

VoLP: 非同期音声コミュニケーション促進のための IoT ノード

橋本 慶紀^{1,a)} 岩井 将行^{1,b)}

概要: 近年、ボイスメッセージの利用が広がりを見せている一方で、モバイル端末に不慣れな人々は非同期音声コミュニケーションを十分活用できていない。この問題を解決するため、本研究では非同期コミュニケーションのための IoT ノード VoLP を提案する。VoLP は録音データへのアクセスを、印刷した QR コードとして提供する。これにより、専用アプリケーションや専用サービスの利用を必要とせず、直感的かつ簡単に非同期音声コミュニケーションを活用できるようになる。本稿では、VoLP の提案と試作 VoLP の実装について述べ、最後に VoLP の利用シナリオを検討する。

キーワード: 非同期, 音声コミュニケーション, IoT

1. はじめに

コミュニケーションは、同期コミュニケーションと非同期コミュニケーションに分けられる。双方が利点・欠点を持っており、我々は日常生活においてコミュニケーションの同期・非同期を使い分けている。

たとえば非同期コミュニケーションには、送信者・受信者の双方がコミュニケーションするタイミングを合わせる必要がないという利点がある。そのため、非同期コミュニケーションが際立って効果的な場面というのがいくつかある。携帯電話・スマートフォンをはじめとする小型端末が普及した現在、非同期コミュニケーションは、以前より頻繁に用いられている。SNS での日常的なテキストチャットや、大学におけるオンデマンド授業がその例である。

非同期コミュニケーションのうち、特に音声コミュニケーションは、メッセージアプリケーションのボイスメッセージとして手軽に利用できる。特に若い世代を中心に積極的な利用の兆しがある。一方で、デジタル機器を利用していなかったり、使い方がわからないなどの理由でデジタル機器の利用が難しい人は、そのような非同期の音声コミュニケーションを手軽に利用することができない状況にある。幼い子供がいる家庭や高齢者施設など、音声のもつ息遣いや抑揚の情報がコミュニケーションに効果的に作用すると思われる場所は多くあるが、前述の理由のため、こ

れらの場所で非同期音声コミュニケーションが十分活用されていない。

この問題を解決するために、著者らは、クラウド上に録音メッセージをアップロードし、そのデータへのアクセスを紙に印刷した QR コードで提供するシステム「空間音声ラベルプリンタ」を考案した。本システムを用いることで、専用アプリケーションの利用や会員登録などの複雑な操作なしに、直感的かつ簡単に非同期音声コミュニケーションを利用することができる。

本稿では、上記システムの実装および、使用シナリオとして、ラベルをボードに複数貼り付ける場合と、物体それぞれに貼り付ける場合について検討した結果を述べる。

2. 関連研究

2.1 非同期コミュニケーション

同期コミュニケーションと非同期コミュニケーションでは、コミュニケーションのありかたが自然と異なってくる。近年では、モバイル端末とインターネットの普及とともに、非同期型のコミュニケーションが急速に広がってきた。その中で非同期コミュニケーションには、同期コミュニケーションとは違った困難さが伴うこともわかってきている。そのためこれまでに、非同期コミュニケーションを支援する研究がいくつかある。時差のある遠隔地の間では、同期コミュニケーションよりも非同期コミュニケーションが適する。辻田らは、時差のある遠隔地の間で、相手の行動を時差の分だけずらして伝達することで、より有効な非同期コミュニケーションを実現する CU-Later を提案した。こ

