

音声認識のブラックボックスシステムにおける簡易的介入法の研究と創作物への応用

藤田 麟太郎^{1,a)} Scott Allen^{1,b)}

概要: 創作行為の発展には、誤用から派生した独自の手法が存在する。その例として、DJ のスクラッチやサーキットベンディングが挙げられるが、それらの共通している点はアナログな手法で既製品に対して行われているもので近年の洗練化されたブラックボックス化された音声認識システムにおいては介入が難しい。本研究では、そのようなシステムに対しても、従来の介入法に近いことが擬似的に実現できる手法を提案する。また、この手法を用いた作品を制作し、作品鑑賞者に対してアンケートを実施して、本研究の手法の有効性を検証した。

1. はじめに

近年、急速に進化する音声認識システムが私たちの日常生活に深く浸透している。これらのシステムは、スマートデバイスや音声アシスタントを含む様々なアプリケーションにおいて、我々の声をテキストやコマンドに変換し、様々なタスクを自動的に処理するプログラムによって処理されている。

一般的な研究では音声認識のブラックボックスシステムを構築する方法やセキュリティ面に焦点が当てられるが、本研究ではそのようなユーザーの信頼性に焦点を当てるのではなく、クリエイターがブラックボックスシステムに対してどうアプローチして創作行為に使用するかが重要なコンテキストとなる。創作行為の発展には、誤用から派生した独自の手法が存在する。例えば、グランドマスター・セオドア氏とグランドマスター・フラッシュ氏らによるレコードプレイヤーの誤用によるスクラッチ奏法の発明 [1] やリード・ガザラ氏によるサーキットベンディングのショートや抵抗値の変動による演奏装置の発明 [2] は、誤用から新たな手法が生まれた事例である。だが、これらの例はアナログな手法で既製品に対して行われているものであり、技術の発展とともに、より洗練化されてきたオープンソースでないプロジェクトにおいては、介入が難しく、それらのシステムを用いて上記に類似した手法で制作された作品は少ないのが現状である。

本研究はブラックボックス化されている音声認識システム

に対して、従来の介入法に近いことが擬似的に実現でき、新たな表現を生み出す可能性を模索する。

2. 関連事例

本システムに関連する先行研究や作品を以下に示す。本システムに近い外部からのシステムへの介入は、GoogleMapsHacks[3]である。GoogleMapsHacks は、ドイツのアーティストである Simon Weckert 氏による作品であり、99 個の中古スマートフォンを手に押し車に載せて運び、Google マップ上に架空の交通渋滞を作り出すというパフォーマンスを行った。この作品は、Google マップという私有プラットフォームでありながら、Google マップの背後にある目に見えないアルゴリズムを可視化するものとして、注目を集めた。

この作品は、音声認識システムのブラックボックス性に対処する上で、重要な示唆を与えるものと考えられる。すなわち、ブラックボックスシステムであっても、外部から介入することができたり、その動作を可視化したり、意図的に操作したりすることができるということだ。

しかしながら、GoogleMapsHacks は位置情報によって渋滞情報を管理しているというアルゴリズムの一点のみにしか介入していない。本研究では、言語の誤認識について、単一なルールではない膨大な数の誤認識する言葉を発見し、多角的に介入している。

3. 提案システム

制作したシステムは大きく二つに分けられる。誤認識させるための音声作成とシミュレートをする PC 側の処理と再生機・スピーカー・iPhone から構成されるモジュール

¹ 京都精華大学

^{a)} f221t157@st.kyoto-seika.ac.jp

^{b)} scottallen@kyoto-seika.ac.jp

である。このシステムを用いることで、意図的に外国語を Siri に日本語と誤認識させる言葉を抽出することができる。

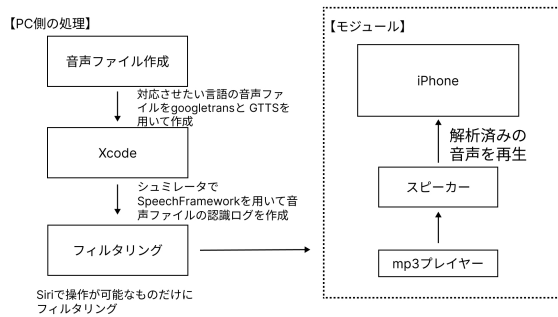


図 1 システム概要図

音声データ作成の際に使用したツールは Google Colaboratory を用いて実行環境を構築し、googletrans を利用して翻訳、取得した文字情報を Google Text to Speech にて発話させる手法で作成した。

