論文 / 著書情報
Article / Book Information

論題(和文)	科学技術振興調整費開放的融合研究推進精度 －大規模コーパスに基づく「話し言葉工学」の構築－
Title	Science and Technology Agency Priority Program -Spontaneous Speech: Corpus and Processing Technology-
著者(和文)	古井貞熙, 前川喜久雄, 井佐原均
Author	SADAOKI FURUI
出典(和文)	日本音響学会誌, Vol. 56, No. 11, pp. 752-755
Journal/Book name	, Vol. 56, No. 11, pp. 752-755
発行日 / Issue date	2000, 11

日本音響学会誌56巻11号（2000），pp. 752-755

小特集—音声研究の新たな方向を探る—

科学技術振興調整費開放的融合研究推進制度

——大規模コーパスに基づく『話し言葉工学』の構築——*

古井貞熙（東京工業大学大学院情報理工学研究科／国立国語研究所）*1
前川喜久雄（国立国語研究所）*2・井佐原均（郵政省通信総合研究所）*3

43.72.Ne

1. はじめに

不特定話者大語彙連続音声認識でも，新聞記事などのあらかじめ用意されたテキストの読み上げ音声なら，かなり高い精度で認識できるようになってきたが，言い直し，言い淀み，繰り返し，間投詞，不正確な発音などの現象を含んだ自然な話し言葉を認識しようとすると，大幅に認識性能が低下してしまうのが現状である[1]。話し言葉でコンピュータシステムと対話をしたり，講演や放送などの話し言葉を文字化し，その内容を把握し，再利用や情報検索に使えるようにしたいという要望が高まっているが，その実現には多くの未解決の課題がある。

本プロジェクトは，このような話し言葉処理に関わる種々の重要な問題を取り上げ，話し言葉からその意味・内容・話し手の意図などを自動的に抽出する技術の基盤を確立することを目的とし，科学技術振興調整費を財源として，平成11年度に開始された[2]。平成15年度まで5年計画で進められる予定である。組織の主体は，国立国語研究所，郵政省通信総合研究所，及び東京工業大学であるが，国内外の大学や研究機関からの研究者の参加も求めている。

2. プロジェクトの概要

プロジェクト全体の概要を図-1に示す。本プロジェクトでは，話し言葉の大規模コーパスの構築と，それを用いた「話し言葉工学」の構築を目標としており，次の三つのサブテーマを持つ。

サブテーマ1：大規模話し言葉コーパスの構築
サブテーマ2：話し言葉を音声認識・理解・要約するための基本技術の構築
サブテーマ3：話し言葉の音声要約プロトタイプシステムの構築

以下にこれらの詳細について述べる。

3. 話し言葉コーパスの構築

3.1 話し言葉コーパスの重要性

現在の音声言語処理（音声認識及び音声合成）技術は，コーパスに基づく統計的手法に基礎をおいている。この手法には大変優れた面があり，特に音声認識に関しては，これによって近年の技術進歩が達成されたと言っても過言ではない。しかし，この手法が機能するためには，大規模なコーパスが必要である。書き言葉に関しては，新聞記事などの大量の原稿を比較的容易に用いることができるので，これまでの音声認識のためのコーパスも，ニュース原稿を含めて，書き言葉をベースとするものが主として用いられてきた。しかしこれでは真の話し言葉をモデル化することができず，話し言葉の音声認識で高い精度を達成することはできない。例えば，アナウンサが発声した放送ニュース音声が90%以上の単語正解精度で認識できる音声認識システムを，そのまま学会の講演や放送の対談番組に適用すると，30ないし40%程度の精度しか得られない。話し言葉のコーパスを構築するためには，音声を録音して書き起こし，それを解析して種々の情報を付与する必要があるため，大量なコーパスの構築は容易ではない。米国，ヨーロッパなどを見ても，まだ話し言葉コーパスは十分とは言えないが，日本では特

* Science and Technology Agency Priority Program—Spontaneous speech: Corpus and processing technology—.

