

# 情報爆発時代におけるヒューマンコミュニケーション —聞き上手な対話システムの実現に向けて—

小島 敬† 川嶋 宏彰† 松山 隆司†

† 京都大学大学院 情報学研究科

## 1 はじめに

情報爆発時代において、求める情報を人がスムーズに得るには、人と場を共有したシステムと、対話を通じて相互に情報を伝え合うことが望ましい。そのためには、システムがユーザの発話（言語、非言語情報）から、ユーザの状態（問題点、要求、心的状態など）を把握し、対話に反映していく必要がある。しかし、従来の人とシステムとの対話というものは、一問一答形式の対話となっており、人はシステムに対して最低限の発話しか行おうとしない。ユーザの状態を把握するためには、システムは間の合った応答によってユーザの発話を促すような、聞き上手になる必要がある。我々は聞くという立場から対話システムの設計にアプローチしており、人が話したいモチベーションを持つ状態での好まれる問のとり方を、カウンセラーの対話や人の主観評価などで分析を行った。本稿では聞き上手の有効な状況として、留守番電話というシナリオでの音声応答が人の発話をどう促すかを検証し、対話の場を共有するシステムの可能性を探る。

## 2 聞き上手な留守番電話

我々は以下の理由により、留守番電話に音声応答が有用であると考えた。

- 話を伝えたいモチベーションが高い
- 話す内容が明確に整理されていない

留守番電話につながった場合、我々は多少なりとも心理的な圧迫感の下でメッセージを残している。音声対話システムが聞き上手になるためには、人がこのような心理的な圧迫感が無くメッセージを残せるような間の合った音声応答を行うことに繋がると考えている。

### 2.1 留守番電話での音声応答の流れ

留守番電話での音声応答の流れを以下の4つの対話フェーズに分けて考える（図1）。

1. コール部...コール音が受話器から聞こえる部分
2. 対話導入部...ゲストユーザの名前を聞き、ホストユーザが電話に出れないことを伝える部分
3. メッセージ入力部...ゲストユーザはメッセージを伝え、システムが相槌をうつ部分
4. クロージング部...ゲストユーザのメッセージを受けたことを伝える部分

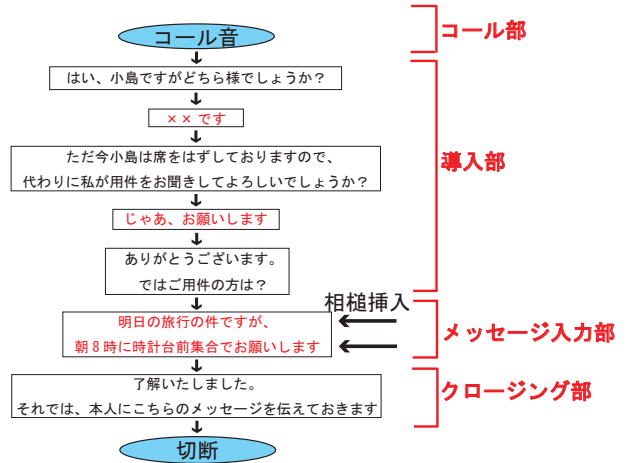


図 1: 音声応答の流れ

なお、電話をかけた人間をゲストユーザ、電話がかかってきた人間をホストユーザと呼ぶこととしている。以上の想定としては、ゲストユーザが個人を特定する電話（固定電話あるいは携帯電話）にかけた場合に、ホストユーザを知る第3者が電話に出て対応するという設定で音声応答の流れを考えている。

### 2.2 留守番電話がユーザに与える心理的影響

原田 [1] は対話相手がいる場合といない場合で、人は Talk 的発話と Speech 的発話の二つのモードを持ち、留守番電話での使いにくさが Speech モードが求められるからと考えている。また、西 [2] らは用件を整理しにくいことが心理的圧迫感の要因として考え、先に名前や電話番号を対話的に聞く留守番電話の有用性を示している（2.1 での対話導入部にあたる）。

原田の研究は発話モードについて考えているが、留守番電話に音声応答を加えた検証は行っていない。また、西らの研究は対話という観点に重点をおいておらず、対話導入部を Speech モードでの発話の負担軽減として考えている。

以上の先行研究をふまえて、聞き上手な留守番電話とは、いかにして人が Talk モードを維持してメッセージを録音できるか、というように解釈できる。本稿では、西らの研究の対話導入部がメッセージ入力部への Talk モード維持につながると考え、対話導入部が有る場合に Talk モードの維持し、無い場合（通常の応答メッセージ）に Speech モードに切り替わることを、メッセージの残しやすさによる主観評価から検証する。

以下では、対話導入部の有無による発話モードの切り替わりを検証するため、Speech モードと Talk モード

† Kei Kojima(kojima@vision.kuee.kyoto-u.ac.jp)

† Hiroaki Kawashima(hiroaki@vision.kuee.kyoto-u.ac.jp)

† Takashi Matsuyama(tm@i.kyoto-u.ac.jp)

