マルチユーザー・マルチモーダル・近距離通信を実現する デジタルサイネージ向け音声コミュニケーションデバイス

後藤 直1 山本 大介1 高橋 直久1

概要:我々の研究室ではパソコン向け音声インタラクション構築ツールキット,MMDAgen を用いて,双方向音声案内デジタルサイネージシステムの研究を行なっている。ユーザに一方的に情報を表示する一般的なデジタルサイネージとは違い,3Dキャラクタが画面に表示され,3Dキャラクタとユーザ間で音声による双方向の情報のやりとりが可能であるといった特徴がある。公共空間における音声入出力デジタルサイネージは,騒音環境での利用や難聴の方の利用など音声情報の伝わりにくい場合があることや,情報が伝わったとしても,録音などをしていない場合情報を確認する方法がないといった問題点が挙げられる。これらを解決するために近接性保証機能,返答文字列取得・送信機能,掲示板形式のインタラクション履歴表示機能を持つ携帯端末とサイネージの連携方式を提案する。本論文では提案システムに基づいて実装したプロトタイプシステムについて述べる。そして,プロトタイプシステムと従来システムを用いて評価実験を行った。その結果 "従来システムより情報を得やすいと感じる"という5段階のアンケート項目について、静かな環境では4.89,騒音環境では5.00という評価が得られた。この提案システムが従来システムよりも情報提供に効果のあるシステムであることが分かった。

Voice communication device for digital signage realizing multi-user, multimodal, short-range communication.

NAOKI GOTO¹ DAISUKE YAMAMOTO¹ NAOHISA TAKAHASHI¹

1. はじめに

近年,多言語化に対応しやすいことや,地域ごとのイベントの最新情報を伝えられることなどを利点としてデジタルサイネージによる観光案内・道案内が普及しつつある.

我々の研究室ではパソコン向け音声インタラクション構築ツールキット、MMDAgent[2][3]を用いて、双方向音声案内デジタルサイネージシステムの研究を行なっている.ユーザに一方的に情報を表示する一般的なデジタルサイネージとは違い、3Dキャラクタが画面に表示され、ユーザとその3Dキャラクタの間で音声による双方向の情報のやりとりが可能といった特徴がある.

公共空間における双方向音声案内デジタルサイネージ (以下, サイネージ) は一般的に騒音環境や難聴の方の利用 時に情報が伝えにくい場合があることや、利用後に情報の 確認が手元でできないと言った問題点がある.

この問題に対して、我々は LINE[1] のインタフェースに着目した。LINEのトーク画面を図1に示す。LINEのトークでは、吹き出しにテキストを表示させることで実際に会話しているかのようなリアルタイムのチャットが掲示板形式で表示される。また、音声や画像、スタンプと呼ばれるトーク用のイラストを送信することができる。さらに、対話の履歴を、画面を上にスライドすることで簡単に確認することができる。

本研究の目的は、双方向音声案内デジタルサイネージとユーザの携帯端末を連携させ、LINEようなの音声入力とテキスト入出力のインターフェースを持つアプリケーションを開発することである。これにより騒音環境などの音声が伝えずらい場合での情報提供や、携帯端末で利用後の情報の確認が可能になるであろう。しかしながら、上記の目的を実現するためには以下の課題がある。

¹ 名古屋工業大学大学院工学研究科 Graduate School of Engineering, Nagoya Institute of Technology



図1 LINE トーク画面の例

課題 1. 双方向音声案内デジタルサイネージによる案内はサイネージの前 1, 2 メートルにいるユーザに対して行っている。このことから施設案内などの多くはサイネージの設置場所を案内の始点としている。しかし,サイネージ-携帯端末間の通信で,サイネージの想定していない遠隔からの利用により,ユーザの場所を考慮していない間違った案内が行われてしまう。よって,サイネージとユーザの携帯端末を近距離に限定できる通信の仕組みが課題である。

課題 2. 多くの双方向音声案内デジタルサイネージは同時に一人のユーザしか利用しないことが基本である. よって対話処理の入力は複数人からの同時の入力に対応していない. しかし, スマートフォンを用いた音声入力を実現した場合, 複数のユーザが同時に利用することが想定される. そこで, 複数のユーザが同時期に発話し, 発話文が対話処理で輻輳した場合にも対応できる機能の実現が課題となる.

