

JST/CREST 発話様式プロジェクトの電話対話データベース Telephone dialogue data base of JST/CREST Expressive Speech Processing Project

芦村和幸*1
Kazuyuki ASHIMURA

ニック・キャンベル*2*1
Nick CAMPBELL

*1 科学技術振興事業団 CREST(JST/CREST)
CREST, Japan Science and Technology Corporation(JST/CREST)

*2 株式会社国際電気通信基礎技術研究所 (ATR)
Advanced Telecommunications Research Institute International(ATR)

JST/CREST Expressive Speech Processing Project has been constructing a large-scale corpus of spontaneous Japanese dialogue under the fund of Japan Science and Technology Corporation. The corpus is designed to contain about 1000 hours of digitized dialogue speech to realize near future friendly computers which can speak like a human being. This paper reports the objective and the data collection method of the telephone dialogue database as a part of this corpus.

1. はじめに

JST/CREST 発話様式プロジェクト(以下、本プロジェクト)では、科学技術振興事業団戦略的基礎研究推進事業の研究資金を得て、2001年7月から2005年3月までの5年間の研究期間で、高度メディア社会のための表現豊かな発話音声処理の研究に取り組んでいる。

本プロジェクトでは、近未来の、人間と同じようにしゃべる親しみやすいコンピュータの実現を目指して、大規模対話データの収集・分析を行なうことにより、状況に応じたさまざまな発話内容および発話様式について、単に定性的な傾向に関する知見を得るのみならず、統計的分析にもとづくモデルを構築することを目標としている。

本稿では、本プロジェクトの一環として構築を進めている対話音声コーパスのうち、特に、電話対話データベースの概要と収集方法について報告する。

2. 電話対話データベース構築

2.1 大規模対話音声コーパスの必要性

日常会話には、発話意図、話し相手、話題、時間、場所などの緒条件に応じて、さまざまな発話内容および発話様式があらわれる。しかしながら、従来の音声合成や音声認識技術は、「スタジオで収録した、高音質だが自発性の低い音声データベース」もしくは「自然対話を収録した、自発性は高いが音質的に劣る音声データベース」を利用しており、[河口他 00][板橋 97][広瀬他 02]において指摘されているように、日常会話においてやりとりされているさまざまな情報の全てを包括的かつ高音質なデータとして扱うことはできていない。

本プロジェクトでは、自然な対話音声を高音質な設定で1000時間規模で収集することにより、今まで定性的にしか分析されてこなかった、日常会話における発話意図や発話環境と、発話様式との関連性を統計的に説明することを目指し、以下の三つの観点で研究を進めている。

大規模自然対話音声コーパスの収集・分析 大量データを効率的に収集・利用するための対話音声データベースを構築

連絡先: 〒 619-0288 京都府相楽郡精華町光台 2-2-2 ATR 内
科学技術振興事業団 CREST キャンベルチーム, E-mail:
ashimura@atr.co.jp

し、発話様式に関する声質分析および音響特徴抽出を行なうことにより、パラ言語情報の分析を行ないながら、韻律モデルを作成していく。

音声特徴自動ラベリング手法の研究 大規模対話音声コーパスの収集・分析を効率的に進めるために、音声データの自動セグメンテーションおよび自動アラインメント手法を検討する。

コーパスベース音声合成技術の改良 自然発話音声データ分析結果にもとづき、音声合成のための、音声単位選択アルゴリズムを高度化する。

2.2 電話対話データベース構築の目的

電話対話データの構築を通して、以下のようなさまざまな発話条件の異なりが、発話内容や発話様式にどのような影響をもたらすかを分析する。

- 被験者ごとの個人差
- 被験者同士の対話と、家族との対話
- 日本語母語話者との対話と、他言語母語話者との対話
- 収録時期の経過

例えば、発話タイミングの分析([芦村他 02]を参照)や、フォルマント情報にもとづく声の固さ(breathiness)の分析などを行なう。

3. データ収集方法

3.1 対話様式

音声収録にあたっては、2人の被験者が別々の会議室に分かれて、電話を介して対話を行なう。発話内容については特に指定せず各被験者に自由に話してもらう。なお、対話は電話を介して行なうが、高音質に収録を行なうために、収録音声は電話とは別に、ヘッドセット型マイクにより録音する。

3.2 書き起こしテキストおよびタグ情報

電話対話データベース中の情報としては、音声および書き起こしテキストが含まれる。

書き起こしテキストの記述にあたっては、音声をできるだけ忠実に漢字かな混じり文に転記するようにしているが、

[前川他 97]で提案されているような「基本形」と「発音形」の併記は行っていない。これは、本データベースが、現在、自然対話の書き起こしに必要なタグを検査しながら、データ作成作業を行なっている段階であり、現時点では、語句の厳密な再現性より、「タグによって表現される各種発話機能」を分析することを主眼としているためである。現在、音声認識や音声合成への応用も踏まえて、音声認識エンジン Julius などを用いた自動音素ラベル付与処理を行なっているが、今後は、感情、意図などのさまざまな発話様式についてもタグ化していく方針である(表 1)。なお、本データベースは関西においてデータ収集を行なっているため、関西弁独自の言い回しを考慮した書き起こしが必要となると考えられる。

