

# Altered auditory feedback に基づく感情誘導における 音声特徴量弁別閾の調査

中村 颯<sup>1</sup> 福田 航希<sup>1</sup> 高道 慎之介<sup>1,2</sup> 大畑 龍<sup>3</sup>

**概要:** リアルタイムに話者の音声を感情的に聞こえるよう変調し、その音声を聴覚フィードバックすることで、話者自身の感情が変調方向へと誘導されることが報告されている。しかし、この感情誘導効果は、話者が変調の存在やその方向、あるいはわずかな違和感であっても検知した場合には生じないことが示されている。したがって、実用的な感情誘導システムの設計には、話者が変調を検知しない範囲、すなわち感情的音声変調に対する弁別閾を定量的に把握することが不可欠である。本稿では、Aucouturier らの happy 変調パラメータに基づくリアルタイム音声変調システムを構築し、大学生を対象とした朗読課題において、感情的変調に対する弁別閾を 2down-1up staircase 法により測定した。あわせて、音楽経験、自己音声への曝露頻度、自己関連指標など複数の個人差要因を収集し、実験条件と併せて弁別閾との関連を検討した。その結果、感情的音声変調の弁別閾は被験者間および条件間で広い分布を示し、各因子との間に有意な関連は認められなかった。この所見は、一律の変調強度に基づくシステム設計による一般化が容易ではないことを示唆すると同時に、実運用を想定した感情誘導システムにおいては、個人ごとの弁別閾に合わせた適応的キャリブレーションが不可欠であることを裏付けるものである。

## 1. はじめに

発話者の音声をリアルタイムに変調し、聴覚フィードバックする技術 (altered auditory feedback; AAF) がある。特に、音声を感情的に聞こえる方向へ変調した場合、発話者自身の感情も同一方向へ変容することが報告されている [1]。精神疾患の多くが感情の問題と関連していると指摘されている現状において [2]、AAF は治療目的での応用が期待されており、オンライン通話における不安の低減などへの応用可能性も示唆されている [3]。

しかし、この感情誘導効果は、発話者が変調の存在やその方向、あるいはわずかな違和感であっても検知した場合には生じないことが知られている [4]。したがって、AAF を実用的に利用するためには、感情的音声変調に対する弁別閾を明らかにし、「気づかれない」変調強度の範囲を定量的に把握する必要がある。また、その弁別閾を決定づける

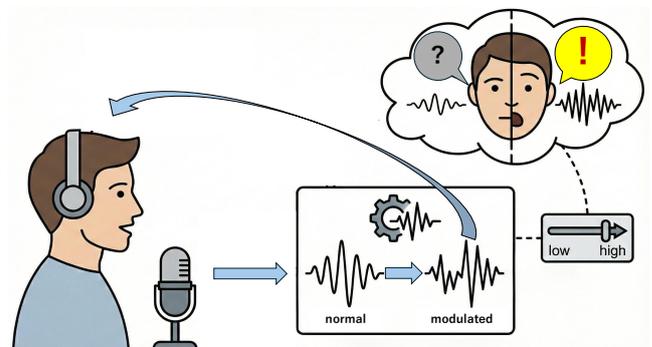


図 1: 本論文の概要。Altered auditory feedback において、どの程度の変調まで発話者に検知されないかを調査する。また、その弁別閾が何の要因に起因するかを調査する。

要因についても明らかにする必要がある。

本稿では、AAF における感情的音声変調に対する弁別閾の分布と、その決定因子の調査を目的とする。先行研究 [1] にて提案されている、F0 などの変調に基づく感情誘導システムを利用し、happy 感情を誘導する変調の弁別閾を測定する。測定と合わせて、音楽経験、自己音声への曝露頻度、感情表出傾向、自己と関連する心理特性についてアンケートを実施し、弁別閾と強い関連を持つ因子を検討する。

<sup>1</sup> 慶應義塾大学  
3-14-1 Hiyoshi, Kohoku-ku, Yokohama-shi, Kanagawa 223-8522, Japan.

<sup>2</sup> 東京大学  
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.

<sup>3</sup> 産業技術総合研究所  
National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8560, Japan

## 2. 関連研究

### 2.1 弁別閾

一般に、弁別閾 (just noticeable difference; JND) とは、標準刺激と比較して差異が知覚される最小の刺激レベル差を指す [5]。本研究のような AAF に基づく感情誘導では、無変調音声 (すなわち、自己発話音声に対する通常の聴覚フィードバック) を標準音声、感情を誘導するように変調した音声を刺激音声として扱う。このとき、変調の方向とは無関係に、感情的な変調の存在あるいは違和感が知覚される最小の変調強度を、本研究における感情的音声変調の弁別閾として定義する。自己発話音声とその変調音声の弁別閾に関し、これまで主として基本周波数や第 1 フォルマントといった単一の音響パラメータに着目した研究が行われてきた [6], [7]。一方で、複数のパラメータを同時に操作した変調を扱った研究は少なく [8]、とりわけ感情的な音声変調に対する弁別閾を直接検討した研究はほとんど存在しない。

### 2.2 AAF による感情誘導効果消失

AAF における弁別閾を検討するうえで重要なのは、感情誘導効果が利用者による音声変調の「検知」に強く依存する [4] という点である。AAF を用いた感情誘導効果と変調検知の関係を検討した研究 [4] では、被験者が明示的または暗黙的に変調を検知した場合、感情誘導 AAF の効果が消失することが示されている。ここで、明示的検出は次のレベル区分においてレベル 4 以下を選択した場合と定義される。