課題 3. 双方向音声案内デジタルサイネージの返答文を携帯端末に表示させるだけでは誰に対しての返答文なのかが分からないといった問題がある。しかし、公共空間に設置される双方向音声案内デジタルサイネージは不特定多数のユーザの利用を前提としているため、ユーザ識別を行なっていないことが多い。よって、返答文が誰のどの発話文に対してのものなのかを識別できる形での通信方式が課題となる。

以上の課題に対し、提案システムでは以下の特徴を持つ、

特徴 1.サイネージと携帯端末の間で BLE 通信による 通信機能を実現する。これにより、サイネージと携帯端末 の間でのテキスト通信が可能になる。また、BLE の受信強度 (RSSI) によりサイネージと携帯端末ユーザとのおおよその距離を把握できるようにする。

特徴 2. 複数のユーザが同時期に発話して輻輳した場合に、サイネージは、携帯端末から届いた発話文を、実行中の対話処理が終わってから受け付ける対話遅延機構を持つ。これにより、発話内容が消失することなく、複数ユーザとサイネージとの対話を可能とする。

特徴 3. サイネージの対話処理からユーザの発話文と返答文を取得し、それに 誰に向けての発話・返答なのかという識別子を付けてサイネージから携帯端末に送信する機能を実現する。これにより、誰の発話文に対し、サイネージが何を返したのかがわかるようにする。

2. 関連研究

近年、デジタルサイネージは、駅、ショッピングモール、ビルの壁面など多くの場所で設置され普及が進んでいる。そういった背景でユーザとの双方向インタラクション可能なデジタルサイネージも多く提案されている。中でもより情報を送信することでユーザの後の行動につなげやすいことで、ユーザの携帯端末と連携することが有効であることが報告されている[7].

宮田ら [8] では、デジタルサイネージ上に複数のコンテンツが表示されている状況において、複数人が同時に異なるコンテンツの詳細の閲覧を可能とするモバイル端末とデジタルサイネージの連携方式を提案している。連携によりユーザーはデジタルサイネージ上に各自のポインタが与えられる。ポインタをモバイル端末の操作により動かし、コンテンツを選択することで、ユーザーのモバイル端末でそのコンテンツの詳細情報を得ることを可能にしている。

また, Kim ら [9] では, デジタルサイネージに映される サムネイルを携帯端末のアプリにも表示し, それを選択す ることによって サイネージの詳細情報を携帯端末で取得す る連携方式を提案している.

これらのように、双方向インタラクション可能なデジタルサイネージと携帯端末の様々な連携方式が提案されている。しかし、今までの研究では画面上にエージェントがいるような双方向音声案内デジタルサイネージと携帯端末の適切な連携方式は明らかにされていない。これに焦点を当てマルチユーザー・マルチモーダル・近距離通信を実現する連携方式の提案に新規性がある。

3. デジタルサイネージ

3.1 MMDAgent

音声インタラクションシステム構築ツールキット MMDAgent とは、音声認識、3 Dモデル制御、3 Dモデル描画、対話制御、音声合成などを高度に統合したシステムである。

MMDAgent は、FST(FiniteStateTransducer) スクリプトに基づいて動作する。FST スクリプトは、音声認識機能などから発生するイベントを入力とし、発話機能や 3D キャラクタの動作機能などへの命令コマンドを出力とする状態遷移機械として記述される。

図2にスクリプトの例を示す. スクリプトの一行は現状態番号, 次状態番号, 遷移条件の入力コマンド, 遷移時の出力コマンドからなっている

例では「こんにちは」を認識すると「こんにちは。」を返してくれるスクリプトになっている.状態1の時に「こんにちは」を認識したというコマンド (RECOG_EVENT_STOP)が入力されると出力なし (<eps>) で状態10に移る.状態10で入力なし(<eps>)で「こんにちは。」の音声合成開始コマンド (SYNTH_START)が出力され,状態11に移る.状態11で音声合成終了コマンド (SYNTH_EVENT_STOP)を受け取ると出力なしで状態1に移るといった動作が記述されている.

MMDAgent の内部メッセージでは FST スクリプトの入出力のコマンド流れるようになっている.



