あいづち生成タイミング制御のための音声対話コーパスの構築と利用

神谷　優貴 †　　大野　誠寛 ‡　　松原　茂樹 †

† 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

Construction and Utilization of Spoken Dialogue Corpus for Timing Control of Back-Channel Feedbacks

Yuki Kamiya†　　Tomohiro Ohno‡　　Shigeki Matsubara†

†Graduate School of Information Science, Nagoya University
‡Graduate School of International Development, Nagoya University

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，**安定性**（作業者による揺れが少ないタグ付けが施されていること）と自然さを備えたあいづちコーパスの構築と，それを利用したあいづち生成タイミングの推定について述べる．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, 2, 3, 4, 5, 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を用いて，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く，あいづち生成タイミングを推定する上で実践的に利用するのに適したデータとはいえない．本研究では，まず，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を与えることにより，タグ付けの揺れが少ないあいづちコーパスを構築する．その上で，構築したあいづちコーパスを用いることにより，適切なあいづち生成タイミングを高精度に推定することを試みる．

本研究では，既存のデータとして，CIAIR 車内音声対話コーパス [7] に収録されているドライバ発話 11,181 ターンを使用した．あいづちタグ付け作業のための Web インタフェースを作成し，5,416 個のあいづちデータからなる大規模なコーパスを作成した．コーパスを評価するために，あいづちタグ付け実験を実施し，本設計によって，高い安定性と自然さを備えたあいづちコーパスを作成できることを確認した．また，あいづちタイミングの推定実験を行い，本コーパスを用いることにより，自然なあいづち生成機能を実現できることを確認した．

本論文の構成は以下の通りである．次の 2 章では，コーパスの設計について論じる．3 章では，あいづちコーパスの構築について述べ，4 章では，構築したコーパスの評価について報告する．5 章では，構築したコーパスの利用について述べる．最後に 6 章で，本論文のまとめと今後の課題について述べる．

2　あいづちコーパスの設計

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り [1, 2, 3, 4, 5, 6]，

- ポーズ中，または，ポーズの後
- 接続詞や終助詞の後
- 音声が弱まったとき
- 末尾に音下げのある節の後

にあいづちが挿入されやすいという傾向が知られている．しかし，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．

そこで本研究では，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，安定性を有するあいづちコーパスを構築する．このために，タグ付け作業の方針として以下を設定した．

- **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．人間同士での対話では，たとえあいづちを打つことが可能なタイミングであっても，あいづちが打たれたり打たれなかったりする [6]．網羅的にタグ付けすることにより，揺れの少ないタグ付け作業が可能となる．

- **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，タグ付けする．オンライン対話環境であいづちを打つのに比べ，打ち忘れやタイミングのずれが少なくなり，安定的なタグ付け作業が可能となる．

- **あいづち発生タイミングの離散化**
  ドライバ発話を時間区分し，各区分に対してタグを付与するか否かを判断する．通常の対話であれば，聞き手は任意の時点であいづちを打つことができるが，自由度が高い分，あいづちタイミングの揺れが大きくなる．タイミングを離散化することにより，タグ付けが安定するとともに，作業を効率化できる．

- **あいづち合成音の再生による推敲**
  作業者は，タグ付けの妥当性を，音声を通してその場で確認する．すなわち，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が自動生成され，作業者はそれを再生することによって推敲する．なお，コーパスはあいづち機能を備えたシステム構築での利用を目的としており，あいづち音声は合成音によって実現する．

3　あいづちコーパスの構築

音声対話コーパスを用いてタグ付け作業を実施し，あいづちコーパスを作成した．

3.1　音声対話コーパス

タグ付けの対象として，CIAIR 車内音声対話コーパス [7] を使用した．このコーパスは，実走行環境下で遂行された，ドライバとオペレータとの間の道路案内や店検索などをタスクとする対話を収集したものであり，音声データとその文字

