

2022 年度 卒業論文  
指導教員 栗原一貴教授

オトノベル：  
感情を含む音声入力を用いたノベルゲーム

津田塾大学  
学芸学部情報科学科 4 年  
G19902 赤田真由

## 目次

<b>概要</b>	3
<b>1. はじめに</b>	4
<b>2. 関連研究</b>	5
2.1. 音声入力でゲーム操作を行うことのできる商品	5
2.2. 音声入力を用いたゲーム没入感向上に関する研究	5
2.3. 音声感情認識に関する研究	5
<b>3. オトノベル</b>	6
3.1. システム概要	6
3.2. システム構成	7
3.2.1. ゲーム開発プラットフォーム「Unity」	7
3.2.2. Unityアセット「Fungus」	7
3.2.3. 音声文字起こしAPI「Google Cloud Speech-to-Text」	7
3.2.4. 感情認識API「Empath」	7
3.2.5. システムの流れ	8
<b>4. 評価実験</b>	8
4.1. 実験手法	8
4.1.1. 仮説	8
4.1.2. 実験概要	9
4.1.3. 実験手順	9
4.1.4. 評価アンケート	10
4.2. 実験結果	10
4.3. 分析	11
4.3.1. 分析方法	11
4.3.2. 分析結果	11
4.4. 考察	13
4.4.1. ゲームに対する没入感についての考察	13
4.4.2. 主人公への感情移入についての考察	14
4.4.3. ゲームの楽しさについての考察	14
4.4.4. 抵抗感についての考察	14
<b>5. 今後の発展と展望</b>	15
<b>6. まとめ</b>	16
<b>7. 謝辞</b>	16

## 概要

良質なゲームというのはゲームに対する没入感が非常に高く、楽しいものである。現在ゲームへの没入感を高めるために様々な工夫が施されているが、その方法の中に自身の身体機能の一部を用いる手法がある。本研究では、会話によるストーリー進行というノベルゲームの特徴に着目し、ゲーム上の会話を自身の声で操作することができる音声入力ノベルゲーム「オトノベル」を提案する。本システムでは音声だけでなくその声色から感情を認識し、その感情によるストーリーの分岐も行えるよう実装した。また、セリフの末尾にどのような感情で読むべきか感情表現テキストによる指示を行い、感情入力を促している。評価実験として、通常のキーボード操作、音声入力操作、そして音声感情入力操作の3種類でノベルゲームをプレイしてもらい、本システムの操作方法がゲームに対する没入感及びゲームの楽しさを向上させられたか、主人公に対して感情移入を促せたか、音声入力及び音声感情入力に抵抗感を感じたかについて検証を行った。その結果、音声入力操作及び音声感情入力操作がキーボード操作よりも没入感を高められるという結果を得られたが、音声入力と音声感情入力の間に有意差は見られなかった。最後に、本システムの感情入力の方法を改良することで、ゲームへの没入感及び感情移入を一層高めることができるような今後の発展について論ずる。

## 1. はじめに

現在、世界のゲーム人口は約 31 億人を越えたと言われている[1]。世界人口が約 80 億人[2]であることを考慮すると、おおよそ 40% の人々が何らかのゲームをプレイしているということになる。スポーツをはじめとするどんな競技においてもこれほど膨大な数のプレイヤー人口を誇るものは存在しない。ゲームに熱中する要因の 1 つとして、ゲーム没入感が挙げられる。ゲーム没入感というのは、「ゲームに没頭すること」、あるいは「実際に自分自身がゲーム内に存在するかのように感じること」である[3]。つまり、ゲームの主人公に対し自己投影をしてしまうほど自分がゲーム内にいるかのような感覚に陥り、没入感が高まると言える。ゲームの多くはプレイヤー自身がその物語の主人公としてその世界を体験するものとなっている。主人公が初めから固定のキャラクタになっているものもあれば、プレイヤー好みに設定できるというものも存在する。多くの人に支持されているゲームというものは没入感が強く、楽しいものであることが多い。その為、様々なゲームでは自分の容姿や性別、名前等を変更することができるという工夫が施されている。自分に近い主人公でプレイできることにより、主人公への共感や感情移入を促しゲームに対する没入感を高めている。

本研究では、ノベルゲームを自分の音声及び感情を用いて入力することができる「オトノベル」を提案する。ノベルゲームの NPC と会話することでストーリーが進んでいくという特徴に着目し、従来の操作方法であるコントローラ及びキーボードを用いるのではなく、自身の声でセリフを読み上げることでストーリーが進行するようにした。また単に音声入力をするのではなく、声のトーンからプレイヤーの感情を検知し、セリフに込められる感情によりストーリーが展開するよう実装した。また、感情を込めてセリフを読むことを促すため、選択肢に感情表現テキストとして込めるべき感情を指定した。そして、従来のコントローラ操作、音声入力操作、本システムを用いた音声感情入力での操作の 3 パターンによるプレイヤーのゲームに対する没入感の違いを探るために、被験者 31 名を 3 グループに分けそれぞれの操作方法でプレイしてもらい、それに対するアンケート調査を行った。本論文ではこれらの結果から本ゲームシステムの有用性についての検証を行う。

本論文の構成は以下のようになっている。第 1 章においては、本研究の背景と目的を示した。次章にて本研究に関連する研究を示し、本研究の位置付けを示す。第 3 章においては、オトノベルの提案手法の設計及びシステム概要を示す。第 4 章においては、本研究の評価実験の手順及び内容を述べ、結果と分析及び考察を示す。第 5 章においては、今後の課題と展望について議論し、第 6 章においては本研究のまとめを示す。

