

音声認識の自己目的的な楽しさ

Autotelic Experiences in Using Speech Recognition Systems

西本 卓也* 新美 康永
Takuya NISHIMOTO and Yasuhisa NIIMI

京都工芸繊維大学 工芸学部 電子情報工学科
Department of Electronics and Information Science, Kyoto Institute of Technology

Abstract For the speech input systems to be commonly used, it is required to build such systems that the user feels like to enjoy them, although the speech recognition is not perfect. Utilization of the playfulness which would exist in the speech recognition itself would make the users willingly learn to use such systems effectively. In this paper, we treated the use of speech recognition as an autotelic experience with its own playfulness, and we studied on the primary factors of it by using speech recognition games. Through our investigation, it turned out that in using the speech recognition occurs the 'flow' status which gives the users some feeling of enjoyment.

If the user feels it to be challenges to make the system correctly recognize his or her utterances, the playfulness is increased. It also turned out that to feel the enjoyment in using speech recognition, it is important that the user perceives to have the enough skill to use speech recognition systems.

1 はじめに

音声認識技術の普及に伴い、これらが広くユーザに受け入れられるかどうかに関心の対象になっている。例えば野村ら [1] は、音声ダイアル機能に慣れることで使用意欲が高まり満足度が上がると報告している。

このような観点からは、たとえ認識性能が完璧でなくてもユーザが使用したいと感じるシステムを構築することが望ましい。例えばユーザが音声認識技術のどのような要因から楽しさを感じるかを分析することにより、音声認識技術の限界を補い、実用的であるとユーザに感じさせるようなアプリケーションの設計が可能になる。

そこで本研究では、音声認識そのものが持つアミューズメント性に注目し、孤立単語の音声認識によるゲームを用いて、音声認識が持つ楽しさの要因を分析した。

2 フロー体験モデル

本研究では「自己目的的な楽しさ」の分析にフロー体験モデルを用いた。Csikszentmihalyi [2][3] によれば、我々が行おうとしている課題のために持っている技能と、その課題を行うことの挑戦がある高いレベルで均衡するとき、調和的で快適に目標を達成できる状態が生じる。この状態は、多くの人が「流れているような感じである」と表現したことから、フロー体験と名付けられている。

このような楽しさは、個人が知覚した挑戦目標が自分の能力と釣り合っている部分で生じる。テニスを例にとれば、もし二人の能力が釣り合っていなければ楽しくない。能力の低い者は不安を感じるし、高い者は退屈するだろう。同様に、聴く人の能力に比べて単純すぎる音楽は退屈であるし、複雑すぎれば聴く気がなくなる。楽しさは退屈と不安の境界、その人の挑戦水準が能力水準とうまく釣り合っている時に現われる (図 1)。

また Hoffman ら [4] はフローモデルをネットサーフィン体験に適用し、ウェブの能動的メディアとしての参加感、ユーザの技能とコンテンツの挑戦が均

* 連絡先：京都工芸繊維大学 工芸学部 電子情報工学科
〒 606-8585 京都市左京区松ヶ崎御所海道町
Tel:075-724-7492 Fax:075-724-7400
<http://www-vox.dj.kit.ac.jp/>
E-mail:mishi@dj.kit.ac.jp

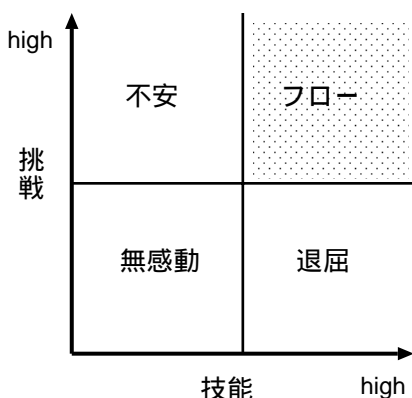


図 1: フロー体験モデル

衡することによってもたらされる、と主張している。

フロー体験による楽しさを構成するものとしては次の要素が挙げられる [3]。

1. 課題を達成できる見通しがあること
2. 自分のしていることに集中できていること
3. 行われている作業に明確な目標があること
4. 行われている作業に直接的なフィードバックがあること
5. 深いけれども無理のない没入状態で行為していること
6. 自分の行為を統制しているという感覚をともなうこと
7. 自己についての意識の消失 (ただし、フロー体験の後では自己感覚はより強く現われる)
8. 時間の経過の感覚が変わること

