5Q-03 音声対話システムにおけるインストラクションによる発話制約効果

中野 晃 † 竹田 孝司 † 村上 仁一 † 池原悟 † 甘粕 哲朗 ‡
† 鳥取大学工学部 , ‡NTT サイバースペース研究所

1. はじめに

コンピュータの普及，高度複雑化に伴って万人が簡単で自由にコンピュータを扱うためのインターフェースの一つとして，音声対話システムの実現が期待されている．しかし従来の研究においては，人間同士の会話に用いられるような自由な発話を対象とされてきたものが多い [1]．これは発話種類を無限に増やし発散傾向にありがちなため，誤認識を招きやすいなどの問題がある．

本研究では事前情報 (インストラクション) を与えることで利用者の発話を制約し，システムの認識率向上を狙うことを考える．この目的でインストラクションの制約効果を調査する．具体的には，音楽 CD 購入をシミュレートする音声対話システムを構築し，男性被験者 12 名をインストラクション有無の 2 グループに分け，音声対話システム利用時の対話データを収集し，インストラクションの有無による発話単語生成種類，発話文生成種類，発話単語エントロピーを評価する．

2. 調査手順

以下にインストラクションの制約効果を調査する手順を説明する．

2.1 実験用音声対話システムの構築

実験用の音声対話システムを計算機上に構築する．ユーザが画面に表示される音楽 CD リストから，音声による対話を行い，希望する CD を購入する．ユーザは音声と画面による入力のガイドを見ながらシステムを利用する．構築するシステムでは音声認識や音声合成は行わず，システム応答を人間により補助する．音声対話システムのイメージを図 1 に，購入に至るまでのタスクの基本的な遷移を図 2 に示す．

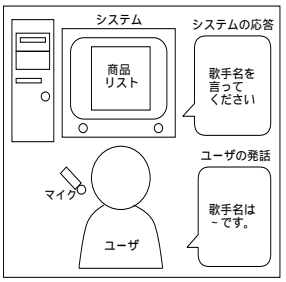

図1. 構築する対話システムのイメージ図

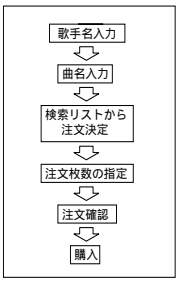

図2. タスクの基本的な遷移の例

2.2 インストラクションのガイド作成

上記の音声対話システム用のインストラクション (デモテープ) を作成する．インストラクションは，システムの流れ (数秒)，システム上の注意点 (数秒) ，各々の場面における発話の例 (約 1 分)，システム利用の対話デモ (約 4 分)，を合計 5 分程度のビデオである．

The restriction effect using the instruction video in a dialog system
Akira NAKANO †Takashi TAKETA †Jin'ichi MURAKAMI †Satoru IKEHARA †Tetsuo AMAKASU ‡
†Tottori University , ‡NTT Cyber Space Laboratories

2.3 被験者実験による比較

2.1，2.2 において作成した音声対話システムおよびインストラクションを用いて，被験者を用いて実験を行う.

1. 被験者が音楽 CD 購入システムを利用する
2. 音声対話システム内での対話音声を録音し，人手により書き起こしを行う
3. 書き起こしたデータに形態素解析をかけ，人手による補正を行う．
4. 単語単位，文単位のデータに変換し，インストラクションの有無で比較する．

なお歌手名，曲名の固有名詞は，制約効果の評価としては 1 つの単語として扱うべきと考え，それぞれを統合する．そして，変換した対話データから単語出現頻度，文出現頻度を調査し，インストラクションの有無で比較する.

3. 実験結果

男性被験者 12 名に実験の主旨を説明した後，インストラクション有り 6 名，無し 6 名に分割し，それぞれが音声対話システムを利用した．対話データ収集は対話の様子をビデオ撮影することで行った．対話平均時間は約 4 分 30 秒であり，ユーザの平均発話数は約 16 回であった．対話文の例を図 3 に示す．図中の S はシステムの発話，U はユーザの発話を表す．