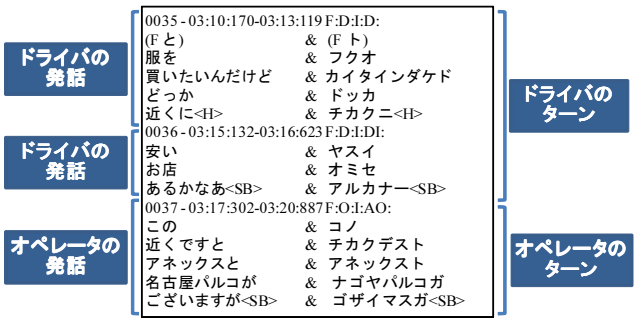

図 1: 書き起こしテキストの例

化データが収録されている．図 1 に書き起こしテキストの例を示す．本研究では，このうちドライバの発話を使用した．形態素解析器 ChaSen[8] を用いて，ドライバ発話を形態素に分割した．なお，辞書には，店名などの固有名詞を独自に追加した Unidic [9] を用いている．また，連続音声認識システム Julius[10] を用いて，各形態素の発声区間の開始・終了時間を推定し付与した．

3.2　対話データへのタグ付け作業

本研究では，2 章で示した設計に従い，ドライバ音声中の適切なタイミングにあいづちタグを付与することにより，あいづちコーパスを作成した．

「網羅的にタグ付け」するため，作業者は，ドライバの対話ターン[1] を最初から聴いていき，あいづちを打てるタイミングが見つかれば，そこにタグ付けする．なお，あいづちを打って不自然となるか否かは，その直前のタグ付け位置も考慮して判断する．

「オフライン環境でタグ付け」するため，作業者は，ドライバの対話ターンの書き起こしテキストに対してタグ付けを与える．

また，「あいづち発生タイミングを離散化」するために，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，基本区間ごとに，それがあいづち挿入可能なタイミングであればあいづちタグを付与する．ただし，無音区間については，その長さが一様でないことを考慮し，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．これにより作業者は，長いポーズ内でのあいづち発生タイミングを，ポーズの冒頭，もしくは，200 ミリ秒以上経過後のいずれかから選択できる．図 2 に，区間分割された対話ターンの例を示す．

さらに，「あいづち音声の再生による推敲」を可能にするために，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ドライバの対話音声を，リアルタイムで生成し，それを再生できる環境を整備した．なお，あいづちにはいくつかの種類があり，場面によって使い分けられることが知られているが，本研究はあいづちの発生タイミングに焦点をあてるものであり，ここでは理解・同意を示す最も一般的な様式として「はい」を採用し，その合成音を生成することとした．

[1] オペレータ発話が終了しドライバ発話が開始されてから，次にオペレータが発話を開始する直前のドライバ発話の終了までを一つの対話ターンと定めた．

基本区間	sp	(Fと)	服	を	sp	買い	たい	ん	だ	けど	どっ	か	近く	に	sp	pause	安い	お	店	ある	か	なあ
開始時間	0.000	0.030	0.090	0.340	0.520	0.610	0.850	1.080	1.150	1.240	1.420	1.670	1.850	2.190	2.880	3.080	4.992	5.362	5.422	5.652	5.832	5.982
終了時間	0.030	0.090	0.340	0.520	0.610	0.850	1.080	1.150	1.240	1.420	1.670	1.850	2.190	2.880	3.080	4.992	5.362	5.422	5.652	5.832	5.982	6.272

図 2: 基本区間への分割

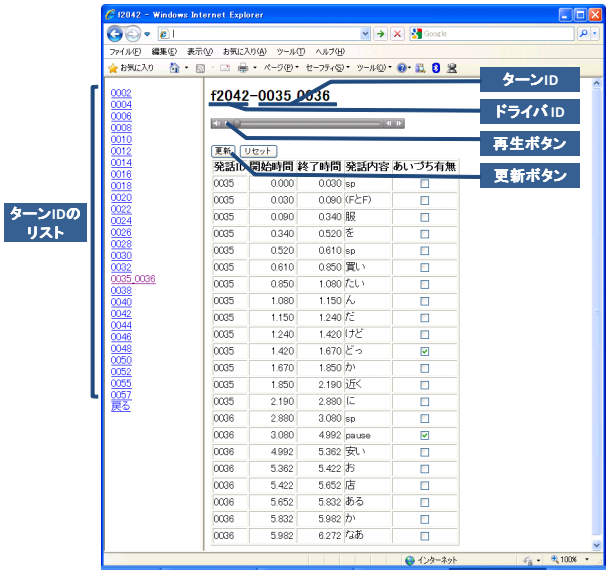

図 3: あいづち収集インタフェース

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[2]．また，あいづち音声「はい」は，HITACHI 製の音声合成ソフトウェア HitVoice を用いて生成した．

