

仮想聴衆における聴衆規模に応じた 適切なうなずき頻度生成のための予備調査

坪内太吾¹ 福地健太郎¹

概要：多数のCGキャラクターにより仮想聴衆を形成し、利用者の話に対してうなずきを行うことで仮想聴衆が利用者に耳を傾けているように感じさせる研究を我々はこれまで進めている。本研究では発話の言語的内容を考慮しない単純な音声駆動型うなずき生成モデルを用いて仮想聴衆をうなずかせているが、仮想聴衆が一人の場合に適切なうなずきを返すパラメータをそのまま複数人数の聴衆に適用するとうなずきが過剰に感じられる。そこで本研究では多人数の仮想聴衆向けに適切なうなずき頻度を探るための予備調査を行った。聴衆規模が15人までの実験の結果、人数が増えるに従ってうなずき頻度を低下させることで適切な頻度に調整できることがわかった。

キーワード：聴衆，うなずき

1. はじめに

人と人との対面コミュニケーションにおいて、うなずきをはじめとしたノンバーバルメディアが果たす役割は大きい。その用途は単に言語情報以外の内容を伝えることに限らず、むしろ言語をはじめとした複数のメディアと協力的に働き、コミュニケーションをより強固で円滑なものとするにある。

我々は、複数の仮想キャラクターがうなずくことで、ユーザーに「自分の話をちゃんと聞いてもらっている」という印象を与えるシステムの作成を目指している。一人一人と複数の仮想キャラクターとで対話コミュニケーションを対象にした研究は多くある。例えばVR曝露療法による不安障害の治療では、社会的不安を再現するために聞き手キャラクターが複数体必要であり[1]、VR内でプレゼンテーションの練習をするシステムでは、人前に立つ経験の再現性を高めるため、大勢の仮想キャラクターを配置し、聞き手の言動に合わせて反応させている[2][3]。

渡辺らは、対面者同士の身体的リズムが次第に同期していく現象である「身体的引き込み」に着目し、ロボットやCGキャラクターが人間の発話に対して身体的な聞き手動作を行うInterActor[4]を開発した。そのInterActorの身体動作生成の中核を担っているのは音声駆動型うなずき生成モデル[5]であり、音声信号のみからうなずき開始のタイミングを予測し、うなずきや他の身体動作を、自然なタイミングで行うことを可能としている。また、渡辺らの他の研究に、うなずき生成モデルを用いて、複数の聞き手キャラクターによる引き込み動作が行われるアプリケーションを開発しているものがある[6][7]。これらの研究ではすべての聞き手キャラクターが引き込み動作を行っており、それによってシステム使用者に臨場感や達成感を与えられると主張している。

しかし、上記音声駆動型うなずき生成モデルを追実装し、

聴衆キャラクターにうなずき動作を行わせる試作を行ったところ、大勢のキャラクターが同時にうなずく様子は大変不自然であり、またその違和感は、単に動作開始タイミングをずらすだけでは改善されなかった。

我々はこの違和感について、複数のキャラクターがうなずく場合、人数に応じてうなずきの頻度を調整することで改善が見込めると考えた。うなずきは、発話者の発話の末尾部分でよく発生する[4]。渡辺のうなずき生成モデルは、その発話末を予測することで、うなずきのタイミングを出している。この特性は一对一の会話においても一对多の会話においても不変であると考えられる。そうであるにもかかわらず、一斉にうなずく様子が不自然であるということから、うなずきが発生するタイミングはそのままに、その頻度を下げることで、適切な反応動作の生成が可能になると考えられる。

そこで本研究では、発話音声のリズムに合わせて複数体の仮想キャラクターがうなずきを返す際の、人数に応じた最適なうなずきの頻度を調査する実験を行った。聴衆規模が15人までの実験の結果、人数が増えるに従ってうなずき頻度を低下させることで適切な頻度に調整できることがわかった。

2. 聴衆うなずきシステム

本節では、発話音声のリズムに合ったうなずきを行う仮想聴衆を呈示するためのシステムを提案する。本システムは、大きく分けて二段階で構成されている。まず一段階目に、渡辺の音声駆動型うなずき生成モデルで、与えられた発話音声のリズムに合ったうなずきのタイミングを割り出す。二段階目に、そのタイミングで、仮想聴衆を構成する内の、どのキャラクターがうなずくのかを決定する。

2.1 音声駆動型うなずき生成モデル

渡辺の音声駆動型うなずき生成モデルは、マクロ層とミ

¹ 明治大学
Meiji University

クロ層からなる階層モデルである。音声データを入力として受け取り、マクロ層からマイクロ層へ順に処理が行われ、音声リズムに合わせて適切なタイミングでうなずきを行うことができる。具体的には、マクロ層では、音声の呼気段落区分での ON-OFF 区間からなるユニット区間（図 1 参照）に、うなずきの開始時点が存在するかを予測する。マクロ層でうなずきが存在すると予測された場合マイクロ層に処理が進む。マイクロ層では、そのうなずきの開始時点がいつであるかを予測する。

