

LINC: 音声及びジェスチャー認識を用いた 住空間情報没入型表示システム

内海和貴^{†1} 金子将之^{†1} 有馬一貴^{†1} 坂本優大^{†1} 今泉優樹^{†1}
三浦翔^{†2} 杉村博^{†1} 一色正男^{†1}

本論文では、住空間情報没入型表示システム LINC (Life-information Natural Connect) を提案する。LINC は、音声やジェスチャー等の自然なアクションを入力として、生活空間に情報を投影し没入させることを可能にするシステムである。本システムはユーザーのジェスチャーを認識するカメラと音声を認識するマイク、情報を投影するためのプロジェクターで構成する。実装したシステムの有用性の確認のため実験を行った結果について記す。

LINC: Living Space Information Immersive Display System Controlled by Natural Actions

KAZUKI UTSUMI^{†1} MASAYUKI KANEKO^{†1} KAZUKI ARIMA^{†1}
MASAHIRO SAKAMOTO^{†1} YUKI IMAIZUMI^{†1}
SHO MIURA^{†2} HIROSHI SUGIMURA^{†1} MASAO ISSIKI^{†1}

In this paper, we propose a system that is living space information immersive display system, is named LINC (Life-information Natural Connect). LINC is controlled by natural actions such as gestures and voice and also displays information on living space for immersive environments. The system is constructed by combination of three input devices which are a microphone for the sound recognition, a camera for the gesture recognition, and a projector to display information. We in this paper describe the entire process, the implementation and the evaluation of a prototype system.

1. はじめに

日常生活のあらゆる場面において、テレビ、スマートフォン、パソコン等の多くのデバイスを利用し、ニュースや天気等の多様な情報を取得している。各機器は専用のリモコンやキーボードで行われるため、家庭内にある複数機器の操作は複雑になる。ユーザーの複雑操作を軽減するために、音声認識やジェスチャー認識を用いた自然な操作を行う PUI (Perceptual User Interface) の研究が行われている。

PUI に関する論文の例として、w3voice では対話やボイスチェンジャーとしての機能を実装している¹⁾。文献(2)では、手の動きに応じてポインターが移動する大画面環境におけるポインティングハンドジェスチャーによるインタラクション手法を提案している。しかし、これらの研究では音声コントロールのみ、ハンドジェスチャーのみといった独立したシステムになっている。

音声、ジェスチャーの双方ともに連携して利用するシステムとして東芝の「てぶらナビ」がある³⁾。しかし、パソコン用アプリケーションとして開発しているため、場所や端末に囚われない入力手法であるにも関わらず、出力先がパソコンモニターとなり表示出来る場所と大きさに制限が

ある。

本論文は、音声とジェスチャーを入力として、情報を生活空間に表示するためのプロジェクターによる投影を出力とする住空間情報没入型表示システム LINC (Life - information Natural Connect) を開発し、実験ではアンケートを利用して有用性を確認した。

2. システム構成

2.1 システム概要

LINC は、音声操作によって部屋の空いている壁や机などに情報を表示し、表示した情報を拡大・縮小・移動するためにジェスチャー操作を利用するシステムである。LINC は図 1 のように、取得した音声情報を文字情報に変換する音声認識モジュール、文字情報から表示する情報を選択するスイッチコントロールモジュール、画像情報から腕や手の動きを認識して状態情報や座標情報に変換するジェスチャー認識モジュール、表示した情報を変化させるパネルコントロールモジュール、ユーザーの近くに適切な大きさで情報を表示する出力コントロールモジュール、Network 上からサービス情報を取得するサービスコントロールモジュール、そしてそれらを繋げる Application 部で構成する。

今回、LINC の有用性について検証するために実機の製作をした。実機は図 2 のように、音声認識用のマイク、ジェスチャー認識用のカメラ、情報出力用のプロジェクター、マイクやカメラ、プロジェクターの制御や Network 上からの情報を取得するパソコンから構成されている。①ジェス

^{†1} 神奈川工科大学 創造工学部 ホームエレクトロニクス開発学科
Kanagawa Institute of Technology
Faculty of Creative Engineering
Department of Home Electronics Development

^{†2} 工学教育研究推進機構 スマートハウス研究センター
Smart House Research Center

チャー入力と音声入力によってユーザーの希望する情報・動きを認識する。②サービスを WAN から取得する。③プロジェクターでユーザーの目の前に出力する。



図 1 ソフトウェアスタック図
 Figure 1 Software stack chart.

