社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

タグ付けの安定性を備えた音声対話コーパスに基づくあいづち生成タイミングの検出

大野　誠寛†　　神谷　優貴††　松原　茂樹††

† 名古屋大学大学院国際開発研究科
†† 名古屋大学大学院情報科学研究科
〒 464-8601 愛知県名古屋市千種区不老町
E-mail: †ohno@nagoya-u.jp

あらまし　本論文では，応答性に優れた車内音声対話システムの実現を目指し，あいづち生成タイミングの分析と検出について述べる．本研究では，あいづち生成タイミングが安定的にタグ付けされた対話コーパスを用いて，あいづち生成タイミングの特徴を網羅的に分析した．また，その分析結果に基づいて，統計的手法により，あいづち生成タイミングを高精度に検出することを試みた．1,219 対話ターンを用いた検出実験の結果，あいづち生成位置の適合率で 82.2%，再現率で 66.1%を達成しており，本手法の有効性を確認した．
キーワード　コーパス，音声言語，対話システム，タグ付け

Detection of back-channel feedback timings based on consistently tagged spoken dialogue corpus

Tomohiro OHNO†, Yuki KAMIYA††, and Shigeki MATSUBARA††

† Graduate School of International Development
†† Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601 Japan
E-mail: †ohno@nagoya-u.jp

Abstract　This paper describes analysis and detection of back-channel feedback timings, aiming at realizing in-car spoken dialogue systems with the high responsiveness. In our research, we analyzed the characteristics of back-channel feedback timings comprehensively by using in-car speech dialogue corpus where back-channel feedback timings were consistently tagged. And then, we tried to statistically detect the timings based on the result of analysis. An experiment using 1,219 dialogue turns, providing a precision of 82.2% and a recall of 66.1%, has shown the effectiveness of our method.
Key words　corpus, spoken language, dialogue system, tagging

1. はじめに

近年，音声対話システムの実用化が進みつつある．なかでも車内音声対話システムは，最も普及している音声アプリケーションの一つであり，その多くは，ナビゲーションや情報検索を主要なタスクとしている．これまで，タスクを確実に遂行できるシステムの実現を目的に，音声認識や発話理解，対話制御などの研究開発が進められ，車内音声対話の技術は大いに向上している．今後は，単にタスクを達成できるだけでなく，ドライバがより快適に対話を進められるように，システムの応答性を高めることが重要となる．

システムの応答性を高めるための一つの方法は，ユーザの発話中においてもシステムが何らかの反応を返すことにより，システムによる認識や理解の状態を適宜開示することである．このような開示は，人間同士の対話の場合，頷きや表情，身振りや手ぶり，あいづちなどの行為を通して遂行される．しかし，実走行車内においては，ドライバの視線は音声対話システムにないため，システムによるドライバへの応答としては，音声による応答，すなわち，あいづちによらざるを得ない．また，ドライバにとって快適な車内対話を実現するためのシステムの応答戦略としては，ドライバの発話中に積極的にあいづちを打つことが望まれる一方，適切なタイミングであいづちが生成され

る必要がある．

そこで本論文では，高い応答性を備えた車内音声対話システムの実現を目指し，対話コーパスを用いたあいづち生成タイミングの検出について述べる．あいづちの発生タイミングに関する研究は既にいくつか存在しており [1]～[6]，人間の間で遂行された対話中のあいづち発生タイミングに基づいて，あいづち位置の分析や推定が行われている．しかし，これらの研究で利用されてきた現存する音声対話コーパスは，あいづち発生タイミングの揺れが大きいため，実践的に利用するのに適したデータとはいえず，あいづち生成タイミングの推定において，十分な精度は達成されていない．本研究では，既存の車内対話データに対して，あいづちの生成に適した位置に網羅的にタグ付けすることにより構築した，あいづちコーパスを用いて，適切なあいづち生成タイミングを高精度に検出する．

本研究では，まず，タグ付けの安定性を備えたあいづちコーパスを分析し，文節境界や節境界，ポーズ，発話速度，ピッチ，パワーと，あいづち発生タイミングの関係を明らかにした．次に，その分析結果に基づいて，統計的手法によりあいづち生成タイミングを検出することを試みた．あいづち生成タイミングの検出実験では，人手による検出結果を若干下回る程度の推定性能（適合率 82.2%，再現率 66.1%）を達成しており，本手法の有効性を確認した．

