東北方言昔話に関する歴史的音声コーパスと機械学習ベース自動音声復元の試み

高道慎之介・丹治尚子・佐伯高明・森松亜依（東京大学 大学院情報理工学系研究科）
庄司潤子・佐藤照一（仙台文学館）
猿渡洋（東京大学 大学院情報理工学系研究科）

概要：本研究では，歴史的音声の復元と保存を目的として，東北方言昔話を例としたディジタル音声コーパスの作成と，機械学習に基づく自動音声復元技術の開発について述べる．アナログ記録媒体に保存された歴史的音声の散逸と劣化を防ぐために，音声のディジタル化と復元が急務の課題である．本論文では，1950年代から1980年代にかけて宮城県と岩手県の一部を中心に収集された昔話をディジタル化し，そのアノテーションと復元の結果を報告する．アノテーション結果を含む音声コーパスと音声復元技術は，オープンソースとして公開している．
キーワード：東北方言，歴史的音源，コーパス，機械学習，音声復元

Historical speech corpus of Tohoku dialect folktales and machine learning-based speech restoration

Shinnosuke Takamichi / Naoko Tanji / Takaaki Saeki / Ai Morimatsu (Graduate School of Information Science and Technology, The University of Tokyo)
Junko Shoji / Shoichi Satoh (Sendai Literature Museum)
Hiroshi Saruwatari (Graduate School of Information Science and Technology, The University of Tokyo)

Abstract: This paper presents developments of 1) the digital speech corpus of Tohoku dialect folktales and 2) machine learning-based automatic speech restoration. In order to prevent the dissipation and deterioration of historical audio stored on analog recording media, digitization and restoration of the audio is an urgent issue. This paper reports the results of digitization, annotation, and restoration of the folktales collected mainly in Miyagi and Iwate prefectures from the 1950s to the 1980s. The speech corpus is open-sourced, including the annotation results and the speech restoration technique.
Keywords: Tohoku dialect, historical audio, corpus, machine learning, speech restoration

1．まえがき

1900年前後から録音技術が普及し，1980年代頃までアナログ記録媒体（アナログテープ，オープンリールテープなど）が使用されてきた．この記録媒体は，過去の音文化を現代に残す重要な資料であり，特に音声言語に関する資料（歴史的音声）は，当時の音声言語の文化と地域性を有する．日本語の資料としては，1900年のパリ万博での録音が初めとされ[1]，約120年ほどの歴史的音声が蓄積されてきた[2]．この資料は，以下の観点から重要な役割を持つ．

- 教育学の観点：消滅の危機に有る地域方言を守るため．佐藤らは，方言研究の目的を，保存継承，共通語教育との対比，言語教育の基礎づくりに分類している[3]．保存継承は，消滅の危機に瀕する地域方言の保存活動を学校教育に取り入れたものを指す．共通語教育との対比は，共通語と方言を比較・対照させながら違いを理解し，それぞれの特質とよさを知ること，基礎づくりは，様々な言語を様々な言語体系として捉える感覚とセンスを磨くことを指す．これらの実施には当該方言の音声資料がしばしば用いられる．
- 人文学の観点：自然言語に現れる，人間の営みの通時変化を明らかにするため．丸山らは，歴史的な書き言葉コーパス「日本語歴史コーパス[4]」，現代の話し言葉コーパス「日本語話し言葉コーパス[5]」に対応させて，昭和話し言葉コーパスを位置づけている[6]．
- 工学の観点：現代音声に限らずあらゆる音声を情報解析できる音声工学技術を開発するため．主要言語における音声認識合成技術の研究が一段落した現在，より多様な環境[7]や地域[8]を対象とする研究が進んでいる．

これらの観点から，歴史的音声に関する資料が散逸・劣化する前にディジタル化およびアノテーションを行い，容易にアクセス可能なディジタルコーパスとして公開すべきである．

本研究で扱う東北方言昔話は，仙台文学館（宮城県仙台市）に保存されている音声資料である．当該資料は，昔話採集家の佐々木徳夫により1950年代から収録され，話者や物語の情報を追跡可能な形で残す，非常に貴重な資料である．本研究では，ディジタル化とアノテーションにより，これをコーパス化する．本研究では更に，機械学習に基づく自動音声復元を試みる．アナログ記録媒体および当時の録音環境は現

在のそれらより良質とは言い難い. また, 記録媒体自体の経年劣化により, 資料の品質が劣化しているものも多い. そこで, 機械学習を用いてこの音声を復元し, 現代音質への回復を試みる. 以上を総括し本研究の貢献を次の2点に定める.