3.3 あいづち収集インタフェース

Web インタフェースを備えたあいづちタグ付け作業環境を開発した．作業画面を図 3 に示す．1 対話ターンに対応する基本区間列が表形式で表示され，あいづち生成可能な基本区間にチェックすることによりタグ付けする．すなわち，1 対話ターンに対するタグ付けは以下の手順で遂行される．

1. 再生ボタンを押し，ドライバ音声を聞く．
2. あいづち生成に適した基本区間にチェックし，更新ボタンを押して，あいづち音声を含むドライバ音声を作成する．
3. 再生し，あいづち音声を含んだドライバ音声を聞く．過不足なくタグが付与されていると判断したら終了．そうでなければ，2 へ．

3.4 あいづちコーパス

CIAIR 車内音声対話コーパス [7] に収録されているドライバ発話 11,181 ターンに対して，訓練を受けた 1 名の作業者によりあいづちタグ付け作業を実施し，5,416 のあいづち

[2] あいづち発生タイミングの調査 [11] では，発話終了からあいづち挿入までのポーズ長が 0 秒から 100 ミリ秒の間に多く分布することが明らかにされており，本研究ではこれを参考に設定した．

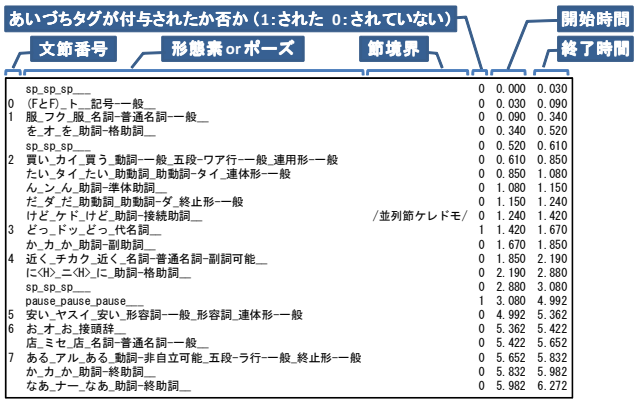

図 4: あいづちコーパスの例

表 1: あいづちコーパスの規模

話者	346
発話ターン	11,181
節	16,896
文節	43,727
形態素	94,030
ポーズ	19,142
あいづち	5,416

からなる大規模なタグ付きコーパスを作成した．表 1 に構築したあいづちコーパスの規模を示す．平均して 26 基本区間に一度の頻度であいづちが発生していた．

構築したあいづちコーパスの例を図 4 に示す．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またはポーズ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，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，節境界情報は CBAP[12] を用いて自動的に付与した．なお，形態素情報は，**3.1** で述べた基本区間への分割と同様に，ChaSen[8] + Unidic[9] による形態素解析結果である．

4 あいづちコーパスの評価

作成した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の実験を実施した．

4.1 タグ付けの安定性

作成したコーパスにおいて，安定的にタグが付与されていることを確認するために，評価実験を実施した．実験では，タグ付けされたデータに対して別の被験者 3 名がそれぞれタグ付けを行った．コーパス構築作業者（A）と被験者（B, C,

表 2: 本コーパスにおける作業者間の一致の程度

	A,D	A,B	B,D	A,C	C,D	B,C
κ	0.763	0.755	0.750	0.727	0.722	0.696
$P(O)$	0.977	0.975	0.975	0.973	0.973	0.969
$P(E)$	0.903	0.898	0.900	0.901	0.903	0.898

表 3: 既存のコーパスにおける作業者間の一致の程度

	a,c	a,d	a,b	c,d	b,c	b,d
κ	0.536	0.438	0.322	0.311	0.310	0.167
$P(O)$	0.974	0.982	0.960	0.969	0.951	0.960
$P(E)$	0.944	0.968	0.941	0.955	0.929	0.952