2. あいづちコーパス

あいづちとは，話し手の発話を受け取ったことを，聞き手が話し手に伝達するサインである [2]．この意味で，車内におけるドライバとシステムとの対話では，ドライバが安心して発話を遂行するために，システムは可能な限り頻繁にあいづちを打つことが望ましい．しかしその一方で，むやみやたらにあいづちを打ち続けたのでは，ドライバはシステムが話を理解しているのかについて疑念を抱くこととなり，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．あいづちタイミングについては，人間による対話を調査した研究がいくつかあるが [1]～[6]，これらの知見だけでもって，あいづちを生成する機構を実現することは容易ではない．というのも，あるタイミングで生成されたあいづちの適切さとは，音響あるいは言語的な諸要因の複合によるものであり，それを統一的に体系化することは困難であるためである．これに対する一つの解決法として，タイミングの適切さを，大規模データに基づいて判定することが考えられるが，現存する音声対話コーパスに収録されたあいづちの発生タイミングは，対話の環境や聞き手の心理状態などによる揺れが大きく，上述の目的に直接利用するためのデータとして適さない．

そこで著者らは，上述の問題を解決し，より実践的なデータを整備するために，以下の 4 つの方針を設けてタグ付け作業を実施し，タグ付けの安定性を備えたあいづちコーパスを構築した [7]．なお，タグ付け作業は，CIAIR 車内音声対話コーパス [8] に収録されているドライバ発話に対して，1 名の作業者により実施した．

1）網羅的なタグ付け: 作業者は，不自然でないあらゆるタイミングにあいづちタグを付与する．

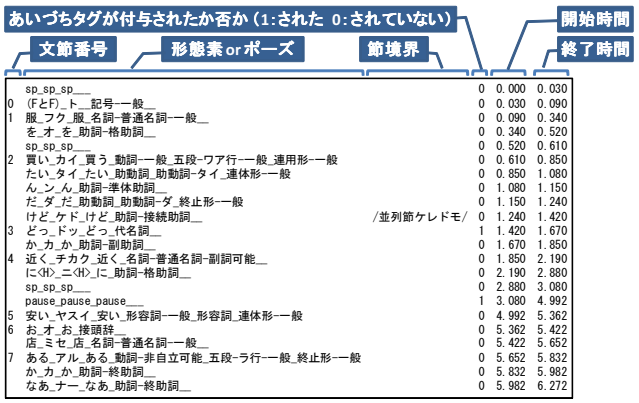

図 1　あいづちコーパスの例

2）オフライン環境でのタグ付け: 作業者は，付与対象の音声を一回以上聞いた上で，ドライバ音声の書き起こしテキストに対してタグ付けする．

3）あいづち発生タイミングの離散化: 対話ターンを，時間軸上に連続した形態素区間または無音区間（以下，**基本区間**）からなる列であるとし，作業者は，基本区間ごとに，あいづち生成タイミングとして適切であるかを判断し，適切であればタグを付与する．なお，無音区間については，200 ミリ秒を超える場合には，200 ミリ秒の無音区間 (sp) とそれ以外の無音区間 (pause) とに分割し，それぞれを基本区間とした．

4）あいづち合成音の再生による推敲: 作業者がタグを付与したタイミングであいづちが発生する対話音声を自動生成し，作業者はその音声を再生することによって推敲する．本研究はあいづちの発生タイミングに焦点をあてるため，いくつか存在する，あいづちの種類のうち，理解・同意を示す最も一般的な様式として「はい」を採用し，その合成音を HITACHI 製の音声合成ソフトウェア HitVoice を用いて生成した．また，あいづち音声は，タグ付けされた基本区間の開始から 50 ミリ秒後のタイミングで発生を開始することとした．

図 1 に構築したあいづちコーパスの例を示す．各行は，基本区間を意味しており，ドライバ発話における形態素区間または無音区間の情報が表記されている．また，開始終了時間，ならびに，あいづちタグが付与されたか否かの情報を与えている．さらに，形態素の場合は，形態素情報や文節境界情報，節境界情報を付与している．ここで，形態素情報は ChaSen [9] + Unidic [10] を，節境界情報は CBAP [11] を，各基本区間の開始・終了時間は連続音声認識システム Julius [12] を用いて自動的に付与した．また，文節境界は CIAIR 車内音声対話コーパスに付与されているものを利用した．なお，表 1 に構築したあいづちコーパスの規模を示す．