- 東北方言昔話に関するディジタルコーパスを構築し, 一般に公開する.
- 機械学習に基づく音声復元技術を当該コーパスに適用し, 音声復元を試みる.

2. 佐々木徳夫の遺した東北方言昔話

対象資料を収集した佐々木徳夫は, 在野の昔話採集家として知られる. 高校教師としての勤務の傍ら, 1957年ごろから昔話(艶笑譚を含む)の聞き取りを始め, 本論文のアノテーションでも参照する「日本の昔話11永浦誠喜翁の昔話」[9]や「遠野の昔話」[10]などを書籍にまとめている. 本人の語るところによれば, 農作業をしている人, 道端で出会った人, バスや列車の中で乗り合わせた人などに話しかけて, 語り手を探したとされている[11].

宮城県と岩手県の一部を中心に収集した物語の数は1万本を超えるとされ, その音声資料が, 3.2節で後述する情報と共に仙台文学館に継承された. 現在判明しているその内訳は, カセットテープ86本, オープンリールテープ88本である.

3. 東北方言昔話に関する歴史的音源コーパスの作成

2節で述べた音声資料のオープンコーパス化の手順を述べる.

3.1 ディジタル化対象の選定

音声資料のうち, 収録時期の古いオープンリールテープを優先して, その音声資料をディジタル化する. これは, 本資料の収録が50年前以上から始まっており, オープンリールテープの経年劣化が進んでいたためである. なお, いくつかのオープンリープテープにディジタル化作業困難なほどのカビの発生が認められた(図1)ため, 事前のクリーニングを施した.

図1:カビ発生の認められるオープンリールテープ
Figure 1: Moldy open-reel tapes.

3.2 アノテーション作業

ディジタル化した資料に対し, 以下の情報をアノテーションする.

- 話者に関する情報
  - 話者の名前
  - 話者の出身地
- 収録に関する情報
  - 収録日
  - 各昔話の語りの開始終了時刻
- 昔話に関する情報
  - 昔話の題目(原題, 改題)
  - 昔話の掲載書籍名
  - 書き起こし

なお, コーパス公開を見据え, 著作権者および話者・ご遺族による許諾見込みのあるものを, 優先的にアノテーションする. また, 公開可能性の不明の昔話(艶笑譚など)についてはアノテーションしない.

3.2.1 話者・収録に関する情報

話者に関する情報については, オープンリールテープに記載されているものを元に作成し, ディジタル化資料を聴取して修正する(各物語の冒頭で, 話者に関する情報が読み上げられている). 収録に関する情報も同様に作成・修正する.

3.2.2 昔話に関する情報:題目・書籍

昔話に関する情報のうち, 昔話の題目にはいくつかの種類が存在する. 昔話にはそもそも題目が存在せず, それを聴取した聞き手が自ら題目を付することが多い. 本論文の音声資料の場合には,

1. 収録直前(物語冒頭)に, 収録者(佐々木徳夫)もしくは話者が読み上げている題目
2. 収録後に, 収録者がオープンリールテープに付した題目
3. オープンリールテープ作成後に, その内容の書籍化にあたり収録者が書籍に記載した題目

の3段階が存在する. 本研究では, 段階3において付された分類的な題目を「昔話(改題)」, 段階2で付された題目を「昔話(原題)」としてアノテーションする. なお, 当該題目に対応するファイルの名称には昔話(原題)を使用し, 昔話(原題)の明らかでない昔話については, 昔話(改題)を使用する.

