同調的対話システムにおけるあいづち挿入タイミング

神谷　優貴[†]　　大野　誠寛[‡]　　松原　茂樹[♭♮]　　柏岡　秀紀[♮]

[†]名古屋大学大学院情報科学研究科　　[‡]名古屋大学大学院国際開発研究科
[♭]名古屋大学情報基盤センター　　[♮]情報通信研究機構

1　はじめに

近年，音声認識技術の進展により，カーナビなど様々な音声対話システムが開発されている．これらのシステムはユーザの要求に的確に応答することを目的とする．しかし，ユーザの発話中にシステムからの反応が全く無いと，システムがユーザの発話を認識しているか分からず，ユーザに不安を与えることになる．

人間同士の対話において聞き手は，話し手の発話終了を待って応答をするだけでなく，相手の発話途中でも，あいづちや笑い，うなずきなどの振る舞いを与えている．特に自動車内での対話では，ドライバの視線は走行中，聞き手に向いていないため，ドライバにとって，音声であるあいづちは重要な役割をもつ．

そこで本論文では，同調的車内音声対話において，ドライバの発話中に適切なタイミングであいづちを打つシステムの開発を目指し，あいづちデータの作成とそれを用いた分析について述べる．これまでにもあいづちを生成するシステムはいくつか開発されているものの [1, 2, 3]，あいづちが打たれるタイミングの網羅的な分析は行われていない．本研究では，話し手の発話中に打つことが可能なあいづちタイミングを網羅的に明らかにすることを試みる．まず，複数の被験者によるあいづち音声を収録し，被験者全体で統合することにより，あいづちデータを作成した．次に，文節や節，ポーズ，発話速度とあいづちタイミングとの関係についてデータ分析を与えた．分析結果に基づいて実験を実施し，あいづち挿入タイミングの推定可能性について検討した．

2　あいづち

あいづちの定義や機能についてはこれまでに多くの議論が行われている．メイナードは，「話し手が発話権を行使している間に聞き手が送る短い表現」と定義し，あいづちの機能を「続けてというシグナル」や「内容理解を示す」などの 6 つに分類している [4]．また，堀口は「話し手の発話が聞き手に届けられたことを話し手に知らせるサイン」であるとしている [5]．しかし，むやみにあいづちを打ち続けたのでは，話し手は聞き手が本当に話を聞き，理解しているのかについて疑問を抱くこととなりあいづち本来の役割を果たせなくなる．このため，あいづちを打つタイミングが重要となる．

一方，あいづちを打つ適切なタイミングは，あいづちを打つ人や状況によって異なる．そのため，対話システムには，話し手やその場の状況に応じて，生成するあいづちを変化させることが求められる．本研究では，不自然なタイミングであいづちを打つことを防ぐために，事前に，可能なタイミングを網羅的に検出し，その中から，ユーザの趣好やその場の状況に合わせて，あいづちを打つタイミングを決定するというアプローチをとる．

表 1: 被験者 4 名のあいづち数

被験者 ID	1	2	3	4
あいづち数	192	180	56	112

3　あいづちデータの作成

あいづちを打つことが可能なタイミングを網羅的に分析するため，複数の話者によるあいづち結果を統合することにより，あいづちデータを作成した．以下では，あいづちデータの作成について述べる．

3.1　あいづち音声の収録とその書き起こし

名古屋大学 CIAIR 車内音声対話コーパス [6] のドライバ音声（60 対話に含まれる 297 ターン）に対して，被験者 4 名があいづちをそれぞれ打ち，その音声を収録した．被験者の発話は「はい」のみとし，不自然ではない箇所に可能な限りあいづちを打つように指示した．また，ドライバ音声をターン単位ごとに流し，ターンの音声中にあいづちを打つように指示した．各被験者のあいづち数を表 1 に示す．

収録した音声に基づいて，実験に使用したドライバ音声と各被験者によりあいづちが打たれた箇所を図 1 のように書き起こした．各行は，ドライバ発話の形態素またはポーズであり，開始終了時間，ならびに，あいづちが打たれたか否かの情報を与えている．形態素の場合

