

ディクテーション作業における楽しさの分析

Factors of Joyfulness in Using Dictation Systems

西本卓也 櫻井晴章 荒木雅弘 新美康永

Takuya NISHIMOTO, Haruaki SAKURAI, Masahiro ARAKI and Yasuhisa NIIMI

京都工芸繊維大学 工学部 電子情報工学科

Department of Electronics and Information Science, Kyoto Institute of Technology

Abstract Applications of speech recognitions are becoming popular, especially for entertainment. However, in business or work oriented usage, they are still unpopular. In prior works, the objectives of the users have not been investigated well. To use the speech recognition system effectively, a positive effort on the part of the user is important. A positive effort stems from the joy of challenge.

We investigated what conditions were necessary for the users of speech recognitions to feel joyfulness. Using the dictation software, we performed an experiment with the following procedure. First, we performed an initial questionnaire and asked the subjects their attitude towards the speech input. The subjects were then divided into four groups to perform the task. Two of the four were given a higher challenge, a more difficult task. The other two were given a lower challenge, an easier task. One group from the higher challenge and one from the lower challenge had much experience. Alternatively, one group from the higher challenge and one from the lower challenge had almost no experience. Finally we performed the follow-up questionnaire for measuring the joyfulness and sense of control of the user.

The results showed that the subjects' initial attitude and interest toward speech recognition determined their sense of self-fulfillment and control in the work. And where the ratio of the task and skill were almost equal, we observed the state of joyfulness and the subjects seemed at ease in the work. These observations indicate the occurrences of 'flow experience' during the work of dictations.

1 はじめに

近年、携帯電話でのサービスやゲーム等において音声認識が用いられるようになったが、PC環境でのビジネス用途においては音声認識は一般的とはいえない。特に、文書作成などにおけるディクテーションへのユーザの要求は高まっているが、音声認識の性能が完璧ではないために敬遠されがちである。

現状の技術では、大語彙連続音声認識の性能は、ユーザが「認識されやすい喋り方」を行うか否かに影響される。音声認識システムをうまく使うためにはユーザの前向きな努力が必要である。

我々は、ディクテーションシステムのユーザに前向きな努力を促すために、ディクテーション作業を自己目的的に楽しむことが重要であると考えた。そこで本研究では、ディクテーション作業の自己目的的な楽しさがどのような要因によってもたらされるか、また、ユーザの態度や認識性能等が楽しさにどのような影響を与えるのかを検討した。

2 関連研究

コンピュータとのインタラクションにおいて内発的動機付けに基づく楽しさを扱った先行研究として、山本ら [1] は、キーボード操作による「しりとり」ゲームを用いた実験を行い、特に「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。

自己目的的な楽しさをモデル化したものとしては、チクセントミハイが提唱したフロー体験理論がある [2][3]。この理論において、自分が望んで行う活動（自己目的的経験）では「楽しさ」という大きな内発的報酬が得られるとし、この自己目的的経験の過程で獲得する感覚を「フロー」の状態としている。

チクセントミハイはフローが成立する条件を次のようにモデル化している。我々が行おうとしている課題に対して持っている技能レベルと、その課題を行うことの挑戦レベルがあるレベルで均衡するときに、快適に楽しく目標を達成できる状態が生じる。例えば、対戦ゲームを例に取れば、挑戦レベルは相手の技能レベ

ルに依存するため、二人の能力が釣り合っていないと楽しくない。能力の低い者は不安や緊張を感じるであろうし、高い者は退屈や不満を感じるであろう。二人のレベルが近ければ近いほど、お互いの技能レベルと挑戦レベルが釣り合うことになり、楽しいと感ずることができる。

フロー体験による楽しさを構成するものとして、以下の要素が挙げられている。

- 課題を達成できるという見通しが立っている
- 自分のしていることに集中できている
- 行っている作業に明確な目標がある
- 作業に直接的なフィードバックがある
- 深いが無理のない没入状態である
- 自分の行為を制御している感覚を伴う
- 自己についての意識の消失
- 時間の経過の感覚が変わる