ただし、フロー状態に入っているかどうかは、その人が挑戦や技能をどう知覚するかに依存している。また、フローが生じ得るように与えられた環境を認知し楽しむことができるかどうかは、個人の経験や能力に依存する。

3 実験の概要

音声認識という体験の基本的な性質を調べるために、本研究で行った実験の概要について述べる。

3.1 被験者

被験者は、理工系の大学生および大学院生 15 名で、全員男性である。日常的にコンピュータおよび

ウェブ環境を使用しており、音声認識の使用経験者を含むが、音声認識を日常的には使用していない。

3.2 音声認識によるゲーム体験

カテナ社製音声認識・合成ソフトウェア「ドキュメントトーカ Ver.3.0」に添付されているゲームのうち、次の 3 種類を用いた。

さんすうのおけいこ 2 桁から 4 桁までの加算・減算などの暗算を出題し、音声認識を用いて回答させる。

発話が正解および対立候補に対して一定の尤度に満たない場合は音声を発話しても無視される。発話がなんらかの候補として認識され、それが正解でなかった場合には誤りとして認識され「え?」という合成音声で再発話を促される。

2 回連続して誤答するか制限時間をオーバーした場合には、その問題は解けなかったものとされる。

あー言えばこー言う 擬人化キャラクターとごく簡単なやりとりを行う「いま何時ですか」「何曜日ですか」といった質問が可能であるが、どのような受け答えが可能であるかはユーザには提示されない。

順に述べよ! 「山手線の外回りの駅名」のように順番が決まっている項目を順に答えさせたり「春の七草を全部述べよ」のように順番のない項目をすべて挙げさせたりするゲーム。音声認識を使って回答する「十二支」「原子番号」「七福神」「政令指定都市 北から順に」「太陽系の惑星」「日本の高い山ベスト 10」などの問題を用いた。

上記のアプリケーションに用いられている音声認識エンジンはそれぞれレベルメータ表示を備えており、音声を受理可能な場合には緑、音声区間を検出した場合には青、音量が大き過ぎる場合には赤で表示される。また、エンジンが音声を受理しない状態では、レベルメータは灰色になる。

実験にはアンチノイズマイクロフォン (Andrea 社製 ANC-500) および外付けスピーカを使用した。実験の詳細は次の通りである。

1. 「ドキュメントトーカ Ver.3.0」添付アプリケーションである「認識のセットアップ」を起動し、

音声入力レベルメータについて説明する。また、マイク音量レベルを調整したり、ヘッドセットの装着方法を工夫するなどして、音声認識に最適な状態にする。

2. 「さんすうのおけいこ」を実行し、2桁の足し算を音声で回答させた。1ゲームは10問であり、すべての被験者が必ず4ゲーム以上行うものとした。被験者にとって足し算が簡単すぎる場合は引き算を行う。正解をリジェクトされる場合や認識率が低い場合には、発話のタイミングや音量、発話の明瞭度やアクセントなどに問題があることを指摘し、認識されやすい発話を行うための助言を与えた。また、各ゲーム試行における音声の受理回数、音声認識および回答の誤り回数、発話のリジェクト回数を計測した。
3. 「あー言えばこー言う」を予備的な説明なしに実行し、自由に話しかけるように被験者に指示をした。被験者が退屈や苦痛を訴えるか、「ばいばい」というキーワードによってアプリケーションが終了するまで対話を続けさせた。
4. 「順に述べよ!」を起動し、被験者が挑戦してみたいと感じたいいくつかの項目について遊ばせた。