## 2. 関連研究

### 2.1. 音声入力でゲーム操作を行うことのできる商品

コントローラ等で操作するのではなく、自身の身体の一部でゲーム操作するという研究は数多く行われ、既にゲームに取り入れられているものも様々存在する。中でも、本研究同様自分の声でゲーム操作する音声入力システムは既に存在している。例えば、Amazon社が発表しているAlexa Game Control[4]等が挙げられる。これは音声アシスタントであるAlexaの技術を活用することで、声だけでゲームや他のソフトとやりとりを行えるようにするものである。対応するゲームソフトでは自身の音声を用いてNPCと会話をを行う、武器を交換するといった行為が行えるようになる。その上、この音声入力操作はゲームの枠内を越え、Alexaが提供するすべての機能で用いることができると言われている。例えば、スマートライトを操作してゲームのムード作りをする機能やゲームの止め時をリマインドするタイマー機能、他にもゲーム中に食事を注文することのできる機能などがある。更にこのシステムはパソコンにマイクやヘッドセットを接続するだけで使用できる為、難しい知識や技術を必要としない。この技術が初めて搭載されたゲームは、ゾンビサバイバルゲームの続編である「Dead Island 2」である。本研究の提案システムではAlexa Game Controlと同様に声でNPCとの会話をを行うことができ、更に声色から感情認識を行えるようにした。感情を込めてセリフを読ませることで、より一層主人公に近い感覚でゲームをプレイしてもらうことができると考える。

### 2.2. 音声入力を用いたゲーム没入感向上に関する研究

本研究同様、ゲームへの没入感を高めるための研究は様々行われている[5][6]。今までに音声入力の技術を用いて作成されたゲームをまとめた研究によると、小川らは、発音の向上に繋がる音声入力ゲームの提案をしている[5]。ゲーミフィケーションの特性を活かしモチベーションを刺激し、楽しみながらゲームをプレイしているうちに発音の向上に繋がるという。このように、音声を用いたゲームというのは発音の練習にも繋がり得ると言える。この研究では感情認識は行っていないが、本研究では音声入力に感情認識を取り入れることにより、発音だけでなく抑揚をつけて発話するためのスキルを身につけることも可能であると考える。また、大山はアドベンチャーゲームにおける選択肢においてプレイヤの発話の音声認識及び表情から感情推定を組み合わせた手法を提案している[6]。この研究ではプレイヤは入力の際に選択するテキストを読み上げると同時に表情を作成し入力を行う。そしてカメラで映る発話時の表情を読み取り、感情推定を行っている。本研究は音声入力とその音声の声色から感情を取得している為、表情からの感情推定という点で本研究とは異なる。

また、音声認識技術を用いたゲームを総合的に評価している研究[7]では、プレイヤやアバタのアイデンティティは、ゲームにおける音声対話とは切っても切れない関係にあり、没入感の向上に貢献している可能性を指摘している。

しかし、今までに音声入力とその音声から感情推定を行うことを組み合わせた入力手法がノベルゲームに用いられたことはない。

### 2.3. 音声感情認識に関する研究

音声に含まれる基本的感情を分析している研究[8]では、平賀らは「怒り」「喜び」「悲しみ」「嫌悪」の4種とし、「平静」音声とピッチ周波数・振幅の変化パターンから比較を行い分析している。同様に、門谷らの研究[9]では、従来の対話システムにおいては話者の心理状態に関わらず、一定調のレスポンスを行うだけであった。音声認識応答システムに人間の能力に近い機能を持たせるよう、まず感情を含んだ音声を従来の連続音声認識システムで認識し、どのような影響があるのかを調べている。本研究では、このような音声感情認識技術をゲームに取り入れている。

### 3. オトノベル

#### 3.1. システム概要

本研究で開発したゲームは全て NPC との会話でストーリーが進行していくノベルゲームとなっている。主人公は登場する NPC 達と会話をしながら、NPC の問い合わせ等に対しセリフを選択しストーリーを進めていく。ストーリーは複雑に分岐が行われる為、プレイヤの選んだ選択によって結末が変わる。本研究ではゲーム上で自分を主人公に重ね合わせやすくする為に、主人公を画面上には表示させず、主人公目線で会話を行えるようにした。また、ゲームに対する没入感を高める為に、音声入力と声色感情認識の 2 つに着目した。ノベルゲームはキャラクタとの会話が主体であるが、通常のゲームでは会話自体はキーボードやコントローラを用いて選択するのが基本であり、会話において最も重要である声を使わない。プレイヤを NPC と実際に会話を交わしている感覚にさせるため、本システムでは音声入力という方法を用いて選択肢のセリフを読み上げるように設計した。これにより、現実と近い感覚でキャラクタとの会話を楽しむことができる。また音声入力という誰でも簡単に操作できる手法を用いることで、アクセシビリティ面の性能を高め、小さな子供や機械に疎いお年寄り、身体の不自由な人でも気軽に操作できるようになることも期待している。

次に、主人公への感情移入をより強く促すために、読み上げた音声から感情を識別できる音声感情認識を取り入れた。本システムでは図 1 のように、セリフの最後にどのような感情で読み上げるべきか感情表現テキストで表示させている。こうすることで主人公と同じ感情でセリフを読み上げさせ、主人公への強い感情移入を促している。また、言い方によっては意味合いや相手の取り方が異なる言葉があることに着目し、図 2 のように同じセリフでも感情表現テキストのみを変えたセリフ選択場面を取り入れ、感情認識の面白さを追求した。