- レベル 1: 「意図的に私の声を操作しましたね」
- レベル 2: 「私の声は時々高くなったり低くなったりして、実験中に変化しました」
- レベル 3: 「私の声は時々奇妙に聞こえて、実験中に変化しました」
- レベル 4: 「私の声は奇妙に聞こえました、それはヘッドホンを通して自分の声を聞くことに慣れていないからだけではありません」
- レベル 5: 「私の声は奇妙に聞こえましたが、それはおそらくヘッドホンを通して自分の声を聞くことに慣れていないからでしょう」
- レベル 6: 「私の声は全く問題ありませんでした」

また、暗黙的検出とはピッチ補償反応を指す。これはリアルタイムに基本周波数が操作された場合、その操作を打ち消す方向に発声が誘導される現象として広く知られている。本研究では、後述する実験設計の都合上、暗黙的検出の指標は扱わず、明示的検出に対応する主観報告のみを対象とする。

### 2.3 弁別能力に影響する要因

音の変調に対する弁別閾が、聴覚経験や自己音声への慣れなど複数の要因に左右されることは、以下に述べる先行研究によって示されている。

純音の周波数を弁別する能力は、音楽経験を有する群がそうでない群よりも高いことが知られている [9]。また、自己音声への曝露頻度 (普段、自分の声を録音や動画を通じて聴く頻度) が高いほど、自己音声と他者音声を区別する課題 (self-other voice discrimination; SOVD) の精度が高くなることも報告されている [10]。一方で、年齢は自身の発話の録音に対する基本周波数弁別閾と関連がないことが示されている [11]。

感情表出性とは、音声・顔・ジェスチャーなどを用いて自分の感情を外向きに表出する傾向の個人差を指し、感情表出性尺度 (emotional expressivity scale; EES) によって定量化される [12]。EES の得点が高い個人は、感情的な自己音声を聴取する機会が相対的に多いと考えられる。実際、自己音声への曝露頻度が高いほど SOVD 課題の成績が高いことから [10]、自己音声に対する感情的な変調に対する弁別閾が低くなることが示唆される。

これらの研究と本研究では刺激音が異なるものの、同様の傾向が認められる可能性があるため、本研究では音楽経験、自己音声への曝露頻度、感情表出性などが感情的音声変調の弁別閾に与える影響を検討する。

### 2.4 自己と関連する属性

感情的に変調された自己音声は「どこまで自分の声として受け取られるか」は、単なる聴覚的弁別能力だけでなく、自己の境界や主体感に関する個人差にも左右されると推測する。したがって、本研究では以下の自己にまつわる心理特性が弁別閾のばらつきとどのように結びついているかを検討する。本研究で対象とする自己にまつわる心理特性を、次に示す。

- **身体所有感**: 身体所有感 (sense of ownership; SoO) とは、自分の身体の部位や感情、思考に対して感じる「これは私のものだ (mineness)」という感覚を指す。古典的には、ラバーハンド錯覚などのパラダイムを用いて知覚・評価されてきた [13]。
- **行為主体感**: 行為主体感 (sense of agency; SoA) は、自らが行為を開始し、その進行を制御しているという現象的な体験を指す。行為主体感において「行為の開始」と「その制御」を区別して論じる立場もあるが [14]、本稿では両者をまとめて SoA として取り扱う。SoA は、ボタン押しなどの実験課題を用いた顕在的な測定と、日常的な経験に関する質問紙による潜在的な測定の両面から評価される。
- **ナラティブセルフ**: ナラティブセルフとは、人間の自己の感覚と人生経験が物語 (ナラティブ) の構造を持

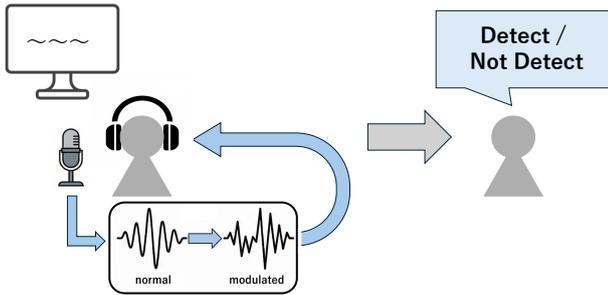


図 2: 本番試行のフロー

ち、そのナラティブが自己形成に大きな影響を与えるとする考え方である。ナラティブは「特定の過去の経験」や「可能な未来の自己像」に関する表象を伴い、自己と他者の分化に深く関与するとされる [15].

- **自己概念の明確性**: 自己概念の明確性は、個人の自己概念の内容や自己についての信念が、明瞭に確信をもって定義され、内的な一貫性を持ち、通時的に安定している程度を表す概念であり、神経症傾向との関連が示唆されている [16].

本研究では、SoO, SoA, ナラティブセルフ、自己概念の明確性といった自己関連の属性が、感情的音声変調の弁別閾の個人差とどのように関連するかを検討する。

### 3. 実験

#### 3.1 実験手続き

- **事前アンケート**: 実験開始前に質問紙による事前アンケートを実施した (詳細は「事前アンケート」節).
- **各セッション**: 朗読する文刺激の感情カテゴリ (happy / sad / neutral) ごとに 1 つのセッションを構成した (詳細は「文刺激」節). 各感情条件セッションは、以下の二段階で構成した。
  - **音量較正・練習**: 音量とピッチ知覚の干渉 [17] を考慮し、フィードバック音声の音量較正と無変調条件での練習試行を行った.
  - **本番試行**: 各セッションでは、提示文ごとに読み直しを認めず、一度だけ自然に朗読させた. 朗読後、「自身の声が加工されたと感じたか/違和感があったか」を尋ね、Yes/No で回答させた. 図 2 は本番試行のフローを表す.
- **事後アンケート**: 全セッション終了後に質問紙による事後アンケートを実施した (詳細は「事後アンケート」節).