Siri に誤認識させる言葉を抽出する過程では、Xcode 上で Apple が提供している SpeechFramework を利用し、音声認識入力システムの構築を行う。その後、ログに出力された音声認識結果をフィルターにかけて Siri で操作が可能なものに絞る。

実機検証で使用している iPhone の iOS バージョンに検証環境を揃え、シミュレーションログをもとに誤認識し、Siri が操作をできるもののみをフィルターをかけることで抽出を行い、最終的に音声を実機に向けて発声させることで実空間での検証を行う。

4. 提案システムの作品制作への応用



図 2 展示の様子

本システムを用いて作品を制作し、2023 年 11 月 25 日・26 日に大阪府大阪市にある Blend Studio で行われた「HOME

WORKS 2023」と 2023 年 11 月 27 日・28 日に大学内オープンゼミにて展示した。その際、鑑賞者の行動を観察した。

作品へ応用するにあたり、システム内で解析した言語はフィンランド語を使用した。使用した理由として、「日本語とフィンランド語は単純な子音構造を持つという点で音節拍リズムをもつ言語に見られる特徴を共有している [4]」とされていたためだ。

表 1 に今回使用したフィンランド語と日本語の誤認識対応表を示す。

表 1 フィンランド語と日本語の誤認識対応表

フィンランド語	フィンランド語の発音	フィンランド語の意味	反応する言葉	動作内容
Joulupukki	ヨーロプッキ	サンタクロース	今日の天気は	今日の天気を教えてくれる。
aamulla	アームラ	アラーム	午前中に	アラームの確認をしてくれる
kahdeksan	カハレクサン	アレクサ	8	気まづがる
neljä	ネレヤ	寝るよ	4	おやすみさいという
penkki	ペンキ	天気	ベンチ	天気を教えてくれる
matalasta	マタラスタ	また明日	低い方から	失礼しますと言う
moi	モイ	もういい	こんにちは	わかりましたと言う
antaa Annan	アンター アンナン	あんな女	アンナにあげる	性別はありませんと言う
Tämä	タアマ	タイマー	これ	タイマーの確認
astua	アストゥフ	明日	踏みます	明日の予定の確認
puhun	プフン	5分	話します	5分のタイマーを開始
tanssi	タンシ	男子	踊ります	性別はありませんと言う
Tietysti	ティエシディ	デートして	勿論	断られる
eläin	エライン	偉い	動物	嬉しいがる
kävellä	カヴァエラ	カメラ	散歩	カメラが起動
asemalle	アスマアレ	明日雨	駅	明日の天気を確認する
onnellisuus	オンネリソウス	女	幸福	性別はないという
nero	ネロ	寝る	天才	仮眠を取っているという

作品中で使用した映像は、同年 11 月 22 日にフィンランド語話者である交換留学生の Marika Schulze さんに協力を依頼し、解析済みの単語が発話された映像を使用した。

展示会場では、ディスプレイを取り付けたモジュールを配置し、観客に対して自由に鑑賞してもらおう形で展示を行った。

5. 実施例に対する評価・考察

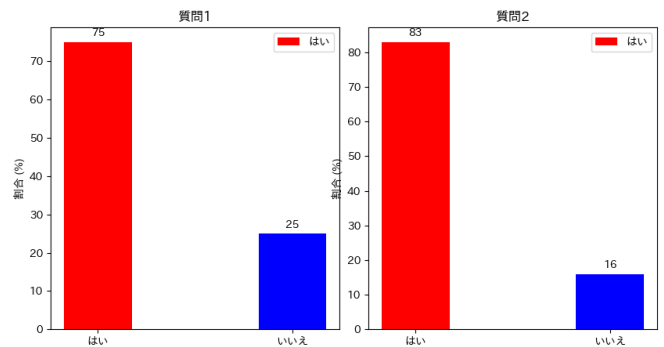


図 3 アンケート結果

5.1 モジュール及び鑑賞者の行動

4日間の展示期間でおよそ200人に鑑賞してもらった。モジュールは時間経過とともに検証時には見せなかった動作を何度かしていた。また、鑑賞者の中には誤認識しているSiriに自らの声を重ねることでさらに干渉していくような行動を見せるものがあった。