図 1 オトノベルセリフ選択画面

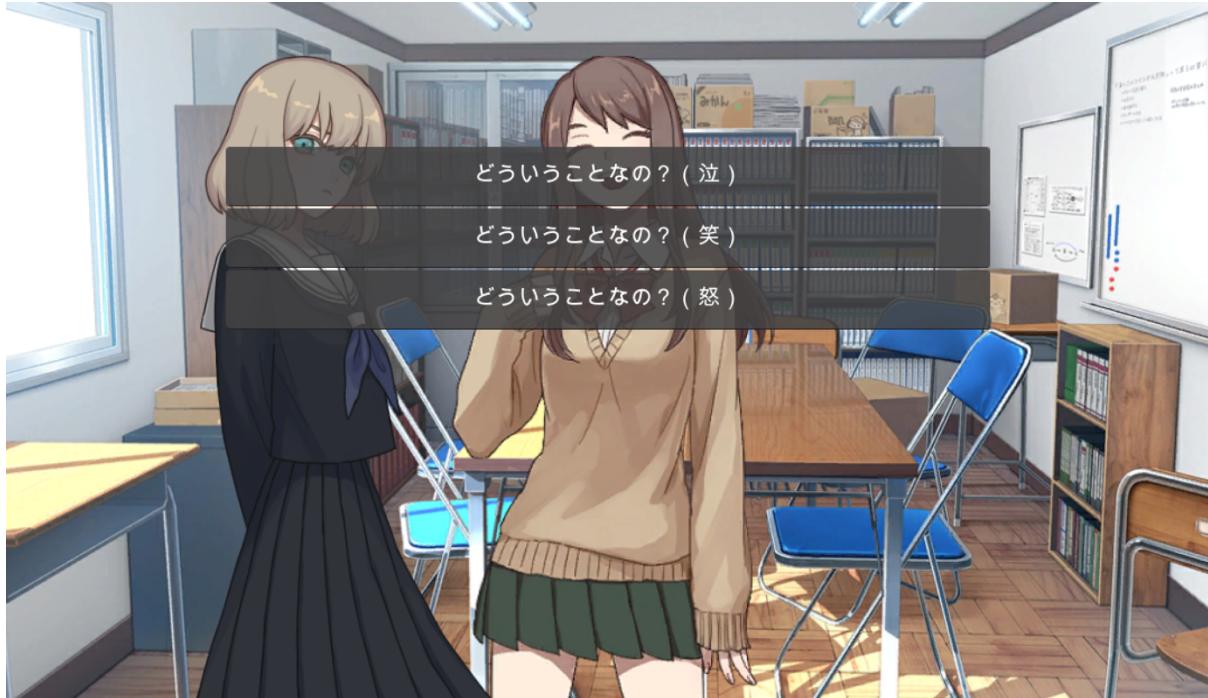


図2 感情表現テキストのみ異なるセリフ選択画面

### 3.2. システム構成

#### 3.2.1. ゲーム開発プラットフォーム「Unity」

Unity[10]とは、Unity Technologies社が開発したゲームプラットフォームである。主にC#を用いたプログラミングが可能であり、PCにUnityをインストールするだけで誰でも簡単に開発を行うことができる。様々なプラットフォームに対応しているが、本研究ではPC向けのゲーム開発を行なった。

#### 3.2.2. Unityアセット「Fungus」

Fungus[11]とはUnityのアセットの1つである。ビジュアルスクリプティングを行うことができ、ノベルゲームにおいて重要なセリフやキャラクターの立ち絵の表示、画面の暗転及び背景の変更、選択肢によるストーリー分岐などのコマンドが豊富に用意されている。本研究で作成したノベルゲームは全てUnityで開発を行っており、キャラクタとの会話場面はUnityのアセットであるFungusを利用し作成した。

#### 3.2.3. 音声文字起こしAPI「Google Cloud Speech-to-Text」

音声入力及び文字起こしに関しては、Google社の提供するGoogle Cloud Speech-to-Text[12]というAPIを用いている。特定のキーが押されるとUnity上でそのAPIにアクセスし音声入力をを行う。本システムでは、現状では発音が英語に聞こえてしまうと入力されないという問題点があるため、全て選択肢は日本語のみとし、英語に聞こえやすいものは選択肢として用意していない。

#### 3.2.4. 感情認識API「Empath」

音声から感情認識を行うAPIはEmpath API[13]を利用した。Empath APIとは株式会社Empathが提供しているAPIで、音声等の物理的な特徴量からその時の気分や気持ちの浮き沈みを判定するプログラムである。数万人の音声データベースを元に喜怒哀楽等を判定している。音声入力時にGoogle Cloud Speech-to-TextのAPIまたはこのAPIにUnity上からアクセスする。図3のように、入力された音声から感情値を「calm(穏やかさ)、anger(怒り)、joy(喜び)、sorrow(悲しみ)、energy(気力)」の5項目で数値化し、その値によるセリフの感情分析を

行う。しかし、本ゲームにおける選択肢の選択は基本的に文字列の一致度判定で行われる。そもそも文字列として異なる選択肢の場合、感情判定は行わない。感情の付与度合いはプレイヤーの自主性に任せている。一方で選択肢が同一で付与すべき感情のみが異なる場合、文字列の一致度判定は行わず感情判定のみを行い、最も近い選択肢が選択される。Empath の音声入力はシステム上最大 5 秒までしか入力を受け付けないので、6 秒以上の入力はエラーとなってしまう。その為、読み上げに 6 秒以上かかるような選択肢は用意しないという配慮、及びセリフの読み上げは 6 秒以下にするよう説明が必要である。

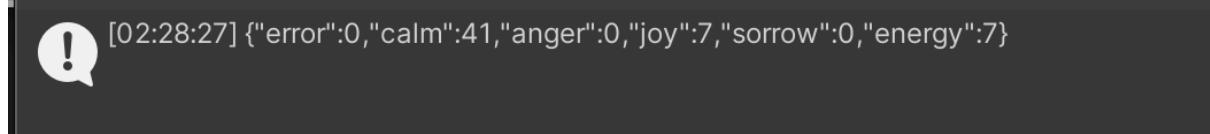


図 3 Empath API を用いて感情値を取得した結果の例

### 3.2.5. システムの流れ

提案システムの構成及び流れについて説明する。ゲームの制作に関しては Unity を用いて製作した。セリフの選択に関しては、特定のキーが押されている間に Google Cloud Speech-to-Text と通信を行うように設計している。キーが押されると音声入力が開始され、キーが離されると入力された音声が Google API 上に送られる。その音声をテキスト化したもの Unity 上に返し、Unity 上でそのテキストと選択肢のセリフテキストと一致しているかの識別を行う。また、プレイヤーの発音によっては音声がセリフと合致しない可能性があるため、テキストの識別時には Google API から Unity 上に返された文字列と選択肢の文字列とのレーベンシュタイン距離を計算し、より小さい方の値が選ばれるようにしている。こうすることで、一字一句合致していない場合でも発言と近い選択肢が選ばれるように実装した。しかし、Google API によりテキスト化した文字列と選択肢の文字列が図 3 のようにひらがな、カタカナ、漢字の変換が一致しない場合、別の文字列として判定されてしまう。この問題点を解決する為、すべての文字列をひらがなに変換し文字列の一致度判定を行う。