#### 3.2 AAF システム

AAF システムでは遅延が数百 ms 程度に増大すると感情誘導が再現できない報告がある [18]. 本研究ではシステ

ム全体の遅延を約 18 ms に抑えた. この遅延は一部で知覚され得るが [19], 発話障害を生じるほどではない [20].

音声変調は、先行研究 [1] が配布する Max 用パッチ (DAVID) をベースに用いた. 本研究では主として発話検出部と実験制御部を拡張し、変調アルゴリズム自体は原則として踏襲した (詳細は付録 A.2.1, 付録 A.2.2, 付録 A.2.3).

本研究では感情的変調を happy に限定した. 先行研究 [1] で happy と sad が有意な効果を示し、特に happy の効果量が大きいため、および後続研究 [4] でも happy の効果が比較的頑健に再現されていることによる. happy 変調は先行研究 [1] に倣い、変調率  $\alpha$  により定義し、本研究では  $\alpha$  のみを変化させて弁別閾を測定した. 変調率  $\alpha$  に基づいて、入力音声に対して以下の変調を加える.

- **基本周波数**:  $+25\alpha$  cent.
- **イントネーション**: 発話開始を  $t = 0$  ms とし、 $t = 0$  で  $-50\alpha$  cent,  $0 \leq t \leq 400$  ms において  $-50\alpha + 0.125\alpha t$  cent,  $400 \leq t$  ms において 0 cent
- **コンプレッサ** ( $0 < \alpha$  の場合のみ適用): 閾値  $-26$  dB を超える音圧を 25% に圧縮. アタックおよびリリースの速度は 10 dB/s.
- **ハイシェルフフィルタ** ( $0 < \alpha$  の場合のみ適用): 8 kHz 以上を  $+10$  dB/octave で増幅.

AAF システムとして、先行研究 [1] の配布する Max 用パッチ (DAVID) に基づく. 具体的なシステム実装内容は付録 A.2 を参照とする.

#### 3.3 文刺激

被験者に提示し朗読させる文刺激は WRIME データセット [21] から選択する. この際、感情的文脈に応じて自然な朗読の F0 や話速が変化する [22] ことをふまえ、happy, sad, neutral 感情のそれぞれについて文を選択する. Happy と sad の文選択時には、データセット中に付与されている感情強度スコアが、全スコア付与者で中程度 (2) 以上のものを採用する. Neutral については、全感情のスコアが全スコア付与者について弱程度 (1) 以下のものを採用する. 使用した文の平均文字数は 26.1 文字であった. 実際に使用した刺激文は付録 A.5 を参照とする.

#### 3.4 試行と応答に基づく弁別閾の測定

弁別閾の測定には、心理物理測定法として調整法の一つである 2down-1up staircase 法 [23] を採用する. これは、被験者が一度弁別に失敗した際に刺激強度 (ここでは変調率  $\alpha$ ) を上げ、連続して二度弁別に成功した際に刺激強度を下げる手続きである. 従来は、刺激強度の上昇ステップ幅  $\Delta_+$  と下降ステップ幅  $\Delta_-$  の比 (ステップ比  $\Delta_+/\Delta_-$ ) を 1 とする設定が慣習的に用いられてきた [7]. しかし、staircase 法におけるステップ比に関するシミュレーショ

ン実験 [24] によると, 2down-1up staircase 法ではステップ比  $\Delta_+/\Delta_- = 0.5488$  とすることで, 各ステップの大きさに依存せず正答率 80.35% に対応する刺激強度をより正確に測定できることが示されている. 本研究ではこのステップ比  $\Delta_+/\Delta_- = 0.5488$  を満たすようステップ幅を設定する.

具体的には, 変調率  $\alpha$  の初期値を 7.0 とし, staircase 開始から最初の反転が生じるまでは下降ステップ幅を  $\Delta_- = 0.80$ , 上昇ステップ幅を  $\Delta_+ = 0.44$  とする. 一度反転が生じた後は, より狭い範囲で閾値を探索するため, 下降ステップ幅を  $\Delta_- = 0.15$ , 上昇ステップ幅を  $\Delta_+ = 0.08$  とした. いずれの段階においてもステップ比  $\Delta_+/\Delta_-$  はおおむね 0.5488 を満たしている. 変調率  $\alpha$  の探索範囲は  $0 \leq \alpha \leq 10$  に制限する.

被験者には, 各試行ごとに「自身の声が加工されたと感じたり, 何か違和感を覚えましたか?」という質問に yes/no で回答するよう求める. 本研究では yes を「変調を検知した」とみなし, staircase の成否判定および弁別閾の算出に用いた. 試行と応答を 75 回繰り返す, あるいは staircase を 15 回反転した時点で終了とする.