本研究では，構築したあいづちコーパスにおけるタグ付けの安定性とタグ付け位置の妥当性を評価するため，(1) 4 名（コーパス作成者 A と被験者 B,C,D）によるタグ付け結果の κ 値 [13] の測定と，(2) タグ付け結果に基づいて生成したあいづち音声の自然さに関する主観的評価，を実施した．その結果，(1) では，κ 値が usable quality（$.67 < \kappa < .80$）[14] を示し，(2) では，全体の 98.5%のあいづちが自然であると被験者により判断されており，本タグ付け作業によって，高い安定性と自然さを備えたあいづちコーパスが構築できることを確認した [7].

表 1　あいづちコーパスの規模

話者	346
発話ターン	11,181
節	14,643
文節	43,723
形態素区間	94,030
無音区間	19,142
あいづち	5,416

表 2　分析データの規模

話者	35
発話ターン	1,219
節	1,421
文節	4,507
形態素区間	9,881
無音区間	1,813
あいづち	546

表 3　文節境界の種類とあいづち発生割合

品詞	割合
助詞	21.29% (329/1545)
記号	3.77% (21/557)
名詞	9.61% (27/281)
接続詞	12.81% (31/242)
感動詞	1.71% (3/175)
助動詞	36.59% (60/164)
副詞	5.45% (6/110)
形容詞	11.59% (8/69)
動詞	15.91% (7/44)
接尾辞	21.62% (8/37)
連体詞	2.70% (1/37)

表 4　節境界の種類とあいづち発生割合

節の種類	割合
感動詞	2.05% (7/342)
主題ハ	10.94% (14/128)
並列節ケレドモ	97.60% (122/125)
談話標識	10.53% (12/114)
引用節	0.00% (0/67)
連体節	11.11% (7/63)
連用節	10.53% (4/38)
理由節ノデ	94.29% (33/35)
理由節カラ	82.35% (28/34)
テ節	35.71% (10/28)

3. あいづち発生タイミングの分析

本研究では，あいづち生成タイミングを推定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，あいづち生成タイミングの特徴分析を行った．分析には，前章で述べたあいづちコーパス [7] の一部を用いた．分析データの規模を表 2 に示す．全基本区間 11,694 個に対して，あいづちは 546 個発生しており，基本区間当たりのあいづち発生割合は 4.7%であった．以下では，文節境界や節境界，ポーズ，発話速度，ピッチ，パワーに着目し，それらとあいづち生成タイミングとの関係について調査した．

3.1 文節境界

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．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，文節と文節の間，すわなち文節境界にあいづちが打たれやすいと考えられる．分析データ中に文節境界は 3,288 個存在しており，そのうち，504 個にあいづちが発生していた（発生割合は 15.3%）．基本区間当たりのあいづち発生割合 4.7%と比べ高い数値を示しており，文節境界にあいづちが発生しやすいということが分かった．なお，次のいずれかの基本区間にあいづちが発生している場合，文節 b_i と b_{i+1} の境界にあいづちが発生していると判定した．

- b_i と b_{i+1} の間に存在する無音区間（sp または pause）
- b_{i+1} の先頭の形態素区間

ここで文節は，先頭と末尾を形態素区間とする，基本区間の列から構成されるものとする．

次に，文節境界をその直前の文節（すなわち，文節 b_i と b_{i+1} の間の文節境界に対して，b_i）の最終形態素の品詞によって分類し，その分類ごとにあいづちの打たれやすさの異なりを分析した．表 3 に，文節境界の種類ごとのあいづち発生割合を示す．直前の文節の最終形態素の品詞が助詞や助動詞，接尾辞，動詞の場合にあいづちが打たれやすいことが分かった．