表 1: タグの例

発話内容に関するタグ	発話様式に関するタグ
時間情報(発話開始・終了)	意図
ポーズ	感情
発話 overlap, 発話連続	発話態度
フィラー(長音)	声の調子
吸排気, 鼻をすする音など	発話機能に関するコメント

なお、上記タグのうち「発話機能に関するコメント」については、[ザトラウスキー 91]などにおいて指摘されているように、一つの発話に複数の発話機能が付与される場合がある。

3.3 データ収録条件

音声収録は、静かな会議室において、マイクロホンと唇の距離を一定に保つためにヘッドセット型マイクロホンを用いて、DAT に記録した。DAT のサンプリングは 48KHz, 16bit であるが、データベース中には、16kHz ダウンサンプリングした音声データを格納している。なお、マイクロホンの装着にあたっては、電話受話器がマイクロホンに干渉しないように、受話器はマイクロホン装着と反対側の手で持つよう指示した。

3.4 収録対話の構成

1 回あたり 30 分の対話を、特に発話内容を規定せずに、10 週間に渡り継続して収録することにより、被験者同士の親密度の変化が発話様式へ与える影響を観察する。具体的な収録対話数と対話時間は被験者の母語ごとに異なり、表 2 の通りである。

表 2: 収録対話数と対話時間

被験者の母語	対話数(対話)	対話時間(h)
日本語 vs. 日本語	80(8 対話 × 10 回)	40
日本語 vs. 英語	20(4 対話 × 5 回)	10
日本語 vs. 中国語	20(4 対話 × 5 回)	10
計	120 対話	60h (2 人対話)

収録話者は、20 代から 40 代の 10 人である。被験者の母語および性別の内訳を表 3 に示す。また、収録の組み合わせを図 1 に示す。

表 3: 被験者の母語と性別

被験者の母語	男性	女性
日本語	3	3
英語	1	1
中国語	1	1
計	5	5

3.5 データの配布方法

音声収録を完了し、現在データベース化を進めているところであるが、具体的なデータ配布方法については未定である。

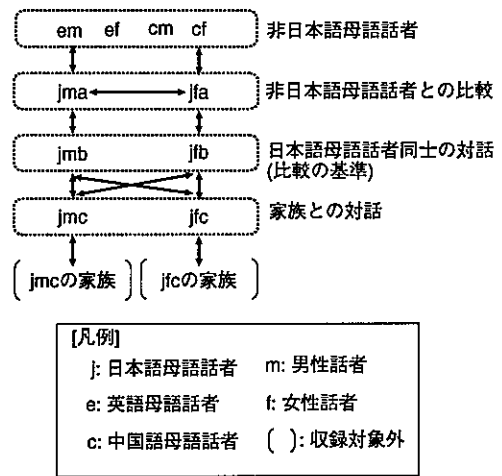


図 1: 収録の組み合わせ

4. おわりに

本稿では、現在本プロジェクトで構築を進めている対話音声コーパスのうち、特に、電話対話データベースの概要と収集方法を報告した。

今後は、完全な自由対話のみならず、例えば「タクシードライバと乗客」のように役割分担が明確な話者間の対話についてもデータ収集・分析を行ないたい。この際、自発性の高い発話とするために、タスクを台本などにより直接指定することはせず、話者間の人間関係を状況に応じて設定することにより、間接的に話題を規定するようにしたい。

なお、本プロジェクトの目指す音声対話データは、1000 時間程度と、今までにない大規模なものであるため、効率的で再利用性の高いデータ管理手法についても検討を進めていきたい。

参考文献

- [河口他 00] 河口信夫, 松原茂樹, 岩博之, 梶田将司, 武田一哉, 板倉文忠: 実走行車内における音声データベースの構築, 情処研報, 2000-SLP-30-12, pp.57-62(2000).
- [板橋 97] 板橋秀一: 音声コーパス, 情報処理, Vol.38, No.11, pp.1012-1018(1997).
- [広瀬他 02] 広瀬啓吉他: 科学研究費補助金特定領域研究(B), 「韻律に着目した音声言語情報処理の高度化」(2002).
- [前川他 97] 前川喜久雄, 籠宮隆之, 小磯花絵, 小椋秀樹, 菊池英明: 日本語話し言葉コーパスの設計, 音声研究, Vol.4, No.2, pp.51-61(2000).
- [芦村他 02] 芦村和幸, ニック・キャンベル, 武田一哉: 日常会話における発話タイミング分析, 情処研報, 2002-SLP-40-19, pp.109-114(2002).
- [ザトラウスキー 91] ポリー・ザトラウスキー: 会話分析における「単位」について, 日本語学, 明示書院第 10 巻, 第 10 号, pp. 79-96(1991).