ただし、フロー状態に入っているかどうかは、その人が挑戦レベルや技能レベルをどのように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。

フロー理論を用いた研究として、例えば Webster ら [4] は、コンピュータとのインタラクションにおける楽しさの構成要素を、自己の制御、注意の集中、好奇心、内発的動機付け、の 4 項目に整理している。さらに、フロー体験がソフトウェアの柔軟性に肯定的な印象を与える、フロー体験によってユーザの探険的操作が促される、フロー体験によって将来の自発的なコンピュータ利用が促される、などの仮説を、表計算ソフトウェアおよび電子メールシステムを用いて検証している。

Wiedenbeck ら [5] は、コンピュータ操作の学習において、対話スタイルと学習経験の差異がフロー体験の有無に影響すること、フロー体験の有無がアプリケーションの使いやすさや有用性の知覚に影響すること、これらの知覚がタスク達成能力に影響すること、などのモデル化を行い、GUI によるワードプロセッサとキーボード操作によるスクリーンエディタを用いて検証している。

また、キーボードによるチャットを用いた研究 [6] においては、目的指向会話を挑戦の高いタスク、自由会話を挑戦の低いタスクとして、挑戦の高いタスクにおけるタイプスキルの高い被験者と挑戦の低いタスクにおけるタイプスキルの低い被験者がそれぞれ会話中にフロー体験をしていたこと、技能に比較して高い挑

戦が与えられた被験者はスキルが向上したこと、などが検証されている。

3 本研究の目的

関連研究 [1][6] は人間同士やそれを模した対話に焦点を当て、[4][5] では機械と人間の関係に注目している。一方我々は、インタラクション様式としての音声認識に注目し、音声認識がたとえ対機械としての利用方法であっても対人間に近い性質を持ち得る、という観点からその自己目的性に注目している。

我々はこれまでに、音声対話ゲームにおける音声認識の自己目的性の検証を試みてきた [7]。その結果によれば、音声認識の使用は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た。また、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的な楽しみの対象になり得ること、音声認識に対する苦手感覚の克服が必要であること、などが確認された。

本研究では、タスク自体が自己目的性を持たないディクテーション作業を用いて、音声認識システム自体が持つ自己目的性についてさらに検討する。

フロー体験モデルに基づく実験構成として、被験者を技能と挑戦それぞれの高低によって 2 要因 4 群にわけ、技能と挑戦のバランスが取れた条件において被験者がより楽しさを得ることを検証するのが本研究の主な目的となる。また、被験者の事前の経験や意識がディクテーション体験の知覚に及ぼす影響をアンケートを用いて検討し、フロー成立の諸要因についても検証する。

4 実験

4.1 実験の概要

被験者は、15~48 歳の社会人および学生で、男性 50 人、女性 37 人の 87 人である。

実験環境を図 1 に示す。防音室に、音声認識ソフト IBM ViaVoice V8 Premium をインストールしたノート型パソコン (Pentium III 600MHz) と、ViaVoice 付属の手持ちマイク (ANC-300-4) とマウスを用意する。また、ノート型パソコンの画面を映す別のディスプレイを用意し、防音室の窓を通して被験者の様子と作業内容が撮影できるようにビデオカメラを設置する。防音室内にマイクを設置し音声も同時に収録可能とする。

