韻律情報を用いた相槌生成システムとその評価

1N−04

竹内 真士　北岡 教英　中川 聖一[1]

豊橋技術科学大学　情報工学系[2]

1　はじめに

対話システムにおいて，機械が人間と同様に相槌を打つことができれば，スムーズな対話を行うことができると期待される．本研究では，リアルタイムでの相槌応答生成を目的として，ピッチ，パワー、ポーズなどの韻律情報のみを用い，相槌を自動生成するシステムを作成する．また被験者実験により，システムが生成する相槌の自然さ，人間との一致度の評価を行う．

2　日本語における相槌の分析

日本語において相槌は重要な役割と持つものとして，その分析を行った例がある．例えば，小磯らは韻律情報が相槌に与える影響を分析しており，1モーラ分のピッチ，パワーがある変動パターンに一致しているときに相槌が打たれるとしている [1].

また，音声対話システムで相槌を用いることで，自然な対話を実現することも試みられている．平沢らは連続音声認識アルゴリズムの中間結果を用いることで，言語情報から相槌を打つことを試みている [2]．西らは，ポーズ情報のみでも適切に応答が返せることを示しており，電話応対のプロンプト音声を送出するタイミングを一定時間のポーズ長で決めている [3].

本研究では，これら分析に基づいて，句音声末のおよそ60ミリ秒（おおよそ1モーラに対応）におけるピッチ，パワーの変動を調べて，相槌を生成することを試みた．

3　相槌生成システム

3.1　相槌生成の韻律情報の変動パターン

小磯ら [1] による分析では，発話の句音声末において，ピッチとパワー両方が，図1のようにいずれかのパターンで変動した場合に，相槌が打たれるとされる．実際の人間の対話を見ると，図2のようにパターンに合う場合に相槌が打たれていることが分かる．そこで，無音区間によって句音声末を検出し，さらにピッチとパワーそれぞれの回帰係数を求め，回帰係数が閾値 α 以上ならば上昇，$-\alpha$ 以下ならば下降，それ以外を平坦とすることとし，パターンに合う場合には相槌を打つとした．

3.2　システムの相槌生成手順

本システムの相槌生成手順は以下のようである．

1. 入力されたユーザの音声に対し，リアルタイムにピッチとパワーを抽出する．

2. 過去の時間幅 T からピッチとパワーの回帰係数を $T/3$ ごとに求め，その変動パターンを調べる．

3. 回帰係数が前節で述べた変動パターンに一致し，時間 τ だけ無音区間が続いた場合，相槌を生成する．

ここではフレーム周期を6[ms]で分析を行い，10フレームから回帰係数を求めた．すなわち，T=60[ms]である．句音声末から相槌を打つまでの無音区間の長さ τ は，90[ms]とした．先行研究 [4] から得られた知見より，最初の相槌が打たれてから14音節から26音節で次の相槌が打たれるということから，一度相槌を打つと2100[ms]は相槌を打たないというルールも用いた．また，システムが返す相槌音声は「はい」のみを用いた．

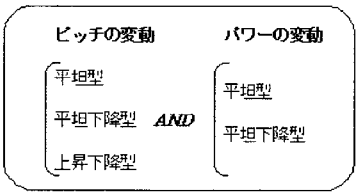


図 1: 相槌を生成するピッチ、パワーパターン

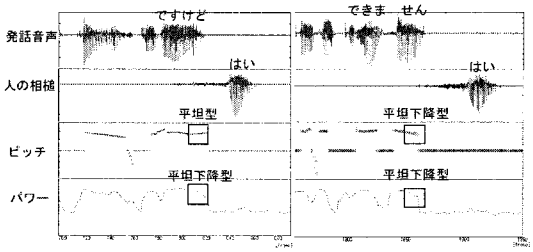


図 2: 発話音声に対する相槌とピッチ、パワー

4　評価実験

評価は3つの方法で行った．1つ目は，人間の相槌との一致度を評価する方法で，評価尺度に検出率，精度を用いた．2つ目は，話者が発話した音声に対して，システムが打った相槌を，第三者が自然であるかどうかの評価を行う方法である．3つ目は，相槌の生成により発話をしやすいかどうかの主観的評価である．

4.1　評価実験におけるタスク

評価は，留守番電話式の音声メッセージ入力のタスクで行った．留守番電話式とは，ユーザが留守番電話

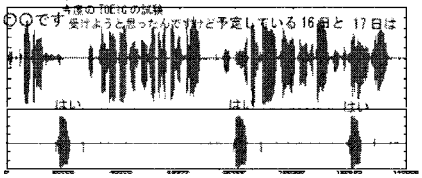
An "aizuchi" generation system using prosodic information and its evaluation

[1]Masashi Takeuchi, Norihide Kitaoka, and Seiichi Nakagawa

[2]Toyohashi University of Technology

に用件を話すように，マイクにより名前，用件などを入力する方式である．

図3に発話の音声波形と本システムにより生成された相槌を示す．



発話音声："○○です．今度の TOEIC の試験受けようと思ったんですけど，予定している 16 日と 17 日は"