3.2.3 昔話に関する情報:書き起こし

方言音声コーパスCOJADS[12]では, アノテーション指針として

1. 方言テキスト:方言の書き起こし, 表記, タグ
2. 標準語テキスト:方言テキストの現代訳
3. 文節認定規則:1, 2 の分かち書き規則

の3項目を定めている. このうち, 項目1は音声の聴取に基づくため, 多大な労力を要する. この簡略化には音声認識エンジンが有用であるため, その学習データとなる書き起こしテキストの作成を, 他の項目に優先して実施する. 近年の音声認識は音素表記と分かち書きを介さず漢字仮名混じり文を直接的に使用するため, それに即する書き起こし文を作成する. また, 同様の理由から, 方言音に即する表記(例えば, ガ行鼻音をカタカナに゜で表記)ではなく日本語共通語で書き起こしを実施する. 表1は, 書き起こしの例である.

表1：方言テキストの書き起こし例
Table 1: Example of dialect transcription.

ルール	例
漢字と読みは”｜漢字[よみ]”で記載する.	｜座頭[ざど]の｜坊[ぼう]っつのは｜眼[まなぐ]の
聞き取れない箇所は”＊”で記載する.	で、＊＊＊つからだいたい、＊＊＊つから
複数の読み候補が存在する箇所は，”(よみ1｜よみ2)”で記載する.	｜人里[(ひ｜し)とざと]

方言音に即した表記と，項目2，項目3，また，音声認識モデルの学習については，今後の予定とする.

4. 機械学習ベース自動音声復元の試み

アナログ記録媒体に保存される歴史的音声は，一般に，現代音声と比較して以下の要素が異なる.

- 音響歪み(チャネル歪み)：記録媒体自体とその経年劣化により，音声に対し乗算的に付与される歪み
- 音声の欠落：記録媒体自体の物理的な欠損により生じる音声欠損
- 背景音の混入：収録時に目的音声以外に加算的に混入される雑音

このうち本研究では，1の要素(すなわち，オープンリールテープそのものとその経年劣化)による音質劣化に取り組む．3の要素については，当該資料について対処不要と判断した．これは，当該資料の大半は十分に静音な環境で録音されているためである．2の要素は，欠損箇所検出のコストが大きいため対象外とした.

歴史的音声の音響歪みは，一般に非可逆の音響的変化であり，その歪みの要因と様子を規則的に定めることは困難である．そのため，機械学習を用いてその歪みを除去する．しかしながら，通常の教師あり学習でこれを実現するのは困難である．教師あり学習の実施には，音響歪みの有無のみが異なる対データが必要であり，歴史的音声についてこれを用意することは不可能である.

そこで本研究では，我々が先だって提案した自己教師あり音声復元アルゴリズム[13]を適用する．図2はその概要である．この手法の音声復元モデル(音声復元のためのニューラルネットワーク)は，前述した対データを必要とせず歴史的音声のみから学習される.

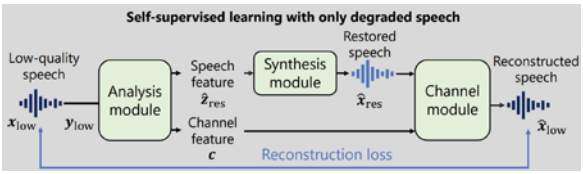

図2：自己教師あり音声復元．[13]より引用.
Figure 2: Self-supervised speech remastering [13].

この手法の機械学習モデルは，

- 分析 **(analysis)** モジュール：歴史的音声から，高品質音声の特徴量とチャネル歪の特徴量を出力する．後段モジュールを駆動する特徴量の抽出を担う.
- 生成 **(synthesis)** モジュール：高品質音声の特徴量から高品質音声の波形を出力する．歴史的音声のうち話者による音声生成に対応する.
- チャネル **(channel)** モジュール：高品質音声波形に対してチャネル歪を付与する．歴史的音声のうち，オープンリールテープとその経年劣化による歪に対応する.