¹ 東京電機大学

^{a)} yoshiki@cps.im.dendai.ac.jp

^{b)} iwai@cps.im.dendai.ac.jp

れは時差をシステムで補正し、別の時間に同じ場所で行われていた行動の映像を表示することで、非同期コミュニケーションを促進するものである。同期コミュニケーションと非同期コミュニケーションの間の感覚的な差異が大きければ大きいほど、コミュニケーションの困難さも増大すると考えられる。音声コミュニケーションについては、文字のコミュニケーションと比べ、同期コミュニケーションと非同期コミュニケーションのギャップが大きいため、スムーズなコミュニケーションが難しくなってしまう。これに対し中茂らは、音声の聞き手役のアバターを配置し、音声情報から表情を自動生成することで、スムーズな非同期音声コミュニケーションが可能なのではないかと考えた。これらの研究は、非同期コミュニケーションがもつ問題を、補完的な映像を使って解決している。しかし映像を使った方法は、システムを構成するノードの操作を煩雑にするほか、コミュニケーションがノードの設置された狭い範囲に限定され、非同期コミュニケーションのもつ場所的・時間的な自由を十分活用できていない。本研究では映像は採用せず、非同期音声を補完するものとして印刷された紙メディアを使用している。紙メディアの印刷は時間が経っても失われにくく、また可搬性も高い。

2.2 音声コミュニケーション

テキストのメッセージと異なり、音声メッセージには以下のような利点がある。

- (1) キーボードや文字入力画面を操作する必要がない
- (2) テキストより、感情を相手に伝えやすい

内平らは利点1に注目して、看護や介護の現場を支援するコミュニケーションシステム「音声つぶやきシステム」を開発した。[5] これはケアスタッフ間の連携の負担を、スマートフォンとサーバを組み合わせたシステムで支援するものである。この研究は行動型サービスを対象としているが、一般的にもっとも簡易で負担の少ないコミュニケーションは音声コミュニケーションであることを示唆している。利点2に注目した幸ら [6] は、音声メッセージの再生中に聞き手の応答音声を記録する非同期型音声メッセージシステムを提案した。このシステムでは声質や感情などの音声ならではの豊かな情報を保存しながら、自然な非同期コミュニケーションの実現を試みている。本研究でも音声メッセージがテキストメッセージに対してもつ利点に注目した。利点1を利用して、文字入力がおぼつかない幼い子供や、高齢者を含むキーボード入力が苦手な人たちに対し、より簡単で自然なコミュニケーションを提供する。また利点2を利用して、家族や親しい人の間でのスムーズなコミュニケーションを支援する。

2.3 二次元コードと紙メディアの利用

デンソーウェーブと豊田中央研究所が共同で開発した

二次元コードは一般に QR コードと呼ばれ、本稿でも QR コードと表記している。QR コードを利用した研究はいくつかある。たとえば古本らは、視覚障害者に音声データを提供する目的で、符号化方式を工夫した多値二次元コードを提案している。この提案は QR コードを通じて音声を提供するという点で本研究と類似しているが、符号化方式が一般に用いられている QR コードと異なるため、QR コードから音声データを引き出すには専用のデコーダが必要という問題がある。ここ数年、決済目的でディスプレイに表示されて使用されることの多い QR コードだが、以前より紙に印刷されて使用されることも多かった。紙に印刷された二次元コードは、電子情報の持つデータ同士の結びつきを保ちながら、可搬性や実在性、高いアクセシビリティを持ち合わせる特異なメディアとなる。脇田らは電子メディアから紙メディアへの変換 (印刷) の過程で失われるリンク情報などを二次元コードで補完できることに注目し、WWW 上の情報と、紙メディア上の情報との融合を試みた。この研究は本研究と発想を同じくするが、本研究ではさらに、モバイル端末とクラウドサービスがもたらした、紙に印刷された二次元コードの持つアクセシビリティにも注目している。

3. 既存の非同期音声システム

3.1 留守番電話

留守番電話は、相手が電話に出られない場合にメッセージを残すことができるシステムであり、非同期コミュニケーションの一例である。留守番電話は電話線を通じてメッセージが送られるため、インターネット接続やスマートフォンが不要であり、幅広い年代の人々に利用されている。しかし、留守番電話のシステムは古いものであり、メッセージの共有や保存が不便である点が課題となっている。