*1 Sadaoki Furui (Tokyo Institute of Technology, Tokyo, 152-8552/National Language Research Institute, Tokyo, 115-8620)
*2 Kikuo Maekawa (National Language Research Institute, Tokyo, 115-8620)
*3 Hitoshi Isahara (Communications Research Laboratory, Kyoto, 619-0289)

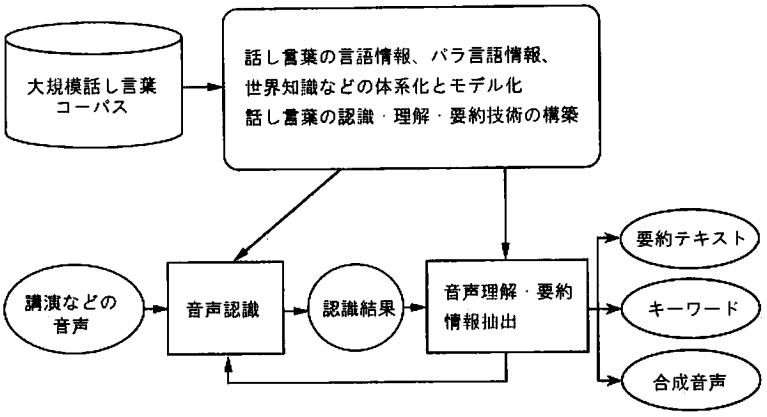

図-1 プロジェクトの概要

に，タスクを限定した対話音声を除くと，まだ比較的小規模なコーパスしかできていない。

3.2 構築するコーパスの概要

本プロジェクトでは，タスクを限定しないモノローグを主体として，のべ約800時間，形態素にして約700万語（形態素）規模の話し言葉コーパスの作成を目標としている[3),4)]。モノローグを基本とするのは，対話音声の処理に関しては別のプロジェクトが活発に活動しており，対話の流れに関する対話音声特有の重要な研究課題があるので，活動をある程度分けて分担した方がよいと考えるからである。対話音声でも，特にコンピュータシステムとの対話などでは，各文を正しく認識するのが基本と考えられるので，このプロジェクトでのモノローグを対象とする研究の成果は，対話音声処理技術の進展にも寄与できると考えている。本プロジェクトで構築するコーパスの中にも，対談などは一部に含む予定である。全体のコーパスの大きさは，音声認識のための統計的言語モデルを構築するために最低限必要なデータ量に基づいて決めた。

コーパスの発声者については，あまり厳格な制限は設けず，文法と分節的特徴が共通日本語であれば，地域差や個人差に起因するアクセントやイントネーションの変動は容認する。音声のソースとしては，国内の種々の学会での講演，大学での講義，国語研究所での設備を用いた模擬講演（一般の話者が簡単なメモを頼りに日常的な話題についてスピーチを行ったもの）などから始め，徐々に対象を広げていく予定である。音声はヘッドセット型の接話型指向性マイクロホンを話者の頭部に装着して，48 kHz，16 bit でDATに収録する。

コーパス全体について，書き起こし，セグメンテーション，読みの付与，形態素解析などを行うが，全体の約10%をコアとし，韻律情報などパラ言語情報まで含めたコーパスとする予定である。話し言葉に関しては，その書き起こし（表記）法や，形態素解析の単位などがまだ確立していない。単に文の切れ目を検出するだけでも，決して容易ではない。間投詞，言い淀み，雑音などをどう表記すべきかという問題もある。話し言葉音声自身の曖昧性のために，真剣に考えるほど解決が難しくなるような問題も多い。これらの基準作りも，このプロジェクトの重要な役割の一つと考えている。作成したコーパスは，できるだけ早く公開する予定である。

4. 話し言葉の音響モデルと言語モデルの構築

本プロジェクトにおける「話し言葉工学」の主たる目標は，話し言葉の音声認識理解と，その自動要約である。話し言葉の音声理解のためには，話し言葉に合った音響モデルと言語モデルを構築することが必要である。このため，コーパスを構成する大量の書き起こしテキストについて，自動的に形態素解析する。このためには，話し言葉に適応した形態素解析ツールを作成することが必要である。これもプロジェクトの重要な目標の一つである。このために，コアの部分については，人