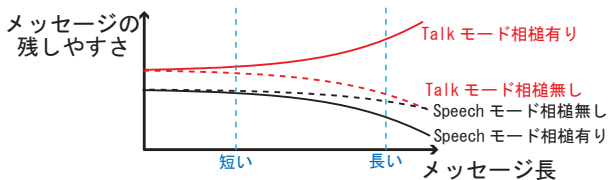


図 2: 発話モードとメッセージの残しやすさ

それぞれの発話モードが、メッセージ入力部での残しやすさにどう影響するか仮説を立て、その仮説をもとにして3節で人の主観評価を行い、対話導入部が人のTalkモードを維持させているかを議論する。

仮説：発話モードとメッセージの残しやすさ

メッセージが長い場合において、Talkモードであれば相槌の挿入によりTalkモードがクロージング部まで維持されメッセージが残しやすくなる。Speechモードであれば相槌の挿入が逆に雑音として捉えられてしまい、メッセージが残しにくくなる。メッセージが短い場合は、相槌の挿入はほとんど起こらず、コール部から発話モードの切り替えが起こらないTalkモードがメッセージを残しやすくなる(図2)

人の主観評価によるメッセージの残しやすさから、以上の仮説通り結果が出るかを評価することで、メッセージ入力部にTalkモードかSpeechモードのどちらのモードで対話導入部から誘導されてきているかを検証する。

### 3 メッセージの残しやすさの主観評価

#### 3.1 実験方法

以上の仮説を検証するために、1:対話導入部無し相槌無し, 2:対話導入部有り相槌無し, 3:対話導入部無し相槌有り, 4:対話導入部有り相槌有り, の4パターンについて実験を行った。

実験に用いた音声応答に関しては、事前に録音した音声を返答する時間長を制御して再生した。なお録音音声に関しては比較的電話対応に慣れている女性の音声を用い、返答するタイミングに関しては伊藤[3]らの発話単位タグごとのタイミングを参考にし、対話導入部での発話を1000ms, メッセージ入力部での相槌を200msとした。またクロージング部のタイミングに関しては、実験者がキー操作により挿入を行った。

実験の評価方法としては、1から4まで、同様のメッセージで順に録音してもらい、1(通常の留守番電話)に比べて2から4が、メッセージを残しやすさを1から7の7段階で、4を基準(特に変わらない)として、上に3段階、下に3段階として評価した。また、順序による影響を無くすため、メッセージを変え、ラテン方格法により順序を入れ替え、合計16回行った。さらに、録音するメッセージの長さによる影響を検証するために、同様の評価手順で、被験者に残してもらったメッセージを長くした場合についても実験を行った。

#### 3.2 実験結果

被験者12人のうち、通常の留守番電話(対話導入部無し相槌無し)よりもメッセージを残しやす(評価が4よりも大きい値)と評価した人数は、メッセージが短い場合においては、導入部有り相槌無しの場合に5人、導入部無し相槌有りの場合に2人、導入部有り相槌有りの場合に1人となった。メッセージが長い場合は、導入部や相槌を入れないことが好まれている。メッセージが長い場合においては、導入部有り相槌無しの場合に8人、導入部無し相槌有りの場合に4人、導入部有り相槌有りの場合に5人となっており、メッセージが短い場合よりも全体として評価は向上している。また現時点で有意差があるとは言えないが、メッセージが長い場合においては、導入部を合わせて相槌を挿入する方が評価が上がっている。

#### 3.3 考察

以上の実験および結果は、現時点では予備的検討の段階であるが、仮説から離れた結果となった大きな要因として以下のことが考えられる。

コール部での発話モード: 被験者は合計で32回メッセージを残しているため、ある程度の回数を重ねると、コール部での発話モードが初めからSpeechモードになってしまい、Speechモードで対話導入部でのやりとりが行われる。そのため、対話導入部でのやりとりを面倒に感じてしまい、相槌の有無に関わらず、メッセージを残しにくいものと感じてしまう。メッセージが長い場合に対話導入部が好まれているのは、Speechモードの際の用件が整理しやすくなったからと考えられる[2]。これらの要因を克服するためには、一人あたりの試行回数を減らすことや、実際に対話相手が出るパターンも加えることで、コール部での発話モードをTalkモードにして、そのまま対話導入部に誘導する必要がある。

また、対話導入部での議論とは離れるが、相槌のタイミングの良し悪しで、評価が大きく変動していたので、相槌の挿入に関しても、再検討の必要がある。謝辞

本研究の一部は、科学研究費補助金18049046の補助を受けて行った。

#### 参考文献

- [1] 原田: 人の視点からみた人工物研究; 共立出版, 日本認知科学会編, pp.101-127 (1997).
- [2] 西, 北井: 蓄積型音声対話システムにおける発話促進手法; 情報処理学会研究会報告. SLP, 音声言語情報処理 Vol194, No.57, pp.67-74 (1994).
- [3] 伊藤, 山田, 荒木: 音声認識・言語理解性能や状況の違いによるタスク指向音声対話の言語的・音響的特徴の比較; 日本音響学会誌, pp. 43-48, (2007).