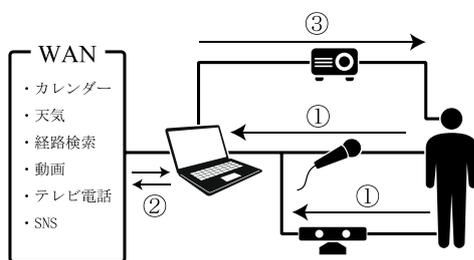


図 2 システムのデータフロー
 Figure 2 Data flow of the system.

2.2 利用イメージ

提案システムの具体的利用イメージは、図 3 に示すように、料理をしながらテレビ電話をしようとするときに、マイクに呼びかけるだけでユーザーの目の前でテレビ電話を行うことができる。音声とジェスチャー操作なので手を洗ったり拭いたりする必要もなく、プロジェクターで前に表示されるためテレビ電話のできる機器の前に移動する必要もなく、機器の防水性考慮も必要なく、料理中の清潔性も保たれる。料理のレシピを参照する場合にはジェスチャーによって機器を触らずに表示画面の拡大も出来る。

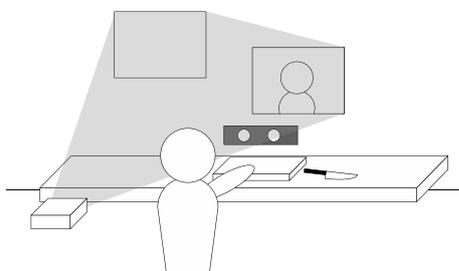


図 3 利用イメージ
 Figure 3 An example of use.

2.3 要求仕様

上述した利用イメージを LINC で実現するには以下のシステム要求 3 つを満たす必要がある。

- ①音声認識によって、ユーザーの音声情報を解析する
- ②音声認識の結果から一致した情報を表示する
- ③表示した情報をジェスチャーで移動・拡大縮小することができる

要求①では、ユーザーの音声情報を解析し、文字情報に変換する。

要求②では、要求①で変換した文字情報の中から、ユーザーに提供する情報を判断し表示する。

要求③では、要求②で表示した情報を操作するために、ユーザーの手の動きをジェスチャー認識で読み取り、情報の移動や拡大縮小を行えるようにする。

以下では、3 つの要求を満たすためのモジュールについて説明する。

2.4 モジュール

(1) 音声認識

音声認識モジュールは、マイクが取得したユーザーの音声情報を文字情報に変換する。主語や動詞をキーワードごとに発した音声情報だけでなく、日常生活で使うような文法に当てはめた音声情報も認識し、文字情報に変換する。文字情報に変換した後は、スイッチコントロールにデータを送る。

(2) スイッチコントロール

スイッチコントロールモジュールは、取得した文字情報を基に Application を経由して、表示命令をサービスコントロールに送る。表示命令はキーワードと関連付けてあり、取得文字情報との一致によって表示命令を選択する。選択した情報はパネルという単位で表示する。

(3) ジェスチャー認識

ジェスチャー認識モジュールは、カメラを用いて手を振る・握る等の自然に行える動きを認識して、ジェスチャーや手の座標情報を取得する。

(4) パネルコントロール

パネルコントロールモジュールは、ジェスチャー認識によって取得された情報を利用して、表示しているパネルを操作する。

(5) サービスコントロール

サービスコントロールモジュールは、表示する情報を取得するために、Network 上からサービス情報を取得する。

取得したサービス情報は、出力コントロールに送る。

(6) 出力コントロール

出力コントロールモジュールは、ユーザーの近くの壁や机、天井等のスペースに、適切な大きさの情報パネルを表示する。

2.5 処理シーケンス

ユーザーが発した音声から、情報を表示するまでのモジュールの流れを図 4 に記す。音声を音声認識モジュールが受け取り、文字情報に変換する。文字情報はスイッチコントロールが受け取り、指定したキーワードがあるか検索する。マッチしない場合は、そこでフローを終了する。サービスコントロールがマッチしたキーワード情報を受け取り、

Web から情報を引き出す。引き出したサービス情報は、出力コントロールを経由してユーザーに出力する。

図 5 にジェスチャーに関するモジュールのシーケンス図を記す。ユーザーの身振り手振りを入力としてジェスチャー認識を行うことで、手の座標情報やジェスチャーの種類を認識する。認識した情報をパネルコントロールへ引き渡したときに、操作の対象となるパネルが存在するかを確認して、存在しない場合は処理を終了し、存在する場合はパネル制御によって移動と拡大縮小を行う。

