

個人に紐付くメディア情報を用いた コミュニケーション可視化ツールの開発

—SignalLog による会話の非言語シグナルの記録と解析—

1. 背景

我々は日々、他者との面と向かったコミュニケーションによって様々な情報を受け取っている。コミュニケーションは、自分が知らなかった情報を得たり、自分が知っている情報を相手に伝えたりするための情報伝達手段として重要である。しかしそれ以上に、コミュニケーションを通じた相互理解や自己の承認感を得ることこそが、コミュニケーションへの欲求を掻き立てる。一方で、コミュニケーションにおける行き違いは人間関係を壊すことにもつながりうる。相手が怒っていたり、もう話に飽きていたりするのに自分の話を続けてしまい、相手の気分を害してしまう、という経験は誰しもあるだろう。このように、他者の心情の推定は人間にとって非常に重要な意味を持つにも関わらず、とても曖昧な形でしか推測することが出来ない。

2. 目的

本プロジェクトでは、このコミュニケーションによって生じる心理的作用の不可視性に着目し、人に紐付けられたメディア情報、及び過去の対話履歴から、コミュニケーションによって生じる心情変化を拡張現実感として付与するシステムを開発することを目的とした。そのために、画像や音声、言語からリアルタイムで心情認識を行いながら、同時にそれらを統合するようなモデルを構築した。本システムにより、これまで人間が気づけなかったコミュニケーションの新しい楽しさが発見され、より良い相互理解を促すことを期待した。

3. 開発の内容

Android上で動作する会話情報収集アプリケーションSignalLogを開発し、得られたデータの解析を行った。SignalLogは、動画情報と音声情報をリアルタイムで処理する機能を持っており、これらについて順に説明する。

まず、動画情報処理について詳説する。処理の流れを図1に示す。動画情報処理では、Androidデバイスによって撮影された動画に対して顔検知と顔追跡を行い、動画に顔が入っていれば顔画像情報を毎フレーム抽出する。顔画像はサーバに送信され、多量の顔画像から学習されたモデルを介して顔認識を行い、会話相手の名前を特定する。顔画像の系列から、口の動きが大きければ相手が話者であると判断することによって話者認識を行う。同時に、顔画像を用いた感情認識システムによって感情値を計算する。顔画像のデータは名前や感情値と紐づき、データベースに保存する。