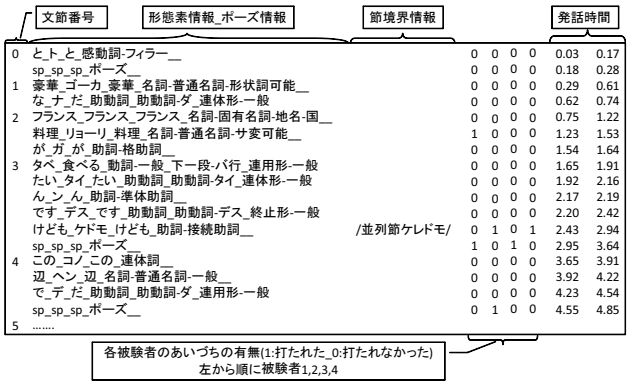

図 1: あいづちタイミングの書き起こしの例

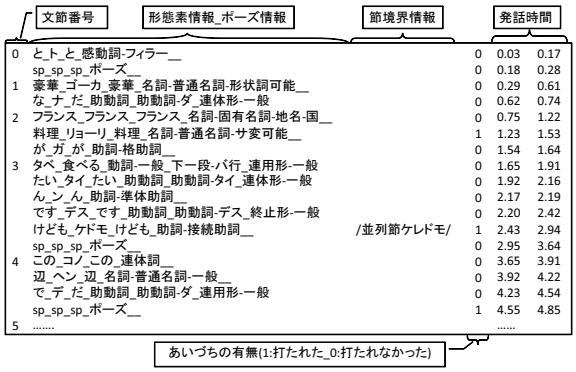

図 2: 統合されたあいづちデータの例

は，品詞情報や文節情報，節境界情報を付与している．本研究では，形態素もしくはポーズを，分析における基本区間と考え，どの区間で被験者のあいづちが打たれたかを分析する．なお，200ms 以上のポーズは，最初の 200ms とそれ以降とに 2 分割し，それぞれを基本区間としている．形態素解析には ChaSen[7] を，節境界解析には CBAP[8] を，基本区間の時間情報付与には Julius[9] を用いた．

3.2 あいづちデータの作成

図 1 及び表 1 が示すように，あいづち回数やあいづちを打つタイミングは被験者ごとに異なっていた．本研究では，被験者 4 人により打たれたあいづち結果を 1 つに統合することにより，あいづちデータを作成した．あいづち結果の統合は以下の手順で行った．

1. 音声として出力されたタイミングが異なっていても，各被験者が打とうと思ったタイミングが同じだと推測されるあいづちについては，あいづちを打った被験者が最も多い 1 箇所に集約する．
2. 上記の処理で残ったあいづちは，いずれも適切なあいづちタイミングであると考えて，和集合をとる．

図 2 は図 1 に示す「各被験者のあいづちの有無」を統合して作成したあいづちデータである．以下では，図 1 と図 2 を用いて，4 人の被験者によるあいづち結果の統合例を説明する．まず，図 1 の「けども (2.43 秒〜2.94 秒)」と「ポーズ (2.95 秒〜3.64 秒)」に打たれたあいづちは，被験者があいづちを打とうと思ったタイミングは同じであると判断し，これら 2 つを，「けども (2.43 秒〜2.94 秒)」へのあいづちとして集約する．次に，図 1 の「料理 (1.23 秒〜1.53 秒)」や「ポーズ (4.55 秒〜4.85 秒)」に打たれたあいづちと先ほど集約した「けども (2.43 秒〜2.94 秒)」に打ったあいづちの和集合をとることにより，図 2 のようなあいづちデータを作成する．統合したあいづちデータの規模を表 2 に示す．

表 2: あいづちデータの規模

対話	60
ターン	297
文節	1,478
形態素	3,179
ポーズ	756
あいづち数	324

4 あいづち挿入タイミングの分析

本研究では，あいづち可能なタイミングを検出するために統計的アプローチを採用する．そのための有効な素性について検討するため，あいづち挿入タイミングの特徴分析を行った．分析には，前節で作成したあいづちデータのうち，50 対話を用いた．50 対話には，基本区間が 3,274 か所 (2,656 形態素 +618 ポーズ) あり，そのうち 265 か所にあいづちが打たれていた．あいづちが挿入された割合 (以後，あいづち挿入割合) は 8.1%であった．なお，この数値はあいづちが平均的に挿入される割合を意味しており，ある特徴をもった基本区間に対するあいづち挿入割合がこの数値より高い場合，その特徴をもった区間にはあいづちが挿入されやすいことを意味する．

4.1 文節とあいづち挿入タイミング

あいづちは，内容理解を示す機能を持つため，話し手の発話をある程度理解したときに打たれると考えられる．一方，文節は，日本語における最小の意味的なまとまりであるため，文節が発話された直後にあいづちが打たれやすいと考えられる．ターン末の文節を除くと文節は 977 個あり，そのうち，その直後の基本区間にあいづちが打たれていたのは，178 個 (あいづち挿入割合：18.2%) であった．文節の直後に対するあいづち挿入割合は，全基本区間のあいづち挿入割合 8.1%と比べて高