手による処理を含めた精密な解析を行い，形態素解析ツールの学習に用いる。

コーパス全体から評価用コーパスを除いた部分の解析結果に基づいて，話し言葉の統計的言語モデルを作成する。上でも述べたように，話し言葉には，言い直し，言い淀み，繰り返し，間投詞，話し言葉特有の言い回しなど，書き言葉にはない多様な変動要因があるので，単に形態素を単位とする統計的言語モデル（バイグラム，トライグラムなど）を作成すればよいというわけにはいかない。話し言葉特有の変動に言語モデルとしてどう対処するか，どのようにデコーディングすればよいかについては，まだ有望な方法が見つかっていない。この解決のためには，これまでの，すべての単語（形態素）を文字化しようとする音声認識のアプローチから，不要な単語は無視し，話し手が伝えようとした内容を抽出することを目標とする音声理解への展開が不可欠である。そのため，コーパスから種々の世界知識，言語的/パラ言語的情報を抽出・体系化し，これらを有効に組み合わせて用いる方法についても検討する。

いくら大量のコーパスを構築しても，将来の音声認識理解の対象となるすべてのタスク（応用場面）をカバーすることはとうていできない。したがって，話し言葉に含まれるタスクに独立な現象と，タスクに依存した現象を分離したモデルを構築し，タスクに依存する部分について，新しいタスクに自動的に適応する方法を作り上げることが不可欠である。

話し言葉の音響モデルの学習も重要な課題である。現在の音声認識システムで用いられている音響モデルは，主として書き言葉を読み上げた音声から学習された環境依存型音素 HMM（hidden Markov model；隠れマルコフモデル）である。これまでの種々の実験で，読み上げ音声と通常の話し言葉音声では，音響的特徴が大きく異なることが分かっている。話し言葉を音声認識するための最適な音響モデルの構造と，その学習法，更にタスクや話者に自動的に適応化させる方法についても，研究を進めていく予定である。

5. 話し言葉の音声理解と音声要約

音声を理解するとは一般的に何を指すかは，まだ必ずしも明確ではない。本プロジェクトでは，音声理解の一つの形態として，音声の自動要約技術の実現を目標とする。最終的な要約結果の出力形態には種々のパターンが考えられるが，正しく要約することができたと判断できれば，ある意味で音声が理解できたと考えることができる。音声認識結果を用いて，音声を自動要約し，要約された内容をキーワード，要約文，その合成音声などで提示する方法について研究し，プロトタイプシステムを構築する。音声要約の結果を音声認識部にフィードバックすることによって，音声認識理解の精度を向上させることもできると期待される。

これまでに，音声から自動的に抽出したキーワード（話題語，重要語）を核として，言語モデルに基づいて，各文を短縮した要約文を自動的に作成する研究を進めている[5]。一方，各文ごとの要約ではなく，複数の文からなる記事を単位とする要約が，自然言語処理の分野で長く研究されており，その基本的方法は，最も重要な文を抽出することである。しかしこれだけでは，抽出されなかった文に含まれる情報が完全に欠落して，重要文だけをつなぎ合わせたものでは意味がとれないことが少なくない，という結果も報告されている。音声認識結果にはどうしてもある程度の認識誤りや不確実性が避けられないので，書き言葉の要約とは本質的に異なる。世界知識や文の意味まで考慮して，全く新しい方法を構築する必要があろう。

本プロジェクトが目標とする技術の応用としては，次のようなものが考えられる。

- ●音声を含む大量のマルチメディア情報の検索のための，インデックスの自動付与
- ●放送などへの字幕付与の自動化など，福祉技術の向上への寄与
- ●会議・講演・講義などの速記や文字起こしの自動化の促進
- ●音声ワープロの高性能化への寄与

6. むすび

本特集で紹介されているように，現在，複数の音声認識関連大型プロジェクトが展開しており，それらは相互に関連している。本プロジェクトを進めていくにあたっても，これら関連プロジェクトと適宜情報交換を行うなど，連絡を密にし，共

有できるデータベース（コーパス）は共有していきたい。音声や言語のデータベースは，幾らあっても十分ということはなく，言葉は常に変化し新しい言葉が作られるので，データベースの構築と維持管理をプロジェクトの期間だけで終わらせず，継続できる態勢を作ることが必須である。これについては，田中穂積先生のGSK構想[6]を軸として，社会的コンセンサスを得るための働き掛けをしていく必要があろう。