一人の被験者がこれらすべてのゲームに費やした時間は、30-50分である。

3.3 アンケート

ゲームで遊んでもらった後、以下のような観点から、アンケート調査を行った。

- 音声ゲームの楽しさに関する質問 (質問 a-c)
- 今回のゲームを含む音声認識システムの自己目的性に関する質問 (質問 d-g)
- コンピュータやウェブの使用における被験者のフロー経験の有無に関する質問 (質問 h-j)
- 被験者にとって音声認識を使うことがどの程度挑戦であり、どの程度の技能を持っていると感じているかを問う質問 (質問 k-l)

アンケートの内容の詳細は次章で述べる。

4 実験結果

4.1 音声認識の自己目的性

アンケートの項目 a), b), c) は、今回の音声ゲームの体験の楽しさに関する質問である。質問文とその回答を以下に示す。

質問 a) どのゲームがどのくらい楽しかったですか?

「非常に楽しい」を5とし、「全然楽しくない」を1としたときの結果の平均を表1に示す。

表 1: 各ゲームの楽しさの評価 (平均)

| ゲーム名 | 内容 | 評価 |
|-----------|-------|-----|
| さんすうのおけいこ | 暗算ゲーム | 3.8 |
| あー言えばこー言う | 会話 | 2.3 |
| 順に述べよ! | 記憶クイズ | 3.7 |

質問 b) ゲームをしていた時間を振り返ってどう感じましたか?

選択肢として「あっという間だった」「とても長く感じられた」「どちらともいえない」を挙げた。各選択肢を選んだ人数を表2に示す。

表 2: 音声ゲームにおける時間感覚

| 選択肢 | 人数 |
|------------|----|
| あっという間だった | 7 |
| とても長く感じられた | 2 |
| どちらともいえない | 6 |

質問 c) 音声認識そのものとゲーム自体の内容の、どちらをより楽しみましたか?

選択肢として「ゲーム自体を大きな目的として楽しんだ」「ゲームより音声認識をうまく使うことを大きな目的として楽しんだ」「どちらともいえない」を挙げた。各選択肢を選んだ人数を表3に示す。

表 3: 音声ゲームにおける楽しみの対象

| 選択肢 | 人数 |
|-----------|----|
| ゲーム自体 | 5 |
| 音声認識 | 6 |
| どちらともいえない | 4 |

4.2 音声認識の自己目的性

質問 d), e), f) は、今回のゲームを含めて、いままで使ったことがある音声認識システムの自己目的性に関する質問である。

以下の質問は、今回のゲームに関する体験を答えてもらっても、いままであなたが使ったことがある音声認識システムに関する経験を思い出して書いてもかまいません。

質問 d) 音声認識システムを楽しいと感じたことがありますか？

「はい」と答えた場合は、具体的に、何をしていた(遊んでいた)ときですか？

また、どのような感じがしましたか？簡単にお書きください。

各選択肢を選んだ人数を表 4 に示す。

表 4: 音声認識における楽しみの経験の有無

| 選択肢 | 人数 |
|-----|----|
| はい | 11 |
| いいえ | 4 |

「はい」と回答した被験者は本実験での「さんすうのおけいこ」や「順に述べよ!」が楽しかったと答えたほか、過去に音声によるアプリケーションの起動や音声コマンド入力を用いたり、音声で動くロボットを使用した経験について「嬉しくなった」「楽しいと感じた」と答えた。

質問 e) 音声認識を使っているときの感じ・気分・フィーリングは、下記の体験にどのくらい似ている(あるいは、似ていない)と思いますか？

自己目的的な活動の特性を知るための質問として、「友情とくつろぎ」「危険と運」「問題解決」「競争」「創造」「目立つ」の6つのカテゴリーに分類される21項目の体験と音声認識ゲームの類似性を問うた[2][5]「非常に似ている」から「非常に違う」までの5段階で回答された結果を、回答者ごとに類似度の高い順に順位づけし、全回答者がつけた順位の平均を各項目について求めた。結果は表 5 に示す。