3.2 節境界

節は文節よりも強い意味的なまとまりであるので，節境界にはあいづちが打たれやすいと考えられる．分析データ中に，節境界は 1,082 個存在しており，このうち，283 個にあいづちが発生していた（発生割合は 26.2%）．文節境界へのあいづち発生割合（15.3%）と比べて高く，文節境界以上に節境界にはあいづちが発生しやすいことが分かった．なお，節 c_i と c_{i+1} の間の文節境界にあいづちが発生しているかどうかの判定は，**3.1** の文節境界の判定時と同様である．ここで，節は，先頭と末尾を形態素区間とする，基本区間の列から構成されるものとする．

次に，節境界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 4 に，節境界の種類ごとのあいづち発生割合を示す．節境界「並列節ケレドモ」,「理由節ノデ」,「理由節カラ」にあいづちが打たれやすいことが分かる．また，節境界の種類によって，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．

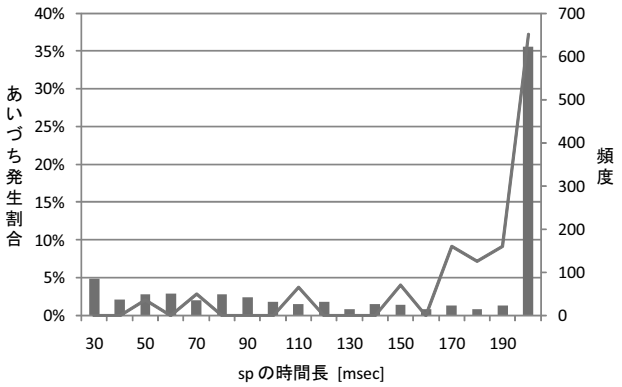

図 2　ポーズ長とあいづち発生割合

3.3 ポーズ

あいづちには相手に発話を促す機能があるため，相手発話中にポーズがあるとあいづちが打たれやすいと考えられる．

本研究では，200ms を超える無音区間は先頭 200ms の無音区間（sp）とそれ以降の無音区間（pause）とに 2 分割し，200ms 以下の無音区間は分割せず sp の 1 区間としている．これは，200ms 程度の無音状態が続くことによって初めてポーズの存在を認識でき，その認識によって，その後，あいづちが打たれることを想定している．調査では，sp の直後の基本区間にあいづちが発生する割合は 20.2%(241/1,196) であった．この割合は，全基本区間のあいづち発生割合 4.7%と比べて高い．なお，pause の直後の基本区間に発生したあいづちはなかった．

次に，sp の長さとあいづちの打たれやすさの関係を分析した．図 2 は，横軸に sp の長さ，縦軸に頻度を棒グラフで，あいづち発生割合を折れ線グラフで示している．sp の長さが短くなるにつれて，あいづち発生割合が低下している．

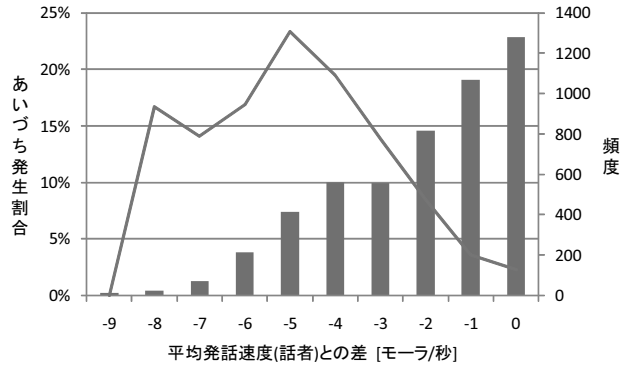

図 3　発話速度と平均発話速度（話者）の差

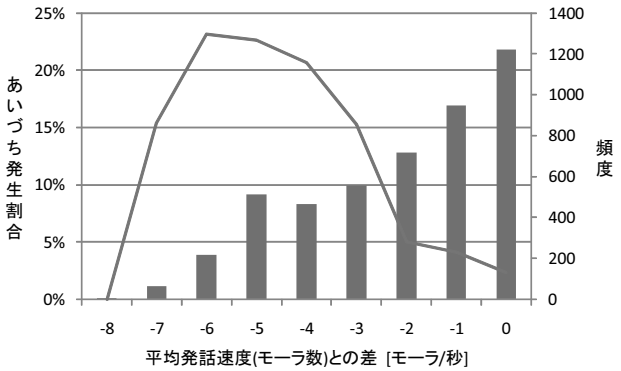

図 4　発話速度と平均発話速度（モーラ数）の差

3.4 発話速度

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，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．平均発話速度として，以下の二つの値を測定し利用した．

