madsen. Forms and prosodic characteristics of backchannels in tokyo and osaka japanese. In *the 14th International Congress of Phonetic Sciences*, pp. 2411–2414, 1999.

[3] R. Kompe, E. Noth, A. Kiebling, T. Kuhn, M. Mast, H. Niemann, K. Ott, and A. Batliner. Prosody takes over: Towards a prosodically guided dialog system. *Speech Communication*, Vol. 15, pp. 155–167, 1994.

[4] Kikuo Maekawa. Phonetic and phonological characteristics of paralinguistic information in spoken japanese. In *1998 International Conference on Spoken Language Processing*, pp. 635–638, 1998.

[5] Tsubasa Shinozaki and Masanobu Abe. Development of cai system employing synthesized speech responses. In *1998 International Conference on Spoken Language Processing*, pp. 2855–2858, 1998.

[6] Wataru Tsukahara. An algorithm for choosing japanese acknowledgments using prosodic cues and context. In *1998 International Conference on Spoken Language Processing*, pp. 691–694, 1998.