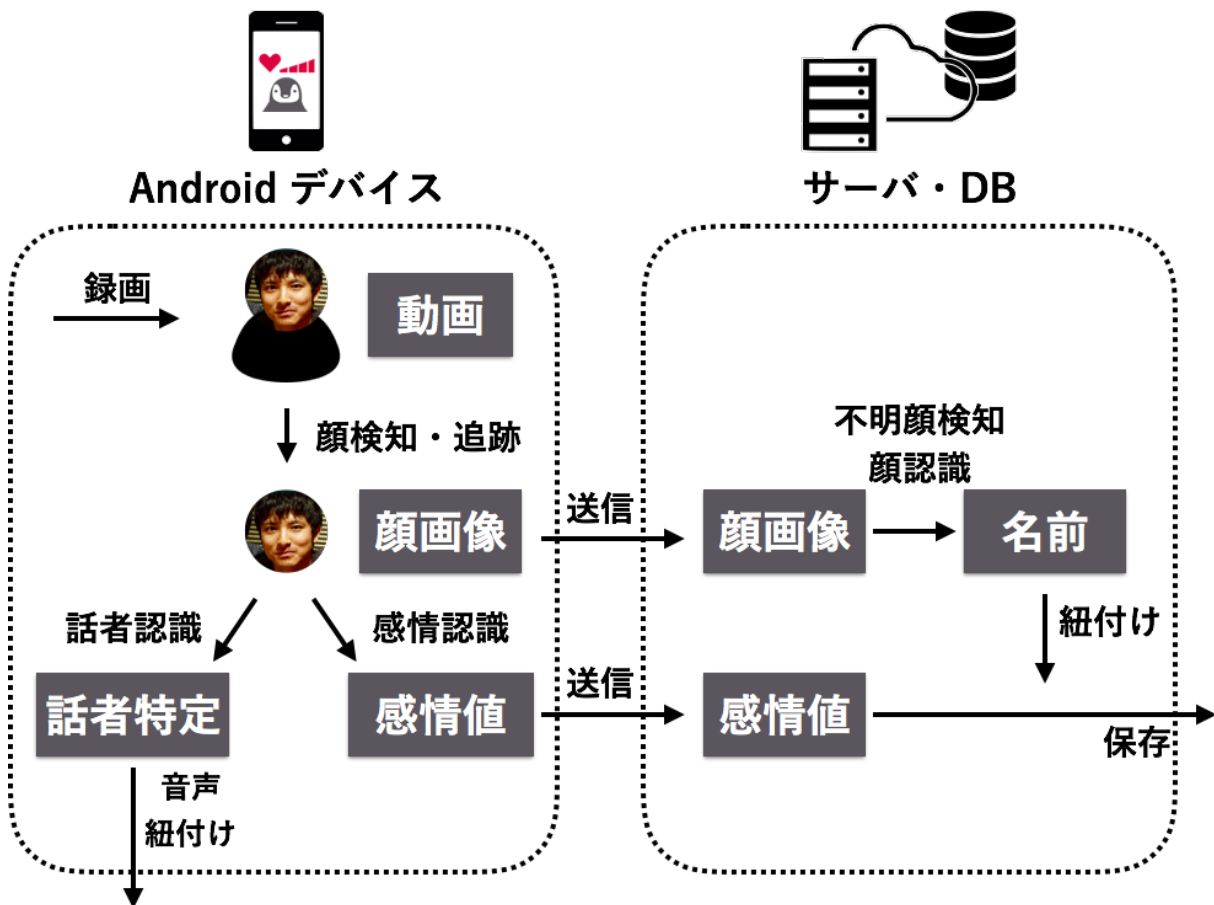


図1: 動画情報処理の流れ

デバイス側で行う処理を左に，サーバ側で行う処理を右にそれぞれ示している。
ここで，デバイス側で行う処理は全て撮影中リアルタイムに行われる。

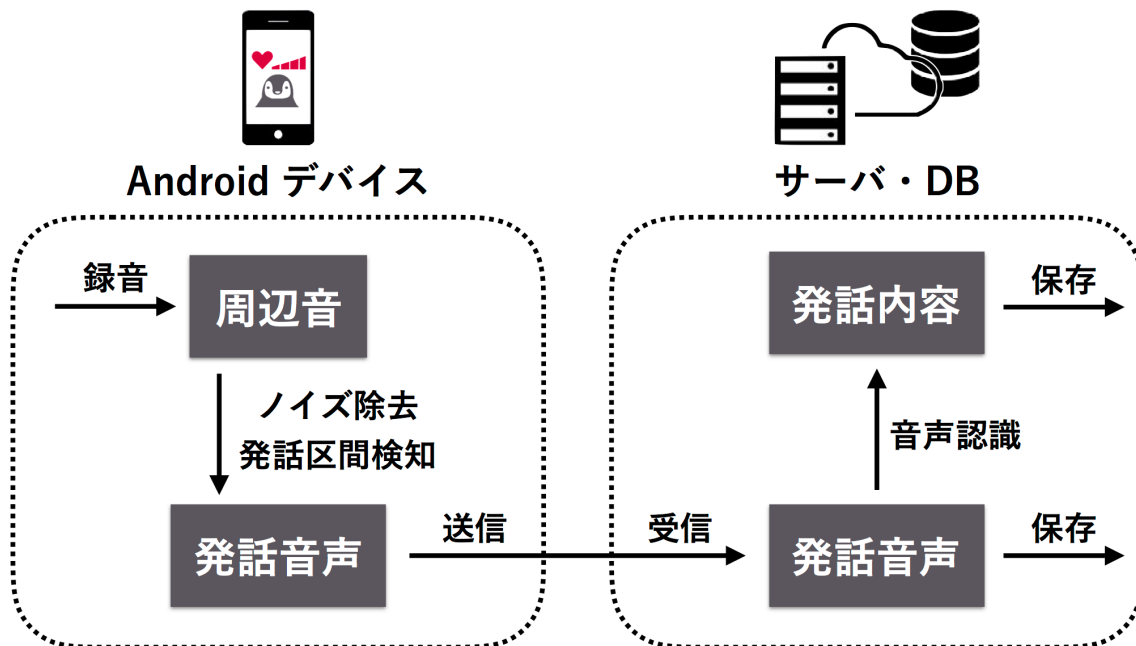


図2: 音声情報処理の流れ

デバイス側で行う処理を左に，サーバ側で行う処理を右にそれぞれ示している。
ここで，デバイス側で行う処理は全て撮影中リアルタイムで行われる。

次に、実装された音声情報処理について詳説する。処理の流れを図2に示す。音声情報処理では、Androidデバイスに入力された音声に対してノイズ除去と発話区間検知を行い、発話区間とみなされた部分については、音声から音量特徴量を抽出し、特徴量と音声そのものの両方をサーバに送信する。サーバに送信された発話音声は逐次的に音声認識処理にかけられ、発話内容のスクリプトを獲得する。

4. 従来の技術（または機能）との相違

本システムで解析の対象とした、対面での会話を撮影することによって得られる表情や発話の時系列的な変化のような話者の内部状態が反映される情報のことを、本プロジェクトでは「非言語シグナル」と呼ぶ。既存の非言語シグナル情報の解析を行うシステムやサービスは、そのほとんどが音声情報のみを主な解析対象としていた。

しかし、本来非言語的なシグナルは表情に現れるものも非常に多い。これについては、認知心理学の観点からの裏付けも存在し、人間の内部状態を知る主要な手がかりとなることも分かっている。本プロジェクトではこの点に着目し、音声特徴量抽出と画像からの表情特徴量抽出がデバイス上でリアルタイムで行える仕組みを開発した。サーバでの計算をデバイス側が部分的に肩代わりできるようにしたことも、副次的な利点である。また、本プロジェクトではAndroidアプリケーションとして開発を行ったが、ライブラリの形式でまとめている機能も多く、ラッパーさえ用意すればiOSやFPGA上など、異なる環境でも動作させることが可能である点もプラットフォームとして優れている点である。