- 平均発話速度（話者）: そのドライバがその形態素より前に発話した全形態素に対する平均発話速度
- 平均発話速度 (モーラ数) : その形態素と同じモーラ数を持つ全形態素に対する平均発話速度

平均発話速度（話者）は，人による発話速度の違いを考慮して設定したものであり，平均発話速度 (モーラ数) は，モーラ数が同じ形態素は同程度の発話速度をもつと仮定して設定したものである．

発話速度が平均発話速度（話者）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 1.4%(88/6,132) であり，遅い場合は 10.6%(458/4,343) であった．また， 発話速度が平均発話速度（モーラ数）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 1.5%(100/6,479) であり，遅い場合は 11.2%(446/3,996) 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．

次に，平均発話速度よりも発話速度が遅くなるほど，その形態素の直後にはあいづちが打たれやすくなると考え，各形態素の発話速度と平均発話速度の差を 1 モーラ/秒ごとに分類し，それぞれ，その直後に対するあいづち挿入割合を調査した．その結果を図 3 と図 4 に示す．いずれの図も，横軸に平均発話速度との差をとり，棒グラフにより頻度，折れ線グラフによりあいづち発生割合を示している．なお，平均発話速度との差は，小数点第 1 位を四捨五入し整数にすることによりクラスタリングしている．図 3 が平均発話速度（話者）と比較した場合，図 4 が平均発話速度（モーラ数）と比較した場合である．全体としては，形態素の発話速度が遅くなるほど，その直後にあいづちが打たれやすくなる傾向があることが分かる．なお，平均発話速度と比較して発話速度が極端に遅くなると，あいづち発声割合が低下している．これは，発話速度が極端に遅い形態素はフィラーや感動詞であることが多く，これらの発話は話し手が考え中であることを伝えるサインとも受け取ることができることから，あいづちが打たれない傾向にあったためだと考えられる．

表 5　各ピッチ変動パターンのあいづち発生割合

ピッチ変動パターン	割合	
平坦-平坦-平坦	5.68%	(189/3,326)
下降-下降-下降	6.22%	(56/901)
平坦-平坦-下降	6.00%	(32/533)
平坦-下降-下降	6.35%	(31/488)
下降-平坦-平坦	5.45%	(22/404)
下降-下降-平坦	6.29%	(21/334)
上昇-平坦-平坦	3.67%	(8/218)
上昇-上昇-平坦	12.15%	(22/181)
平坦-上昇-上昇	13.19%	(19/144)
平坦-下降-平坦	6.36%	(7/110)

3.5 ピッチとパワー

あいづち発生タイミングは，話者音声のピッチやパワーと関係があり，ピッチ変化が尻上がり型や上昇型である場合やパワーが弱まったときに，あいづちが打たれやすいといった知見が報告されている [2]．本研究では，ピッチやパワーの変化とあいづち発生タイミングの関係を網羅的に調査するため，各基本区間ごとに，その直前の形態素区間までの 100ms 区間からピッチとパワーの変動パターンを算出し，それらのパターンごとにあいづち発生割合を分析した．表 5 に，頻度が多い上位 10 個のピッチ変動パターンのあいづち発生割合を示す．ピッチ変動パターンが「上昇-上昇-平坦」や「平坦-上昇-上昇」であるときは，その直後にあいづちが発生しやすいことがわかる．次に，表 6 に，頻度が多い上位 10 個のパワー変動パターンのあいづち発生割合を示す．パワー変動パターンが「平坦-下降-下降」または「下降-下降-下降」であるときは，その直後にあいづちが発生しやすいことがわかった．

なお，ピッチとパワーの変動パターンは，文献 [5] と同様に算出した．具体的には，まず，注目している基本区間の直前の形態素区間の終了時間からさかのぼって取ってきた 100ms 区間を 3 つの区間に分け (窓枠は 50ms，フレームシフトは 25ms)，各区間のピッチとパワーの 1 次回帰係数を求める．次に，その値に基づいて，各区間を 3 分類 (上昇，平坦，下降) し，3 つの区間の分類クラスを順に列挙した列 (例えば，上昇-平坦-下降) を変動パターンとした．ここで，ピッチとパワーの値は，音声分析ツール Praat [15] を使用し，5ms ごとに解析したものを用いた．