### 3.5 キャッチ試行

すべての試行で何らかの変調が付与される実験設計では, 被験者が「常に変調が存在する」と学習し, 回答にバイアスが生じるおそれがある. このような被験者のバイアスを緩和するため, キャッチ試行として変調率  $\alpha = 0$  (無変調) の試行を, 各試行に対して 10% の確率でランダムに挿入する. キャッチ試行における正答は「No」とした.

### 3.6 被験者

対象者は, 聴覚に異常がないと自己申告した慶應義塾大学に所属する大学生  $N = 16$  名 (平均年齢  $M = 21.62$  歳,  $SD = 1.02$ , 男性:女性 = 13:3) であり, 機縁法により収集した. 被験者には, 本研究を「自己音声に付与された音声変調の有無を判別するタスク」と説明し, 変調が感情的なものであることは明示しなかった. 実験前には, 学内倫理審査を経た説明文書を配布し, 口頭で補足説明を行ったうえで, 同意書への署名を得た.

### 3.7 事前アンケート

実験開始前のアンケートでは, 質問紙を用いて, 実験前の感情・気分を基本 9 感情 [25] で測定する. あわせて音楽経験 [10] と自己概念の明確性 [16] を被験者に尋ねる. 具体的なアンケート項目は付録 A.1.1 を参照されたい.

### 3.8 事後アンケート

実験終了後のアンケートでは, 質問紙を用いて, 自己音声への曝露経験 [26], 自分の声を「見失った」と感じた度

合い (7 段階), 判別の精度に対する自信 (7 段階), 感情発露傾向 [12], 日常生活における SoO, SoA およびナラティブセルフの強さ [27] を評価した. あわせて, 実験全体に関する感想や気づいた点について自由記述で回答を求める. 具体的なアンケート項目は付録 A.1.2 を参照されたい.

## 4. 分析手法と結果

### 4.1 観測データの棄却

Staircase 法における弁別閾は, 最終  $R$  回の反転点における刺激強度の平均値として定義され,  $R$  は研究によって 4~7 程度の値が用いられる [24]. 本研究では  $R = 3\sim 8$  の各値について検討し,  $R = 6$  に設定した. 詳細を付録 A.3 に記載した.

あわせて, 以下のいずれかの条件を満たすデータは信頼性を欠くと判断し, 解析から除外した.

- 反転回数が  $R$  未満である条件
- 弁別閾の  $Z$  スコアの絶対値が 3 以上である外れ値
- キャッチ試行に対する false alarm 率<sup>\*1</sup>が 90% 以上

### 4.2 弁別閾の個人差を説明する因子の検討

感情的音声変調に対する弁別閾の個人差を説明し得る要因として, 表 1 に示す諸因子を検討対象とした. これらの因子を説明変数, 感情的音声変調に対する弁別閾を従属変数として統計的解析を行い, AAF による感情誘導システムの設計において考慮すべき個人差要因を検証する.

### 4.3 統計解析の手順

弁別閾の個人差要因を検討するために, 線形混合効果モデルを用いた [29]. ただし, サンプル数に対して候補となる因子 (固定効果) が多いため, すべての因子を一度に投入すると過学習や推定不安定性の問題が生じる. そこで, 本研究では先行研究 [30] と同様に, 次の二段階アプローチを採用した.

まず, 意図的に実験条件として設定した因子, あるいは理論的背景に基づき, 弁別閾に影響すると予測される因子を主要因子として選定し, これらのみを固定効果に含むベースモデルを構築する. 次に, それ以外の因子を候補因子と位置づけ, L1 正則化の lasso [31] を用いて係数推定を行い, 非ゼロの係数を持つ因子を絞り込む. その後, 選択された候補因子を加えた線形混合効果モデルを構成し, 最終的な解析を行う.

本稿では線形混合効果モデルの収束性を優先するため, ランダム効果は被験者ごとのランダム切片のみとする. また, lasso の実装には scikit-learn [32] の LassoCV<sup>\*2</sup> を用

\*1 無変調の音声に yes (変調あり) と回答した割合

\*2 [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LassoCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LassoCV.html)

表 1: 各変数の概要

Label	Factor	Description
<i>Categorical factor</i>		
Gender	性別	自己申告による性別カテゴリ.
Music genre	主たる音楽ジャンル	主に経験している音楽ジャンルを「クラシック」または「それ以外」に分類したカテゴリ変数 [10].
Emotion category (happy, neutral, sad)	朗読文の感情カテゴリ	朗読に用いた文に付与された感情カテゴリ.
Music experience	音楽経験	音楽経験の有無.
<i>Continuous factor</i>		
Session order	セッション順	実験内における各感情条件セッションの提示順序を表す指標 (1-3).
Emotion (positive, negative)	感情・気分	基本 9 感情を用いた質問紙に基づく感情状態指標 [25]. ポジティブ・ネガティブ感情についてそれぞれスコアが算出される.
Music years	音楽経験年数	演奏経験の年数. 音楽経験に関する質問紙に基づく指標 [10].
SCC	自己概念の明確性	自己属性に関する認知の一貫性・安定性を測定する質問紙による指標 (self concept clarity) [16].
SVE	自己音声曝露経験	日常的に録音などにより自己音声を聴取する頻度. 1 週間あたりの SNS への音声付き動画投稿回数と, プレゼンテーション練習目的で実施した録音回数の合計により定量化される [26].
EES	感情発露傾向	感情表出性尺度 (emotional expressivity scale) に基づく, 感情表出のしやすさの指標 [12].
Ownership	身体所有感	身体や感情を「自分のもの」と感じる程度を測定する SoO 質問紙に基づく指標 [27].
Agency	行為主体感	行為の開始および制御に対する主体感を測定する SoA 質問紙に基づく指標 [27].
Narrative	ナラティブセルフ	一貫性を持った存在としての自己感を測定するナラティブセルフ質問紙に基づく指標 [27].
Mean F0	朗読 F0 平均	各朗読音声から pyWorld [28] を用いて抽出した 70-300 Hz の基本周波数の時間平均値を, セッション単位で平均した指標.
Discrimination confidence	弁別の自信	判別課題に対する自己評価. 7 段階リッカート尺度.
FA rate	False alarm 率	キャッチ試行 (変調率 $\alpha = 0$ ) に対して「変調あり」と誤答した割合.
Self voice difference	自己聴取音との差異	無変調のフィードバック音声と実際の自己聴取音との程度異なるかに関する 7 段階リッカート評価. 初回セッションの練習時に評価する.
Voice self-loss	自分の声の見失い	実験中に自分の声をどの程度見失ったかに関する 7 段階リッカート評価.