質問 f) 音声認識に対して「前向きな挑戦」を感じたことがありますか？

「はい」と答えた場合は、具体的に、何をしていた(遊んでいた)ときですか？

また、どのような感じがしましたか？簡単にお書きください。

表 5: 音声認識と各体験の類似性得点の平均順位

| 体験 | 平均順位 |
|----------------------|------|
| 問題解決 | |
| 将棋・碁・オセロなどのゲームをする | 2.8 |
| プラモデルなどを組み立てる | 4.8 |
| クイズを解く | 1.3 |
| トランプやマージャンをする | 5.2 |
| 知らない場所を探検する | 2.9 |
| 創造 | |
| 何か新しいものを設計したり発見したりする | 4.3 |
| 目立つ | |
| テレビ番組に出演する | 7.2 |
| 大勢の観客の前で演技をする | 3.7 |
| 友情とくつろぎ | |
| いい音楽を聴く | 13.1 |
| 恋人と過ごす | 11.9 |
| すばらしい映画を見る | 11.3 |
| 親友と一緒にいる | 11.1 |
| 面白いマンガや本を読む | 8.5 |
| 危険と運 | |
| 墜落する飛行機からスカイダイビングする | 12.6 |
| 宝石をさがしてサメのいる海にもぐる | 10.1 |
| 猛スピードで車を運転する | 10.5 |
| 家族を助けるために火事の家に飛び込む | 13.8 |
| 思い切って遠くの沖のほうまで泳ぐ | 8.7 |
| 競馬に大金をかける | 9.5 |
| 競争 | |
| 競争して走る | 6.9 |
| 競争でやるスポーツをする | 6.0 |

各選択肢を選んだ人数を表 6 に示す。

表 6: 音声認識における前向きな挑戦の知覚の有無

| 選択肢 | 人数 |
|-----|----|
| はい | 8 |
| いいえ | 7 |

「はい」と回答した被験者は、主に音声認識がうまくいかない場合に前向きな挑戦を感じている。アンケート回答の一部を以下に引用する。

- なかなか認識してくれないとき、発音やアクセント、リズム、高さを変えてみた。
- 何度も誤認識されたときの再入力。だっ子に言い聞かせるような感じがしました。
- 正しく認識してくれない一般単語をなんとかしてわからせようとした時。ものわりの悪い人、もしくは異国の人に向かって話し掛けている感じがします。

音声認識はユーザに楽しさと同時に失望や怒りなどの感情を促すことも考えられる。質問 g) ではそのような経験について問うた。

質問 g) あなたがいままで音声認識を使って、がっかりしたり腹が立ったことがありますか？もしあればお書きください。

この結果、15人のうち13人が「ある」と答えた。アンケート回答の一部を以下に引用する。

- ちっとも認識してくれないとき・反応がないとき
- 同じ言葉を何度言っても、認識しなかったとき
- 発音が悪いのでうまく認識してくれない
- 何度喋っても認識してくれなければ腹が立つ。
- 自分の喋り方の認識されにくさにつながりした。
- 機械の声で「え？」ときき返されるとき

4.3 被験者のフロー経験

質問 h), i), j) は、コンピュータやインターネットの使用における被験者のフロー経験の有無に関する質問である。こういった経験の有無は、音声利用によるフローの起きやすさに影響すると考えられる。

