報道関係各位

Press Release

すべての産業の新たな姿をつくる。

Laboro

オーダーメイド AI ソリューション、
『カスタム AI』開発

株式会社 Laboro.AI

2020 年 11 月 19 日

TV 録画から自動構築した音声コーパス 『 LaboroTVSpeech 』を開発＆公開

～　日本語音声コーパスとして最大規模　2,000 時間の音声データから構成　～

株式会社 Laboro.AI は、ワンセグ TV 録画から抽出した約 2,000 時間の音声データから構成される音声コーパス『LaboroTVSpeech』を開発し、学術研究用に無償公開いたしました。

＜今回のポイント＞

- ✓ 日本語音声コーパスとしては最大規模の**約 2,000 時間のデータ**
- ✓ TV 番組に含まれる音声と字幕データから、**音声コーパスを自動構築**するシステムを開発
- ✓ 既存の音声コーパスより**優れた誤認識率**を達成し、商用の音声認識 API にも匹敵する精度を確認

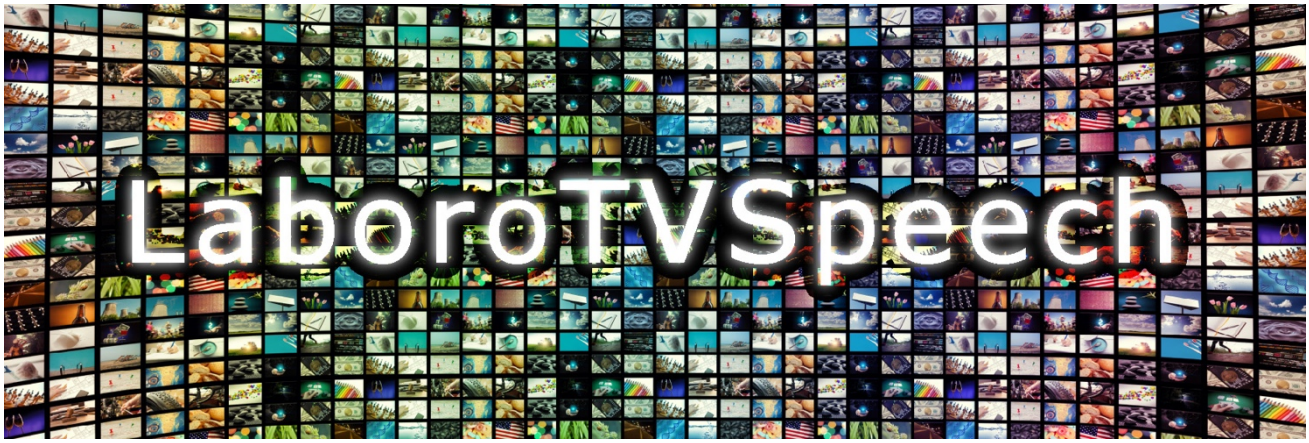
株式会社 Laboro.AI
代表取締役 CEO　椎橋徹夫・代表取締役 CTO　藤原弘将

オーダーメイドによる AI・人工知能ソリューション『カスタム AI』の開発・提供およびコンサルティング事業を展開する株式会社 Laboro.AI（ラボロエーアイ、東京都中央区、代表取締役 CEO 椎橋徹夫・代表取締役 CTO 藤原弘将。以下、当社）は、当社の研究開発として、TV 録画から長時間音声と字幕テキストを抽出して音声コーパスを自動構築する独自システムを用い、約 2,000 時間に及ぶ音声データから構築した日本語音声コーパス『LaboroTVSpeech（ラボロティービースピーチ）』を開発し、学術研究用に無償公開いたしました。

LaboroTVSpeech は、B-CAS カードによるアクセス制限がないワンセグ放送を利用しており、複数ジャンルの計 9,142 番組の TV 録画から抽出した約 2,000 時間の音声データから構成されています。研究用途として代表的な日本語話し言葉コーパス（CSJ：約 600 時間）や新聞記事読み上げ音声コーパス（JNAS：約 90 時間）など、これまで公開されている日本語音声コーパスと比較しても最大規模のものです。当社比較実験の結果では、LaboroTVSpeech で構築した音声認識モデルが従来の研究用日本語音声コーパスで構築したモデルを凌ぐ誤認識率となり、さらに商用で提供されている主要な他社製クラウド音声認識 API にも匹敵する誤認識率を確認いたしました。

なお、本年 12 月 2 日（水）・3 日（木）に開催される（一社）情報処理学会 第 246 回自然言語処理・第 134 回音声言語情報処理合同研究発表会にて、LaboroTVSpeech についての報告を実施予定です。

本件について詳細は、次ページ以降にてご確認ください。



＜ 背景 ― これまでの音声認識モデルと音声コーパス※1＞

一般的に音声認識モデルの性能は、その学習データの量に大きく左右され、高品質な音声認識システムを構築するためには大規模な音声コーパスが必要とされています。そのため、英語モデルの開発の場合には、商用目的では数千～数万時間を超える音声データが用いられることもあり、研究目的でも SwitchBoard-Fisher データセット（約 2,000 時間）や LibriSpeech（約 960 時間）などの大規模な音声コーパスが公開されています。