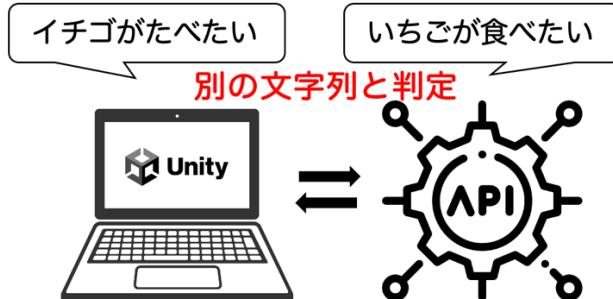


図 3 文字列の一致度判定

## 4. 評価実験

### 4.1. 実験手法

本システムをキーボード操作、音声入力操作、音声感情入力操作の 3 つの操作方法によって比較する実験を行う。

#### 4.1.1. 仮説

ゲームに対する没入感、主人公への感情移入、ゲームの楽しさ、音声入力及び音声感情入力の抵抗感の 4 項目について、本研究で作成したノベルゲームにおいてキーボード操作、音声入力操作、音声感情入力操作の 3 種類で操作した場合どのような差が出るか仮説を立てた。

ゲームに対する没入感において音声感情入力操作、音声入力操作、キーボード操作の順に優位になる（仮説 1），主人公への感情移入において音声感情入力操作、音声入力操作、キーボード操作の順に優位になる（仮説 2），ゲームの楽しさにおいて音声感情入力操作、音声入力操

作，キーボード操作の順に優位になる（仮説3），抵抗感においては音声入力と音声感情入力の間に有意差はない（仮説4）の4点が本実験における仮説である。

これらのことと検証する為，以下の実験を行う。

#### 4.1.2. 実験概要

被験者31名をAグループ（キーボード操作10名），Bグループ（音声入力操作11名），Cグループ（音声感情入力操作10名）の3つに分け実験を行った。図5のように本研究で開発したゲームをそれぞれの操作方法でプレイしてもらい，ゲームの没入感，主人公に対する感情移入，ゲームの楽しさに関してアンケートを行った。グループBに対しては，音声入力に対する抵抗感，グループCに対しては音声入力の抵抗感と感情を込めて読みことに対する抵抗感に関する項目も回答してもらった。実験時間は実験者がスタートボタンを押してから20分間であり，終了時にはゲームの進行状況に関係なくアンケートに移るものとした。また，感情入力を行わないAグループ及びBグループに対しては感情表現テキストによる感情指定は行わなかった。

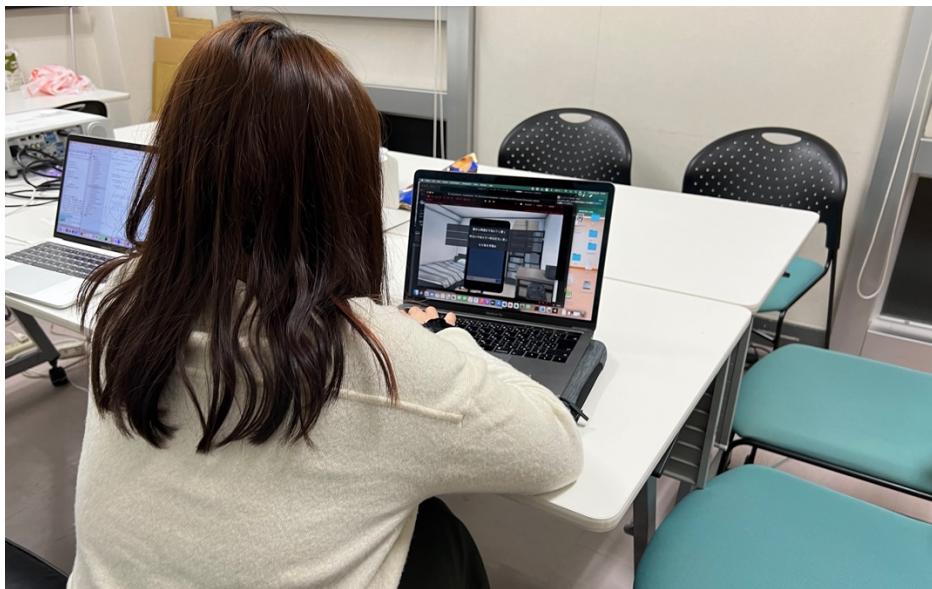


図5 プレイしている様子

#### 4.1.3. 実験手順

グループAには実験を始める前に，被験者にオトノベルの概要及び操作方法を説明した。キーボード操作でゲームをプレイするよう指示を行い，音声入力及び音声感情入力に関する説明は一切行わなかった。その後，ゲームを20分間プレイし，アンケートに回答してもらった。

グループBには，グループA同様ゲームの概要の説明を行った後，音声入力の操作方法についても説明を行い，基本的な操作は全て音声入力で行うよう指定した。その後，20分間ゲームをプレイし，アンケートに回答してもらった。

グループCには，グループA，B同様にゲームの概要及び音声入力操作方法に関する説明を行い，加えて感情表現テキストによる感情指定の説明を口頭で行った。感情によるセリフ選択に関する説明も行い，なるべくセリフに感情を込めて読み上げるよう求めた。その後ゲームを20分間プレイし，アンケートに回答してもらった。なお，本実験においては感情表現を強制された際のプレイヤの心理にのみ興味がある為，実際の感情認識機能は活用していない。選択肢はすべて十分異なる文字列のセリフとなっており，音声認識のみで実験協力者の意図するものを識別可能にした。実験において，全ての実験協力者は指示された感情を込めた発話をを行っていたことを確認した。