質問 h) 音声認識にかぎらず、コンピュータやインターネットを使っていて時間の流れが普段より早く感じられたり、逆に遅く感じられたりしたことがありますか？

質問 i) 音声認識にかぎらず、コンピュータやインターネットを使っていて「自分自身をうまくコントロールしている」という感じがしたことがありますか？

質問 j) 音声認識にかぎらず、コンピュータやインターネットを使っていて自分の操作に対して明確なフィードバックが得られていると感じたことがありますか？

各選択肢を選んだ人数を表7に示す。

表7: フロー特有の経験の有無 (人数)

| | はい | いいえ |
|-----------------|----|-----|
| h) 時間経過の感覚の変化 | 13 | 2 |
| i) 自分の行為のコントロール | 2 | 13 |
| j) 直接的なフィードバック | 9 | 6 |

質問 h) に「はい」と回答した被験者は、時間の流れが普段より速く感じられた体験として、興味のあるウェブ情報の閲覧、文章の作成、掲示板やチャットへの参加、プログラミングやダウンロードなどを挙げた。被験者の一人は「底なし沼に入り込んでいく感じで抜けだせない気持ち」と回答している。

質問 i) において被験者の一人は、自分自身をコントロールしている体験として「こんな風に発音した方が認識されやすいだろうなと意識して喋って、それが当たっていると感じたとき」を挙げた。

質問 j) に「はい」と回答した被験者らは、明確なフィードバックを得ていると感じた体験として、ゲームや検索エンジン、ウェブでのショッピングやアンケートなど、音声とは無関係な対話的システムの利用経験を挙げた。

4.4 音声認識率とゲームの楽しさ

暗算ゲームの4~8回の試行(1試行は10問)において、3桁までの数字音声の最高認識率と平均認識率を表8に示す。ただし、アンケートにおいて「音声認識を楽しみと感じたことがある」と答えた被験者を+印で示す。

表8: 暗算ゲームにおける音声認識率

| 被験者 | 最高 (%) | 平均 (%) |
|-------|--------|--------|
| s1 + | 57 | 41 |
| s2 + | 73 | 65 |
| s3 + | 100 | 74 |
| s4 + | 100 | 84 |
| s6 + | 54 | 41 |
| s7 + | 82 | 57 |
| s10 + | 100 | 89 |
| s11 + | 91 | 79 |
| s12 + | 64 | 55 |
| s13 + | 91 | 80 |
| s14 + | 91 | 75 |
| s5 | 91 | 71 |
| s8 | 69 | 53 |
| s9 | 88 | 52 |
| s15 | 69 | 52 |

+印は音声認識を楽しみと感じた被験者 G1。無印は G2。

4.5 音声認識における挑戦と技能

音声認識における挑戦や技能を、被験者がどのように自覚しているかを問う質問を行った。質問項目は音声認識に関する項目群およびコンピュータやウェブの操作に関する項目群から構成される。回答は「とても得意」を5とし「とても苦手」を1とする5段階評価である。

「音声認識を楽しみと感じたことがある」と回答したグループ G1 (11人) および、ないと回答したグループ G2 (4人) のそれぞれについて回答を集計し、2つの正規母集団の母平均が等しいという仮説で片側 T 検定を行った。その結果を表9に示す。

表 9: 被験者が自覚する技能の平均 (G1:11 人, G2:4 人)

| 項目 | G1 平均 | G2 平均 | |
|-------------------------------|----------|----------|----|
| (音声認識に関して) | | | |
| 適切なタイミング 声の大きさを発声すること | 2.64 | 1.75 | * |
| 入力したい項目の読み方を 正しく判断すること | 2.91 | 2.25 | |
| 音声がうまく認識されるように 発音を工夫して喋ること | 3.00 | 2.00 | |
| 不要な声に反応しないように 独り言を我慢すること | 2.36 | 1.50 | |
| どの操作が音声で行えるか 正しく意識すること | 3.27 | 2.50 | * |
| 平均 () | 2.84 | 2.00 | ** |
| (コンピュータ操作に関して) | | | |
| マウスとキーボードを 効率よく併用すること | 3.36 | 3.50 | |
| 画面のスクロール操作を 手際よく行うこと | 3.27 | 3.50 | |
| ウェブ画面の「戻る」「進む」 操作を混乱せず行うこと | 3.55 | 4.00 | |
| 平均 () | 3.39 | 3.67 | |
| (-) | -0.56 | -1.67 | * |

* は 0.05 水準で有意, ** は 0.01 水準で有意.