基本的には上記の渡辺らの提案に準じた追実装を行ったが、特にマクロ層において、技術的な理由により、元のシステムを一部簡略化した箇所がある。

元システムのマクロ層では、音声の呼気段落区分での ON-OFF 区間からなるユニット区間にうなずきの開始が存在するかを予測するが、技術的な理由で実装が困難であった。本システムのマクロ層では実装を簡略化することとし、各ユニット区間には必ず一つうなずきが含まれるという仮定の上で、ユニット区間の識別のみを行うこととした。

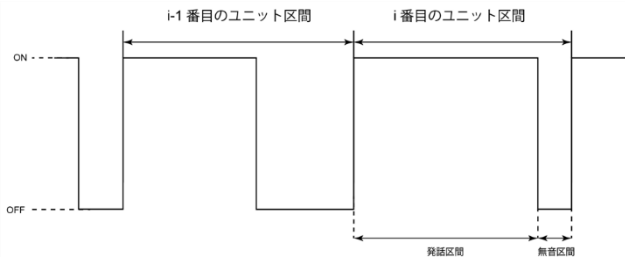


図 1 ユニット区間の並び方

図 1 に示すように、ユニット区間は前半の発話区間と後半の無音区間で構成されるため、音量がこのような並び方をしていることを検出できれば、ユニット区間を検出できたことになる。そこで、過去 2 秒間の音声データ配列 (60Hz, 120 個) と、式(1)の配列との内積をとり、得られた値がそのユニット区間で初めて 0 を下回ったら、そこでそれまでのユニットは終了し、新しいユニット区間が始まったとみなす。新しいユニット区間が始まると、マイクロ層の処理へと進む。

$$[60, 59, 58, \dots, 1, 0, -1, \dots, -59] \quad (1)$$

ただし、この実装法は音量の変化に過敏に反応してしまうという問題があった。そのため、破擦音や発話音量の強弱などの影響を受け、誤った判定をするケースが多発した。具体的には、ユニット区間を一度検出した直後、数フレームの無音のみの区間を 5 回から 10 回ほど続けざまに検出する誤りである。この誤認識を排除するため、一度新しいユニット区間を検出したら、直後の 60 フレームはユニットの区切りが検出されてもマイクロ層へと処理を移行させないものとした。

マイクロ層は渡辺らの手法をそのまま踏襲している。具体的には、音声の ON-OFF データ (60Hz, 120 個) をマイクロ層

への入力とし、(2)式を用いて、無音区間が始まるタイミングを検出する。音声データは、音量が閾値より大きければ 1、そうでなければ 0 が割り当てられる二値データである。音声データ $V(i)$ とは、 i フレーム時点のデータを指す。予測値 $M(i)$ が閾値を下回った場合、無音区間が始まったとみなし、この時点をうなずき開始のタイミングとする。これにより、ユニット区間内の発話末を予測する。なお、(2)式内の係数 64 は、予測値 $M(i)$ の振れ幅を拡大し、閾値で判別可能にするために必要となった。

$$M(i) = \sum_{j=1}^{120} 64 \times b(j)V(i-j) \quad (2)$$

$b(j)$: 線形予測係数

$V(i)$: 音声データ

2.2 複数キャラクターへの適用

各キャラクターに対して「うなずき確率」を設定する。上記うなずき生成モデルでうなずき開始のタイミングが検出されたとき、各々のキャラクターはそれぞれの「うなずき確率」でうなずく。例えばあるキャラクターのうなずき確率を 100% に設定したならば、そのキャラクターはうなずきタイミングで確実にうなずく。50% に設定したならば、概ねうなずきタイミング 2 回に 1 回、うなずくことになる。0% ならば全くうなずかない。

この処理はキャラクターごとに行われるため、うなずき確率が 100% や 0% でない限りは、多くの場合、うなずきタイミングでうなずくキャラクターとうなずかないキャラクターが混在することになる。

3. 講義音声に対する仮想聴衆うなずき実験

3.1 実験概要

本実験では、仮想聴衆のうなずきの頻度が現実の聴衆のうなずいている様と同様であると感じられるとき、その値がどのような傾向をとるのかを調査する。

実験に際して、被験者は実験プログラム・音声ファイル・実験手順書を Google ドライブからダウンロードし、実験手順書に沿って各々で実験を行い、実験結果を Google フォームで送信した。

実験プログラムは、2 節で示した聴衆うなずきシステムを実装している。入力音声としてコンピューターに出力されている音声のループバックを受け取り、画面に複数キャラクターがその音声に合わせてうなずく様子を表示する。実験時にコンピューターに出力すべき音声は、実験プログラムと同時にダウンロードされる音声ファイルとして配布した。被験者は実験プログラムが動作している間、音声ファイルを再生させておき、実験プログラムが音声ファイル