実験は以下の手順で行う。

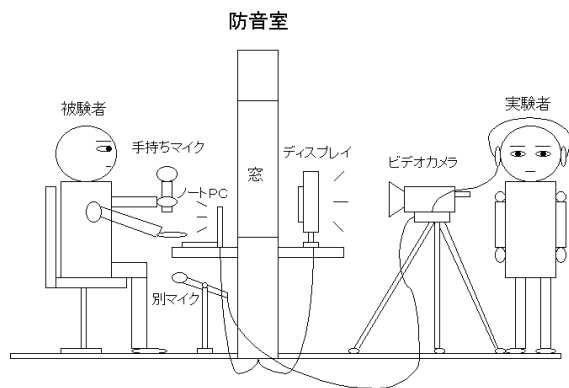


図 1: 実験環境

1. 被験者の音声認識に対する事前の関心を調べるため、被験者に事前アンケートに回答させる。
2. 事前説明を行い、実験の手順、注意事項などを簡単に説明し、ひとりずつ防音室に案内する。
3. 防音室では、まず、クイックエンロール（ディクテーションシステムの話者への適応）を行わせる。また、マイクの操作方法、誤り入力をした場合の修正方法などを詳しく説明する。
4. 各々の条件で課題テキストを与え、音声入力アプリケーション（スピークパッド）を用いて約 10 分のディクテーション作業を行わせる。
5. 課題終了後に事後アンケートに回答させる。

課題テキストは新聞記事のコラムなどから、記事自体に楽しさを感じさせないもので、未登録語や入力しにくい語を含まず、文字数や文字の難易度が同程度のもの 3 種類を用意する。

実験中に誤認識があった場合は先に進まずその場で修正させる。また、誤認識の修正においても、マウスによるカーソル移動とキーボードによる削除の操作のみを許し、文字入力には音声を用いるように指示する。被験者が 10 分以内に課題を終えた場合は、時間になるまで同じ課題を繰り返すように指示する。

4.2 アンケート

被験者の前向きな態度が、作業における楽しさの知覚やその他のフロー体験に結びつく、という仮説を検証するために、以下のようなアンケートを行う。

まず、事前アンケートにおいては以下の質問と選択肢を用意する。

質問：キーボード入力のかわりに音声入力を使うことについて、当てはまるものを選んでください：

A 音声入力ができることを知らない

B 音声入力を使いたくない

C 音声入力を使ってみたいが使ったことがない

D 音声入力を試したが満足できなかった

E 音声入力を使用している

特に C と回答した被験者は音声認識に前向きな態度を持っており、B および D と回答した被験者は音声認識を否定的に捉えている、と考えられる。

事後アンケートでは、前向きな態度な態度の持ち主がディクテーション作業をより「楽しかった」と回答することを期待し、被験者の作業中の意識や感情の状態について内観させる質問を行う。文献 [4] を参考に、コンピュータ操作における楽しさを構成する 12 項目の要素を挙げ、「まったくそう思わない」から「とてもそう思う」までの 7 段階で回答をさせる。

4.3 技能と挑戦に関する構成

本研究では、被験者の技能と挑戦のそれぞれに高低の差を設ける。技能レベルの低い群を X, 高いレベルの群を Y として、以下のような条件に設定する。

- 条件 X：音声認識の使い方に関する説明や練習は行わず、課題のみを 10 分間行わせる
- 条件 Y：音声認識の上手な使い方について、ViaVoice の解説書を抜粋した説明を提示して、5 分間の自由練習時間を与えた後、課題を 10 分間行わせる

上手な使い方としては以下を提示する。

- まとまった長さの文章を、普段話すときと同じペースで発音する
- 単語を一つ一つ区切らない（3 つ以上の単語を続けて話す）
- 先頭の単語をはっきり発声する
- 文章の修正は、単語で区切らず、ある程度長い文を読み直す
- 中断したいときは手持ちマイクのスイッチをオフにする

フロー体験モデルにおいては、必ずしも技能そのものの高さが重要視されておらず、被験者が自らの技能をどのように知覚しているかが重要であるとされている。従って、条件 Y での「練習」に期待する効果も、認識性能が上がるのではなく、被験者の「技能の知覚」が高まることである。