い, パラメータは以下のように設定した.

- `alphas` (正規化強度の候補集合):  $10^{-4}$  から  $10^2$  の範囲から 200 個の点をランダムに抽出
- `cv_fold` (交差検証分割数): 3
- `max_iter` (最適化の最大反復回数): 50000

Lasso に投入する前に, すべての因子を標準化した. 線形混合効果モデルの推定には `statsmodels.formula.api` [33]

の `mixedlm`<sup>\*3</sup> を使用し, パラメータ `re_formula="1"` としてランダム効果に切片のみを指定した.

主要因子としては, 次の 3 因子を採用した.

- 朗読文の感情カテゴリ: 意図的に実験条件として設定した要因であるため.
- セッションの順番: 意図的に実験条件として設定した

\*3 <https://www.statsmodels.org/stable/generated/statsmodels.formula.api.mixedlm.html>

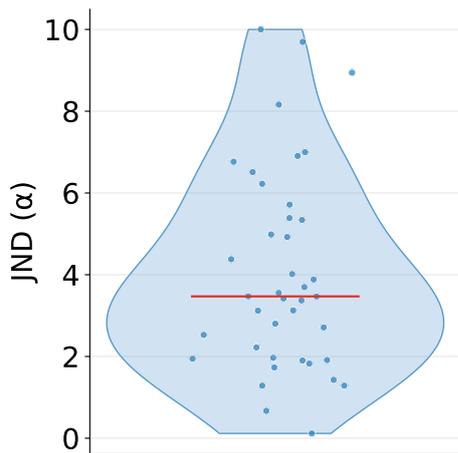


図 3: 弁別閾分布図. 点は被験者, 赤線は中央値を表す.

要因であるため.

- 朗読平均 F0 : F0 の変調量は元音声の F0 によらず一定である (3.2 節) が, 特に, 中程度 (数百 Hz) の周波数帯における F0 知覚および弁別閾は対数スケールで変化することが知られており [34], 元音声の F0 の違いが弁別閾に影響しうると考えられるため.

#### 4.4 弁別閾の全体分布

弁別閾の分布を図 3 に示す. また, 平均と標準偏差は  $4.060 \pm 2.486$  であり, 中央値は 3.470 であった.  $\alpha = 4$  が 100 cent の基本周波数変調に対応することから, 平均的には半音程度の変調が検出されるといえる.

#### 4.5 候補因子の選択

L1 正則化の結果, 全候補因子の係数が 0 に縮退した. この結果は, 本研究のデータにおいて, 主要因子を除く候補因子が弁別閾の変動を安定的に説明する予測変数として採択されなかったことを意味する. したがって, 以降に示す線形混合効果モデルでは, 主要因子のみを固定効果として用いるものとする. 一方で, lasso はサンプルサイズや共線性, 効果量の大きさに強く依存し, 小標本では真に関連のある変数が選択されない可能性もある. そこで, 選択されなかった候補因子についても, 弁別閾との関連の可能性を探索的に把握する目的で, 各候補因子と弁別閾との単相関係数を算出し, 表 2 に示す. 係数の絶対値が大きい順に self-voice difference (無変調のフィードバック音声と自己聴取音の差異), positive emotion (実験前における被験者のポジティブな感情), gender と続き, これらの候補因子には中程度の相関が観察される.

#### 4.6 弁別閾を説明する要因

線形混合効果モデルを用いて推定された, 各主要因子の

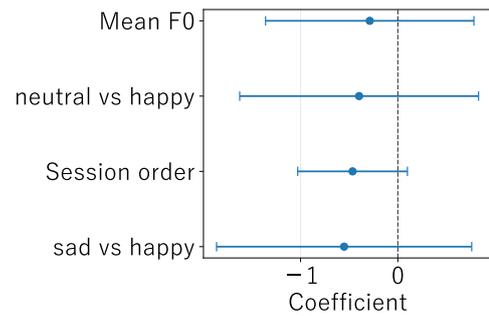


図 4: 線形混合効果モデルにおける主要因子の回帰係数と 95%信頼区間. いずれの信頼区間も 0 をまたいでおり, 有意ではないことを図示している.

表 2: 候補因子と弁別閾の相関係数. どれも強い相関を持たないことが示される.