の音声に合わせてうなずき様子を観察する。なお、本プログラムはゲームエンジン Unity[8]を用いて制作され、被験者が簡単に起動できるように、exe ファイルとしてビルドしたものを配布した。

音声ファイルには、明治大学総合数理学部先端メディアサイエンス学科の、ある講義動画から抽出した 30 分程度の音声を使用した。

音声ファイルの選定基準を述べる。まず音声は、話者一人が複数の聞き手に向けて(あるいはそのように想定して)話しているものである必要があり、その発話は被験者にも聞き取ることができるように、明瞭に話されているべきである。また、発話区間と無音区間が区別できるように、BGM や効果音、発話者の身体動作による雑音など、発話音声以外の音声は可能な限り収録されていないものが望ましい。本実験では、この基準を満たした音声ファイルを用いている。

3.2 実験プログラム

実験プログラムの画面構成は図 2, 3, 4 のようになっている。

左上のドロップダウンメニューはキャラクターの数を変更する際に用いる。「15 体」「5 体」「1 体」の内から一つを選択し、対応する数のキャラクターを表示するように切り替えることができる。

右上のスライダーでキャラクターのうなずき確率を増減させる。左端は 0, 右端は 100 である。初期値は 50 に設定されており、アプリケーション起動時やキャラクター表示数を切り替えた時には、Nod Probability=50 である。

右下のボリュームメーターは、本プログラムが認識している音量を可視化している。実験前の音量調節に用いる。

本プログラムの入力音声は、コンピューターに出力されている音声のループバックである。したがって、発話音声は記録されている音声ファイルをバックグラウンドで再生しながら、本プログラムを起動することで、その発話音声に合わせてうなずき聴衆キャラクターを表示することが可能である。なお、ループバックのゲインは高めに設定しており、発話音声の些細な強弱の影響を無視し、無音区間と発話区間を明瞭に分けることができるようになっている。

表示されているキャラクターは、2 節で解説した聴衆うなずきシステムにより、入力音声に合わせてうなずき動作を行う。聴衆うなずきシステムがうなずきタイミングを検知すると、画面内の各キャラクターは右上の「Nod Probability」を「うなずき確率」として、その確率でうなずきを行う。本システムでのうなずきは、キャラクターの頭部を 12 フレームかけて下に移動させ、次の 8 フレームで元の高さまで上に移動させることで表現した。

キャラクターの外見は、丸や四角といった単純な図形の組み合わせで表現した。

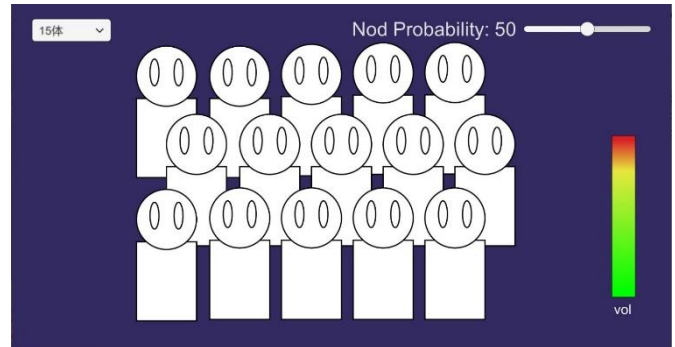


図 2 実験プログラムのキャプチャ
(表示キャラクター数 15 体)

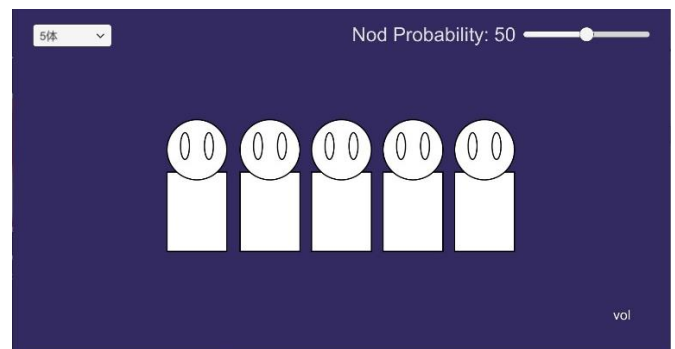


図 3 実験プログラムのキャプチャ
(表示キャラクター数 5 体)

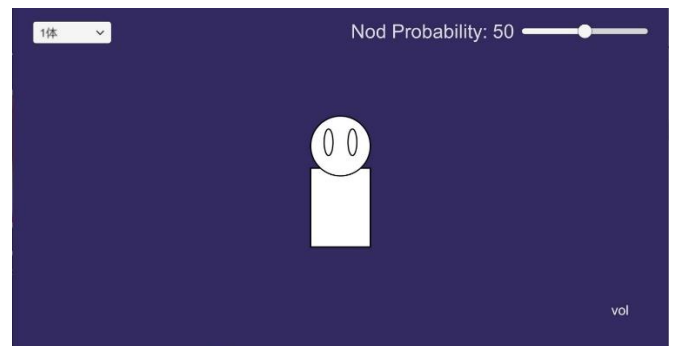


図 4 実験プログラムのキャプチャ
(表示キャラクター数 1 体)