#### 4.1.4. 評価アンケート

本研究の評価アンケートは全て7段階で評価を行った。アンケート内容は①ゲームに対する没入感を感じたか、②主人公に対して感情移入をしたか、③ゲームの楽しさを感じたかの3項目について回答してもらった。また、入力方法について抵抗感を探る為グループB、グループCに対しては音声入力及び音声感情入力に対して抵抗感の有無についてのアンケートも行った。

#### 4.2. 実験結果

アンケートの結果を表1、表2、表3に示す。

表1 グループAのアンケート結果

被験者ID	Q1.ゲームに対する没入感を感じたか	Q2.主人公に感情移入したか	Q3.ゲームは楽しかったか
1	3	5	7
2	5	6	7
3	4	5	7
4	4	5	7
5	5	7	7
6	4	5	7
7	4	5	7
8	6	4	7
9	6	4	7
10	4	5	7
平均	4.5	5.1	7
分散	0.85	0.69	0
標準偏差	0.921954446	0.830662386	0

表2 グループBのアンケート結果

被験者ID	Q1.ゲームに対する没入感を感じたか	Q2.主人公に感情移入したか	Q3.ゲームは楽しかったか	Q4.音声入力に対する抵抗はあったか
11	6	6	7	5
12	5	6	7	5
13	6	6	7	2
14	7	6	7	6
15	7	7	7	4
16	7	7	7	1
17	7	7	7	1
18	5	7	7	1
19	6	5	7	1
20	7	7	7	2
21	7	7	7	1
平均	6.363636364	6.454545455	7	2.636363636
分散	0.595041322	0.429752066	0	3.504132231
標準偏差	0.771389216	0.655554777	0	1.871932753

表3 グループCのアンケート結果

被験者ID	Q1.ゲームに対する没入感を感じたか	Q2.主人公に感情移入したか	Q3.ゲームは楽しかったか	Q4.音声入力に対する抵抗はあったか	Q5.感情を込めて読むことに抵抗はあったか
22	6	5	7	1	1
23	4	6	7	1	7
24	5	5	6	3	2
25	7	5	7	5	1
26	7	7	7	1	2
27	5	5	7	1	3
28	7	5	6	1	2
29	4	4	3	5	5
30	6	4	7	6	6
31	6	5	7	2	7
平均	5.7	5.1	6.4	2.6	3.6
分散	1.21	0.69	1.44	3.64	5.24
標準偏差	1.1	0.830662386	1.2	1.907878403	2.289104628

以下、自由記述をそれぞれまとめる。

グループAでは、「主人公が可哀想で同情した」、「主人公が男にも関わらず女である自分も楽しめた」等、ゲームの内容に関する感想が殆どであった。

グループBではグループA同様ゲームの内容に関する感想に加え、音声入力及びそれに対する抵抗感に関する感想が多く見られた。音声入力に関する感想では、「自分の声で選択肢を読み上げるため、通常のゲームよりも没入感が得られると感じた」、「声で操作することにより、主人公になったような気持ちになれた」、「声に出すことでのその言葉を相手に伝えるという意識が強くなった」、「主人公のすべてのセリフを音読してプレイした為、非常に没入感があった」、「キーボード操作であれば選ぶ選択肢も、自分が読み上げるとなるとリアルな自分が言いそうな選択肢を選んでしまった為、主人公がより自分に近く感じた」等の意見があった。抵抗感に関しては、「音声入力の抵抗は全く感じなかった」、「音声入力に対する抵抗の部分に関してはあまり感じられなかつたが、親しくない人や大勢の中等で行えば抵抗感を感じると思う」といった意見が見られた。他にも、「ストーリー的に感情移入しやすかったのか、音声入力だから感情移入しやすかったのか正直分からなかつた」といった指摘もあった。

グループCでは、感情入力に関するコメントが多く見られた。「感情を込めてセリフを読まなければならなかつた為、声優になった気分だった」「最初は抵抗があつたが、言っているうちに楽しくなってきて非常に没入感があつた」という前向きな感想があつた一方で、「感情を込めるのが恥ずかしかつた」、「音声入力は良いが感情を込めるのは恥ずかしい」、「見られていると思うと恥ずかしかつた」等の感想も多かつた。

#### 4.3. 分析

##### 4.3.1. 分析方法

ゲームに対する没入感、主人公への感情移入、ゲームの楽しさの3項目に対し平均値と標準偏差を調べた。クラスカルウォリス検定を行い、各操作方法での感じ方を比較した。またライアンの方法で多重比較を行つた。

##### 4.3.2. 分析結果

ゲームに対する没入感、主人公への感情移入、ゲームの楽しさの3項目に関するアンケートに対し、それぞれ分析を行う。分析の結果を図6、図7、図8に示す。図6は、ゲームに対する没入感の質問に対して、0を「全く感じなかつた」、7を「とても強く感じた」としたときの平均値と標準偏差のグラフである。グループAの平均値は4.5、グループBは6.4、グループCは5.7である。3つのグループの平均値についてクラスカルウォリス検定を行うと、その結果3つのグループの間に有意差が見られた( $p<0.001$ )。ライアンの方法で多重比較をしたところ、グループAとグループBの間には有意差が見られた( $p<.01$ )。また、グループAとグループCの間にも有意差が見られた( $p<.05$ )。これは、音声入力を用いた操作方法はゲームに対する没入

感を高められた為であると考えられる。一方で、グループBとグループCの間には有意差は見られなかった ( $p=0.168$ )。グループCの自由記述において、感情を込めて話すことには恥ずかしさを感じたという内容のコメントが多く見られたことから、音声入力時にゲームに集中できていなかつたことが原因だと考える。