タスク順序	対話文	タスク順序	対話文
1	S：注文開始で注文受付を開始します	11	S：御希望の曲はこちらでよろしいですか
2	U：注文開始	12	U：はい
3	S：御希望のCDの歌手名を入力して下さい 曲名だけで検索するときは無しと言って下さい	13	S：注文する枚数を1~9の枚数で指定してください
		14	U：1枚
4	U：ビーズ	15	S：御注文はこちらでよろしいですか
5	S：御希望のCDの曲名を入力して下さい 歌手名だけで検索するときは無しと言って下さい	16	U：はい
		17	S：注文を受け付けました、他の御注文はありますか
6	U：無し	18	U：いいえ
7	S：検索したリスト中に御希望のCDはありますか	19	S：御注文の最終確認をします 御注文はこちらのリストで全てですか
8	U：はい		
9	S：リスト中の御希望のCDを歌手名、曲名の順で言って下さい	20	U：はい
10	U：ビーズのリングを注文	21	S：御利用ありがとうございました

図 3．実際の対話文の例

収集した対話データを単語種類，文種類，単語エントロピーを用い評価した．結果を表 1，図 4，図 5 に示す．表 1 に示されるように，インストラクション有りの場合，出現単語種類は約 1/3 に減少した．出現文種類は約 1/2 に減少した．単語エントロピーは 1.85 ビット減少した．逆に出現文総数は無しの 82 から 104 と増加した．また図 4，図 5 から，インストラクション有りの場合，単語や文の大部分が上位頻度に分布し，頻度 1 の単語や文があまり発生していない．逆にインストラクション無しの場合は頻度 1 の単語や文が多く，発散の傾向があることが読み取れる．

4. 考察

本研究におけるインストラクションは，ビデオによる例示で利用例を示す程度と簡易的であり，インストラクションの時間は 5 分とタスクの平均達成時間 4.5 分と同等なので，ユーザにかかる負荷は低いと考えられる．それに対してユーザの発話は制約されるので，システムの認識率向上に有効であると言える．

表 1: インストラクションの有無による比較

インストラクション	有り	無し
出現単語総数	141	164
出現単語種類数	17 種類	62 種類
出現文総数	104	82
出現文種類数	18 種類	34 種類
単語エントロピー	3.28	5.13

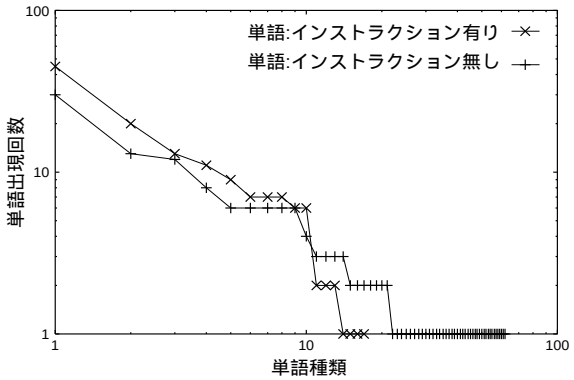

図 4．インストラクションの有無による
出現分布の違い (単語)

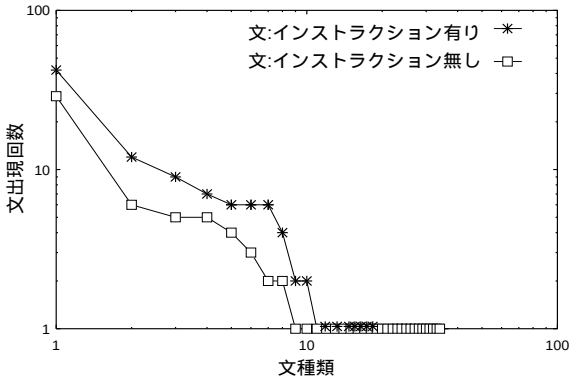

図 5．インストラクションの有無による
出現分布の違い (文)

5. まとめ

音声対話システムにおいてインストラクションが発話に与える制約効果について調査した．インストラクションによって発話単語種類，発話文種類がそれぞれ約 $1/3$ の減少した．単語エントロピーは約 2 ビットの減少した．よって制約効果が示された．今後の課題として，実際のシステムにおいてインストラクションを与え，ユーザの発話を制約することでシステムの認識率が向上するかを調査する必要がある．

参考文献

[1] 堂下： 音声・言語・概念の統合処理による対話の理解と生成，人工知能学会誌 Vol12(1997)

[2] 堂下 新美 白井 田中 溝口: 音声による人間と機械の対話，オーム社 (1998)