3.3 実験手順

被験者はまず、実験を行うにあたり必要である、実験プログラム・音声ファイル・実験手順書を Google ドライブから各々の PC にダウンロードする。

実験開始に先立ち、被験者は音量調整を行う。実験プログラムと音声ファイルを開き、実験プログラムの画面右下にあるボリュームメーターが、発話区間では赤色から黄色を、無音区間では緑色を指すように音量を調節する。

音量調整ができ次第、実験を開始する。実験プログラムを開いた最初の画面での表示キャラクター数は 15 体である。講義音声に合わせてうなずきキャラクターを観察しな

がら、画面右上のスライダーで「Nod Probability」を増減させ、キャラクターのうなずきが「普段の講義のような」印象となるように調節する。調節が終わったら、実験結果送信用 Google フォームの質問の「キャラクター15 体のときの Nod Probability はいくらでしたか」に回答する。回答したら、左上のドロップダウンメニューでキャラクター表示数を 5 体に変更する。以下同様の手順で Nod Probability の調整、Google フォームへの回答を行い、キャラクター表示数を 1 体に切り替えて再び同じ手順を踏む。

以上の手順を終えたら、Google フォームで以下の質問に回答する。

3.3.1 質問項目

質問 1. あなたの学年を選択してください

質問 2. キャラクターのうなずくタイミングは適切でしたか

質問 3. あなたは授業中、よくうなずく方だと思いますか

質問 4. Nod Probability を調整した結果、うなずきを普段の講義のような感じに調整できましたか

質問 5. 講義中に教員の話をしている際、あなたはどのようなときによくうなずきますか

質問 6. 講義中に教員の話をしている際、あなたはどのようなときによくうなずきませんか

3.3.2 各質問への回答方式

質問 1 は学部一年から修士二年までの中から選択して答える。

質問 2 から 4 は五段階リッカート尺度で答える。1 が「そう思わない」、5 が「そう思う」である。

質問 5 と 6 は自由記述で答える。

3.4 実験結果

実験には大学生・大学院生計 9 名が参加した。

3.4.1 Nod Probability 値の分布

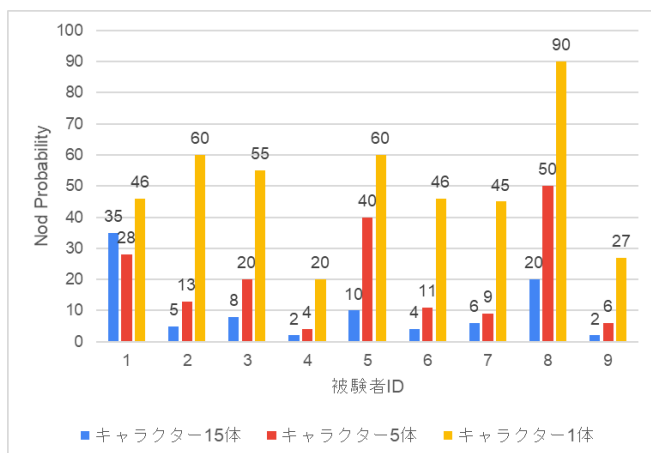


図 5 各被験者の設定した Nod Probability

図 5 は、各被験者が調整した後の Nod Probability 値を示したものである。被験者 1 番（被験者 ID が 1 の者）以外、

表示キャラクター数が減るたびに Nod Probability が増えている。このような傾向がある被験者に限っても、各被験者が設定した Nod Probability 値にはかなりのバラつきがある。具体的には、例えば被験者 3 番のキャラクター 5 体の Nod Probability と、被験者 4 番のキャラクター 1 体の Nod Probability、被験者 8 番のキャラクター 15 体の Nod Probability は等しく 20 である。

図 6 は、被験者が調整した後の Nod Probability 値の分布を箱ひげ図で示したものである。箱の内部にあるバツ印は平均値を示す。キャラクター 15 体の部分には、ひげの外に出ている Nod Probability 値の 35 が、外れ値としてプロットされている。この図からも、表示キャラクター数が減るたびに Nod Probability が増える傾向が見て取られる。また、キャラクター表示数ごとの Nod Probability もバラつきが大きいことがわかる。特に、キャラクター 15 体の Nod Probability の最小値は 20、最大値は 90 である。システムが許容する最小値が 0 で最大値が 100 であり、ほとんどこの範囲全域に渡って散らばっていることになる。