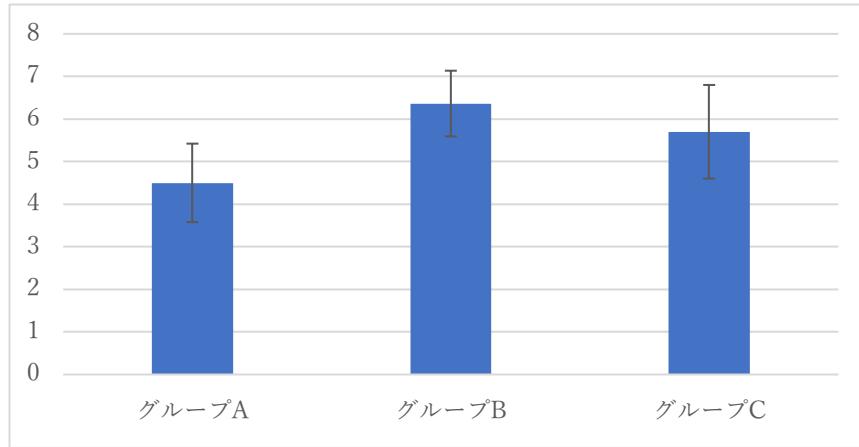


図6 「ゲームに対する没入感」の平均値と土標準偏差

図7は、主人公への感情移入の質問に対して、0を「全く感情移入しなかった」、7を「強く感情移入した」としたときの平均値と標準偏差のグラフである。グループAの平均値は5.10、グループBは6.46、グループCは6.40であった。3つのグループの平均値についてクラスカルウォリス検定を行うと、その結果3つのグループに有意差が見られた ( $p=0.005$ )。ライアンの方法で多重比較をしたところ、グループAとグループB ( $p<.01$ )、及びグループAとグループC ( $p<.01$ ) の間にも有意差が見られた。故に音声認識を用いる方法は、主人公への感情移入をより強く促すことができると考えられる。一方で、グループBとグループCの間には有意差は見られなかつた ( $p=0.567$ )。感情を込めてセリフを読み上げることによる主人公への感情移入の度合いの差も見られなかつた。

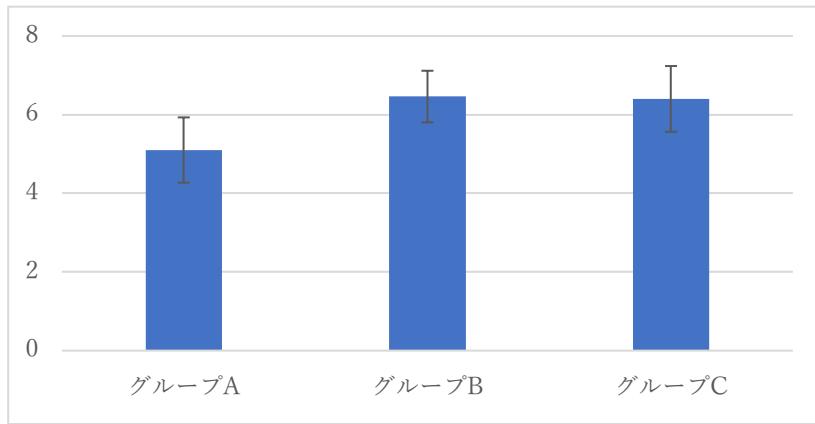


図7 「主人公への感情移入」の平均値と土標準偏差

図8は、ゲームの楽しさに関する質問に対し、0を「全く楽しくなかつた」、7を「とても楽しかった」としたときの平均値と標準偏差のグラフである。グループAの平均値は7.0、グループBは7.0、グループCは6.2となった。3つのグループの平均値についてクラスカルウォリス検定を行うと、その結果有意差は見られなかつた ( $p=0.61$ )。従って、ゲームの楽しさは操作方法にさほど左右されないと考えられる。

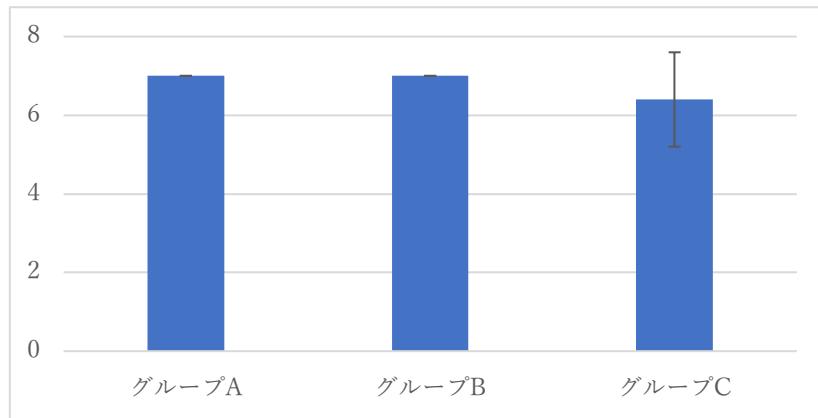


図8 「ゲームの楽しさ」の平均値と標準偏差

図9は、音声入力及び音声感情入力への抵抗感に関する質問に対し、0を「全く感じなかつた」、7を「強く感じた」としたときの平均値と標準偏差のグラフである。グループBの平均値は2.64、グループCは3.60であった。2つのグループの平均値についてマン・ホイットニー検定を行うと、その結果2つのグループに有意差が見られなかった( $p=0.247$ )。音声入力と音声感情入力の間に抵抗感の差はないといえる。