挑戦レベルについては高い群を 1, 低い群を 2 として、以下のような条件に設定する。

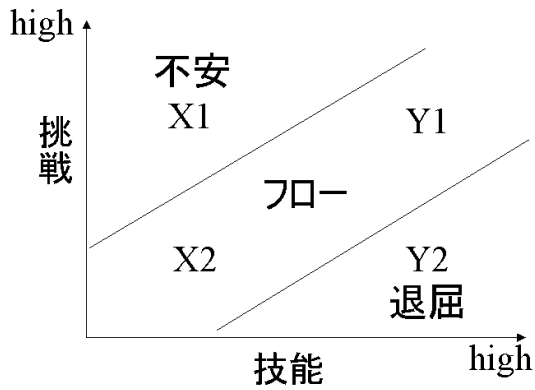


図 2: フロー理論モデルと実験条件との対応

- 条件 1: 句読点や記号はキーボード入力させ、その他の文字は音声で入力させる

例: 音声「このほどもめでたく」/ キーボード「、」
/ 音声「電子メールの送受信ができるようになった」/ キーボード「。」

- 条件 2: すべての文章を音声で入力させる

例: 音声「このほどもめでたく てん 電子メールの送受信ができるようになった まる」

条件 1 ではキーボードと音声の併用が増えるため、被験者が作業に集中しにくくなる。また ViaVoice が本来推奨している使い方は条件 2 の使い方であるため、条件 1 においては認識性能が低下し、被験者にとっての高い挑戦になると予想される。

このように X/Y および 1/2 の 2 要因を設定し、X1, X2, Y1, Y2 の 4 群で実験を行う。

各群とフロー理論モデルとの対応を図 2 に示す。X2 および Y1 がフローをもたらす状態（フローチャンネル）と定義される。

5 実験結果

5.1 単語認識率

実験条件における挑戦（条件 1/2）の差を、各被験者の単語認識率によって検証した。ただし、ディクテーションソフトの使用経験があった被験者や、著しく認識率が悪かった被験者は除外して、83 人のデータを用いた。

課題テキストを ChaSen Ver.2.2.8 を用いて形態素解析し、単語に分割した。被験者の作業を録画したビデオから、被験者の発話と画面に表示された認識結果を書き起こして、ChaSen による単語単位での対応付けを行い、各被験者ごとの単語認識率を求めた。

表 1: 各条件における単語認識率 (%) の平均

条件	認識率	条件	認識率
X1	86.1	Y1	89.7
X2	92.5	Y2	93.7

表 2: 事前アンケート結果

回答	態度	人数
A	中立的	5
B	否定的	15
C	肯定的	55
D	否定的	7
E	検討外	1

各条件における単語認識率を表 1 に示す。これは、誤認識の修正のための言い直しを含めたすべての発話に対する認識率である。

分散分析の結果、X1-X2 間に 5%水準の有意差が、Y1-Y2 間に 1%水準の有意差が認められた。条件 X および Y の間には有意差は認められなかった。

5.2 アンケート結果

事前アンケートの結果を表 2 に示す。

また、事前アンケートの回答ごとに事後アンケートにおける 7 段階の各回答の平均を求めたところ、表 3 に示す各項目において有意差または有意傾向が認められた。

これらの項目（ア）～（エ）の具体的な質問事項は以下の通りである。カッコ内は質問に対応する楽しさの構成要素を示す。

- （ア）スピークパッドを使っていたら、このソフトに興味がわいてきた（好奇心）
- （イ）スピークパッドを使うことは私にとって退屈だった（内発的動機・反転項目）
- （ウ）スピークパッドは私の好奇心を刺激してくれた（好奇心）
- （エ）スピークパッドは使っていて楽しかった（内発的動機）

5.3 実験条件の差異の影響

事前アンケートで A および C（肯定的または中立的態度）と回答した被験者について、2 要因 4 群の実

表 3: 被験者の事前の態度 (A,B,C,D) と事後アンケート回答の間の分散分析

質問項目	フロー構成要素	事前の態度との関連
(ア)	好奇心	B が有意に低 (5%水準)
(イ)	内発的動機 (反転)	B が有意に高 (1%水準)
(ウ)	好奇心	B が有意に低 (1%水準)
(エ)	内発的動機	A・C が高 (有意傾向)

