

Establishing Spatiotemporal Linguistics: A New Approach to the
Interdisciplinary Study of Language History and Human Movement

Spoken Language Technologies for Historical Speech and Audio (歴史的音声言語を分析する音声処理技術)

Shinnosuke Takamichi (Keio University)
(高道 慎之介 (慶應義塾大学))



Takamichi Laboratory
慶應義塾大学 高道研究室

Self introduction (自己紹介)



@forthshinji

Name

Shinnosuke Takamichi (高道 慎之介)

Position

Associate Prof. of Keio University
(慶應義塾大学 准教授)

Research interests

Speech processing (音声処理)
Machine learning (機械学習)

My talk introduces our study for Tohoku (northeastern Japan) dialects.
(本講演では、東北地方の方言に関する我々の研究を紹介します。)

Collaborative Research Projects with NINJAL (国語研 共同研究プロジェクト)



大学共同利用機関法人 人間文化研究機構

国立国語研究所

National Institute for Japanese Language and Linguistics

English

| 交通案内

| サイトマップ



Google 提供



国語研案内

研究活動

大学院教育

データベース・
コーパス・資料

催し物

Development of Open Corpus and AI-based Speech Restoration System for Archiving Historical Audio

(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)

The aim of this research is to develop an open speech corpus and automatic speech restoration system for archiving historical audio.

Several historical spoken languages have deteriorated over time, and they contain many acoustic distortions.

In this project, a speech corpus for the analysis of the Japanese language and a deep-learning technique for speech restoration are developed using the dialect of the Tohoku region.

After the end of the project period, the corpus and restoration techniques will be made publicly available free of charge for research purposes only.

1. Development of Tohoku folktale corpus (東北地方民話コーパスの構築)

Sendai Bungakukan stores folktale audio data. (仙台文学館が民話音声を保存している)

あ あ Q

仙台文学館

ことばの
杜を
あるこう

2025.2.21(金)

お知らせ

『仙台文学館ニュース』45号
に関する訂正とお詫び
2024年11月15日(金)

写真でおしゃべり「ひらり、
ふわり～政宗さまの桜めぐ

むずかしいことをやさしくやさしいことをふか



その他

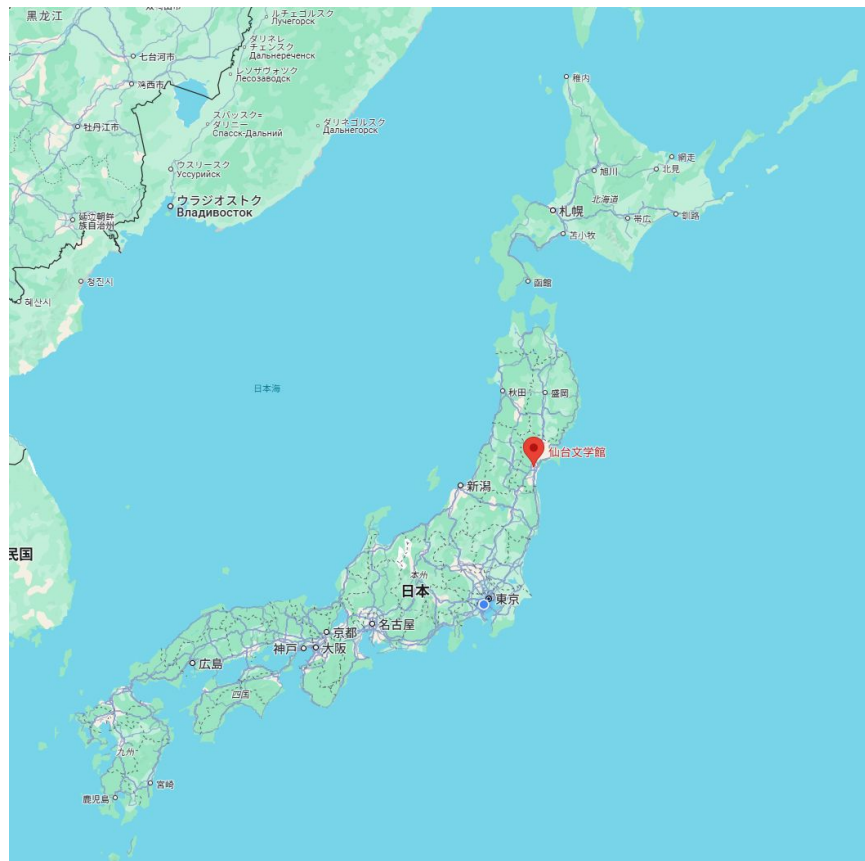
2024年度(～2025年3月)のスケジュール



開館25周年記念特別展
「詩人・石川善助をたずねて～
北方への道のり」
4月27日(土)
～6月30日(日)