キャラクター表示数ごとの Nod Probability の平均値の差が統計的に有意であるかを確かめるため、Bonferroni 法による多重比較を行った。キャラクター表示数 15 体と 5 体との間では $t(8) = -2.36, p = 0.06$ 、15 体と 1 体の間では $t(8) = -6.33, p < 0.001$ 、5 体と 1 体の間では $t(8) = -8.01, p < 0.001$ であった。よって 15 体と 1 体、5 体と 1 体の間で、有意差が認められた。

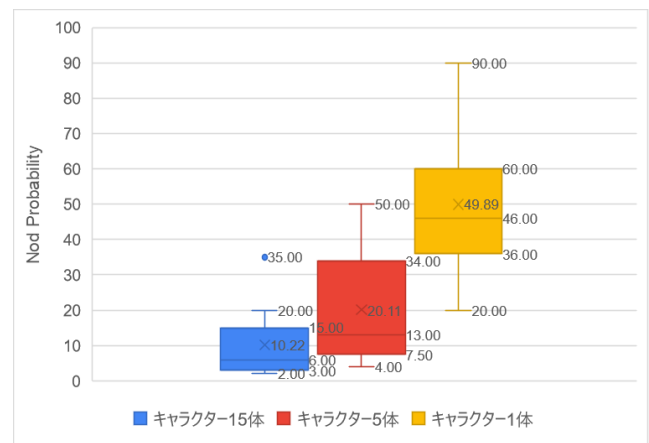


図 6 被験者が設定したキャラクター表示数ごとの Nod Probability の箱ひげ図

3.4.2 質問 1 から 4 の回答と Nod Probability との関係

表 1 は、各被験者の質問 1 から 4 への回答である。

質問 1 の回答から、被験者の学年の内訳は全 9 名中、学部四年が 7 名、学部三年が 1 名、修士二年が 1 名である。想定より学部四年に人数比が偏っているため、学年データによる分析は行わなかった。

質問 2 で、うなずきのタイミングを適切/不適切だと判断した被験者の割合は半々である。また、質問 4 の回答から、

うなずきを普通の講義のような印象に調整できたか否か、判断した被験者の割合は半々である。

表 1 各被験者の質問 1 から質問 4 の回答

被験者ID	質問1	質問2	質問3	質問4
1	学部四年	4	3	5
2	修士二年	2	4	1
3	学部四年	4	2	2
4	学部四年	4	5	4
5	学部四年	2	1	3
6	学部四年	4	1	4
7	学部四年	2	5	2
8	学部三年	2	2	4
9	学部四年	4	3	3

質問 2 の回答と、各表示キャラクター数の Nod Probability との散布図を図 7 に示す。キャラクター表示数が減るごとに、うなずきタイミングを適切でないとして評価する被験者は、Nod Probability の設定が高めである傾向が見られる。

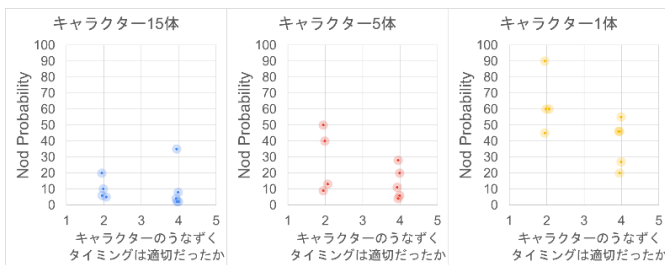


図 7 質問 2 の回答と

各表示キャラクター数の Nod Probability との散布図

質問 3 の回答と、各表示キャラクター数の Nod Probability との散布図を図 8 に、質問 4 の回答と、各表示キャラクター数の Nod Probability との散布図を図 9 に示す。ここからは明確な関係性は認められなかった。すなわち、自身を頻繁にうなずくものだと自認していること、および実験プログラムでの調整に満足していることと、設定するうなずき確率について、関連性は不明なままとなった。

表 2 は、各被験者の質問 5 と質問 6 を記したものである。質問 5 に多かった回答は、興味関心を示したとき、あるいは同意を示したときである。質問 6 にはその逆に、無関心である時という回答が多い。しかし、それに加えて、理解できないときという回答も見られる。理解できないということは、理解しようと試みてはいるということだから、興味はある程度持っているものと考えられる。これらのことから、うなずきをよく行うときは、話し手の話題に興味・関心があり、かつ理解しており、それを示したいときであると言える。