Candidate factor	Corr. coefficient
Self voice difference	0.417
Positive emotion	0.346
Gender	0.301
FA rate	0.231
Self concept clarity	0.184
Voice self-loss	0.123
Discrimination confidence	0.080
Ownership	-0.027
Narrative	-0.054
Self voice exposure	-0.055
Music experience	-0.056
Music years	-0.098
Emotion expression scale	-0.141
Negative emotion	-0.156
Agency	-0.196

係数と 95 % 信頼区間を図 4 に示し, 詳細な推定結果 (係数, 標準誤差,  $p$  値) を表 3 に示す. いずれの主要因子についても弁別閾に対する有意な効果は認められなかった (いずれも  $p > 0.05$ ). すなわち, 本研究の範囲では, 朗読文の感情カテゴリ, セッション順序, 朗読 F0 平均のいずれも, 弁別閾の系統的な変動を説明する強い要因とはなりにくいことが示される.

セッション順序については, 有意ではないものの  $p$  値が比較的小さい. この結果を考察するために, 図 5 にセッション順序別に見た弁別閾の分布を示す. 図 5 では, セッションを重ねるほど弁別閾が低下する方向の傾向が視覚的には示される. この傾向は, 変調提示の反復により, 同一方向の変調の手掛かり (特に F0 上昇やイントネーション変化) へ注意を向ける学習が進んだ可能性により説明される.

#### 4.7 False alarm が頻発する群に特有の因子

前節までの実験では, 3.5 節で述べたように, キャッチ

表 3: 線形混合効果モデルにおける固定効果の回帰係数.

Factor	Coefficient	p-value
Sentence: neutral vs happy	-0.398 ± 0.626	0.525
Sentence: sad vs happy	-0.552 ± 0.669	0.409
Session order	-0.465 ± 0.288	0.106
Mean F0	-0.289 ± 0.546	0.597

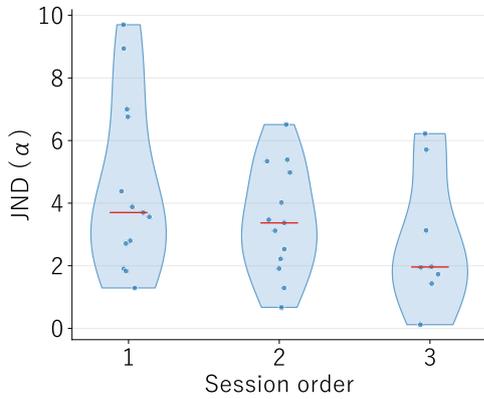


図 5: セッション順序別にみた弁別閾の分布. セッションを重ねるごとに弁別閾が低下する傾向がある.

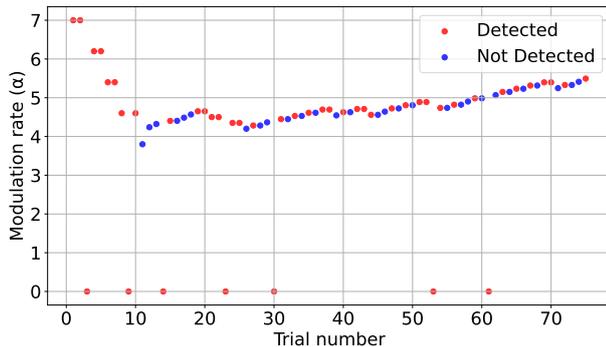


図 6: 無変調の自己音声と変調された自己音声を取り違えた応答パターンの例.  $\alpha = 0$  の試行ですべて変調を検出したと回答している.

試行における false alarm 率が 0.9 以上の被験者を除外している. 除外された被験者の応答結果例を図 6 に示す. この例では, 全てのキャッチ試行, すなわち無変調音声に対して「変調あり」と回答している. 本節では, 除外したこの群を self-loss 群, それ以外の被験者を normal 群と称し, 両群間で各因子がどのように異なるかを検討する. self-loss 群, normal 群の人数はそれぞれ 5 名, 11 名であった. 連続変数については welch の t 検定, カテゴリ変数についてはサンプル数が少ないことを考慮し,  $\chi^2$  検定ではなく fisher の正確確率検定を用いた.

結果, すべての因子について, self-loss 群と normal 群の間に有意差は確認されなかった (図 7, 表 4, 5). したがって, 本研究で収集した因子の範囲では, self-loss 群を

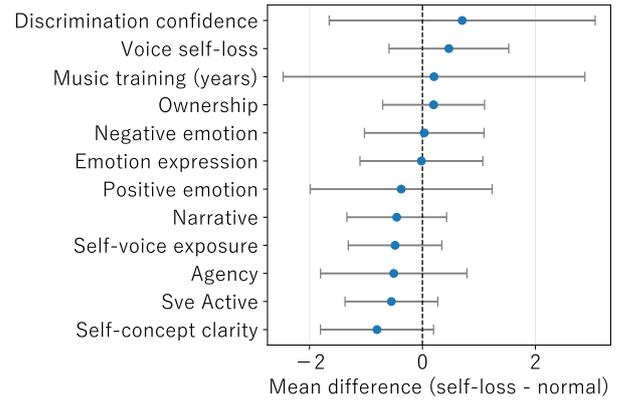


図 7: Self-loss 群と normal 群における各連続因子平均の差と 95%信頼区間. いずれも信頼区間が 0 をまたいでおり, 有意差を持たないことが示される.

表 4: Fisher の正確確率検定の結果.  $p$  がどれも 1.000 となり, 有意差を持たないことが示される.