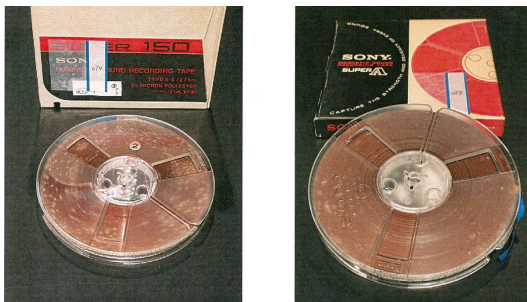


夏休み企画 こども文学館えほんのひろば
「せとうちたいこさんに あいたーい!」
長野ヒデ子 絵本と紙芝居
7月20日(土)
～9月8日(日)

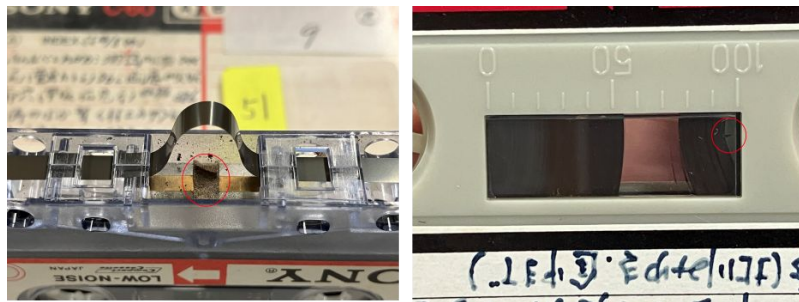


We digitized 400-hour audio recorded in 1960s-1970s. (1960~1970年代に録音された400時間の音をデジタル化)

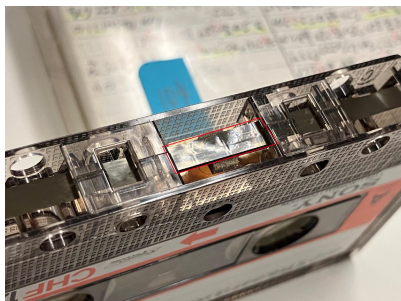
Open reel



Cassette tape
(some were damaged)



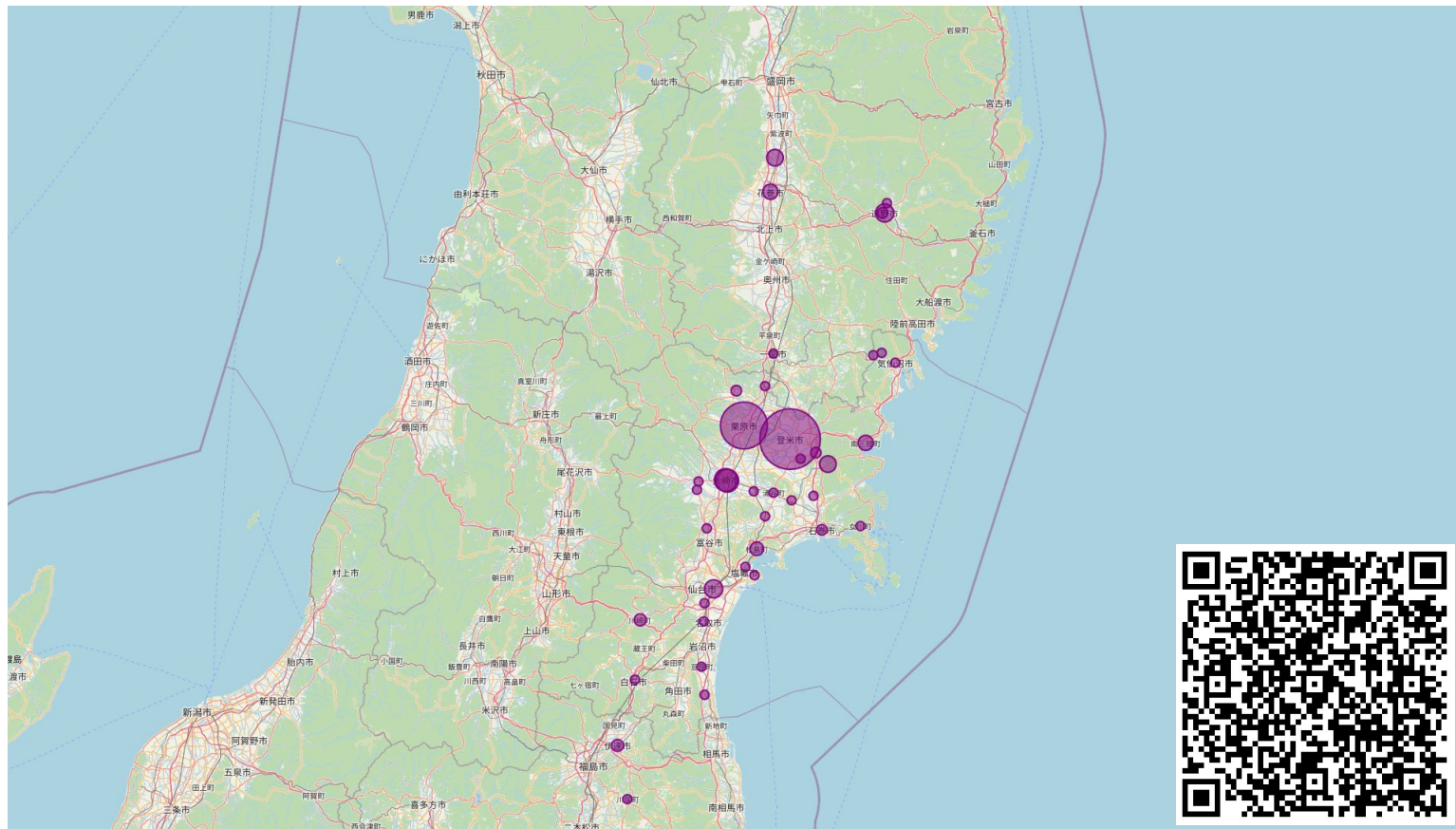
Cassette tape
(some were moldy [カビ])