表 6　各パワー変動パターンのあいづち発生割合

パワー変動パターン	割合	
平坦-平坦-平坦	4.49%	(72/1,602)
平坦-平坦-下降	9.45%	(151/1,598)
平坦-下降-下降	11.85%	(139/1,173)
上昇-上昇-平坦	1.44%	(12/836)
下降-下降-下降	9.87%	(74/750)
上昇-平坦-下降	2.33%	(14/601)
上昇-平坦-平坦	1.67%	(8/478)
上昇-上昇-上昇	0.87%	(4/460)
下降-上昇-上昇	2.08%	(8/384)
上昇-上昇-下降	1.31%	(4/305)

4. あいづち生成タイミングの検出

構築したあいづちコーパスを利用し，統計的手法によるあいづち生成タイミングの検出を試みた．

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，あいづちを生成できるか否かを Support Vector Machine(SVM) を用いて推定する．

ある基本区間 m_i に対してあいづち生成タイミングを推定する際に，SVM で利用した素性を表 7 に示す．これらの素性はすべて，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，その基本区間の入力が終わり次第得られるものとする．ここで，m_j とは，m_i の直前に位置する形態素区間のことであり，m_{i-1} が形態素の場合は m_{i-1}，m_{i-1} が sp の場合は m_{i-2}，m_{i-1} が pause の場合は m_{i-3}，となる．素性 9 と 10 の平均発話速度（話者）と，素性 11 と 12 の平均発話速度 (モーラ数) は，**3.4** で述べた定義である．素性 13 は，直前にあいづちが生成されていない場合は使用しない．素性 14 と 15 における変動パターンは，**3.5** と同じ方法で求める．

5. 検 出 実 験

あいづちコーパス [7] を用いてあいづち生成タイミングの検出実験を実施した．

5.1 実 験 概 要

構築したあいづちデータのうち，分析に用いた 1,219 対話ターンをテストデータとして，残りの 9,962 対話ターンを学習データとして使用した．SVM のツールとして LibSVM [16] をデフォルトのオプションのまま使用した．

評価には以下の指標を用いた．

$$\text{適合率} = \frac{\text{正しく生成されたあいづちの数}}{\text{生成されたあいづちの数}}$$

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

コーパス構築作業者のタグ付け結果を正解データとし，その正解データとの間であいづち生成箇所が一致すれば正しく生成されたと判定した．なお，比較のために，**2.** のあいづちコーパスの評価において被験者 B，C，D が実施した，あいづちタグ付け作業の結果を用いて，それぞれの適合率と再現率を測定した．

表 7　SVM で用いた素性

1.	m_j(m_i の直前に位置する形態素区間) が文節の最終形態素であるか否か
2.	1 が真の場合，m_j の品詞
3.	1 が真の場合，m_j が属する文節が名詞，動詞，形容詞のいずれかを含むか否か
4.	m_j が節の最終形態素であるか否か
5.	4 が真の場合，m_j が属する節の種類
6.	m_{i-1} が sp であるか否か
7.	6 が真の場合，m_{i-1} のポーズ長 α(秒) が以下の 4 分類のいずれであるか．（「$\alpha \leqq 0.1$」,「$0.1 < \alpha \leqq 0.17$」,「$0.17 < \alpha < 0.2$」,「$\alpha = 0.2$」）
8.	m_{i-1} が pause であるか否か
9.	m_j の発話速度が平均発話速度 (話者) より遅いか否か
10.	9 が真の場合，m_j の発話速度と平均発話速度 (話者) の差 β (モーラ/秒) が以下の 3 分類のいずれであるか．（「$\beta < 2$」,「$2 \leqq \beta < 6$」,「$6 \leqq \beta$」）
11.	m_j の発話速度が平均発話速度 (モーラ数) より遅いか否か
12.	11 が真の場合，m_j の発話速度と平均発話速度 (モーラ数) の差 γ(モーラ/秒) が以下の 3 分類のいずれであるか．（「$\gamma < 2$」,「$2 \leqq \gamma < 6$」,「$6 \leqq \gamma$」）
13.	直前のあいづち生成時間から m_{i-1} の終了時間までの時間長 δ(秒) が以下の 4 分類のいずれであるか．（「$\delta \leqq 0.6$」,「$0.6 < \delta \leqq 1.4$」,「$1.4 < \delta \leqq 2.9$」,「$2.9 < \delta$」）
14.	ピッチ変動パターン
15.	パワー変動パターン

