

非言語情報による計算機と人間の長時間音声コミュニケーション

副田 俊介 ナイジェルワード
東京大学 工学部 機械情報工学科

1 はじめに

人間が作業しながら使用する通信手段として音声を利用する場合、長時間 (200ms 程度) で対応することが特に重要となる場合がある。しかし現在の音声認識は単語や発話の終了から処理を開始するうえに、その処理にも数秒程度かかるため長時間での処理には向いていない。

一方で人間同士のコミュニケーションでは韻律によって相手の感情を推定することが確認されている。韻律の処理による情報の抽出は発話の途中でも実時間で処理することができるのでマン・マシンインタフェースの構築にも有効であると考えられる。

本研究では「人間が作業をしながらその作業と協調して動作するマシンへ指示を送る」のための「長時間の音声によるコミュニケーション」を調べることが目的である。

2 評価手法

2.1 タスクの説明

本研究では長時間での複数の操作が必要なタスクドメインが望ましい。アクションゲームがタスクドメインとして適している。そこで月面に宇宙船を着陸させることを目的としたゲームをタスクとして選んだ。

2.2 評価システム

評価のためのシステムは大きく 3 つのモジュールに分かれている (図 1)。

ゲームモジュール 宇宙船を月面に着陸することが目的のゲーム (xlander) で本来は全ての操作をキーボードで行う。本システムでは前後左右はユーザが

Hard-real-time communication between human and computer using non-verbal information, Shunsuke Soeda and Nigel Ward, Department of Mechano-Informatics, School of Engineering, University of Tokyo

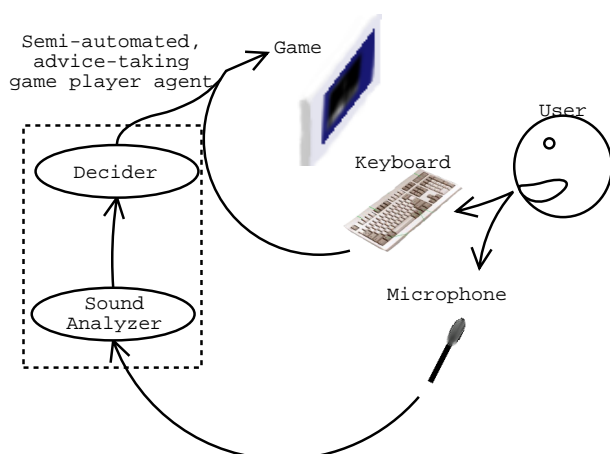


図 1: システムの概要

キーボードから操作し、上下方向についてはコンピュータに音声で命令する形になっている。

音声分析モジュール ユーザの発した音声の基本周波数とエネルギーを計算し、主に基本周波数の変化の割合に注目してユーザの伝えたかったことを推定する。

判断モジュール 有限状態オートマトンになっており、現在の自分の状態に応じて宇宙船の上下方向をコントロールする。宇宙船の状態に関する情報は一切使わず、ユーザから与えられた命令のみを元に判断する。

3 共同作業における非言語情報の分析

3.1 コーパスの収集

人間同士が音声をどのように解釈しているかを調べるために被験者 2 人を用いての予備実験を行った。この実験では被験者の片方 (ユーザ役) がゲームの画面を見ながら宇宙船の前後左右方向を操作し、上下方向に関してはもう一人の被験者 (エージェント役) に音声で指示を出した。エージェント役の被験者はユーザ役の被験者からの音声のみを元に判断し宇宙船の上下方向の操作を行った。

表 1: 韻律を元に指示の内容を推定するルール

ルール略称	ルールの例	ルールの意味	条件	判断モジュールの対応
「悲鳴」ルール	「あがれ~!」	せっぱつまった声	エネルギーの変化が大 ($mPower > 1.0$)	このルールにあてはまる場合には墜落を回避するために全力で上昇しようとする。
「少し行動」ルール	「ちょっと上がって」	弱い指示	基本周波数が急激に下がる ($mF_0 < -0.6$)	このルールにあてはまる場合には状態が遷移する。例えば「上昇中」なら「下降中」に状態が遷移する。
「行動」ルール	「あがって」	普通の指示	基本周波数が上がる ($mF_0 > 0.3$)	このルールにあてはまる場合には現在とは違う状態に以降するようになる。
「現状維持」ルール	「いいよいいよ」	現在の状態を肯定する発話	それ以外で発話がある時	このルールにあてはまる場合には現在の状態の持続時間を延長する。

ここで、ユーザ役の伝えたい内容は指示の声に含まれる韻律に反映されていると仮定し2つの実験を行った。1回目はユーザ役の被験者からエージェント役の被験者への指示をそのまま送り、エージェント役の被験者の対応を記録した。2回目はLPCによって韻律のみを抽出して送り、エージェント役の被験者の対応を記録した。その結果韻律のみでもエージェント役の被験者はほぼ正しい判断(30回中25回)を下すことができることが確認された。

4 コーパスの分析

4.1 分析方法

主として韻律のみの実験のコーパスを分析した。指示に含まれる韻律とそれに対するエージェント役の被験者の反応の関係を調べた。

4.2 分析結果

コーパスを分析した結果、表1のように4つのルールを提案する。

なお $mPower$ はエネルギーの変化の割合、 $mF_0[Hz/ms]$ は基本周波数の変化の割合を示す。

5 関連研究

韻律のみを用いて実時間音声コミュニケーションシステムを構築しようとしている研究はない。

6 おわりに

韻律のみを用いて人間と同じ程度の時間で音声に対応できるシステムを構築し、マン・マシンインタフェースへの応用の可能性を示した。今回は *xlander* というゲームに関して有効な判断モジュールとルールのセットを用いて評価したが、表1のルールのセットがどこまで普遍的なものであるかを調べたい。

チューリングテスト(人間らしくふるまうことができるか)とアンケートによる調査(どれだけ満足できたか)を行う予定だが、どちらもまだ行っていない。

参考文献

- [1] Tsukahara, Wataru. An Algorithm for Choosing Japanese Acknowledgments Using Prosodic Cues and Context, In *Proc. 5th International Conference on Spoken Language Processing*, pp.691-694, 1998.