の3モジュールから成る．各モジュールはニューラルネットワークで記述され，そのパラメータは歴史的音声から学習される．具体的には，分析モジュールとチャネルモジュールはU-Net構造[14]を，生成モジュールはHiFi-GAN構造[15]を持つ．この生成モジュールを現代音声(現代環境で収録された高品質音声)を用いて事前学習することで，高品質音声の生成を担う．また，歴史的音声と現代音声を併用した双方向学習により，分析モジュールが，高品質音声の特徴量とチャネル歪の特徴量を高精度に抽出できる.

5. 実施結果

3節と4節の内容について実施した結果を述べる.

5.1 コーパス作成の結果

仙台文学館に保存されていたカセットテープ86本とオープンリールテープ88本のうち，ディジタル化した内容を表1に示す．表の通り，現在89時間のディジタル化を完了した.

表2：ディジタル化した資料
Table 2: Contents of digitized speech materials.

項目	数値・内容
収録期間	1967年～1983年
ディジタル化したオープンリールテープ	44本
話者数	80名
時間数	89時間

この内容のうち，男性話者1名と女性話者1名について，アノテーションを実施した．作業結果を表3に示す．本コーパスは，https://sites.google.com/site/shinnosuketakamichi/research-topics/tohoku-dialect_corpus にて公開している．また，アノテーションした昔話の題目と内容の例を付録Aに，書き起こしたテキストの例を付録Bに記載する.

表3：アノテーションした資料の内容
Table 3: Contents of annotated materials.

音声データ	2名分 合計約14時間
	(1)明治24(1891)年生・女性 宮城県仙台市

	約54話，約6時間27分
	(2)明治42(1909)年生・男性 宮城県登米市 約56話，約6時間49分
テキストデータ	題目，各話書き起こし
メタデータ	収録日，各話の開始終了時刻，掲載書籍名

5.2 音声復元の結果

学習データ，評価データは，表3音声データ(1)から選択した，それぞれ1,500発話，100発話とした．1発話あたりの平均時間長は7秒である．サンプリング周波数は22.05kHzとした．それ以外のハイパーパラメータについては，https://github.com/Takaaki-Saeki/ssl_speech_restoration を参照とする．

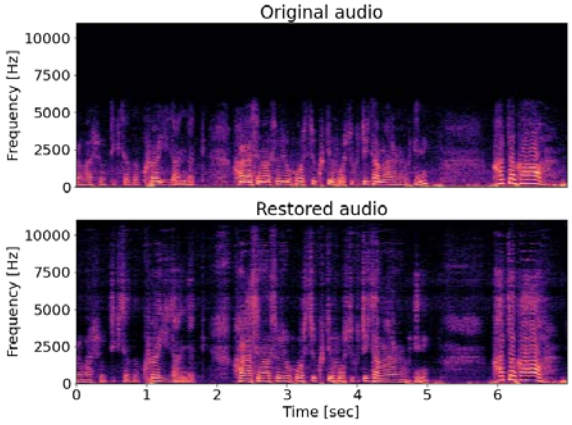

図3：音声復元前(上)と後(下)のスペクトログラム
Figure 3: Spectrograms before (above) and after (below) speech restoration.

図3に，音声復元前後のスペクトログラムを示す．高周波数帯域が復元されているとともに，音響的な歪が低減されていることが分かる．

クラウドソーシングサービス Lancers を利用して主観評価を実施した．評価対象の音声は，ディジタル化のみを施した元音声と，音声復元を施した音声の2種類である．各評価者は，ランダムに抽出された1発話を受聴し，その音質を1（非常に悪い）～5（非常に良い）の5段階で評価した．評価者あたりの回答数は20回答であり，評価者数は70名である．

評価の結果，元音声の平均評価値（mean opinion score: MOS）は2.779，復元音声の値は2.934となった．両者の間にはp-value < 0.001 で有意差が認められた．以上より，音声復元により歴史的音声の音質を有意に改善できることが明らかになった．一方，その値は未だ低いため，今後の改善が必要である．

