

5 S-4 ワードスポットティング音声認識のメニュー選択への応用

吉田 慎介

NTT情報通信処理研究所

北井 幹雄

NTTヒューマンインタフェース研究所

1. まえがき:

近年、機械-人間の会話実現に向けた会話分析が行なわれ⁽¹⁾、著者らも電話取り次ぎ実験に基いて会話のパターン分析、発話内容理解の検討を行ってきた。本報告では、機械-人間の会話において、利用者が、システムが次に実行すべき処理を指示するフェーズ(以後、メニュー選択と称する)に注目し、単語に余剰語(抽出すべき単語の前後に発声される単語)が付いて発声された場合の認識結果を分析し、音声会話におけるユーザの操作性向上のための方法を提案する。

2. メニュー選択:

メニュー選択とは、ユーザがシステム操作中に、システムが次に実行すべき処理を決定するための指示を行うフェーズである。ユーザがメニューを発声する場合には、メニューの前後に余剰語が自然に付随する場合が多い。⁽²⁾

余剰語が出現する状況でメニュー単語を抽出する場合は、抽出すべき単語と余剰語との区切りが不明確となるので、認識性能の劣化(候補単語の認識距離の増大、等)が起こり、これを補うための処理が増加し、音声会話としてのユーザの操作性低下が予想される。

そこで、余剰語の付加の有無による認識結果の変化を実験システムを作成して調査した。

3. 実験方法:

図1の実験システムを用い、余剰語の有無、余剰語の付与位置を変えて、被験者にメニュー単語を発声して貰い、その時の認識結果を分析した。図1の実験システムは被験者の発声する音声を認識する認識装置(同時認識語彙数48単語の連続

DPマッチングによる特定話者用ワードスポットティング音声認識装置を使用)、被験者に発声を促すためのガイダンスを送出する再生装置、電話網からの着信を検出し被験者と認識装置及び再生装置を接続するスイッチングユニット(SWU)、さらに、前記3装置を制御して被験者とシステムの音声会話を実現するワークステーションから構成される。

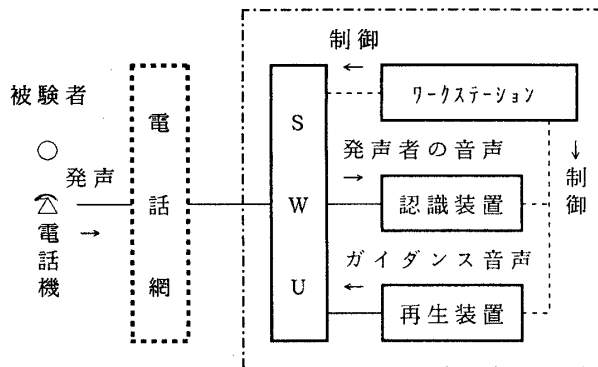


図1 実験システムの構成

被験者が電話機から発信すると、システムは「サービス名をどうぞ」のガイダンスを送出する。ガイダンス終了後に被験者が発声すると音声認識装置が認識処理を行い、事前に登録した辞書単語との認識距離から認識結果(候補単語とその確からしさ)を評価する(図2)。

確からしさは3レベルに分類し、認識尺度は、筆者が十数回程度、余剰語を付けずに単語発声し

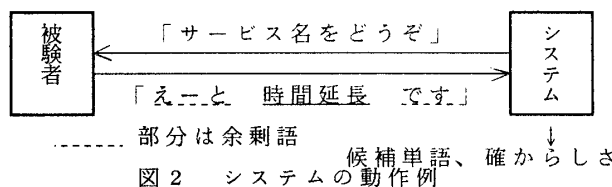


図2 システムの動作例

An application of word-spotting-type voice recognition for menu selections.

Shinsuke YOSHIDA*¹, Mikio KITAI*²

*1 NTT Communications and Information Processing Laboratories

*2 NTT Human Interface Laboratories

た結果を元に経験的に決定したものを使用した。

4. 試験条件：

実用に近い環境で測定するため、電話網を経由し、マシン室内で実験を行った。

また、メニュー単語としては、6種類のサービス名、メニューの前段に付与する付加語（前置）及びメニューに後続する付加語（後置）としては一般的に良く使われる2単語を用いた（表1）。

表1 試験条件

試験環境		マシン室内（ファン騒音下）	
発声者		男性2名	
使用単語	メニュー	「時間延長」、等の6種類	
	余剰語	前置	「えーと」、「あー」
		後置	「お願いします」、「です」
発声条件		電話網を介して電話機で発声	

5. 試験結果：

被験者に以下の4個のパタンの発声をして貰い、その時の認識結果（候補単語とその確からしさ）を調査した（表2）。

- ① 余剰語を付与しないで発声（サービス名のみ）
- ② サービス名の前に余剰語を付与して発声
- ③ サービス名の後に余剰語を付与して発声
- ④ サービス名の前後に余剰語を付与して発声

一方、確からしさについては、
 （レベル0）候補単語が正しい可能性が低い
 （レベル1）候補単語が正しい可能性が中程度
 （レベル2）候補単語が正しい可能性が高い
 に分類した。

尚、今回の試験ではレベル0と評価された候補は無かった。また、候補単語が正解でなく、且つ、レベル2と評価され誤操作となる事例も無かった。

6. 考察：

表3から、サービス名の前に余剰語を付与した場合はと余剰語を付与しなかった場合とでは確からしさの分布には殆ど差が無いのに対して、サービス名の後に余剰語を付与した場合及び前後に付与した場合と余剰語を付与しなかった場合とでは確からしさの分布に大きな違いが存在する。

これは、サービス名の前に来る「えーと」、

「あー」、等の付加語は、サービス名と分離して発声され易いのに対し、後続する、「お願いします」、「です」、等の付加語はサービス名と連続して発声され易いためと考えられる。

ところで、レベル1と評価される候補が多くなることは、その後に候補単語をユーザに確認する動作が必要となるため、ユーザの操作回数が増加し、操作性の低下につながる。一方、サービス名の後に余剰語を付与した場合でも候補単語の正解率は高いことから、余剰語が付加した場合でも付加しない場合程度の確からしさを得る様な評価関数を用いることにより、余剰語がサービス名の前に出現する場合と同程度の操作性を確保できる可能性があることを示している。

表2 認識結果

		余剰語を付与しない	余剰語を付与		
			前置	後置	前後置
サンプル数		50	100	100	200
候補単語の正解率		100	100	100	99
確からしさ	レベル2	76	81	41	32
	レベル1	24	19	59	68

7. あとがき：

本報告では、電話サービスのメニュー選択にワードスポッティング音声認識を導入した場合、余剰語の出現状況に応じて候補単語の確からしさを決定するための評価関数を変更することにより音声会話の操作性向上を行える可能性が有ることを示した。今後は、今回の結果を元に確からしさの評価関数を決定する予定である。

最後に、日頃適切な御指導をいただく当研究所の川野辺リーダー、並びに関係各位に感謝します。

[参考文献]

- (1)有田、他：メディアに依存する会話の様式
 （情処学会 研究会 61-5 S62.5）
- (2)大内、他：電話取次時の会話構成
 （信学会 全国大会 521 S62.11）
- (3)吉田、他：取り次ぎ会話に於ける間投詞
 （信学会 全国大会 D-272 S63.3）