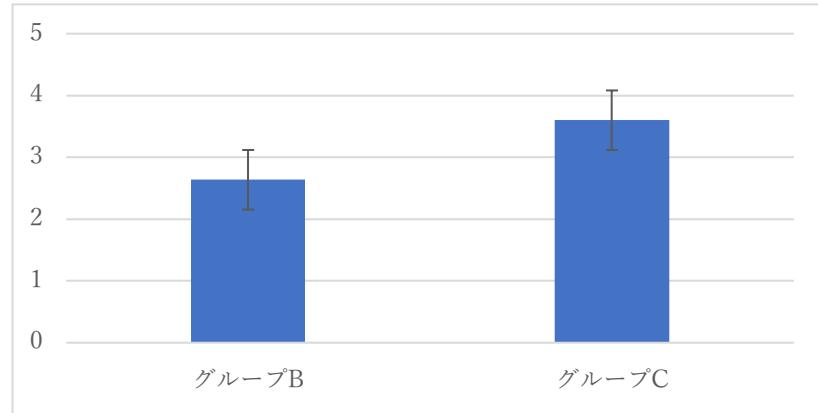


図9 音声入力及び感情入力への抵抗感

#### 4.4. 考察

##### 4.4.1. ゲームに対する没入感についての考察

評価実験の結果、ノベルゲームにおいて自身の声で操作することでゲームに対する没入感を高めることができると分かった。ノベルゲームで展開されるストーリーは、実際の世界では会話によって行われる行為である為、音声入力の方がNPCと会話している感覚をより実感することができるのかもしれない。しかし、音声感情入力は音声入力より優位になると予想していたが、仮説1「ゲームに対する没入感において音声感情入力操作、音声入力操作、キーボード操作の順に優位になる」は支持されなかった。音声感情入力に関しては抵抗感があったと回答する人が多かったことや、自由記述に恥ずかしさを感じたという感想が多く見られたことから、そのような抵抗感や羞恥心はゲームへの没入感を妨げてしまう要因になり得ると考える。また、「見られていると思うと恥ずかしかった」、「音声入力に対する抵抗の部分に関してはあまり感じられなかったが、親しくない人や大勢の中等で行えば抵抗感を感じると思う」という感想から、実験の際に被験者の隣に実験者がいたことが羞恥心や抵抗感を生んでしまったとも考えられる。その為、他人がいない個室で実験を行う、またはオンラインでPC越しに実験を行う等の対策が必要である。

#### 4.4.2. 主人公への感情移入についての考察

ゲームの世界では、世界各地を一瞬で飛び回る、サバイバル生活を楽しむ等、現実の世界ではあり得ない体験をすることが可能である。その為、ゲーム内では性別を変える、自分とは全く違うキャラクタになりきるという楽しみ方もできる。ノベルゲームでは現実とは異なる世界で自分が選ばないような言動をすることができる。しかし、アンケートの自由記述では「キーボード操作であれば選ぶ選択肢も、自分が読み上げるとなるとリアルの自分が言いそうな選択肢を選んでしまった為、主人公がより自分に近く感じた」という感想があった。つまり、いざ口に出すとなると現実で自分が言いそうな選択肢を選ぶ傾向があると考えられる。また本システムの主人公が男性であったことから、選択肢のセリフが男性口調であるものが度々存在した。実験後に女性の被験者から、「男性口調のものを読み上げるのはやや抵抗感があった」という意見をあったことから、やはり音声入力において現実の自分が言いそうなセリフを選んでしまいがちであることが分かる。主人公の言動が自分に近くなる為、自分がその世界にいるかのような感覚に陥ることができ、キャラクタと自分自身を重ね合わせることから、感情移入を促せると考察できる。また評価実験の結果からグループAとグループB及びグループAとグループCに有意差が見られたことからも、音声入力を用いることで通常のキーボード操作よりも感情移入を促すことができたと分かる。一方で、ゲームに対する没入感と同様に感情移入に関しても音声感情入力操作が最も優位になると予想をしていたが、こちらも予想とは反する結果となつた。グループBとグループCの間に有意差が見られず、仮説2「主人公への感情移入において音声感情入力操作、音声入力操作、キーボード操作の順に優位になる」は支持されなかつた。本システムでは感情入力を感情表現テキストによって指示していた為、自分の感情と異なつていた場合に感情移入しづらい部分があつたことが原因ではないかと推測する。更に、没入感同様、実験者が被験者の隣にいたことによる羞恥心も影響を及ぼしている可能性がある。

#### 4.4.3. ゲームの楽しさについての考察

ゲームの楽しさには有意差が見られなかつた為、仮説3「ゲームの楽しさにおいて音声感情入力操作、音声入力操作、キーボード操作の順に優位になる」は支持されなかつた。今回の実験では単一のストーリーを用いており、そのストーリーの性質がゲームの楽しさに強い影響を与えてしまつた為、入力手法の差が出づらかった可能性が考えられる。内容の異なるストーリーを複数作成し、実験をし直す必要がある。

#### 4.4.4. 抵抗感についての考察

音声入力というものは、評価実験アンケートからも分かるように比較的抵抗感を感じやすい。しかし、感情音声入力という自分の感情を声で表現するという行為には、音声入力とさほど抵抗感の差がないことが分かつた。自由記述から、「最初は抵抗があったが、言っているうちに楽しくなってきた」とあるように、最初は抵抗感があるが発話しているうちに慣れてくるという場合も考えられる。以上より、仮説4「抵抗感においては音声入力と音声感情入力の間に有意差はない」は支持された。ここで、ジェンダー差による音声入力と音声感情入力の抵抗感に着目した。音声入力を行ったグループAとグループBの21人を男性9人、女性12人で分けて平均値と標準偏差を取ったところ図10のようになった。全体的に、男性よりも女性の方が音声入力に対する抵抗感を感じたことが分かる。また、音声感情入力を行ったグループCの10人を男性5人、女性5人に分けて平均値と標準偏差を取ったところ、図11のようになつた。感情音声入力も同様に、女性の方が抵抗感をより感じたことが示された。発話及び感情表現に対する抵抗感はジェンダー差による影響があると推測できる。