3.2 電子メール

電子メールはインターネットを利用した文字によるコミュニケーションツールで、音声データの添付も可能である。これにより、非同期で音声コミュニケーションを行うことができる。電子メールのメリットは、全世界で広く利用されており、遠隔地にいる相手とも容易にコミュニケーションを取ることができる点である。しかし、電子メールは主にテキストベースのコミュニケーションツールであるため、音声メッセージを利用することは一般的ではない。また、音声ファイルの容量が大きい場合、メールの送受信に時間がかかることがあり、これがユーザーにとっての障壁となる可能性がある。

3.3 携帯型録音機

携帯型録音機は、音声メッセージを録音し、保存することができる携帯型のデバイスである。携帯型録音機のメ

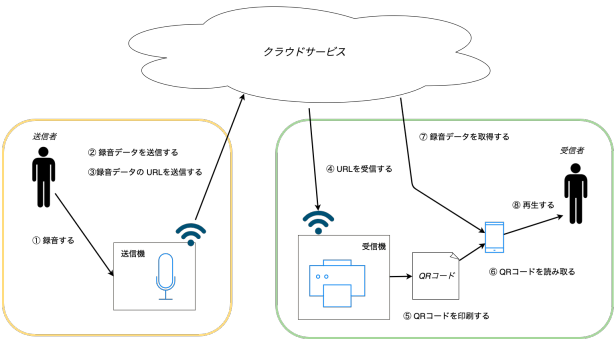


図 1 提案手法の概要図

表 1 ノード一覧

音声送信側	音声受信側
送信機	受信機
	スマートフォン

リットは、専用の機器を用いるため、簡単に高品質な音声録音ができる点である。また、録音したデータはその場で再生できるため、非同期コミュニケーションが容易である。しかし、携帯型録音機で録音したデータを共有する際には、別のデバイスやメディアに移動する必要がある、これが手間となる。

3.4 SNS 上のボイスメッセージ

近年、多くのソーシャルネットワーキングサービス (SNS) では、ボイスメッセージ機能を利用することができる。これにより、ユーザーは簡単に音声メッセージを録音し、共有することができる。また、SNS 上のボイスメッセージはインターネットを通じて共有されるため、遠隔地にいる相手とも簡単にコミュニケーションをとることができる。しかし、SNS のボイスメモ機能を利用するためには、スマートフォンや専用アプリケーションが必要であり、これがモバイル端末に不慣れな人々にとってはハードルとなる可能性がある。

4. 提案手法

4.1 概要

本稿で提案する IoT ノード VoLP は、録音データへのアクセスを印刷した QR コードとして提供することで、すべての人が直感的かつ簡単に、非同期音声コミュニケーションを活用できるようになることを目的とする。この提案手法全体の概要図を図 1 に示す。

4.2 ノード一覧

本手法は大きく送信側と受信側に分けられる。ノード一覧を表 1 に示す

表 2 ハードウェア一覧

送信機	受信機
MCU/MPU	MCU/MPU
マイクロフォン	プリンター

4.3 データの流れ

図 1 中に、本手法におけるデータの経路を囲み数字で示した。経路は時系列順に以下ようになる。

4.3.1 送信側

- ① 音声を録音する
- ② 録音した音声データを送信する
- ③ 録音した音声データの URL を送信する

4.3.2 受信側

- ④ URL を受信する
- ⑤ URL を QR コードに変換し、印刷する
- ⑥ 受信者所有のスマートフォンで QR コードを読み取る
- ⑦ 送信側で録音した音声データを取得する
- ⑧ 音声データを再生する

4.4 通信プロトコル

本手法で通信プロトコルは特に定義されないが、4.3 中の ②録音データの送信、⑦録音データの取得 では HTTP が、③録音データへのリンクの送信、④リンクの受信 では MQTT が使用されることを想定している。

4.5 ハードウェア

送信機受信機それぞれを構成するハードウェアの一覧を、表 2 に示す

5. 試作

4 章で提案した IoT ノード VoLP を、以下の観点で考慮して試作した。本章では試作の詳細を述べる。

- 屋内の電力が供給できる場所に設置することを想定
- 非同期コミュニケーションの場所的な自由度を高めるため、据え置き型のプリンターではなく感熱プリンターを採用

5.1 概要

試作するにあたり、図 1 をもとに具体的なノードの構成や使用するハードウェア、通信インタフェースを決定した。試作の概要図を図 2 に示す。なお、スピーカーの図柄は音声データを、図中の@は音声データの URL を示す。

