

国立国語研究所 異分野融合型共同研究
2024年度 ワークショップ (2025/02/27)

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 プロジェクト概要

高道 慎之介^{1,2}, 佐伯 高明¹, 中田 亘¹, 丹治 尚子¹, 庄司 潤子³, 佐藤 照一⁴
(1: 慶大, 2: 東大, 3: 仙台文学館, 4: 昔話採集家)



Takamichi Laboratory
慶應義塾大学 高道研究室

自己紹介



@forthshinji

名前

高道 慎之介 (たかみち しんのすけ)

現職

東京大学 講師 (-2023) → **慶應義塾大学 准教授 (2024)**

経歴

離島→熊本高専→長岡技大→奈良先端大. 2016年に博士

兼務

東京大学 特任准教授, {理研, 産総研, 海洋機構} 客員研究員,
IEEE Speech and Language 専門委員会 理事 など

受賞

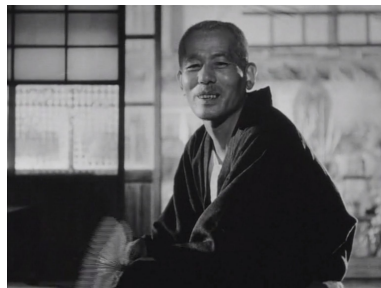
IEEE SPS Young Author Best Paper Award
文部科学大臣表彰 若手科学者賞

歴史的音源アーカイブ



録音風景 (1950s)

<https://www2.ninjal.ac.jp/conversation/showaCorpus>



小津安二郎
”東京物語“ (1953)



Aerosmith
“Dream on” (1973)



NASA
Apollo 11 (1969)

録音は歴史のスナップショットとしての音メディア。
音声・言語・音響を通時的に扱い、歴史・文化を識ることができる。

(何もしなければ) 歴史は消えていく



デジタルメディアの前の収録機器は劣化していく。廃棄されていく。

プロジェクト目標

- **目標1：AI 音声復元技術の開発**

- 実際の歴史的音源に耐える音声復元技術の開発
- オープンソース化

- **目標2：東北方言昔話コーパスの開発**

- 東北方言昔話音源のデジタル化
- デジタル化音源のアノテーション
- オープンソース化

プロジェクトメンバ

目標1：機械学習G

目標2：コーパス開発G

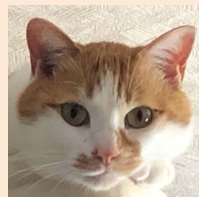
研究参加者
(コアメンバ)



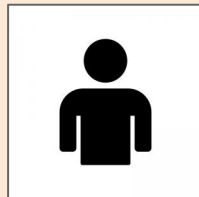
佐伯 高明
(東大)



高道 慎之介
(東大)



丹治 尚子
(東大)



庄司 潤子
(仙台文学館)

研究協力者



金森/兵藤
/中田 (東大)



森松
(東大)



佐藤
(昔話採訪家)



田村
(方言アドバイザー)



佐々木
(仙台)

本セッションの内容

- 高道, ”プロジェクト概要”
- 佐伯, 高道, ”歴史的音源のみから学習可能な音声復元技術”
 - 機械学習を用いた音声復元技術
 - 方言音声を含む, 音声基盤モデル用データセットの策定
- 丹治, 庄司, 佐藤, 高道, ”東北方言昔話のオープンコーパス化”
 - 東北方言音声データのデジタル化
 - 音声データのアノテーション, 分析

研究代表者の大学異動 & 主宰研究室の立ち上げにより,
本年度はあまり進捗がありません. ご容赦ください.