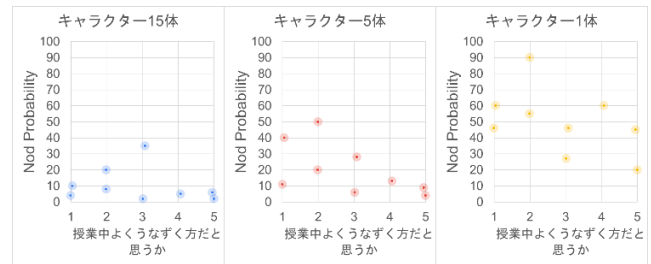


図 8 質問 3 の回答と

各表示キャラクター数の Nod Probability との散布図

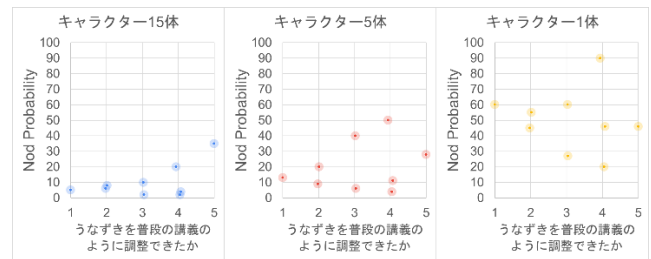


図 9 質問 4 の回答と各表示キャラクター数の Nod Probability との散布図

表 2 各被験者の質問 5 と質問 6 の回答

ID	質問5	質問6
1	暇な時	ノートをとっているとき
2	疑問形で話を投げかけられた時	板書をうつしているとき、うなずくよりも授業の内容がわからなくて、それをもう少し説明してもらいたくて首をかしげることのほうが多い
3	話が面白く、意識して聞いている時	退屈な時
4	相手の話に納得した時や、相手の話に凄いと感じたとき	相手の話が理解できないときや、相手の話に面白みを感じない時
5	なるほどなと思ったとき	話に興味を持っていないとき
6	自分の知らなかったことで興味・関心のある内容を聞いたとき	普通の内容、自分の知らなかったことでも関心のないテーマについて話しているとき
7	同意を求められたとき (~ですよね?、でしたよね?) や、講義において軸となる話の終わりにかけのタイミングなど	あまりにも既知すぎる内容、本筋とはあまりにもずれている話のとき
8	講義している先生が、聞き手に確認を求めてくる意図の間を作ったと分かるとき。	講義している先生が話している時。
9	教員の話と話の間。教員の話していることを自分が理解した時。	教員の話していることを理解できない時。

3.4.3 考察

実験結果から、キャラクター表示数が減るにつれて、うなずき確率が高く設定された。この結果が示唆するのは、仮想聴衆のうなずきが普段どおりの印象を持つためには、聴衆の人数を増やすにつれてうなずく人数を減らす必要が

あるということである。ただし、本実験では最大の人数でも 15 人しか取り扱っていない。仮想聴衆には 1000 人ほどの大人数が必要とされる例もある[2][3]。15 人よりも多い人数でもこの傾向が見られるかどうかには更なる調査が必要である。

また、キャラクター表示数が減るごとに、うなずきタイミングを適切でないと評価する被験者は、Nod Probability の設定が高めである傾向が見られた。これは、本システムにおける渡辺の音声駆動型うなずき生成モデルの再現度によって引き起こされた可能性がある。マクロ層の再現が困難であったことにより、全ユニット区間にうなずきタイミングが存在することを仮定したため、本来よりも過分にうなずきが行われていると判断されたと思われる。これにより、うなずき確率を低く設定しなかった被験者は、高く設定した被験者よりも、うなずきのタイミングが合っていないと感じたと思われる。

4. 自由発話音声に対する仮想聴衆うなずき実験

4.1 実験概要

3 節の「講義音声に対する仮想聴衆うなずき実験」(以下、講義音声実験とする)では、あらかじめ録音された講義音声に合わせてうなずくキャラクターを、被験者が観察していた。4 節で扱う本実験(以下、自由発話音声実験とする)では、被験者自身の発話を音声入力として、それに合わせてキャラクターがうなずく様子を、被験者が観察する。講義音声実験と同様なプログラムを用いて、被験者はキャラクターのうなずきを観察しながら、スライダで Nod Probability 値を操作し、キャラクターのうなずきが「普段の講義のような」印象となるように調整する。また、被験者は実験中の自由発話や実験後の聞き取り調査にて、聴衆うなずきシステムによるうなずきから受ける印象や、Nod Probability 値を決める際に取りうる戦略について回答する。これにより、仮想聴衆によるうなずきの頻度から受ける印象について、より詳細な記述を得ることを目的とする。

実験中、被験者の話題は主に、聴衆うなずきシステムによるうなずきから受ける印象や、Nod Probability 値を決める際に取りうる戦略としている。総実験時間は 30 分程度におよぶため、被験者単独では自由発話を継続して行うことに支障がある。これを補助するため、実験は被験者と実験者の対話形式で行われる。本研究では、実験者は著者が担当した。

講義音声実験とは異なり、本実験は被験者と実験者が対面するかたちで行う。マイク入力を用いている関係で音量調整が難しく、被験者単独に任せるには負担が大きいと判断したためである。