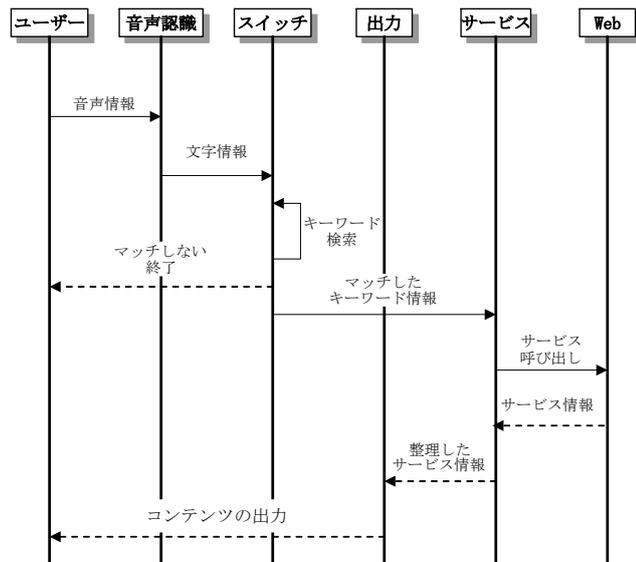


図 4 情報表示の流れ
 Figure 4 Data flow of the voice recognition.

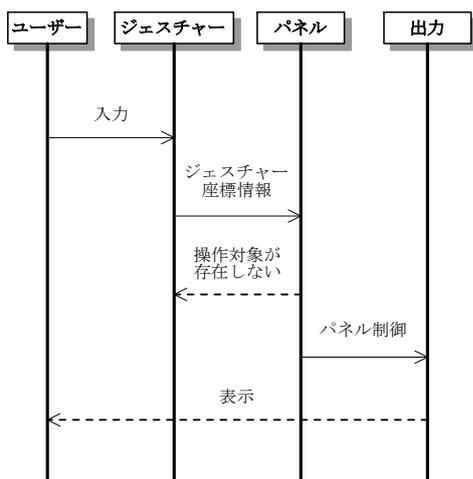


図 5 パネル操作の流れ
 Figure 5 Data flow of the hand gesture manipulation.

3. 実装

3.1 音声認識

本研究では、音声認識モジュールに Google Inc. が提供し

ている JavaScript API の Web Speech API⁴⁾ を利用する。マイクが取得した音声情報は、Web Speech API を利用してサーバーに送信し、文字情報として返す。取得した文字情報は、スイッチコントロールモジュールに送る。

3.2 スイッチコントロール

取得した文字情報に対して、下記表 1 のキーワードを部分一致検索することで、対応する表示命令をサービスコントロールモジュールに送る。

表 1 表示命令の対応表

Table 1 Table of keywords and commands.

キーワード			表示命令
カレンダー	予定	用事	Google Calendar
ツイッター	つぶやき		Twitter
地図	マップ	経路	経路(Google Maps)
天気	Weather	天候	天気(Google Maps)
動画	YouTube		YouTube
通話	チャット	電話	テレビ電話 (WebRTC)

3.3 ジェスチャー認識

本研究では、手の座標及びジェスチャーを検出するためのセンサーデバイスとして Kinect⁵⁾ を使用した。手の座標として、ユーザーのスケルトンデータから手の XYZ 座標を検出する。手の状態として Grab, Release の 2 つのジェスチャーを認識する。手の座標と状態は、Node.js で作成したローカルサーバーを経由して、ブラウザに送信する。

3.4 パネルコントロール

取得した手の座標はカーソルに反映させ、手の状態はパネルの操作に反映させる。

Grab を検出した場合は、検出時点で手のカーソルに重なっているパネルとカーソルの座標を関連付けることで、手を握っている間はカーソルに対してパネルが追従する。手を握っている間は、手の Z 座標の変化に応じてパネルが拡大縮小するため、手を手前に引けば拡大、奥に押せば縮小させることができる。

Release を検出した場合は、カーソルとパネルとの関連付けを削除することで追従を終了し情報を配置する。

3.5 サービスコントロール

本研究では、カレンダー機能として Google Calendar、地図情報として Google Maps、リアルタイム情報収集として Twitter、テレビ通話として WebRTC、動画コンテンツとして YouTube のサービスを利用する。