5 考察

被験者 15 人のうち 11 人が音声認識を楽しんでいるが、ゲーム自体のみを楽しんだ被験者は多くない(表 3)ことから、音声認識そのものがゲーム性を持ち、体験の楽しさにくらか貢献していると考えられる。また、今回の被験者の多くは時間感覚の変化を報告しており(表 2)、個人差はあるものの、音声認識に関して「自分の行為の統制」や「前向きな挑戦」が自覚されている。これらフロー特有の現象は、音声認識によるフロー体験の存在を示唆している。

各被験者の音声認識率とアンケート回答からは、認識率が悪くてもシステムを楽しんで使ったという被験者がいる(s1,s6,s12) ことが分かる。

表 5 によれば、音声認識の自己目的性は「問題解決」「創造」「目立つ」の 3 つのカテゴリと特に類似している。つまり、音声認識の使用が創造的で知的な問題解決、あるいは演技などの表現活動と類似した体験として知覚されている。一方で「友情とくつろぎ」に関する体験との類似度は低く、音声認識はシステムを親しみやすくする手段としては知覚されていない。

被験者からは、発話がうまく認識されない場合に行う試行錯誤を楽しんだ、認識されやすい喋り方を

できない自分に対して不満を感じた、などの意見があった。これらは音声認識が「自己目的な問題解決」であったことを反映している。

表 9 によれば、特に音声認識に関する技能 5 項目の平均点において有意な差がみられ、音声認識に関する操作を苦手だと知覚することと、音声認識を楽しめないこととの関連が示された。この他に音声認識に対する挑戦の知覚についても質問をしたが有意な違いはなかった。

山本ら [6] は、コンピュータとのインタラクションにおいて「相手が人間であると思うこと」が被験者に楽しさを与えると指摘している。本実験の結果をこれと比較すると、特に質問 f)などで、被験者が音声認識システムに対して、相手が子供や外国人であるかのように感じたと報告している点が興味深い。

6 まとめ

音声認識という体験が持つ性質を簡単なゲームを用いて分析した。その結果、認識率を高めようとするユーザの問題解決的な努力が特に自己目的な楽しみの対象となり得ることや、音声認識を使用するための各種の技能において苦手感覚の克服が必要であることなどを確認した。

今後は音声認識のこのような性質を各種システムについて調べると同時に、より幅広いユーザに対して検討を行う必要がある。また、ユーザに挑戦的要素を与えるだけでなく、ユーザに親しみやすさをもたらすような音声システムのあり方についても検討したい。

参考文献

- [1] 野村 和也, 大蜘蛛 篤: 携帯電話音声ダイヤルのモニタ評価, 音講論集 1-R-5 (1998-9).
- [2] Mihaly Csikszentmihalyi: Beyond boredom and anxiety, Jossey Bass Publishers (1975). (M. チクセントミハイ, 今村浩明訳: 楽しむということ, 思索社 (1991)).
- [3] Mihaly Csikszentmihalyi: Flow, Harper & Row Publishers (1990). (M. チクセントミハイ, 今村浩明訳: フロー体験喜びの現象学, 世界思想社 (1996)).
- [4] Donna L. Hoffman and Thomas P. Novak: Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations, Journal of Marketing, Vol. 60, pp. 50-68 (1996).
- [5] 佐藤郁哉: 暴走族のエスノグラフィー—モードの叛乱と文化の呪縛—, 新曜社 (1984).
- [6] 山本吉伸, 松井孝雄, 開一夫, 梅田聡, 安西祐一郎: 計算システムとのインタラクション—楽しさを促進する要因に関する一考察—, 認知科学, Vol. 1, No. 1, pp. 107-120 (1994).