4.2 実験プログラム

講義音声実験で用いた実験プログラムの音声入力を、コンピュータに出力された音声のループバックから、マイクの音声入力に変更する。それ以外の構成は講義音声実験のものと同様である。

4.3 実験手順

音量調整ができ次第、実験を開始する。実験プログラムを開いた最初の画面での表示キャラクター数は 15 体である。実験者は被験者に自由発話をうながす。その際には、実験者の発話音声になるべく被験者の発話をさえぎらないこと、またマイクに拾われないことに極力留意する。被験者の話題は主に、聴衆うなずきシステムによるうなずきから受ける印象や、Nod Probability 値を決める際に取りうる戦略としている。被験者が自由発話に苦勞している様子が見られたら、実験者は話題に沿った質問を投げかけ、被験者の発話を補助する。被験者は、自身の発話音声に合わせてうなずくキャラクターを観察しながら、画面右上のスライダで「Nod Probability」を増減させ、キャラクターのうなずきが「普段の講義のような」印象となるように調節する。調節が終わったら、実験者が「キャラクター15体のときの Nod Probability」を記録する。その次はキャラクター表示数を 5 体にして同様の手順を踏み、それも終わればキャラクター表示数を 1 体にして同様に実験を進める。

以上の手順を終えたら、実験者は実験プログラムを終了し、引き続き被験者に質問を行う。

4.4 実験結果

被験者として講義音声実験にも参加した情報系大学生 2 名(学部 4 年生 1 名、学部 3 年生 1 名)が実験に参加した。以下、彼らを被験者 A、被験者 B と呼称する。

4.4.1 被験者 A (学部 4 年生) の場合

被験者 A の講義音声実験の結果、キャラクター表示数 15 体のとき nod probability=35、5 体のとき nod probability=28、1 体のときの nod probability=46 である。

被験者 A の自由発話音声実験の結果、キャラクター表示数 15 体のとき nod probability=34、5 体のとき nod probability=45、1 体のときの nod probability=75 である。

以下、被験者 A の評価戦略と自由発話の概要について記述する。

まず、被験者 A は表示キャラクター数が複数体である場合には、うなずいているキャラクターの配置がバラバラになることを重視していた。(縦横問わず)同じ列に並んでいる全キャラクターがうなずいた場合、即座に Nod Probability を下げていた。うなずいているキャラクターの人数をある程度確保しつつ、その配置がばらけるような、Nod Probability の均衡点を探る戦略を取っているように見受けられた。表示キャラクター数が減るほど、必然的にうなずくキャラクターの配置がかたまるケースは減るため、Nod Probability はそれに伴って上がっていった。

表示キャラクター数が1体のときとプログラム終了後には、話者の注意が如何に聞き手に向いているかがうなずきの頻度を変えると指摘していた。一対一でほとんど顔を突っつき合わせている場合には、ずっとうなずいていても違和感はないと述べており、また、教師の視線が生徒に向いている場合はよくうなずき、そうでない場合はうなずく意味はないとも述べている。

うなずきタイミングが適切であるかどうかという質問には、うなずき確率の調整をすれば違和感を無くすことができるのでそこまで意識していないという旨を答えた。

4.4.2 被験者 B (学部3年生) の場合

被験者 B の講義音声実験の結果、キャラクター表示数 15 体のとき nod probability=20, 5 体のとき nod probability=50, 1 体のときの nod probability=70 である。

被験者 B の自由発話音声実験の結果、キャラクター表示数 15 体のとき nod probability=40, 5 体のとき nod probability=40, 1 体のときの nod probability=50 である。

以下、被験者 B の評価戦略と自由発話の概要について記述する。

実験の全進行段階に渡って、被験者 B は Nod Probability 値を設定する際、まず Nod Probability を高く設定し、そこから徐々に値を下げて行った。

表示キャラクター数が 15 体のときは、Nod Probability が 70 以上の場合、「そろいすぎて気持ち悪い」印象だった。60 から 50 になると、現実的に有り得そうだと判断し始めた。40 と 30 を比較した時点で、二つの印象が全く違くと述べた。30 以下は現実的ではあるとしながらも、全然聞いていない印象になってしまうと判断した。そして 40 から 60 の範囲で Nod Probability を動かし、最終的に nod probability を 40 に設定した。その判断過程には、50, 60 は「違和感がなく」「よく聞いてくれている」印象があるが、「実際の授業の印象」となれば 40 が適している、という考えがあったとした。

キャラクター数が 5 体のときは、Nod Probability が 100 の時点で現実味を帯びるとしながらも、あくまで普通の講義のような印象を求めた結果、15 体のときと同じく Nod Probability を 40 に設定した。30 以下を選択しなかった理由は 15 体のときとやや異なっていた。被験者 B 曰く、30 以下になるとうなずくキャラクターが 1 体だけとなるケースが出始め、その様子が変であったという。うなずく人数がこのように少ない場合、本来は誰か一人よくうなずく人がいて、他の一人程度がちらほらうなずくものであるが、プログラム上ではキャラクターの中でうなずく一人が順繰りに回っていくようになってしまっていると指摘した。そして 40 と 50 を比較するとき、同時にうなずくキャラクターの数を基準に判断していた。50 は多くて 4 人、少なくとも 2 人 3 人であり、40 だと 2, 3 人がうなずいていて、普通の授業はこの様子が合っているとした。