このプロジェクトの目標達成のためには，情報科学・工学系だけでなく，音声学，言語学などに代表される人文系言語諸分野を含めた多面的かつ総合的なアプローチが必要である。幸いプロジェクトが多様な研究者から構成されているので，この面でも新しい環境を作っていきたい。

平成11年の秋に，本プロジェクトに関する意見交換などのために，米国の主要関係機関（NIST，LDC，MIT，AT＆T研究所，Bell研究所など）を訪問したが，幸いどこでも大きな関心を得ることができた。

本プロジェクトでは，この分野の国内外の10名のリーダの方々に評価委員（評価委員長は京都大学の長尾真総長）をお願いしており，平成12年2月28日，29日に各評価委員の講演とプロジェクトの紹介を中心とした国際シンポジウムを開催すると共に，3月1日に評価委員会を行った。これらを通じて，本プロジェクトの進め方に関する貴重なサジェスチョンや提案を多数得ることができ，今後のプロジェクトの活動に生かしていく予定である。本プロジェクトが，真の意味での話し言葉の認識・理解技術の構築に，役に立てれば本望である。

謝　辞

客員研究員として本プロジェクトに参加し，貴重な意見やコメントを提供して下さる大学や研究機関の方々に感謝する。末筆ではあるが，本プロジェクトの立ち上げのためにご尽力いただいた多数の方々のご努力に心から感謝する。

文　献

1) B.-H. Juang, "Progress & challenges in automatic recognition and understanding," Proc. Int. Symp., Toward the Realization of Spontaneous Speech Engineering, Tokyo, 9-13 (2000).
2) S. Furui, "Steps towards natural human-machine communication in the 21st century," Proc. Int. Symp., Toward the Realization of Spontaneous Speech Engineering, Tokyo, 1-8 (2000).
3) K. Maekawa and H. Koiso, "Design of spontaneous speech corpus for Japanese," Proc. Int. Symp., Toward the Realization of Spontaneous Speech Engineering, Tokyo, 70-77 (2000).
4) 前川喜久雄，小磯花絵，籠宮隆之，古井貞熙，井佐原均，"共通日本語話し言葉コーパスの設計，" 音講論集 1-Q-31 (2000. 3).
5) 堀　智織，古井貞熙，"話題語と言語モデルを用いた音声自動要約法の検討，" 信学技報 SP 99-110 (1999).
6) "言語資源共有機構（GSK）活動報告，" 自然言語処理システムに関する調査報告書，5章，249-292 (2000)；http://tanaka-www.cs.titech.ac.jp/gsk/.

古井　貞熙

昭43東大・工・計数卒。昭45同大学院修士課程了。同年NTT研究所入社。以後，同研究所において，音声認識，話者認識，音声知覚などの研究に従事。昭和53～54米国ベル研究所客員研究員。昭和61 NTT基礎研究所第四研究室長。平1音声情報研究部長。平3古井特別研究室長。平6東京工業大学客員教授。平9同大学院情報理工学研究科計算工学専攻教授。工博。科学技術庁長官賞受賞。IEEE，電子情報通信学会，日本音響学会から論文賞など受賞。著書「音声情報処理」など。IEEE及び米国音響学会Fellow。日本音響学会技術委員長。

前川喜久雄

昭和59年上智大学大学院（言語学専攻）博士後期課程中退。鳥取大学講師を経て国立国語研究所言語行動研究部勤務。現在，同第2研究室長。専門は音声学，言語学。最近は音声によるパラ言語情報の伝達機序の解明と音声データベースへの韻律ラベリングを中心的な研究テーマとしている。

井佐原　均

昭和53年京都大学工学部電気工学第二学科卒業。昭和55年同大学院修士課程修了。博士（工学）。同年通商産業省電子技術総合研究所入所。平成7年郵政省通信総合研究所関西支所知的機能研究室室長。自然言語処理，機械翻訳の研究に従事。言語処理学会，情報処理学会，人工知能学会，日本認知科学会，ACL各会員。