図 3: 発話音声波形とそれに対する本システムの相槌

4.2 人間の相槌との一致度を用いた評価

被験者3人の発話について，人間とシステムの両方に相槌を打たせ，その一致度を調べることにより評価を行った．評価には検出率，精度を用いた．

$$\text{検出率} = \frac{\text{人間とシステムの相槌の一致回数}}{\text{人間の相槌回数}}$$

$$\text{精度} = \frac{\text{人間とシステムの相槌の一致回数}}{\text{システムの相槌回数}}$$

結果を表1に示す．

表 1: 検出率と精度の割合

	検出率 [%]	精度 [%]
被験者 1	29	41
被験者 2	33	25
被験者 3	25	47

検出率25%のとき，精度が最大で47%となった．また，被験者により結果に違いがあることから，被験者の声の大きさ，話し方などにも影響されると考えられる．相槌の箇所は，正解というものが存在せず非決定的であり，検出率と精度は単なる目安に過ぎない．

4.3 第三者による評価

相槌は人それぞれに個人差があり，人間との一致度で評価することが必ずしも適切とは限らない．そこで，ユーザ3名の発話に対してシステムが打った相槌（61個の相槌）を，被験者以外の3名が聴取により評価するという方法を用いた．個々の相槌に対して，「遅すぎる」から「早すぎる」までを5段階もしくは「相槌とは言えない」（すなわち，不適切である）のいずれかを選択してもらった．その結果を表2に示す．

表 2: システムの相槌の第三者による評価

1:遅すぎる　2:やや遅い　3:適切　4:やや早い　5:早すぎる
×:相槌とは言えない
(システムの 61 回の相槌に対する評価値の頻度)

評価	1	2	3	4	5	×
聴取者 1	4	11	28	6	0	12
聴取者 2	2	21	24	4	4	6
聴取者 3	1	8	39	0	0	13
合計	7	40	91	10	4	31
割合 [%]	3.8	21.9	49.7	5.5	2.2	16.9

表2の結果より，適切な相槌は全体の49.7%，やや速いもしくはやや遅いが，相槌の箇所としては自然であるものを含めた場合は，77.1%が適切な相槌であるとされた．

4.4 相槌生成による発話のしやすさの評価

被験者4人にシステムに音声メッセージを入力してもらい，その発話に対して相槌がない場合，Wizard of Oz法で人間が相槌を生成する場合，システムが相槌を生成する場合，での対話のしやすさを評価し，相槌がある場合にはその相槌の自然性の評価も行った．表3，4がその結果である．

表 3: ユーザによる対話のしやすさの評価値の頻度

A:話しやすい　B:まあまあ話しやすい　C:どちらともいえない
D:少し話しづらい　E:話しづらい

評価		A	B	C	D	E
相槌の生成法	相槌なし	0	2	0	1	1
	Wizard of Oz 法	0	4	0	0	0
	システム	0	1	2	1	0

表 4: ユーザによる相槌の自然性の評価値の頻度

A:自然　B:どちらともいえない　C:不自然

評価		A	B	C
相槌の生成法	Wizard of Oz 法	4	0	0
	システム	1	2	1

表 3,4 の結果では，現段階でのシステムの性能は十分であるとは言えない．被験者からは，相槌が全体的に遅い，発話内容に合わない相槌音声が出ることがある，といった意見があった．不適切な相槌は，時にはユーザ発話の流れを止めてしまうこともあり，改善が必要である．

5 まとめ

本稿では，韻律情報を用いて相槌を挿入するシステムの構築を行った．評価実験の結果，人間の相槌との比較による評価では検出率が25%のときに，精度が最大で47%となった．第三者の聴取による評価では，77.1%がほぼ自然であると判定された．ユーザによる評価では，現段階でのシステムは対話のしやすさ，自然性の面でまだ十分でなく，改善が必要であることが分かった．

今後は，句音声末を正確に検出することで相槌の自然性を上げること，相槌の出現確率を導入することで，より人間の相槌に近づけることを考えている．

参考文献

[1] 小磯，堀内，土屋，市川：「先行発話断片の終端部分に存在する自発発話者に関する言語的・韻律的要素について」，信学技報，NLC95-72，pp.25-30(1996).

[2] 平沢，川端：「音声対話システム Noddy −ユーザ発話途中でのうなずき・相槌生成−」，情処研報，SLP-20-4，pp.51-52(1998).

[3] 西：「音声対話システムにおけるプロンプト音声送出タイミングの評価と制御法」，電子情報通信学会論文誌，D-II，Vol.J79 − D-II，No.12，pp.2170-2175(1996).

[4] 岡登，加藤，山本，板橋：「韻律情報を用いた相槌の挿入」，情報処理学会論文誌，Vol.40，No.3，pp.469-477(1999)