6. まとめ

本論文では，東北方言昔話を例としたディジタル音声コーパスの作成と，機械学習に基づく自動音声復元技術の開発について述べた．以降に今後の予定を示す．

- コーパス規模の拡大：本論文でディジタル化したのは，現存する174本のうち44本のみである．今後は，残る音声資源のディジタル化とアノテーションを進める．
- 音声復元技術の改善：本論文中で音声復元の効果は認められたものの，その程度は限定的である．この改善にはより大規模コーパスの利用が期待されるため，前述した音源ディジタル化と共に改善を測る．
- 音声認識技術の開発：アノテーションの効率化を効率化するため，方言音声認識技術を開発する．現在，https://github.com/sarulab-speech/whisper-asr-finetuneにて音声認識モデル Whisper [16] のファインチューニングコードの整備を進めている．COJADSコーパスにて予備実験を実施したところ，約20分の単一方言データにより学習下モデルが，カタカナ表記でCER（character error rate）50%を達成した．今後は，本論文のコーパスに適用する．
- 音声合成技術の開発：方言音声の音韻と韻律を計算機が学習できるかを検証するために，歴史的音声を合成する音声合成技術を開発する．これまで開発した方言適応モデル[8]，方言アクセント推定[17]を利用して開発する．

謝辞

本研究は，国立国語研究所 異分野融合型共同利用型「歴史的音源アーカイブに向けたオープンコーパスの整備とAI音声復元技術の開発」の下で実施された．本研究の実施にあたり有意義なコメントをくださった田村文子さま，川口里比さま，ならびに国立国語研究所 五十嵐陽介先生に感謝する．

参考文献

[1] 清水 康行，“百年前の日本語を聴く”，日本女
[2] 丸山 岳彦，小磯花絵，西川賢哉，“「『昭和話し言葉コーパス』の設計と構築”，国立国語研究所論集，Vol. 22，pp. 197–221，2014.
[3] 佐藤 髙司，“言語教育の基礎としての方言教育，”共愛学園前橋国際大学論集，2015.
[4] 国立国語研究所，“日本語歴史コーパス”，2022. https://clrd.ninjal.ac.jp/chj/
[5] K. Maekawa, “Corpus of Spontaneous Japanese: Its design and evaluation.” Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition, Tokyo, 2003.
[6] 丸山 岳彦，“『昭和話し言葉コーパス』の設計・構築と分析”，言語処理学会 第26回年次大会 発表論文集，2020.
[7] J.H. Hansen et al., “Fearless Steps: Apollo-11 Corpus Advancements for Speech Technologies from Earth to the Moon,” Proc. Interspeech, 2018.
[8] T. Akiyama, S. Takamichi, H. Saruwatari, “Prosody-aware subword embedding considering Japanese intonation systems and its application to DNN-based multi-dialect speech synthesis,”

Proc. APSIPA, 2018.
[9] 佐々木 徳夫, “日本の昔話 11 永浦誠喜翁の昔話”, 日本放送出版協会, 1975.
[10] 佐々木 徳夫, “遠野の昔話”, 桜楓社,1985.
[11] 佐藤 一子, “昔話の口承と地域学習の展開 : 岩手県遠野市の「民話のふるさと」づくりと語り部たちの活動”, 法政大学キャリアデザイン学部紀要, 2013.
[12] N. Kibe, T. Otsuki, K. Sato, “Intonational Variations at the End of Interrogative Sentences in Japanese Dialects: From the “Corpus of Japanese Dialects””, Proc. LREC, 2018.
[13] T. Saeki, S. Takamichi, T. Nakamura, N. Tanji, H. Saruwatari, “SelfRemaster: Self-Supervised Speech Restoration with Analysis-by-Synthesis Approach Using Channel Modeling,” Proc. Interspeech, 2022.
[14] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. MICCAI, 2015.
[15] J. Kong, J. Kim, and J. Bae, “HiFi-GAN: Generative adversarial networks for efficient and high fidelity speech synthesis,” arXiv preprint arXiv:2010.05646, 2020.
[16] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever, “Robust Speech Recognition via Large-Scale Weak Supervision,” OpenAI paper, 2022.
[17] K. Yufune, T. Koriyama, S. Takamichi, Hiroshi Saruwatari, "Accent Modeling of Low-Resourced Dialect in Pitch Accent Language Using Variational Autoencoder" Proc. The 11th ISCA SSW, 2021.