表 4: 各条件における笑いの回数の平均

条件	人数	回数	条件	人数	回数
X1	15	2.00	Y1	19	2.68
X2	16	4.88	Y2	10	0.90

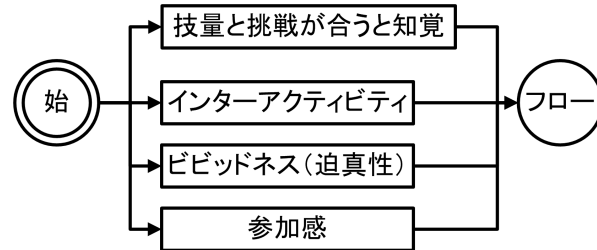


図 3: Hoffman らのフローモデルの一部

験条件ごとの集計を行ったところ、事後アンケートの回答には特に有意な影響は与えていなかった。

しかし、ビデオに録画した被験者の様子を視察したところ、技能と挑戦が釣り合うと仮定した条件 X2 および Y1 では、楽しさの表情が多く見られた。また、技能と挑戦が釣り合わないと仮定した条件 X1 および Y2 では、淡々と作業をする被験者が多く緊張や委屈の表情が見られた。

そこで、被験者が作業中に笑った回数を集計したところ表 4 の結果が得られた。分散分析では有意性が示され、Fisher PLSD 検定の結果、X1-X2、X2-Y1、X2-Y2 において 5%水準での有意差が認められた。

6 考察

6.1 単語認識率に関する考察

表 1 および分散分析の結果より、条件 1 の被験者の単語認識率は条件 2 に比較して低く、本実験に用いた条件設定のうち、行うべき作業が与える挑戦の高低については、仮定通りの実験が行われたと考えられる。

また、条件 X と比較して条件 Y の単語認識率はやや高いが、この差は棄却されており、練習を行った被験者が必ずしも有利な実験条件とはなっていないことを示している。つまり、条件 X と条件 Y の差は技能そのものには有意な影響を与えておらず、被験者自身の技能に対する知覚にのみ影響を与えたと考えられる。

なお、練習と同時に示した音声認識の上手な使い方のヒントについても、それを実践した被験者が必ずしも効率的に入力できていたわけではなかった。被験者によっては、例えば 1 単語の誤認識をフレーズで言い

直したことによって誤認識が増加してしまった場合もあった。

6.2 アンケート結果に関する考察

表 2 および表 3 から、実験前に「音声入力を使いたくない」と回答していた被験者は他の被験者と比較して、ディクテーション作業への興味や、作業への内発的動機付けの程度が低かった。

一方、「音声認識ができることを知らない」「音声認識を使ってみたいが使ったことがない」という被験者は、音声入力に対する事前の前向きな態度を持っており、この態度が楽しさにつながる感覚をもたらしていた。

チクセントミハイによれば、フローが生じ得る環境が与えられたとしても、それを楽しむためには個人の経験や能力が重要であるとされている。また、フロー体験モデルをネットサーフィン体験に適用した Hoffman らのモデルによれば、フローの要因としては「技量と挑戦が合うこと」に加えて図 3 のように「インターアクティビティ」「ビビッドネス(迫真性)」「参加感」なども重要とされている(文献 [8][9])。

本項で述べた結果は、ディクテーション作業の自己目的的な楽しさが、単に技能と挑戦のバランスによって得られるのではないこと、実験前の前向きな態度による作業への参加感が、好奇心や内発的動機付けなどフロー特有の感覚をもたらす前提となっていること、などの点でフロー理論を支持するものと考えられる。