Factor	Odds ratio	$p$	$n_{SL}$	$n_N$
Gender	0.889	1.000	5	11
Music experience	0.563	1.000	5	11
Genre: classic	2.000	1.000	3	8

表 5: Welch の t 検定結果. どれも  $p \geq 0.05$  となり有意差を持たないことが示される.

Factor	Mean(SL)	Mean(N)	$p$
Self concept clarity	-0.553	0.252	0.105
Self voice exposure	-0.375	0.171	0.170
Narrative	-0.313	0.142	0.287
Voice self-loss	0.368	-0.100	0.337
Agency	-0.349	0.159	0.391
Discrimination confidence	0.553	-0.151	0.407
Positive emotion	-0.259	0.118	0.587
Ownership	0.136	-0.062	0.645
Music training (years)	0.149	-0.056	0.823
Negative emotion	0.022	-0.010	0.949
Emotion expression	-0.013	0.006	0.971

一貫して特徴づける説明変数は同定できなかった. この所見は, 質問紙で捉えにくい要因, 例えばヘッドホン提示への不慣れ, 期待や不安に基づく応答方略 (「変調があるはず」という前提), 遅延や音量バランスに起因する知覚的違和感への感受性などに左右されている可能性がある.

## 5. 考察と本研究の限界

### 5.1 考察

**弁別閾の個人差.** 本研究で得られた弁別閾の分布では, 変調率が 0 付近から 10 付近まで幅広い値をとり (図 3), 標準偏差がおおよそ 2.5 程度であったことから, 感情的音声変調に対する弁別閾はケースに応じて大きく変化し, 被験

者間および条件間で顕著な個人差が存在することが示唆される。

**朗読時の F0, 性別の影響.** 主要因子に含まれる朗読 F0 平均については、有意な効果は確認されなかった。しかし、理論的には変調前の朗読 F0 が弁別閾に影響を与える可能性がある。効果が検出されなかった理由として、被験者の性別の偏り（男性：女性 = 13：3）に起因する朗読 F0 平均の分布の偏りが挙げられる（図 8）。より性比の均衡したサンプルや、広い F0 範囲を持つ被験者を含めた検証が必要である。なお、gender が弁別閾と中程度の単相関を示した点についても、性別それ自体の因果効果というより、性別比の偏りを介して F0 分布や発話特性が偏ったことを反映している可能性がある。

**自己暴露経験の影響.** 自己音声への曝露経験を定量化するために、一週間における自分の声の入った動画を SNS にアップロードする本数および、一週間における自分の声を録音して確認する頻度の和を、先行研究 [26] に基づいて使用した。しかし、被験者の回答分布は一部の値に大きく偏っており（図 9）、曝露頻度の個人差を十分に捉えきれていない可能性がある。尺度設計やサンプルの多様性の観点から、今後の検討が必要である。

**無変調音声に対して変調ありと回答した被験者.** 4.7 節で触れた通り、デバイスを通すことによって無変調を変調された音声と取り違えた self-loss 群が約 30%程度存在したことから、実際のシステム利用場面においても、デバイスを通した無変調の自己音声を変調された音声と取り違えるケースが一定の割合で生じることが示唆される。本研究では staircase 法の都合上、試行の初期段階は明らかに変調が知覚されるほど強い刺激から始めたが、刺激が知覚されないレベルから始めた場合には、この傾向がより強く現れる可能性がある。そのような変調強度の時間的ダイナミクスも、今後検討すべき重要な課題である。

**Self-voice difference (無変調フィードバックと自己聴取音の差異).** 弁別閾との単相関が中程度であった self-voice difference は、無変調条件におけるフィードバック音声、被験者の主観的な自己聴取音とどの程度異なるかを表す指標である。無変調の時点で当該指標が高い被験者では、ヘッドホン提示や収録・再生系そのものに起因する違和感がベースラインとして存在している可能性がある。この場合、「変調の有無」を判断する際の基準点が不安定化し、結果として変調検知の一貫性が低下することで、弁別閾推定値に影響を与え得る。すなわち、変調強度そのものの問題というより、フィードバック提示系に対する適応度や自己音声の同一性判断の基準が、弁別課題に混入している可能性が示唆される。一方で、本因子は lasso による変数選択では採択されなかった。この背景として、サンプルサイズの制約により推定が不安定であった可能性に加え、効果が線形・加法的に現れず、他の個人差要因や実験条件