Cassette tapes
(some required the cases to be destroyed)

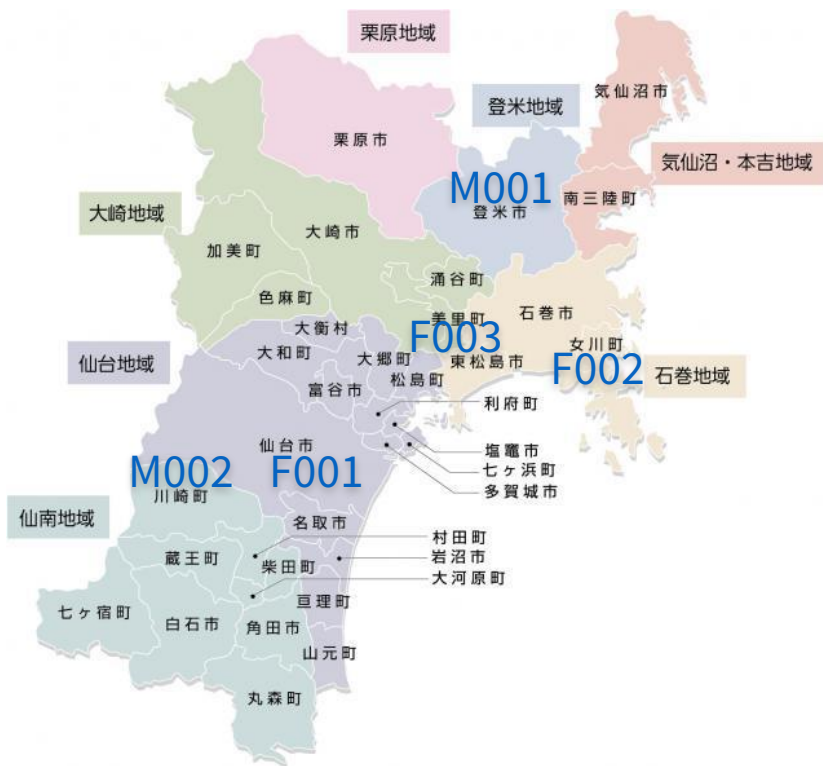


Map of the digitized audio (デジタル化した音源の地図)



https://ss-takashi.sakura.ne.jp/corpus/tohoku-folktale/number_of_tales.html

Audio data we transcribed and open sourced. (書き起こして公開した音声データ)



Spkr	M/F	Birth	Place (Miyagi pref.)	#tales	Duration
F001	F	M24 (1981)	Sendai (仙台市)	74	9h15m
F002	F	M40 (1907)	Onagawa (牡鹿郡女川町)	23	2h25m
F003	F	T13 (1914)	Misato (遠田郡美里町)	24	1h36m
M001	M	M42 (1909)	Tome (登米市)	74	9h18m
M002	M	M29 (1896)	Kawasaki (柴田郡川崎町)	35	2h22m

2. deep learning-based speech restoration (深層学習に基づく音声復元)