4. アンケート結果

図 6 のように LINC⁶⁾ を実装し、評価のためにアンケートを実施した。10~50 代の 54 名 (男性: 22 名, 女性: 32 名) を対象に以下の流れで行った。

1. 研究概要を簡潔に口頭説明
2. LINC の特徴や機能を口頭説明
3. アンケート回答者に LINC を実体験
4. アンケート用紙記入

アンケート回答者の年代内訳を図 7 に示す。主に大学生以下にアンケートを実施した。

男性、女性、全体のアンケート結果を図 8 ~ 10 に示す。男性は 96%，女性は 100%，全体は 98%が「とても面白い」「面白い」という回答となった。



図 6 実装した LINC

Figure 6 A prototype of the system.

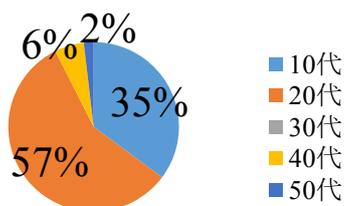


図 7 アンケート回答者年代内訳

Figure 7 Result of each age.

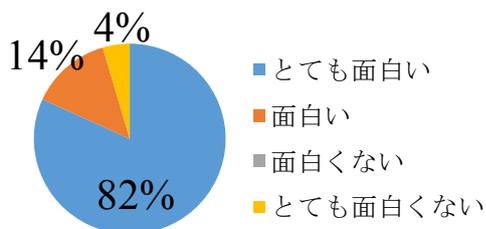


図 8 男性のアンケート結果

Figure 8 Result of men.

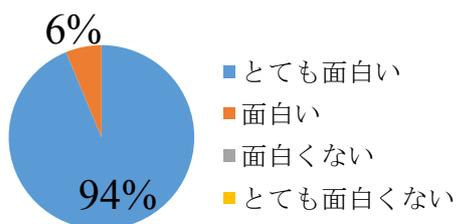


図 9 女性のアンケート結果

Figure 9 Result of women.

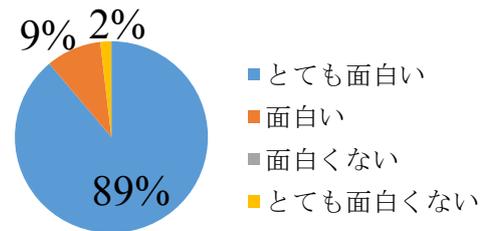


図 10 全体のアンケート結果

Figure 10 The distribution of the interest.

5. まとめ

本研究では、家庭内における自然な情報取得方法として、没入型情報表示システム LINC を提案した。従来の情報取得方法であるテレビやスマートフォン、パソコンと比べ、音声認識やジェスチャー認識を利用した自然な入力手法を用いているため、ユーザーが特別な操作方法を覚える必要がない。

LINC の有用性を確かめるため、音声認識、スイッチコントロール、ジェスチャー認識、パネルコントロール、サービスコントロールの 5 つのモジュールを実装し、アンケートを行った。

アンケートを通じて、LINC がユーザーの興味を引くシステムであることがわかった。また、体験した人は、音声入力によって情報を呼び出し、目の前に手をかざして情報を掴んで操作する直感的なシステムを面白く良いと感じる結果となった。

今後の課題としては、投影する場所を指定するための出力コントロールモジュールの作成と提供するサービスの拡張等が挙げられる。さらに、アンケートによって介護利用という新たな可能性があったため、多方面分野にわたる応用を研究してゆくつもりである。

参考文献

- 1) 西村竜一, 三宅純平, 河原英記, 入野俊夫: 音声入力・認識機能を有する Web システム w3voice の開発と運用, 2007 - SLP - 68 - 3, pp.13-18 (2007)
- 2) 中村卓, 高橋伸, 田中二郎: 大画面環境におけるハンドジェスチャーの選択手法 : ダブルクロッシングの提案と他の選択手法との比較(ヒューマンコンピュータインタラクション)電子情報通信学会論文誌. D, 情報・システム J96-D(4), pp.978-988 (2013)
- 3) 池司, 中洲俊信, 岡田隆三: 自然な手振りによるハンドジェスチャー ユーザーインタフェース, 東芝レビュー, Vol.67, No.6 (2012)
- 4) Web Speech API Specification
<https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html>
- 5) Kinect for Windows | Voice, Movement & Gesture Recognition Technology
<http://www.microsoft.com/en-us/kinectforwindows/>
- 6) Smart-ilab
<http://smart-ilab.sakura.ne.jp/>