D）の計 4 名の間の一致の程度を測るため，各 2 者間の一致度を測定することとし，指標として，Cohen の κ 値 [13] を使用した．これは，観測された一致率を $P(O)$，期待される一致率を $P(E)$ とするとき，

$$\kappa = \frac{P(O) - P(E)}{1 - P(E)}$$

で計算される．実験では，ランダムにドライバ 10 名を選択し，この 10 名によって発話された 358 対話ターンを使用した．また，各被験者は，コーパス構築作業者と同様に，事前にタグ付け練習を実施した．評価の結果を表 2 に示す．2 人の作業者の一致率について，文献 [14] では，$.80 < \kappa$ を good reliability，$.67 < \kappa < .80$ を usable quality としており，作業者間で実質的な一致が見られることが示された．

比較対象として，既存のあいづちコーパス [15] における κ 値を使用した．このコーパスでは，収録されたドライバ音声 297 対話ターンを再生し，4 名の被験者が同一のデータに対して独立にあいづちを発声するという設定で収集されている．すなわち，本研究の設計とは異なり,「オンラインでのタグ付け」「連続的なタイミングでのタグ付け」「推敲なしでのタグ付け」という条件下でコーパスが作成されている．ただし，被験者に対しては，不自然でないタイミングにできる限りあいづちを打つように指示している．本コーパスと同様にドライバ音声を基本区間に分割し，各被験者のあいづち音声がどの基本区間で開始されているのかを対応づけた上で，被験者（a, b, c, d）間の κ 値を測定した．その結果を表 3 に示す．最も κ 値が低い被験者間で 0.167，最も高い被験者間でも 0.536 であり，本実験の作業は，いずれの被験者間の κ 値よりも高い値を示しており，本コーパスの高い安定性が示された．

4.2 タグ付け位置の妥当性

本コーパスでは，タグ付けの安定性を高めるためにタグ付け位置を離散化したが，その一方で，そのような制約がタグを付与する位置を制限し，不自然なタイミングにタグが付与される可能性がある．そこで，あいづちタイミングの自然さを評価するために，被験者実験を実施した．実験では，あいづちコーパスの音声を被験者 1 名が聴取し，各あいづちの自然さを主観的に判定した．なお，実験に用いた音声は，タグつけで生成したものと同様であり，あいづち音声は基本区間の開始から 50 ミリ秒後に発生する．

実験では，131 のあいづちを含む 345 対話ターンを使用した．このうち，不自然であると判定されたあいづちは 2 つであった．自然なあいづちが全体の約 98.47%に相当しており，このことからタグ付け位置を離散化したことの妥当性が示された．なお，2 か所のあいづちが不自然であると判定した理由を被験者に聴取したところ，その直前に打たれたあいづちとの間隔が近すぎることが挙げられた．コーパスの構築では，タグ付け作業者は，直前に付与したあいづちタグとの間隔を考慮した上で自然だと感じられるあいづち挿入位置を特定し，網羅的にタグを付与している．本実験においては，直前のあいづち位置からの時間的な間隔に基づく自然さの許容度において，作業者と被験者との間で若干の差があったため，このような判定結果が生じたものと推測される．

5 あいづちコーパスの利用

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推定を試みた．

5.1 あいづち生成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 4 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）とは，その話者が m_j 以前に発話した全形態素の平均発話速度を意味し，素性 11 と 12 の平均発話速度 (モーラ数) とは，学習データ中に存在する m_j と同じモーラ数を持つ全形態素の平均発話速度を意味する．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共に，次のように求める．まず，文献 [5] と同様に，m_j 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 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) を変動パターンとする．なお，ピッチとパワーは音声分析ツール praat[16] を用いて 5ms ごとに解析している．