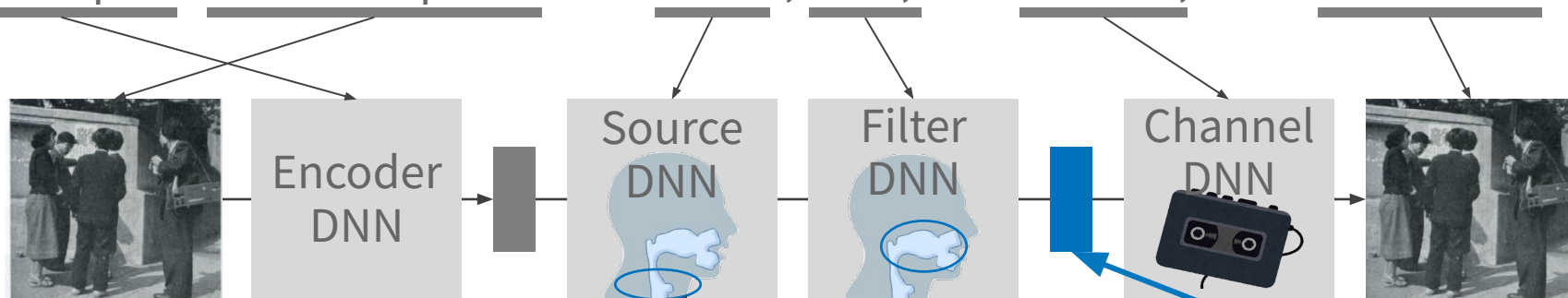
Let's restore low-quality to modern audio quality.
(劣化音声を現代音質に復元しよう)

Self-supervised learning based on source-filter-channel decomposition (ソース・フィルタ・チャネル分解に基づく自己教師あり学習)

[Training (学習)]

(劣化音声を要素分解して、ソース・フィルタ・チャネル過程を通して再構成)

Decompose distorted speech into source, filter, and channel, then reconstruct it.



Decompose speech into the source and filter, then construct **high-quality speech**.
劣化音声を要素分解して、ソース・フィルタ過程だけを通し**高品質音声**を復元

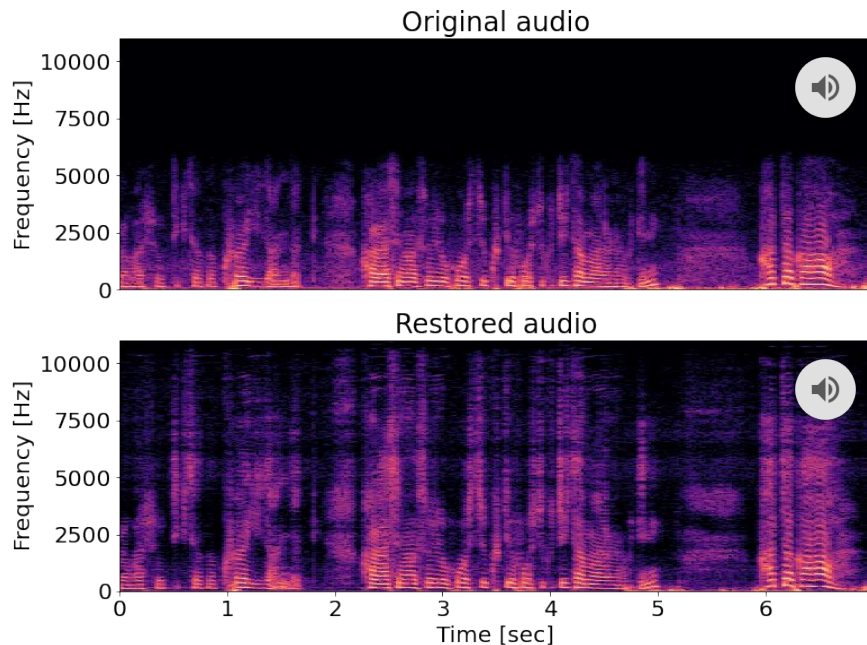
[Restoration (復元)]

Restoration example (復元例)

Training data (学習データ)

Tohoku-folktale [Takamichi22]	Tohoku folktale recorded in 1960-170s.(1960~1970年代に録音された東北地方民話)
CPJD Kyokotoba [Takamichi18]	Crowdsourced Kyoto dialect (クラウドソーシングで収集した京ことば)
tri-jek [Takamichi21]	Ja-Kr-En speech by one speaker (日韓英のトリリンガル話者の声)
Tono [Sasaki85]	Tono folktale recorded in 1960-1970s (遠野の昔話. 1960-1970年代に収録)
Total [hours]	8

Restored data (復元データ)