図2 FST スクリプトの例

3.2 しゃべるバス停

MMDAgent を利用した音声案内デジタルサイネージとしてしゃべるバス停 [4] がある。これは地域の登録された、バス停名を音声入力するとそのバス停までの経路を、施設名を音声入力すると、その施設の最寄りのバス停までの経路を音声と地図画面で案内してくれるシステムである。

3.3 本研究にて対象とするデジタルサイネージ

本研究で対象とするデジタルサイネージは上記のしゃべるバス停のような公共空間での使用を想定したものであり、ユーザーの発話を入力とし、発話に対応したコンテンツの音声・画像情報を出力とするものである。

4. 提案システムの概要

4.1 提案システムの構成図

提案システムのインタラクションに関する機能の構成図



図3 しゃべるバス停システム

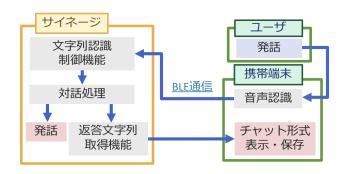


図 4 インタラクションに関する機能の構成図

を図4に示す.

提案システムのアプリケーションを通したインタラク ションの動作順序を説明する.

提案システムではユーザの携帯端末で音声認識行い, BLE 通信を用いて認識した発話文をサイネージに送信する. サイネージでは文字列認識制御機能によって対話制御が可能な状態のときに対話制御部に発話文を認識させる. 対話制御によって返答文が決定したときに返答文字列取得・送信機能によって返答文を取得する. 発話文と返答文に近接無線通信を用いて,携帯端末に送信する.携帯端末にて掲示板形式にして表示を行う.

4.2 提案システムの機能

ここでは提案システムの各機能について説明を行う.特 徴1は近接性保証機能,特徴2は文字列認識制御機能,特 徴3は返答文字列取得・送信機能,掲示板形式のインタラ クション履歴表示機能によって実現する.

4.2.1 近接性保証機能

近接性を考慮した通信を行うために、BLE 通信を用て RSSI を取得し、サイネージの想定のしていない範囲から の利用を制限する機能である.

BLE 通信は数十メートルから数百メートル間での無線通信である。また、サーバクライアントシステムを採用している。サーバ側はクライアントの電波受信強度 (RSSI) を取得することができ、RSSI からサーバとクライアント間のおよその距離を算出することができる。

提案システムではサイネージをサーバとし、ユーザの携帯端末をクライアントとする。RSSIを1秒ごとに取得して閾値よりも低くなったとき、サイネージの通信処理部でその携帯端末との通信を切断する。これによりサイネージの想定のしていない範囲からの利用を制限する。

4.2.2 文字列認識制御機能

携帯端末から送られてきた文字列を対話制御が文字列認 識可能な状態のときに認識させる機能である.

FST スクリプトに基づく音声対話システムはスクリプトの繁雑性を少なくするため、コマンドの出力をする状態などでは対話処理の入力を受け付けていないことがほとんどである。

提案システムは一つのサイネージに複数の携帯端末が接続されるため、複数のユーザからの発話文が同時に入力される場合がある。複数の発話文をそのまま対話処理に入力すると、初めに入力された発話文の認識で状態が遷移し、発話文の認識で遷移しない状態になる。この状態で次の発話文の認識が行われても遷移先が用意されていないので遷移は起こらず、結果的に無視をする動作をする。

これを防ぐためにサイネージの状態番号の取得する機能と、発話文とその入力元携帯端末 ID を一時的に保存するキューを作成する。入力元携帯端末 ID は 4.2.3 章で説明を行う。

サイネージに送られて来た発話文からキューに入れ、発話文の認識による遷移が可能な状態番号になったとき、キューから発話文を取り出して対話処理に入力する. これにより発話文が対話処理で輻輳した場合でも、サイネージが複数の入力に対応できるようになる.

4.2.3 返答文字列取得・送信機能

携帯端末で誰からの発話に対して何を返答したのかを判別可能な表示をするため、各データを取得し、接続されている全携帯端末に送信する機能である.

取得するデータはサイネージ ID, 入力元携帯端末 ID, 発話文, 返答文である. ここで, サイネージ ID, 入力元携帯端末 ID はサーバとクライアントをそれぞれ識別するための識別子を用いる.

発話文と入力元携帯端末 ID は対話処理が終わるまで保

持し、返答文は対話処理の出力から取得する。対話処理により返答文が出力されたとき、各データを取得し一組みにする。それを接続されている全携帯端末に送信する。

4.2.4 掲示板形式のインタラクション履歴表示機能

返答文字列取得・送信機能よりデータ組を携帯端末で受信し、そのデータを図5のような掲示板形式に表示する機能である.