一方、日本語の音声コーパスについては、研究用として代表的な日本語話し言葉コーパス（CSJ※2）で約 600 時間、新聞記事読み上げ音声コーパス（JNAS※3）で約 90 時間など、英語と比較すると十分なデータの音声コーパスが存在しているとは必ずしも言えないのが現状です。

この背景としては、音声コーパスの構築に際して書き起こしや録音作業など、人手による手間やコストがかかることが理由として挙げられます。その対応として、これまで人手の作業を伴わず自動的にデータ収集を行う手法が模索されてきました。その一つの方法として挙げられるのが、テレビ放送を用いた音声コーパスの自動構築です。多くのテレビ番組には字幕情報が付与されているため、音声と字幕のテキスト情報を時間的に紐付けることで、コーパスを自動構築できる可能性が示されてきました。

しかしながら、字幕は視聴者にとっての読みやすさを優先して作成されているため、必ずしもテレビ音声の正確な書き起こしにはなっていません。そのため、音声コーパス構築の自動化を行うためには、以下をはじめとする点に留意する必要があり、その実現ハードルは高いのが実際です。

- 字幕の表示時間と実際に音声が発生される時間にはズレがあり、特にニュースなどの生放送番組では音声が発せられてから字幕が表示されるまで 10 秒以上の遅れを伴う場合がある。
- コマーシャルなど、全ての音声に字幕があるわけではなく、また字幕が実際の発話とは異なる形で整形されたり短縮されたりすることがあり、忠実な書き起こしとは限らない。
- バラエティ番組などでは発話の一部が字幕ではなく，いわゆるテロップとして映像上に付与される場合があるが、テロップはテキストとして情報を取得できないため、データとして取得される字幕テキスト上では複数の単語や文章が不規則に削除されているような状態となる。

※1　コーパスとは、言語情報を大量に集積したデータベースのことを指します。
※2　国立国語研究所、情報通信研究機構（旧通信総合研究所）、東京工業大学が共同開発した話し言葉データベース。
※3　日本音響学会が公開している新聞記事を読み上げた音声コーパス。

＜ LaboroTVSpeech について ＞

今般当社で構築した音声コーパス LaboroTVSpeech は、B-CAS カードによるアクセス制限がないワンセグ放送を利用し、2020 年 2 月～9 月にかけて放送された、12 ジャンル（デジタル放送規格の番組種別）計 9,142 番組のテレビ番組の録画データを用いており、2,049 時間に及ぶ音声データから構成された大規模音声コーパスです。当社では、LaboroTVSpeech をアカデミア領域での AI 技術研究に広く活用いただくことを目的に、学術研究用に無償で公開することといたしました。

番組ジャンル	音声の長さ（時間）
ニュース／報道	767.0
バラエティ	316.4
情報／ワイドショー	323.3
ドラマ	206.0
ドキュメンタリー／教養	175.6
趣味／教育	101.0