SignalLogによって実際にデータ収集を行い、限られた状況下ではあるがその有効性を検証した。収集したデータは、チームメンバーのうち内橋、高濱、寺田の3人の間に上下関係を仮定した上で行った会話データと、内橋が自身の交際相手（彼女）と会話したデータと女友達との会話データである。特に、上下関係を仮定したデータでは、収集した特徴量に各メンバーで明確な違いが見られ、「非言語シグナルが関係性を反映する」という仮説が確かめられた。また、内橋が彼女と会話したデータと女友達と会話したデータを比較し、その違いを観察することによって、「親密さ」とも呼ぶべき関係性が音量や発話スピード、表情変化に主に現れることが確認できた（図3）。

さらに、上記のような機能を実現するため、デバイス上で高速に動作する深層学習モデルのデプロイを簡単に行える仕組みを開発したことも貢献の一つである。これは、従来サーバで行うことが常識であった深層学習による計算をデバイス上で行うための、高速なI/Oとデプロイシステムを構築することによって実現した。

5. 期待される効果

非言語的なシグナルが解析対象となることによって、人間の関係性がより本質的な意味で明らかになる可能性がある。これまで行われてきたコミュニケーションデータからの関係性抽出は、職場や病院といった狭いコミュニティ内で収集したデータ

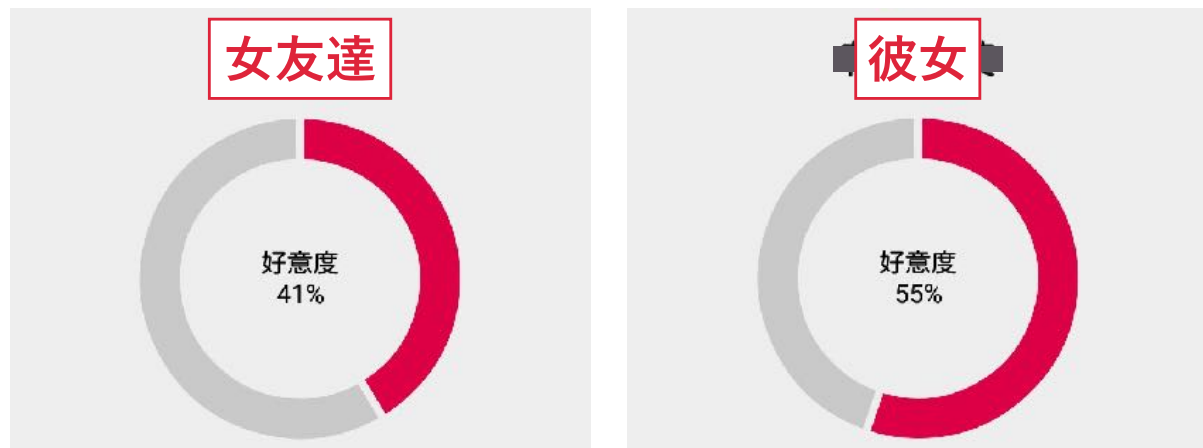


図3: SignalLogによる非言語シグナルの解析結果の一例
内橋が自身の彼女および女友達と会話した際の高感度。
女友達より彼女の方が高感度が高くなっていることが分かる。

が中心であったが、本アプリが示したARとしてユーザに示す機能によって、一般に広く利用されるライフログ収集アプリとなる可能性が示された。

さらに、上記機能を実装する過程で開発したリアルタイムで映像解析を行うプラットフォームとしての機能は、今後広がっていくであろうIoTプラットフォームと親和する形で提供していくことによって、非常に高度な映像解析をIoTデバイス上で行うことも可能になる。

6. 普及（または活用）の見通し

SignalLogは現在はAndroid上での実装であり、会話の撮影方法でその使いづらさがあるため、普及のためにはそのUI/UXを改善する必要がある。例えば、ARデバイスへの実装を行うことでUI/UXを改善することで、ライフログ収集アプリとしての普及の可能性がある。ユーザ数はARデバイスの普及度合いに比例すると考えられ、今後Google GlassやHoloLensなどのデバイスが、どのように普及していくかを注視していく。

SignalLogをリアルタイムで映像の解析が行えるプラットフォームとして見た場合、映像をサーバに送信しないことや解析のリアルタイム性などの利点から、今後広がっていくであろうIoTの煽りを受け、活用される可能性は大いにある。

7. クリエータ名（所属）

内橋 堅志 （京都大学 大学院情報学研究科 システム科学専攻）

宮戸 岳 （株式会社Preferred Networks）

高濱 隆輔 （京都大学 大学院情報学研究科 知能情報学専攻）

寺田 凜太郎 （京都大学 大学院情報学研究科 知能情報学専攻）