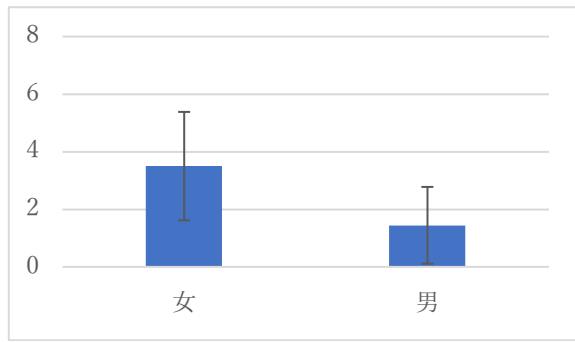


図 10 「音声入力への抵抗感」の男女差

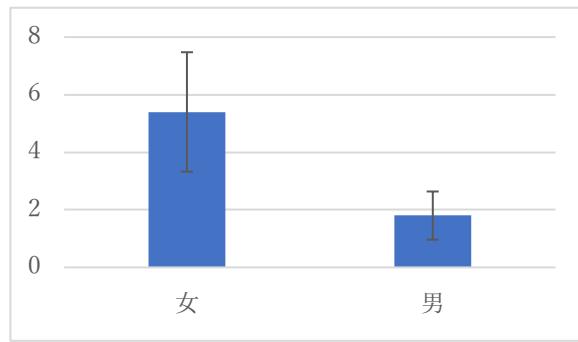


図 11 「音声感情入力への抵抗感」の男女差

## 5. 今後の発展と展望

実験における被験者の感想から「ストーリー的に感情移入しやすかったのか、音声入力だから感情移入しやすかったのか正直分からなかった」という指摘があった。本システムは現状プレイできるストーリーは1つとなっている為、複数のストーリーを作成し感情移入及び没入感を促すことが可能であるか調べる必要がある。

また、音声感情入力による没入感や感情移入、楽しさへの有意差が見られなかつた為、感情入力の方法に改善が必要であると考える。感情表現テキストによって感情を指定するだけではなく、自然と感情を込めてセリフを読み上げることをプレイヤに促すことにより、没入感を更に得られるのではないかと考える。例えば、自分自身の感情による声色でストーリーが分岐する、あまり小さい声で発言するとNPCに「聞こえない」と指摘される等のような改善を行う。これにより、指定された感情で発言するのではなく、プレイヤ本人の感情をゲーム操作に使うことができ自然に音声感情入力を行うことができる為、音声入力のみよりも没入感が高まることが期待できる。また、本システムでは感情指定に感情表現テキストを用いたが、絵文字や顔文字などのエモティコンで指示を行うことで、指示できる感情の幅が広がると考える。

最後に、音声入力は未だ社会受容性が低いという現状がある。その為本システムは室内でプレイすることを予想しPC向けに開発を行ったが、様々なプラットフォームで開発を行うことができるUnityの特徴を活かし、スマートフォンでもプレイできるように改善を行うことも視野に入れていきたい。

## 6.まとめ

本研究では、自身の声でストーリーを進めることのできる音声感情認識ゲーム「オトノベル」の提案と実装を行った。本システムの入力手法がゲームに対する没入感及び主人公に対する感情移入、ゲームの楽しさを向上させ、入力方法による抵抗感が変わらないかを調べる為、評価実験を行った。同時に本ゲームの有用性も検証した。評価実験では計31名にオトノベルを体験してもらい、アンケートを行った。アンケートの結果、音声入力はゲームに対する没入感および主人公への感情移入を強く促すことができるという結果が得られたが、音声感情入力は音声入力との有意差が見られなかった。原因として、周りの環境が音声感情入力への羞恥心生んでしまったことや、システム上プレイヤの自然な感情を入力に用いることができなかつたことが原因であると考えられる。また、今回の実験では入力方法によるゲームの楽しさにあまり差は見られなかつた。おそらくゲーム内容が楽しさに影響を与えてしまったことが原因であると推測する。入力方法に対する抵抗感については、音声入力と音声感情入力の間に有意差は見られなかつた為、本システムの入力方法の抵抗感は既存音声入力手法と著しく異なると考察する。今後は、実験方法及び感情入力方法を再検討し、没入感及び感情移入、楽しさを高められることを目指す。

## 7. 謝辞

本研究を行うにあたり、ご指導を頂いた栗原一貴教授に深く感謝を申し上げます。また、実験を実施するにあたり実験協力者を快く引き受けてくださった研究室の皆様、津田塾大学の皆様、他大学の皆様、その他の皆様にも心より感謝申し上げます。

## 参考文献

- [1] DFC INTELLIGENCE (2022) : Global Video Game Consumer Segmentation. 入手先<  
<https://www.dfcint.com/product/video-game-consumer-segmentation-2/>>
- [2] ELEMINIST, 【2022年】最新世界人口ランキング 日本は11位、前年より50万人減, <  
<https://eleminist.com/article/2084>> 参照日 (2023/01/06)
- [3] Banos, R. M., Botella, C., Alcaniz, M., Liano, V., Guerrero, B., & Rey, B. (2004). Immersion and emotion: Their impact of sense of presence. *Cyberpsychology & Behavior*, 7, 734-741.
- [4] Amazon : Alexa Game Control,  
<<https://www.itmedia.co.jp/news/articles/2208/25/news108.html>> (参照日 :  
2022/12/22)
- [5] 小川直希, 平賀瑠美 : ゲームによる発音の向上を目的とした音声を用いたゲームの開発, 研究報告アクセシビリティ (AAC), pp1-4 (2019)
- [6] 大山智弘 : アドベンチャーゲームにおけるユーザ体験向上のための感情推定技術を用いた入力手法, 研究報告ヒューマンコンピュータインタラクション (HCI), pp1-8 (2022.4)
- [7] Marcus Carter, Fraser Allison, John Downs, and Martin Gibbs. Player Identity Dissonance and Voice Interaction in Games. In Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15). Association for Computing Machinery, New York, NY, USA, pp265-269 (2015) .
- [8] 平賀裕, 斎藤 善行, 森島 繁生, 原島 博 : 音声に含まれる感情情報抽出の一検討, 進学技報 93, pp1-8 (1994)
- [9] 門谷信愛希, 阿曾弘具, 鈴木基之, 牧野正三 : 音声に含まれる感情の判別に関する検討, 情報処理学会研究報告音声言語情報処理 (SLP), pp37-42 (2000)
- [10] Unity Technologies : Unity, 入手先 <<https://unity.com/ja>> (参照日 : 2022/12/23)
- [11] : Fungus, 入手先 <<https://github.com/snozbot/fungus>> (参照日 : 2022/1/6)
- [12] Google : Speech-to-text, 入手先 <<https://cloud.google.com/speech-to-text/>> (参照日 : 2022/12/26)
- [13] 株式会社 Empath : Empath API, 入手先 <<https://webempath.net/lp-jpn/>> (参照日 2022/12/23)