5.2 実験

本コーパスを用いてあいづち生成タイミングの推定実験を実施した．

表 4: 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. (「$\alpha \leq 0.1$」,「$0.1 < \alpha \leq 0.17$」,「$0.17 < \alpha < 0.2$」,「$\alpha = 0.2$」)
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. (「$\beta < 2$」,「$2 \leq \beta < 6$」,「$6 \leq \beta$」)
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. (「$\gamma < 2$」,「$2 \leq \gamma < 6$」,「$6 \leq \gamma$」)
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. (「$\delta \leq 0.6$」,「$0.6 < \delta \leq 1.4$」,「$1.4 < \delta \leq 2.9$」,「$2.9 < \delta$」)
14.	ピッチ変動パターン
15.	パワー変動パターン

5.2.1 実験概要

構築したあいづちデータのうち，9,962 対話ターンを学習データとして，残りの 1,219 対話ターンをテストデータとして使用した．SVM のツールとして LibSVM[17] 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，4.1 節の被験者 B，C，D によるあいづちタグ付け結果の適合率と再現率を測定した．

5.2.2 実験結果

表 5 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推定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 であり，表 2 に示した作業者間の κ 値と同程度であった．

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．

表 5: あいづちタイミング推定実験の結果

	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

5.2.3 あいづち生成タイミングの推定誤りの分析

本手法によるあいづち生成タイミングの推定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，本手法の推定結果と 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安定性を備えた対話コーパスの設計・構築・評価について述べた.「網羅的なタグ付け」「オフラインでのタグ付け」「発生タイミングの離散化」「音声再生による推敲」という設計方針のもと，タグ付け作業を実施した．CIAIR 車内音声対話コーパスのドライバ発話を使用し，11,181 対話ターンに対する 5,416 個のあいづちを含む大規模なコーパスを作成した．評価の結果，本

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

コーパスが高い安定性と十分な自然さを有していることを確認した．また，本コーパスを用いたあいづち生成タイミングの推定実験を行った．推定実験により，自然なあいづち生成の実現性を確認した．

今回の推定実験は先行研究による知見を基に，SVM で用いる素性を決定した．今後は，データの詳細な分析を行い，あいづちタイミングの推定に有効な素性を調査する予定である．また，その分析結果に基づいて，より高精度なあいづちタイミングを検出するシステムを開発することを予定している．

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．

参考文献

[1] Nicola Cathcart, Jean Carletta, and Ewan Klein. A shallow model of backchannel continuers in spoken dialogue. In *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] Senko 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] Masashi Takeuchi, Norihide Kitaoka, and Seiichi Nakagawa. Timing detection for realtime dialog systems using prosodic and linguistic information. In *Proceedings of the First International Conference on Speech Prosody (SP2004)*, pp. 529–532, 2004.

[6] Nigel Ward and Wataru Tsukahara. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, Vol. 32, pp. 1177–1207, 2000.

[7] Nobuo Kawaguchi, Shigeki Matsubara, Kazuya Takeda, and Fumitada 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.

[8] Yuji Matsumoto, Akira Kitauchi, Tatsuo Yamashita, and Yoshitaka Hirano. *Japanese Morphological Analysis System ChaSen version 2.0 Manual*. NAIST Technical Report, NAIST-IS-TR99009, 1999.

[9] 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.

[10] Akinobu Lee, Tatsuya Kawahara, and Kiyohiro Shikano. Julius – an open source real-time large vocabulary recognition engine. In *Proceedings of the Seventh European Conference on Speech Communication and Technology (EUROSPEECH2001)*, pp. 1691–1694, 2001.

[11] 岡登洋平, 加藤佳司, 山本幹雄, 板橋秀一. 韻律パターンの認識を用いた相槌挿入とその評価. 情報処理学会研究報告, 96-SLP-10-7, pp. 33–38, 1996.

[12] 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.

[13] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, pp. 37–46, 1960.

[14] Jean Carletta. Assessing agreement on classification tasks. *Computational Linguistics*, Vol. 22, No. 2, pp. 249–254, 1996.

[15] Yuki Kamiya, Tomohiro Ohno, Shigeki Matsubara, and Hideki Kashioka. Construction of back-channel utterance corpus for responsive spoken dialogue system development. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC2010)*, pp. 2414–2419, 2010.

[16] Paul Boersma and David Weenink. *Praat: doing phonetics by computer (Version 5.1.05)*, 2009. Software available at `http://www.praat.org/`.

[17] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.