番組ジャンル	音声の長さ（時間）
スポーツ	66.6
アニメ／特撮	39.7
音楽	17.6
福祉	20.1
映画	6.2
劇場／公演	10.4

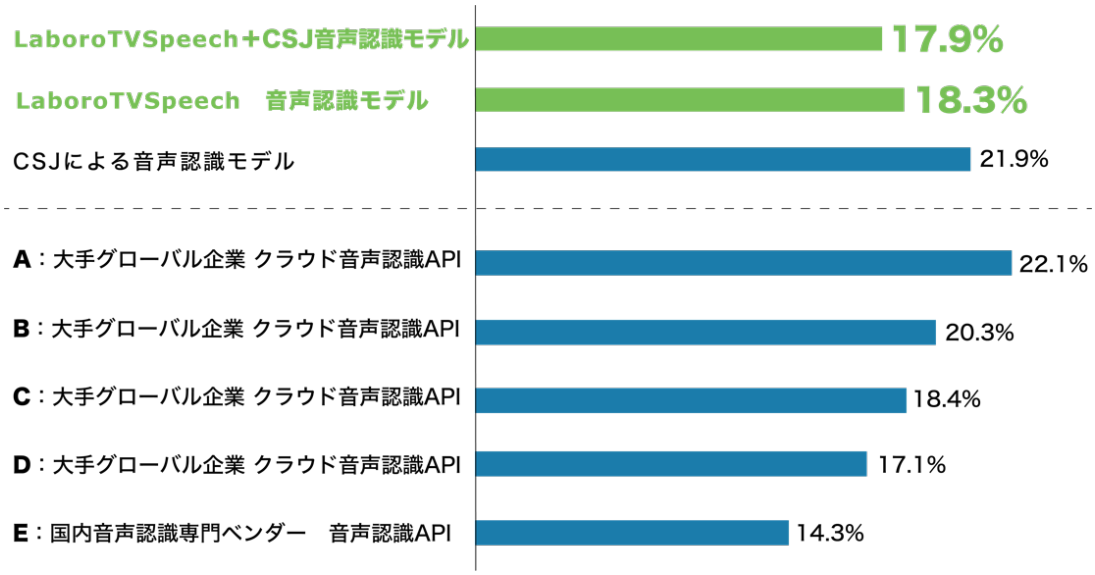
LaboroTVSpeech を構成する番組ジャンルと音声の長さ

LaboroTVSpeech は、当社が独自開発したシステムにより構築しています。具体的には、テレビ番組の長時間の音声データと、その不完全な書き起こしである字幕データの時間的な対応関係を抽出する手法である準教師付きデコーディング（lightly-supervised decoding）と呼ばれる手法をベースとしています。これにより、本来であればテレビ番組のデータから音声と字幕がセットになって抽出されるべきところ、先のような何らかの問題で対応した情報として取得できなかった場合に、準教師付デコーディングによる音声と字幕の対応関係の抽出を繰り返し行うことで、一度対応が取れなかった区間からも可能な限りデータ抽出を行う仕組みを採用しています。

なお、LaboroTVSpeech については、本年 12 月 2 日（水）・3 日（木）に開催される（一社）情報処理学会 第 246 回自然言語処理・第 134 回音声言語情報処理合同研究発表会にて報告を予定しています。

＜ LaboroTVSpeech の比較実験について ＞

LaboroTVSpeech を用いたモデルの音声認識の性能を確認するため、日本語の TEDx を用いて構築した独自の音声認識システム評価用データセット[※4]を用意した上で、従来の日本語音声コーパス CSJ で学習されたモデル及び、国内外の主要なクラウド音声認識 API との比較実験を行いました。その結果、当社開発モデルが従来の研究用日本語音声コーパスで構築したモデルを凌ぐ誤認識率を達成しました。さらに商用で提供されている他社製クラウド音声認識 API との比較では、LaboroTVSpeech 用いることで主要な大手グローバル企業が提供する API にも匹敵する誤認識率を確認いたしました。



日本語 TED コーパスに対する誤認識率（CER：Character Error Rate）の比較※5

※4　YouTube 上のプレイリスト「TEDx talks in Japanese」に含まれる動画から、1 万発話文の音声とその字幕データを取得した上で、人手で修正を加えたものです。
※5　クラウド音声認識 API の評価は、全てデフォルトの音響モデル及び言語モデルを用いて実施しています。上記の結果は、実環境での音声認識システムの性能とは異なる場合があります。

＜ LaboroTVSpeech の今後の可能性＞

LaboroTVSpeech の利点は、テレビ放送をデータソースとしていることからデータ量を絶えず増加させることが可能な点にあります。当社では LaboroTVSpeech のデータ量を増強するために定期的なアップデートを実施し、公開することで、国内における AI 音声認識分野の技術力向上に寄与してまいります。

＜ LaboroTVSpeech の利用について ＞

LaboroTVSpeech に含まれる音声及びテキストデータの権利は、元のテレビ放送の著作権者に帰属していますが、著作権法 30 条の 4 に基づき、情報解析等の用途のために、大学等の学術研究機関における非商用利用を対象に当社ホームページ（https://laboro.ai/column/eg-laboro-tv-corpus-jp/）にて無償で公開いたします。ただし、元のテレビ番組の音声を再構成し鑑賞する事を防ぐために、発話単位でランダムに並び替えられており、かつ番組名や放送局等の付加情報は含まれておりません。

営利企業における研究開発用途や商用目的での利用をご希望の場合は、当社ホームページのお問い合わせフォーム（https://laboro.ai/contact/other/）からのご相談をお願いしております。

株式会社 Laboro.AI について

(株)Laboro.AI は、オーダーメイドの AI ソリューション『カスタム AI』の開発・提供を事業とし、アカデミア（学術分野）で研究される先端の AI・機械学習技術をビジネスへとつなぎ届け、すべての産業の新たな姿

をつくることをミッションに掲げています。業界に隔たりなく、様々な企業のコアビジネスの改革を支援しており、その専門性から支持を得る国内有数の AI スペシャリスト集団です。

＜会社概要＞

社　名：株式会社 Laboro.AI（ラボロ エーアイ）
事　業：機械学習を活用したオーダーメイド AI 開発、
　　およびその導入のためのコンサルティング
所在地：〒104-0061 東京都中央区銀座 8 丁目 11-1
　　GINZA GS BLD.2 3F
代表者：代表取締役 CEO　椎橋徹夫
　　代表取締役 CTO　藤原弘将
設　立：2016 年 4 月 1 日
ＵＲＬ：https://laboro.ai/

＜本リリースに関するお問い合わせ＞
株式会社 Laboro.AI リードマーケター 和田 崇
Mail：press@laboro.ai　Tel : 03-6280-6564（代表）

※昨今の新型コロナウイルス感染拡大の状況を鑑み、当社では、当プレスリリース発信時点で原則、全社的なリモートワーク体制を敷いております。当社代表電話番号へのご連絡をお受けできない可能性がございますため、その際は、メールにてご連絡いただけますようお願いいたします。関係各位にはご不便をお掛けいたしますが、何卒ご理解賜れますようお願い申し上げます。