ユーザの携帯端末でテキスト表示することによって音声の伝わりにくい状況でも情報を伝えることが可能になる。また、情報を忘れてしまった場合でも、携帯端末で掲示板形式に履歴が残っていることで、どんなインタラクションを行っていたのかを新しいものから順に確認できるようになる。

自分→サイネージ 名古屋工業大学に行きたい

サイネージ→自分

栄18号、妙見町行きにお乗りください

名大病院にて下りて、東方向に6分ほど歩くと到着します

自分→サイネージ ありがとう

サイネージ→自分

どういたしまして

いつでも話しかけてくださいね

図 5 栄から名古屋工業大学までバスで行こうとした時の対話履歴 を掲示板形式に表示した例

5. 実現方法

ここでは提案システムの機能の実現法を述べる。はじめに近接性保証機能について述べる。次に、インタラクションの流れに沿って、文字列認識制御機能、返答文字列取得・送信機能、掲示板形式のインタラクション履歴表示機能について述べる。

5.1 近接性保証機能

近接性を考慮した通信を行うために、BLE 通信を用いてサイネージとユーザの携帯端末を接続する。BLE 通信ではサーバ側であるセントラルとクライアント側であるペリフェラルとを役割を分けて実装する必要がある。

今回一サーバに対して複数のクライアントを繋げる形を 取るので、サイネージをセントラル、携帯端末をペリフェ ラルとして実装する.

接続後、1秒ごとに RSSI を取得し更新させる。これより、サイネージは各携帯端末のおおよその位置を把握できるようにする。また、サイネージの利用を想定していない範囲の接続を切るために、閾値を決める。今回の実装では-85 とした。閾値よりも低い価の RSSI が検出された場合に切断を行う。

5.2 文字列認識制御機能

複数入力に対応するためにサイネージの状態番号の取得する機能と、発話文とその入力元携帯端末 ID を一時的に保存するキューを作成する.

状態が遷移したときと、発話文を受信してキューに入れたときに状態番号の取得を行う。発話文の認識による遷移が可能な状態番号だったとき、キューからデータを取り出して対話処理に認識コマンドを入力する。取り出した発話文と入力元携帯端末 ID は返答文字列取得・送信機能で必要になるため保持しておく。

5.3 返答文字列取得·送信機能

送信するデータ組の取得を行う。取得するデータはサイネージ ID,入力元携帯端末 ID,発話文,返答文である。

サイネージ ID, 入力元携帯端末 ID は BLE 通信のサーバとクライアントをそれぞれ識別するための識別子, サービス UUID・ペリフェラル UUID を用いる.

サイネージ ID は静的なデータなので取得する必要はない. 入力携帯端末 ID と発話文は文字列認識制御機能より取得する. 返答文は対話処理から音声合成機能への出力コマンド (SYNTH_START) から取得する.

取得したデータ組を一文にして接続されている全携帯端末に送信する.

5.4 掲示板形式のインタラクション履歴表示機能

返答文字列取得・送信機能により得られるデータ組から 図5に示したような掲示板形式の表示を行う.

サイネージからデータ組を貰うと、入力元携帯端末 ID がその携帯の ID だった場合は"自分"と表記してその下に発話文を表示する。他人の ID だった場合はその ID を表示する。また、サイネージ ID があらかじめ登録してあるものについてはそのサイネージの名前を表記し、その下に返答文字列を表示するようにする。登録していない ID についてはその ID を表示する。

新たなデータ組を貰うたびにサイネージへの入力と出力を下に順次追加を行っていくことで下に行くほど最新の、上に遡るほど過去のインタラクションの履歴を確認できるようにする.

6. プロトタイプシステム

5章の実現方法にしたがってプロトタイプシステムを作成した。プロトタイプシステムの画面を図6に示す。

接続していないとき、左下のボタンを押すと接続を開始し、接続されているとき、左下のボタンを押すと切断処理を行う.

右下のボタンの押下中に音声認識を行い,離すと音声認識が終了する。音声認識処理が終わり次第,接続されているサイネージに発話文を送るようにした。

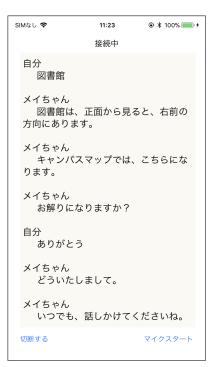


図 6 アプリケーション画面

7. 評価実験

7.1 実験の目的

実験の目的は二つある。一つ目は双方向音声案内デジタルサイネージはインタラクションシステムであることから、反応時間が遅いとユーザビリティに影響を与える。反応時間を測定し、その変化がユーザビリティに影響を与えているかを検証することである。二つ目は、提案システムが騒音環境においてユーザーの情報取得が容易になるのかどうかを検証することである。

以上から,反応時間測定実験を行うことで反応時間の変化を明らかにし,また,評価実験にて使用した時の反応時間への評価と情報取得の容易さについての評価を行う.