表 8　あいづちタイミング検出実験の結果

	適合率	再現率	F 値
本手法	82.2% (361/439)	66.1% (361/546)	73.3
被験者 B	74.0% (168/227)	78.9% (168/213)	76.4
被験者 C	74.4% (157/211)	73.7% (157/213)	74.0
被験者 D	79.0% (162/205)	76.1% (162/213)	77.5

5.2 実 験 結 果

表 8 に本手法ならびに各被験者のタグ付け結果に対する，再現率，適合率及びこれらの調和平均である F 値を示す．F 値において，本手法は，被験者によるタグ付け結果を若干下回る程度の検出性能を達成している．なお，本手法によるあいづちタイミングの推定結果と本コーパスの間の κ 値は 0.728 であった．**2.** のあいづちコーパスの評価で測定した，コーパス作成者 A と被験者 B, C, D の間の κ 値（0.755, 0.727, 0.763）と同程度であった．

また，本手法による推定結果と正解データが全ての基本区間で一致した対話ターンは 1,023 ターン存在し，全対話ターンの 83.9%を占めた．ただし，正解データにおいてあいづちタグが 1 回以上付与された対話ターンは 349 個あり，そのうち，本手法が全ての基本区間で正解した対話ターンは 48.1%(168/349) であった．1 対話ターンの全基本区間であいづちタイミング推定が成功した例を図 5 に示す．1 行目は話者音声の基本区間列を，2 行目は正解データ上のタグ付けを，3 行目は本手法の推定結果をそれぞれ表す．2,3 行目において，1 が記された基本区間はあいづち生成箇所を表す．この例では，形態素区間「郵便」「近く」へのあいづち生成や，sp と pause の連続個所におけるあいづち生成箇所の選択が成功している様子がわかる．

基本区間	(F えっと)	年賀	状	の	はがき	sp	買い	たい	から	郵便	局	に	行き	たい	ん	だ	けど	近く	で	sp	pause	いちばん	近い	とこ	で	sp	pause	郵便	局	ある	か	なあ
正解										1								1			1						1					
実験結果										1								1			1						1					

図 5　あいづちタイミング推定の成功例

基本区間	(F んー)	雰囲気	の	いい	お	店	が	いい	けど	sp	pause	どっち	が	いい	か	な
正解										1						
実験結果											1					

図 6　本手法が誤ってあいづちを生成した例 (1)

基本区間	すし	の	こうずし	って	いう	の	sp	pause	を	sp	pause	お	sp	pause	願い	し	ます
正解								1									
実験結果								1		1							

図 7　本手法が誤ってあいづちを生成した例 (2)

基本区間	(F えーと)	sp	ファースト	フード	sp	みたい	な	sp	お	店	sp	どっ	か	sp	ある	か	なあ
正解					1												
実験結果																	

図 8　本手法が検出できなかったあいづち発生位置の例

5.3　検出誤りの分析

本手法によるあいづち生成タイミングの検出誤りについて分析した．

本手法があいづちを生成できると推定した基本区間のうち，正解データのあいづち発生位置と異なるものは 78 箇所存在した．そのうちの 17 箇所は，図 6 のように，正解データのあいづち発生位置と 1 区間だけずれたものであり，すべて sp と pause の選択に失敗したものであった．残りの 61 箇所は正解データのあいづち発生位置と 1 区間以上ずれていた．このうちの 39 箇所は，無音区間（sp または pause）へ誤ってあいづちを生成したものであった．図 7 に，本手法が無音区間（sp または pause）へ誤ってあいづちを生成した例を示す．この例では，「を」の後の sp に誤ってあいづちを生成している．