6.3 実験条件の差異に関する考察

表 4 より、被験者は技能と挑戦のバランスが取れた条件 X2 および Y1 においてより多く笑ったことによって楽しさを示していた。

本実験での被験者を笑わせた要因はディクテーションシステムの応答である。誤認識など被験者が予期しない応答は、一般的にはユーザに不快感を与えられ考えられるが、フローチャンネルに属する X2 および Y1 の被験者はこのような状況に対して寛容であり、システムの反応をより多く楽しんでいたと考えられる。

音声認識を単に意外な面白さをもたらすものと考えれば「認識率が低いほど作業を楽しめた」という対立仮説が挙げられる。しかしこの仮説は X1-X2 間の有意差によって棄却される。

他の対立仮説として「認識率が高いほど作業を楽しめる」を考えたときに、この意味での楽しさは笑うという行為につながるとは限らないが、前述したアンケートからこの対立仮説を支持する結果は得られなかった。

なお、X2 および Y1 の 2 条件の間においても笑いの回数に有意差があった。これは Y1 の被験者が練習を行ったことにより、誤認識に被験者が慣れてしまったものと考えられる。

7 まとめと課題

本研究では、ディクテーション作業における楽しさの分析をするため、市販のディクテーションソフトを用いて被験者にテキスト入力作業を行わせる実験を行った。

実験はフロー体験モデルに基づいて、技能と挑戦それぞれの高低によって 2 要因 4 群に分けて行った。キーボード併用の有無を挑戦の高低とし、事前の練習の有無を挑戦の高低と仮定した。ディクテーションにおいてキーボード操作の併用を強制した条件では、認識率が有意に低く、被験者にとって高い挑戦となっていた。

各被験者は実験前と実験後にアンケートに回答した。この結果を分析したところ、実験前に音声認識に否定的な先入観を持っていた被験者よりも、中立的あるいは肯定的な被験者の方が、作業によって好奇心を刺激されたり内発的動機付けを促されたりしており、音声認識に対する前向きな興味がディクテーション作業を楽しめると自覚するための前提となっていた。

被験者の作業過程を視察で分析したところ、技能と

挑戦が釣り合う 2 群では、他の 2 群と比較して笑う回数が有意に多く、誤認識などのシステム応答を楽しみやすい状況となっていた。これらの結果は目的指向の作業における自己目的性の存在を裏付けており、ディクテーション作業におけるフロー体験モデルを支持するものであった。

今後の課題としては、事前の興味と技能・挑戦の均衡、という 2 つの要因の位置づけをより明確にする必要がある。また、音声認識に否定的な先入観を持つユーザへの動機付けをいかに行うか、楽しさの感覚を音声入力技能の向上やディクテーション作業の効率アップにいかにつなげるか、といった点についても検討したい。

謝辞 助言を頂いた細馬宏通氏、羽尻公一郎氏、小磯花絵さんに感謝します。また、本研究は科学研究費補助金・奨励研究 (A) 12780279 を受けた。

参考文献

- [1] 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, 日本認知科学会, 認知科学, 1 巻 1 号, pp.107-120, (1994)
- [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] Jane Webster, Linda K. Trevino, Lisa Ryan: The Dimensionality and Correlates of Flow in Human-Computer Interactions, Computers in Human Behavior Vol.a, pp.411-426, (1993)
- [5] Susan Wiedenbeck, Sid Davis: Intrinsic Motivation, Ease of Use and Usefulness Perceptions as Mediators in Computer Learning, Proceedings of HCI International 2001, Vol.1, pp.1553-1557 (2001)
- [6] 羽尻公一郎: チャットにおけるフロー体験, 社会言語科学 Vol.4 No.1 pp.17-23 (2001)
- [7] 西本卓也, 新美康永, 音声認識の自己目的的な楽しさ: 人工知能学会研究会, SIG-SLUD-9804, pp.13-18, (1999)
- [8] Donna L. Hoffman and Thomas P. Novak: Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations, Journal of Marketing, Vol. 60, pp. 50-68 (1996)
- [9] 赤木昭夫: インターネット・ビジネス論, 岩波書店, (1999)