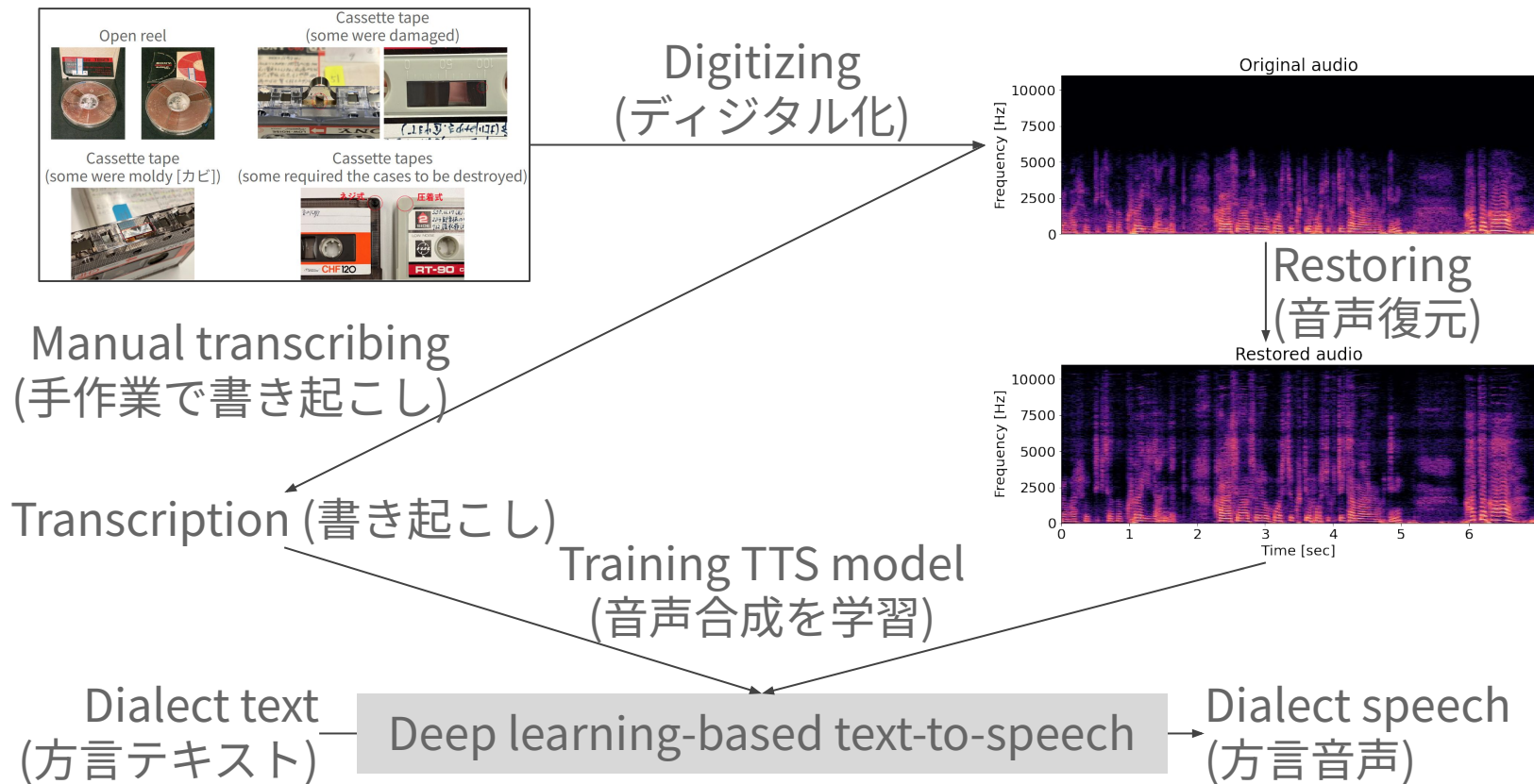


Successfully restored historical audio! (歴史的音源の復元に成功！)

3. dialect speech synthesis (方言音声合成)

Let's build AI that speaks the dialect!
(方言でしゃべる AI を作ろう)

Building dialect text-to-speech system (方言音声合成システムを構築する)



Folktale movie with synthesized dialect speech (合成音声を用いた民話)

雪女郎 (第1章 途中まで)

Future direction (今後の予定)

Future direction (今後の予定)

- 90% of digitized data (360 hours) has no annotation.
 - Need AI-powered annotation, e.g., automatic speech recognition
(さらなるアノテーションのため、音声認識などの技術が必要)
- Apply speech restoration to a variety of distorted speech.
(音声復元技術を様々な劣化音声に適用する)