5.2 鑑賞者からのフィードバック

実施例に対するアンケート結果を図3に示す。展示を見た鑑賞者を対象に、以下の2点を10代から30代の音楽家、学生の男女合計13人に確認した。

- (1)「音声認識システムの専門家でなくても意図的にSiriを操作ができていると思うか」[5]
- (2)「それまで隠されていた未知の機能を発掘し、そこから予想外の使用法を発見したり、新たな美学を創造していると思うか」[6]というアンケートをGoogleFormにて実施した。このアンケートでは、冒頭で述べた、従来の介入法に近いことが擬似的に実現でき、新たな表現を生み出す可能性があるということを検証すること目的とし、2つの論文からサーキットベンディングの特徴を捉え、提案システムが従来の介入法を擬似的に実現できているかを計るものである。

アンケートの結果、「音声認識システムの専門家でなくても意図的にSiriを作者が操作できていると思うか」という問いに対しては75%が「はい」25%が「いいえ」と回答した。「それまで隠されていた未知の機能を発掘し、そこから予想外の使用法を発見したり、新たな美学を創造していると思うか」という問いに対しては83%が「はい」16%が「いいえ」と回答した。

5.3 考察

表1で示した対応表からは、フィンランド語の発音と反応する言葉の発音の類似度が低いものもある。しかし、類似度が低くなくてもSiriに誤認識させられている。どのような基準で言葉が反応しているかはブラックボックス化されているため断定することは難しいが、そのような状況でこそ本研究で提案しているシステムはクリエイターが創作行為に使用する際に有効であると考察できる。

図3で示したアンケート結果から、被験者の殆どは2つの質問に対して、「はい」と答えており、従来の介入法を擬似的に実現できていることがわかる。

質問1で否定的な回答をした被験者の「無意識から全く別のものが生まれるのは面白いAIのよう、というかAI」、「初めはただ上手いこと聞き取れないSiriかと思ったのですが、それを前提に(?)した作品なの面白かった」等の自由回答から、今回展示した作品の解釈が個人の経験に左右されるような質問であったことが考えられる。

質問2で否定的な回答をした被験者は「Siriには何とな

くの愛着が出てしまうのが楽しいところですよ」と自由回答しており、Siriの機能をそのまま使っている点が質問内容である未知の機能を発掘していると言えないと捉えたのではと考察することができる。

6. まとめ・今後の展望

本研究では、音声認識のブラックボックス化されているシステムに対しても、サーキットベンディングのような従来の介入法に近いことが擬似的に実現できる手法を提案し、擬似的に実現できているということがアンケート結果によって証明された。

また、実際に本システムを用いた作品を制作し、展示で動作を確認できた。今後は、提案システムを用いた多言語による誤認識を用いた作品の実現を計画している。

今回使用したシステムはApple社のものだが、似たようなシステムを構築することでGoogle HomeやAlexaといった音声認識システムでも応用できると考えられる。

今回提案したシステムを用いた制作を続けることでオープンソースでない状況でもクリエイターが表現する余地を見出し、新たな表現のジャンルとして確立していきたい。

参考文献

- [1] Grand Wizard Theodore accidentally invents scratching (or does he?)(online), 入手先 (<https://www.theguardian.com/music/2011/jun/13/grand-wizard-invents-scratching>) (2023.12.22).
- [2] Reed Ghazala: Circuit-Bending: Build Your Own Alien Instruments(2005).
- [3] Garnet Hertz and Jussi Parikka: Zombie Media: Circuit Bending Media Archaeology into an Art Method(2012).
- [4] 堀 誉子美: 音節構造と聴覚印象—日本語とフィンランド語の対照研究—(1995).
- [5] ‘Google Maps Hacks(online), 入手先 (<https://www.simonweckert.com/googlemaphacks.html>) (2023.12.22).
- [6] 井藤 雄一: 情報メディアの利用によるアート表現—メディアの変則的利用がもたらす可能性—(Art Expression Created through Anomalous Use Information Media)(2014).