表 3: 文節の最終形態素の品詞とあいづち挿入割合

助詞	25.2%	(102/405)
感動詞	8.5%	(21/246)
助動詞	27.5%	(28/102)
名詞	24.0%	(18/75)
接続詞	5.5%	(3/55)
副詞	9.5%	(3/42)
形容詞	0.0%	(0/17)
動詞	8.3%	(1/12)
接尾辞	9.1%	(1/11)
代名詞	0.0%	(0/10)

い数値を示しており，文節の直後にはあいづちが打たれやすいということが分かった．

次に，文節の種類による，あいづちの打たれやすさの異なりを分析した．表 3 に，文節をその最終形態素の品詞によって分類し，その分類ごとに算出したあいづち挿入割合を示す．文節の最終形態素の品詞が助詞や助動詞，名詞の場合にあいづちが打たれやすいことが分かった．

4.2 節とあいづち挿入タイミング

節は文節よりも強い意味的なまとまりであるので，節の直後にはあいづちが打たれやすいと考えられる．ターン末の節を除き，節は 341 個存在しており，そのうち，その直後の基本区間にあいづちが挿入されていたのは 66 個 (あいづち挿入割合：19.4%) であった．節の直後に対するあいづち挿入割合は，全基本区間のあいづち挿入割合と比べて高く，節の直後にあいづちが打たれやすいといえる．また，文節の直後へのあいづち挿入割合と比べても高く，節の直後には，あいづちがより打たれやすいことが分かった．

次に，節の種類によるあいづちの打たれやすさの違いを分析した．表 4 に，節の種類ごとに算出した，その直後に対するあいづち挿入割合を示す.「並列節ケレドモ」の直後であいづちが打たれやすいことが分かる．また，節の種類によって，その直後に対するあいづちの打たれやすさが異なることが分かった．

4.3 ポーズとあいづち挿入タイミング

あいづちには相手に発話を促す機能があるため，相手発話の無音区間であるポーズにはあいづちが打たれやすいと考えられる．ポーズ 618 区間のうち，102 区間にあいづちが挿入されていた．ポーズに対するあいづち挿入割合 16.5%は全基本区間のあいづち挿入割合 8.1%と比べて高く，ポーズにあいづちが打たれやすいことが分かった．

表 4: 節の種類とあいづち挿入割合

感動詞	5.1%	(6/118)
引用節	8.6%	(3/35)
並列節ケレドモ	67.6%	(23/34)
談話標識	7.4%	(2/27)
主題ハ	12.0%	(3/25)
テ節	42.9%	(6/14)
連用節	7.1%	(1/14)
間接疑問節	40.0%	(4/10)

4.4 発話速度とあいづち挿入タイミング

あいづちには相手の発話を促す機能があるため，相手の発話がゆっくりになったときに，あいづちが打たれやすいと考えられる．そこで，ドライバ発話における各形態素の発話速度 (モーラ/秒) を計測して，1 形態素の平均発話速度より早い場合と遅い場合とに分類し，それぞれ，その直後にあいづちが挿入される割合を調査した．なお，1 形態素の平均発話速度は，人による違いの大きさを考慮し，ドライバごとに算出したものを利用した．発話速度が平均発話速度より速い形態素の場合，その直後の基本区間に対するあいづち挿入割合は 6.3%(114/1,813) であり，遅い場合は 16.5%(162/984) であった．このことから，発話速度が平均発話速度より遅くなると，その形態素の直後にあいづちが打たれやすいことがわかった．

5 あいづち挿入タイミングの推定

4 節の分析で明らかにした特徴をふまえ，統計的手法によるあいづち挿入タイミングの推定可能性について検討した．

5.1 あいづち挿入タイミングの推定方法

本研究では，1 対話ターンの基本区間列 $m_1 \cdots m_n$ 中の基本区間が連続して入力されることを想定し，基本区間が 1 つ入力されるごとに，その入力基本区間に対して，あいづちを挿入することができるか否かを Support Vector Machine (SVM) を用いて推定する．ある基本区間 m_i に対してあいづち挿入タイミングを推定する際に，SVM で利用した素性を表 5 に示す．これらはすべて，入力基本区間 m_i の直前までの基本区間列 $m_1 \cdots m_{i-1}$ から得られる素性である．なお，各基本区間の形態素情報や節情報，時間情報などは，その直後の基本区間が入力されたときに得られるものとする．