5.2 ハードウェア

試作物の送信機・受信機それぞれを構成するハードウェアの一覧を、表 3 に示す

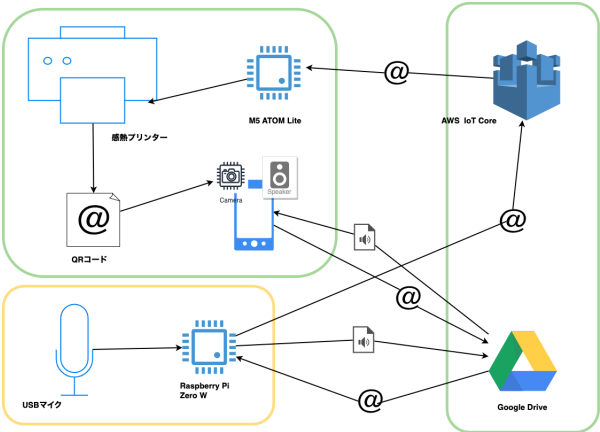


図 2 試作物の概要図

表 3 試作物のハードウェア一覧

名称	用途
マイクロフォン	集音
RaspberryPi Zero W	マイクロフォンの駆動, データの送信
M5 ATOM Lite	プリンターの駆動, データの受信
プリンター	QR コードの印刷

表 4 試作物の通信インタフェース

用途	プロトコル	URI など	ペイロード
録音	USB	—	WAV(F32)
音声アップロード	HTTP	GoogleAPI	WAV(I16)
レスポンス	HTTP	—	JSON1
URL 送信	MQTT	volp/share/link	JSON2
URL 取得	MQTT	volp/share/link	WAV(F32)
音声取得	HTTP	\$SHAREDLINK	—
取得レスポンス	HTTP	—	音声データ

5.3 通信インタフェース

試作にあたり、通信インタフェースを具体的に定義した。それらを表 4 にまとめる。また、表中にある GoogleAPI は

[https://www.googleapis.com
/upload/drive/v3/files](https://www.googleapis.com/upload/drive/v3/files)

であり、アップロードへのレスポンスに含まれる JSON1 の定義は

```
{ "kind": "drive#file", "id": "FILEID",  
  "name": "FILENAME", "mimeType": "MIMETYPE" }
```

JSON2 の定義は

```
{ "link": "$SHAREDLINK" }
```

である。

6. 利用シナリオの検討

本研究の IoT ノードは、音声データへの URL を紙に印刷する。これは 2 関連研究で調査した既存の提案にはない特徴的な性質である。この性質から、本 IoT ノードは以下

のように利用方法が拡張される

(1) 声ラベルの余白に手書きすることで、音声データにメタデータを後から付与できる

例: 声ラベルの余白に、誰に向けたメッセージなのか書き込む

(2) 声ラベルが印刷された紙を、アノテーションしたい対象にテープなどで貼り付けることで、音声の説明する対象や音声の文脈などを、現実世界に表示できる

例: 声ラベルが保持する音声の中で指示語を使い、何を指しているかをラベルを張ることで表示する

(3) 声ラベルをホワイトボードや黒板に多数貼り付けることで、スペースのかぎり、音声データを間接的に一覧できる

7. おわりに

本研究では、非同期コミュニケーションを促進するための IoT ノードを提案した。そして、そのプロトタイプを設計・実装した。最後に、IoT ノードの利用シナリオについて検討した。本 IoT ノードにより、非同期音声コミュニケーションの応用範囲が広がることが期待できる。今後は、この IoT ノードを一般家庭や教育施設、高齢者施設に設置し、長期実証実験を行い評価することで、有効性を検証する予定である。