7.2 実験方法

7.2.1 反応時間計測実験

従来システムと提案システムの反応時間を測定する. 計測にはボイスレコーダーにて録音し計測する. ユーザーの発話終了からサイネージの発話開始までの時間を反応時間とする.

7.2.2 従来システムと提案システム

従来システムは、今回実装した機能を持たない MMDAgent を用いた双方向音声案内デジタルサイネージのことを指す. 入力はサイネージに接続したマイクへの発話である. 提案システムはプロトタイプシステムを使用する. 入力はユーザの携帯端末への発話である.

7.2.3 実機を用いたユーザによる評価実験

聴覚の健常な被験者(本学学生9名)が3人1組となり、従来システムと提案システムを静かな環境と騒音環境で使用して比較評価を行ってもらう。また従来システムにない機能については、提案システムのみで評価を行う。被験者にシステムの使い方を一通り説明した後、指定した対話を行い、以下のアンケートに5段階評価と自由記述で回答してもらう。1を「そう思わない」、5を「そう思う」とし、5に近づくほど評価が高くなる。

ここでの静かな環境と騒音環境は以下の環境とする. 静かな環境:50~60dB, 騒音環境:70~75dB

実験環境は、デジタルサイネージを置き、その正面から 1m 離れた位置から利用してもらう。従来システムのマイクを利用する際は発話時にマイクの前に移動し、発話してもらう。騒音環境は被験者の後ろからスピーカーで都市の騒音 [5] を流し、擬似騒音環境において実験を行った。騒音計には Decibel X dBA デシベルテスター [6] を用いて計測した。

7.3 アンケート

アンケートは5段階評価と自由コメントを用意した.アンケートの項目は表1の通りである.番号1~3と番号4の音声・映像については、従来システムと提案システムに、その他の項目については提案システムのみ、それぞれの項目の評価をしてもらった.評価は5段階で行う.評価基準は「1:あてはまらない、2:ややあてはまらない、3:どちらともいえない、4:ややあてはまる、5:あてはまる」とした.アンケート項目を以下に示す.

表 1 アンケート項目

後 1 / 2 / 1 / 3		
番号	質問内容	
1	応答時間は早い	
2	話しかけやすい	
3	対話を自然に行える	
4	情報取得に役に立った	
4.1	音声	
4.2	映像	
4.3	返答文表示	
4.4	インタラクション履歴	
5	従来システムより情報が得やすい	

7.4 実験結果と考察

7.4.1 結果

反応時間測定の結果を表 2 に示す.

表 2 反応時間測定結果

TO A DOCUMENT OF THE PARTY OF T		
システム	時間 [s]	
従来システム	0.524	
提案システム	0.990	

被験者9名にアンケートで5段階評価を行った結果の平均を図7,図8に示す。

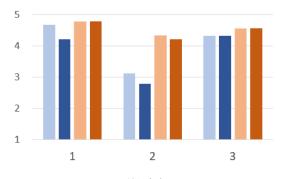


図 7 質問内容 1~3

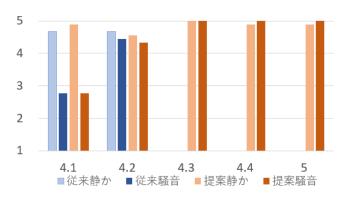


図 8 質問内容 4~5

7.4.2 考察

反応時間について

表2の反応時間より、従来システムの反応時間より0.476[s] 遅くなっていることが分かる。しかし、図7の質問番号1 "応答時間は早い"の項目ではどちらの環境においても提案システムの方が評価が高くなっている。このことから平均0.476[s] の反応時間の低下があったが提案システムの出力に対しユーザは遅いとは感じないことが分かり、このプロトタイプシステムでは反応時間の低下はユーザビリティに影響は与えていないと言える。

使用方法の変化について

提案システムでは使用状況や使用方法が変わったが、図7の質問番号2、3から提案システムとの対話において従来システムより話しかけやすく感じ、また対話を自然に行えることがわかった。

音声・映像について

音声の結果から騒音環境では音声による情報提供は40%ほど情報提供能力が落ちていることが明らかになった.サイネージの映像の結果はどちらも提案システムの方が評価を落とした.「見る画面がサイネージと携帯端末の二つになることでサイネージの映像の情報を見逃した」と言う

コメントもあった. これは携帯端末を注視しているときに サイネージに地図画像が表示されても見逃してしまうこと などが起ったことが原因である. 携帯端末でもサイネージ の画像を表示するなどの改善が考えられる.