キャラクター数が 1 体のときは、Nod Probability を高く設定できる可能性が、5 体のときよりもさらに高いとしながら、結局は 50 に設定した。その判断をする根拠は、表示キャラクター数が複数体であった前二段階のときとはかなり異なっていた。うなずきタイミングが発話末にきて、そのタイミングでキャラクターがうなずき確率に従ううなずくことを自分で気づいた上で、相手に確認や同意をうながす発話末はおおよそ 2 分の 1 の確率でやってくるから 50 が適しているのではないかと述べた。

プログラム終了後、二つの実験を通しての被験者 B の戦略の振り返りを求めた。講義音声実験ではキャラクター表示数が減るにつれ Nod Probability を高く設定していたが、自由発話音声実験ではあまり変化がなかった件について、前者は生徒目線で、後者は教師目線で行う実験であり、その影響によるものだと述べた。講義音声実験の結果は自分のうなずきの特性が反映されるため、人数が減るほどうなずきやすくなる自分の癖が表れているとし、一方で自由発話音声実験では、実際に教師の立場に立つ想定をするため、自分の特性があまり出なかったためであるとした。ただし図 8 にある通り、よくうなずく人間であると自認していることと、Nod Probability との間には関係が認められなかったことを付記しておく。

5. 今後の発展

本研究で、仮想聴衆の人数が減るにつれてうなずきの頻度が増えることを示したが、実験した最大のキャラクター数は 15 体であった。実際に仮想聴衆を用意する場面では、その数が 1000 体を超えることもある[2][3]。仮想聴衆の人数とうなずきの頻度の関係の、15 体以上のキャラクターへの適用可能性は、追って調査が必要である。

本研究で用いたモデルによるうなずきタイミングの生成は、あまり高い評価を得られなかった。4.4.2 節で示した意見を鑑みると、原因はマクロ層の再現不十分である可能性がある。今後改善が必要である。

本研究で、仮想聴衆の人数が減るにつれてうなずきの頻度が増える傾向が発見できたが、うなずくキャラクターの分布はランダムで制御不可能であった。4.4.1 節にあるように、うなずくキャラクターの分布によって印象が変わる可能性は十分にある。どのような分布に沿ってうなずくキャラクターを限定すればよいかの判断基準はまだ得られていない。これを明らかにするためには、実際の聴衆を観察する必要がある。また、社会心理学の研究において、参与人数が増えるほど課題遂行量が減る「社会的手抜き」と呼ばれる現象が知られている。本研究で見られた、人数増加に伴う、うなずき頻度の低下と社会的手抜きとの関係を調査することで、うなずきの再現の改善が期待される。

参考文献

- [1] F. Mostajeran, M. B. Balci, F. Steinicke, S. Kühn and J. Gallinat. The Effects of Virtual Audience Size on Social Anxiety during Public Speaking. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). (2020) pp. 303-312. doi: 10.1109/VR46266.2020.00050.
- [2] 和田毬那, 福地健太郎. VR プレゼンカラオケ:プレゼンテーション練習のための高臨場感 VR システムの提案. <https://www.wiss.org/WISS2018Proceedings/demo/3-A08.pdf>
- [3] 和田毬那, 福地健太郎. 大人数仮想キャラクターの表情デフォルメによる視認性向上手法の提案およびプレゼンテーション練習システムへの応用. 第 24 回日本バーチャルリアリティ学会大会論文集 (2019 年 9 月). <http://conference.vrsj.org/ac2019/program/common/doc/pdf/1D-03.pdf>
- [4] WATANABE T, DANBARA R and OKUBO M. InterActor: Speech-driven embodied interactive actor. Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communication. (2002) pp.430-435. doi: 10.1109/ROMAN.2002.1045660.
- [5] WATANABE T. A Voice Reaction System with a Visualized Response Equivalent to Nodding. Advance in Human Factors/Ergonomics, A. (1989) vol.12A, pp.396-403.
- [6] 小畑淳, 渡辺富夫, 大久保雅史. 音声駆動型引き込みコミュニケーションシステム"SAKURA"ヒューマンインタフェース 2000 論文集, pp.331-334(2000)
- [7] 田中一也, 渡辺富夫, 石井裕. 全観客キャラクターが講演者発話にうなづく没入型講演体験システムの開発. 情報処理学会 第 79 回全国大会講演論文集. 2017, p. 187-188.
- [8] Unity, <https://unity3d.com/jp>