付録A. アノテーション済み昔話の例

表4:アノテーション済みの昔話(女性話者)
Table 4: Annotated folktales (female speaker)

昔話(原題)	内容
吉野の石落し	盗みを生業にする宿に泊まった旅人が子守歌を聞いて難を逃れる話
オンチョロチョロ参られ候	旅の坊さんに習ったお経を婆が唱えて泥棒が慌てて逃げていく話
一本橋	継母の仕打ちで死んだ三人姉弟が小鳥になり父親に知らせる話
タコの昼寝	野良猫に七本の足を食われたタコが敵をとろうとする話
長い長い名前	長生きを願って付けた長い名前が仇になってしまう子供の話

表5:アノテーション済みの昔話(男性話者)
Table 5: Annotated folktales (male speaker)

昔話(原題)	内容
座頭の坊が井戸に入った話	座頭を助けて運を授かる親切な家と, それを真似た欲深な家の話
牛になった和尚さん	旅の僧が牛になり福を授かった夫婦と, それを真似た欲深な男の話
折ればかり	商売にとって縁起の悪い出来事を妻の機転で福へと変える話
トラヤー, トラヤー	飼い猫が世話になった和尚に不思議なお経を教えて恩返しする話
所風(沢庵風呂)	洗足鉢の湯が熱いので沢庵でかき回して飲んだ愚かな婿の話

付録B. 書き起こしテキストの例

表6:書き起こしテキスト. 女性話者による「吉野の石落し」.
Table 6: Transcribed dialect text (“Yoshino no Ishiotoshi” by female speaker)

｜昔々[むかしむかし]ね、あのー、ある｜山[やま]のふもとにね、｜一軒[いっけん]の、｜宿屋[やンどや]があったんだど。
でそごーそのー｜山越｜[やまご]えする｜人[ひと]の｜恰好[かっこう]な｜旅路[たンびじ]の｜場所[ばしょ]なもんだから｜大変(たいへん)｜繁盛[はんじょう]**してねー、たくさんの｜泊[とま]りが、｜客[きゃぐ]が**あったそうです。
である｜時[とぎ]ぬその｜若[わが]い｜男[おどご]の｜人[ひと]が、｜泊[と]まったの。で、そこへ、あのーあー、そごの｜宿屋[やンどや]はね、あのー｜金[かね]のありそうな｜人[ひと]を、その｜[とぐべつ]の｜部屋[へや]に｜移[うづ]して。そして｜夜中[よなが]に｜殺[ころ]しておし｜金[かね]やら、｜持[も]ぢ｜物[もの]を｜盗[と]ってたんだそうです。

表7:書き起こしテキスト. 男性話者による「座頭の坊が井戸に入った話」.
Table 7: Transcribed dialect text (“Zatoh no bou ga ido ni haitta hanashi”)

｜昔[むかす]あっどごぬねェ、んー、とっても｜親切[しんしぇづ]だげっつも、｜貧乏[びんぼう]たが

りなお｜百姓[ひゃくしょう]さんあったっだど。
この｜[家]いえさ、あー、｜晩方[ばんがだ]この、｜一人[しとり]の、｜座頭[ざど]｜坊の[ぼ]ぁ｜来[き]て、｜座頭[ざど]の｜坊[ぼう]っつのは｜眼[まなぐ]の｜見[み]えねぇ｜人[しと]ね。「どうぞ｜泊[と]めてけらいん。」て｜語[かだ]っだんだと。
ンで、｜親切[しんしぇづ]なひとだから、「いがすいがす、｜泊[と]まらいん。」てまず｜泊[と]めで、そすていろいろ｜御馳走[ごっつぉう]をすて、えーー、｜休[やす]ませだんだどっしゃ。