返答文表示・インタラクション履歴について

どちらの機能についても高い評価を得ることができた. 一方で,「自分の発話に対応するまでの間,自分の発話した文字列が表示されなかったため,正しく認識してデータが送られているのかどうかが不安になる時があった」というコメントがあった. サイネージが送信データに対応するまでの間,別のテキストボックスに認識した文字列を表示し,データが送信できていることを分かりやすくすることで不安をなくすといった改善が考えられる.

"従来システムより情報を得やすいと感じる"について

この項目はシステムを使用してもらい総合的に情報が得やすくなっているかどうかについて質問した。静かな環境では 4.89, 騒音環境では 5.00 という評価を得られた。この提案システムが騒音環境によらず情報提供の向上に役立つことが分かった。

8. おわりに

本研究の目的は、公共空間での利用を想定し、携帯端末と双方向音声案内デジタルサイネージの近距離無線通信による連携によって、近接性を考慮しつつ、音声情報が伝わりにくい状況でもテキストにより情報を伝え、また、インタラクションの履歴を表示することで情報の確認がしやすくなるシステムを提案し、その実現法について述べた。また、提案システムを実装し、実装したプロトタイプシステムの評価実験を行った。その結果"従来システムより情報を得やすいと感じる"という5段階のアンケート項目について、静かな環境では4.89、騒音環境では5.00という評価が得られ、この提案システムが従来システムよりも情報提供に効果のあるシステムであると言える。

今後の課題として、携帯端末を注視しているとデジタルサイネージの画像を見逃してしまう問題に対し、携帯端末にもその画像を表示させる機能の実装。また、サイネージにデータが送られているのかが不安になる問題に対し、サイネージが送信データに対応するまでの間、別のテキストボックスに認識した文字列を表示し、データが送信できていることを分かりやすくすることで不安をなくすといった改善が考えられる。

謝辞 本研究は JSPS 科研費 25700009, および, 総務省 SCOPE の助成を受けたものです.

参考文献

- [1] LINE 入手先 (https://line.me/ja/)
- Akinobu Lee, Keiichiro Oura, Keiichi Tokuda, MMDAgent - A fully open-source toolkit for voice interaction

- systems, Proceedings of the ICASSP 2013, pp. 8382-8385, 2013.5.
- [3] MMDAgent, 入手先 (http://www.mmdagent.jp/) (参照 2018-01-29).
- [4] 山本 大介, 竹内 健祐, 林 晃大, 福安 浩明, 加藤 りか, 田中 亮佑, 高橋 直久, , オープンデータに基づくしゃべるバス 路線案内システムの実現, DEIM forum 公演論文集 E6-5 , 2018
- [5] 環境音 都会の雑踏 入手先 〈https://www.youtube.com/watch?v=n3tRTyH601U〉 (参照 2018-01-29)
- [6] Decibel X dBA デシベルテスター 入手先 (https://itunes.apple.com/jp/app/id448155923?mt=8) (参照 2018-01-29)
- [7] She, J. et al.: Convergence of Interactive Displays with Smart Mobile Devices for Effective Advertising: A Survey, ACM Trans. Multimedia Comput. Commun. Appl., Vol. 10, No. 2, pp. 17:1-17:16, DOI:10.1145/2557450 (2014).
- [8] 宮田 章裕, 瀬古 俊一, 青木 良輔, 橋本 遼, 石田 達郎, 伊勢崎 隆司, 渡辺 昌洋, 井原 雅行, デジタルサイネージとモバイル端末を連携させた複数人同時閲覧のための情報提示システム, 情報処理学会論文誌 Vol. 56 No. 1 p. 106-117 , 2015.
- [9] Gisung Kim, Soobin Lee, Joonyoung Park, Sungkwan Jung, Sangsik Kim, Yong-chul Shin, Thumbnail Selection: Delivering Digital SignageContents to Mobile Phone, Proceedings of the 15th International Conference on Advanced Communications Technology,