表 5: SVM で用いた素性

1.	m_{i-1} が文節の最終形態素であるか否か
2.	1 が真の場合，m_{i-1} の品詞
3.	m_{i-1} が節の最終形態素であるか否か
4.	3 が真の場合，m_{i-1} が属する節の種類
5.	m_{i-1} が 200ms のポーズであるか否か
6.	m_{i-1} の発話速度が平均発話速度より遅いか否か
7.	6 が真の場合，m_{i-1} の発話速度と平均発話速度の差
8.	直前のあいづち挿入時間 (あいづち未挿入の場合は m_1 の開始時間) から m_i の開始時間までの時間長

表 6: 実験結果

	適合率	再現率	F 値
提案手法	54.2% (13/24)	22.0% (13/59)	31.3
被験者	63.6% (14/22)	23.7% (14/59)	33.8

5.2 実験と検討

あいづち挿入タイミングの推定実験を実施し，推定可能性について検討した．

3 節で作成したあいづちデータのうち，4 節の分析に用いた 50 対話を学習データとして，残りの 10 対話をテストデータとして使用した．SVM のツールとして LibSVM[10] をデフォルトのオプションのまま使用した．また，比較のため，データ作成時の被験者とは別の被験者 1 名によるあいづちデータを収集した．

評価には，以下の指標を用いた．

$$\text{適合率} = \frac{\text{正しく挿入されたあいづちの数}}{\text{挿入されたあいづちの数}}$$

$$\text{再現率} = \frac{\text{正しく挿入されたあいづちの数}}{\text{正解データにおけるあいづちの数}}$$

正解データとの間で基本区間が一致すれば正しく挿入されたと判定した．

結果を表 6 に示す．被験者による結果は，F 値で 33.8 にとどまっており，これは正確な挿入タイミング推定の難しさを示唆している．本手法は，被験者による結果を若干下回る程度の推定性能を達成している．

一方，音声対話システムにおけるあいづち挿入タイミングに多少のずれが生じることは許容できると考えられる．そこで正解データのあいづち箇所の前後 200ms 以内であれば正解として評価した．結果を表 7 に示す．適合率は 70.8%と比較的高い数値を示しており，今後，学習データの拡充を進めることにより，ユーザが許容できる程度のタイミングでのあいづち挿入性能の実現可能性が示された．

表 7: 実験結果 (前後 200ms 以内を正解とした場合)

	適合率	再現率	F 値
提案手法	70.8% (17/24)	28.8% (17/59)	41.0
被験者	72.7% (16/22)	27.1% (16/59)	39.5

6 おわりに

本論文では，まず，50 対話 250 ターンのドライバ音声に対して，被験者 4 人があいづちを打った結果を統合し，計 265 個のあいづちタイミングが付与されたあいづちデータを作成した．次に，作成したデータを用いて，あいづち挿入タイミングの特徴分析を行い，節境界の直後やポーズ，発話速度が低下した箇所であいづちが打たれやすいことを明らかにした．分析に基づき，SVM を用いてあいづちタイミングの推定を行ったところ，適合率 54.2%，再現率 22.0%であった．

今後は，韻律情報を用いた分析を行い，学習データの拡充や韻律情報を素性に追加するなど，あいづち挿入タイミングの推定精度の向上を図る予定である．

謝辞 本研究は一部，科研費挑戦的萌芽研究「音声対話システムの個性化に関する基礎的研究」による．

参考文献

[1] 西村, 北岡, 中川: 応答タイミングを考慮した雑談音声対話システム, 人工知能研資, SIG-SLUD-A503-05, pp.21-29 (2006).

[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，1997-SLP-020, pp.51-52 (1998).

[3] N. Ward: In Japanese a Low Pitch Means “Back-Channel Feedback Please,” 情処研報, 1996-SLP-011, pp.7-12 (1996).

[4] メイナード: 会話分析, くろしお出版 (1993).

[5] 堀口: 日本語教育と会話分析,くろしお出版 (1997).

[6] N. Kawaguchi, S. Matsubara, K. Takeda, F. Itakura: CIAIR In-Car Speech Corpus -Influence of Driving Status-, *IEICE Trans. Inf. Sys.*, Vol.E88-D, No.3, pp.578-582 (2005).

[7] 松本，高岡，浅原: 形態素解析システム『茶筌』, version 2.4.0，使用説明書 (2007).

[8] 丸山，柏岡，熊野，田中: 日本語節境界検出プログラム CBAP の開発と評価, 自然言語処理，Vol.11, No.3，pp.39-68 (2004).

[9] 河原，李: 連続音声認識ソフトウエア Julius, 人工知能学会誌，Vol.20，No.1，pp.41-49 (2005).

[10] C. Chang and C. Lin: LIBSVM: a library for support vector machines, `http://www.csie.ntu.edu.tw/~cjlin/libsvm` (2001).