一方，正解データのあいづち発生箇所のうち，本手法であいづちを生成できなかったのは 185 箇所であった．前述した適合率を低下させる誤りの場合と同様に，このうちの 17 箇所は，本手法の検出結果と 1 区間だけずれたものであり，すべて sp と pause の選択に失敗したものであった．残りの 168 箇所は，その前後 1 区間以内に本手法であいづちを生成できなかった箇所であり，このうちの 103 箇所は，無音区間 (sp または pause) であいづちが発生したものだった．図 8 に，本手法があいづちを生成できなかった無音区間の例を示す．この例では,「フード」の後の sp にあいづちを生成することができなかった．

6. おわりに

本論文では，あいづち発生タイミングのタグ付けの安定性を備えた大規模コーパスに基づいた，あいづち生成タイミングの分析と検出について述べた．分析では，文節境界や節境界，ポーズ，発話速度，ピッチ・パワーに着目し，あいづち発生タイミングの特徴を明らかにした．この特徴分析に基づいた統計的手法により，あいづち生成タイミングを検出する手法を提案した．検出実験の結果，適合率で 82.2%，再現率で 66.1%（F 値：73.3%）を示しており，本手法の有効性を確認した．

今後は，あいづち生成タイミングの検出に有効な素性の選択をより詳細に行った上で，各素性の効果について調査する予定である．また，実験結果に対して，被験者による主観的評価を実施し，本手法により生成したあいづちの自然さを評価する予定である．

謝辞　本研究は一部，科研費挑戦的萌芽研究 (No. 21650028)「音声対話システムの個性化に関する基礎的研究」による．

文　献

[1] N. Cathcart, J. Carletta, and E. Klein, “A shallow model of backchannel continuers in spoken dialogue,” Proceedings of the Tenth Conference on European Chapter of the Association for Computational Linguistics (EACL2003), pp.51–58, 2003.

[2] 堀口純子，日本語教育と会話分析，くろしお出版，1997.

[3] S.K. Maynard, Japanese conversation : self-contextualization through structure and interactional management, Ablex, 1989.

[4] 水谷信子，“日本語教育と話ことばの実態 -あいづちの分析-,” 金田一春彦博士古稀記念論文集, 第二巻，言語学編，pp.261–279, 三省堂，1984.

[5] N. Kitaoka, M. Takeuchi, R. Nishimura, and S. Nakagawa, “Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems,” Journal of the Japanese Society for Artificial Intelligence, vol.20, no.3, pp.220–228, 2005.

[6] N. Ward and W. Tsukahara, “Prosodic features which cue back-channel responses in English and Japanese,” Journal of Pragmatics, vol.32, pp.1177–1207, 2000.

[7] Y. Kamiya, T. Ohno, and S. Matsubara, “Coherent back-channel feedback tagging of in-car spoken dialogue corpus,” Proceedings of the 11th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL2010), pp.205–208, 2010.

[8] N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura, “CIAIR in-car speech corpus –influence of driving status–,” IEICE Transactions on Information and Systems, vol.E88-D, no.3, pp.578–582, 2005.

[9] Y. Matsumoto, A. Kitauchi, T. Yamashita, and Y. Hirano, “Japanese morphological analysis system ChaSen version 2.0 manual,” NAIST Technical Report, NAIST-IS-TR99009, 1999.

[10] 伝　康晴，小木曽智信，小椋秀樹，山田　篤，峯松信明，内元清貴，小磯花絵，“コーパス日本語学のための言語資源：形態素解析用電子化辞書の開発とその応用,” 日本語科学，vol.22, pp.101–122，2007.

[11] 丸山岳彦，柏岡秀紀，熊野　正，田中英輝，“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.

[12] A. Lee, T. Kawahara, and K. Shikano, “Julius – an open source real-time large vocabulary recognition engine,” Proceedings of the Seventh European Conference on Speech Communication and Technology (EUROSPEECH2001), pp.1691–1694, 2001.

[13] J. Cohen, “A coefficient of agreement for nominal scales,” Educational and Psychological Measurement, vol.20, pp.37–46, 1960.

[14] J. Carletta, “Assessing agreement on classification tasks,” Computational Linguistics, vol.22, no.2, pp.249–254, 1996.

[15] P. Boersma and D. Weenink, “Praat: doing phonetics by computer (version 5.1.05),” 2009. Software available at `http://www.praat.org/`.

[16] C.-C. Chang and C.-J. Lin, “LIBSVM: a library for support vector machines,” 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.