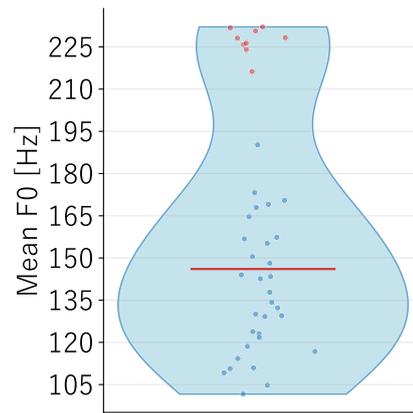


図 8: 被験者ごとの朗読平均 F0 の分布

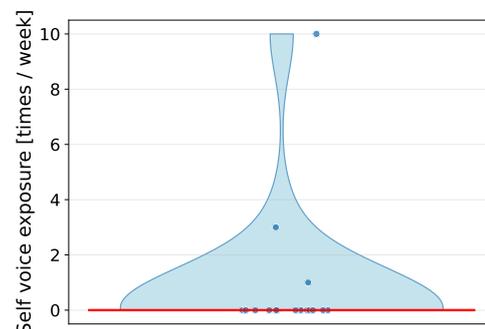


図 9: 自己音声への曝露頻度 (self voice exposure) の分布

との交互作用としてのみ現れる可能性が考えられる。

**Positive emotion (実験前のポジティブ情動).** 同じく弁別閾との単相関が中程度であった positive emotion は、実験開始前の情動状態のポジティブさを表す指標である。事前の情動状態が happy 方向に偏っている場合、提示される happy 変調による主観的印象変化が相対的に小さくなる、あるいは変化が「自然な範囲」として解釈されやすくなる可能性がある。その結果、変調による違和感が目立ちにくくなり、弁別閾が高く見積もられる方向に寄与し得る。ただし、本研究では実験中に逐次的な感情評価を行っていないため、この解釈はあくまで事前状態と検知判断の関係に基づく推測である。また、本因子も lasso では採択されておらず、self-voice difference と同様に、小標本による推定不安定性や、交互作用的な効果構造により線形モデル上で寄与が顕在化しにくかった可能性がある。

## 5.2 本研究の限界と今後検討すべき点

**主観指標のみに基づく弁別閾評価の限界.** 本研究には、いくつかの方法論的限界が存在する。第一に、本研究で用いた弁別閾はすべて主観報告に基づくものであり、客観的指標を併用していない点である。弁別に関する客観的指標としては、脳波計測による指標 (mismatch negativity; MMN) [8], [35] や、暗黙的な検知指標としてのピッチ補

償反応 [1], [4] が挙げられるが、本研究ではこれらの計測を行っていない。

**ピッチ補償反応計測の困難性・主観的弁別指標の妥当性。** 本研究でピッチ補償反応を扱わなかった背景には、以下の二点がある。(1) 実用的な音声変調システムの利用状況に近い課題設定を優先し、短音ではなく文朗読を用いたこと、(2) 文朗読中の連続発話では、基本周波数操作と発声制御が密接に干渉するため、補償反応を安定して抽出することが難しいことである。

もっとも、MMN と主観的弁別判断の指標が大きく乖離しないこと、および音の弁別に関しては主観指標のほうが相対的に寛容であることが報告されている [6]。したがって、本研究の主観指標を用いた弁別タスクは、一定の妥当性を有すると考えられる。

**staircase 法による弁別閾推定精度の制約。** 第二に、本研究で用いた staircase 法は、弁別閾の推定に広く用いられる一方で、個々の被験者における推定値の精度や不確実性(分散)を詳細に評価するには限界がある。ベイズ推定に基づく実験設計法 [36] では、あらかじめ閾値分布の仮定を置く必要があるが、感情的変調に対する閾値分布について十分な知見がない現状では、その仮定が成り立たない可能性もある。本研究では、まず staircase 法により分布の概形を把握することを優先した。

**自己音声同一性の揺らぎと弁別判断の混同。** 第三に、一部の被験者からは、「無変調の自分の声」と「変調された自分の声」のどちらが本来の自己音声に近いかわからなくなったという報告が得られた。アンケートでは差異自体は知覚しているものの、自己同一性の観点では判断が難しいというケースが存在しており、この点は自己認知と弁別閾の関係を考える上で重要な論点である。

**短期的訓練効果が弁別閾に及ぼす影響。** 第四に、短期的な訓練効果の影響が挙げられる。短期間の訓練によって音の弁別能力が向上し得ることが報告されている [37]。本研究では、変調の方向が happy 条件において基本周波数(F0)の上昇を伴うことから、試行を重ねるにつれて弁別閾が低下する傾向が一部の被験者で観察された。ただし、この傾向が課題構造の理解に起因するものであるかどうかは、本研究のデザインからは直接的には判断できない。したがって、観察された弁別閾の変化については、訓練効果を含む複数の要因が寄与した可能性を踏まえて解釈する必要がある。

**呼吸状態および反応様式に関する未統制要因。** さらに、呼吸状態や反応様式に関する要因も本研究では十分には統制できていない。例えば、吸気時のほうがモーフィング刺激に対する SOVD タスクの精度が高いことが報告されている [38] が、本研究では呼吸位相を明示的には制御していない。また、自他声モーフィング課題では、回答に用いる手の左右差が成績に影響する可能性も指摘されている [39]。

**明示的弁別課題と暗黙的自己声処理の乖離。** さらに、人が自分の声であるかどうかを明示的に判断する状況では、暗黙的な状況と比較して識別精度が低下し得ることが報告されている [40]。本研究で用いた課題は、被験者に対して「提示された音声は自分の声か否か」を判断するよう明示的に指示する明示的弁別課題である。一方、ここでいう暗黙的な検出とは、「自分の声であるかどうか」を事前に説明されず、被験者が単に話者の弁別を行うタスクを指す。このため、本研究で測定した弁別能力は、暗黙的な自己声処理とは異なる処理過程を反映している可能性がある点に留意が必要である。

**より効果的な感情誘導に向けた変調設計の示唆。** 先行研究 [1] では、朗読開始後の最初の 2 分間は無変調とし、その後 5 分間かけて変調度を  $\alpha = 1$  まで漸増させている。一方、本研究では、2down-1up staircase 法の収束点が検知率 80.35% に対応する設定であり(第 3.4 節参照)、弁別閾の全体平均は  $\alpha = 4.060$  であった。したがって、朗読開始直後から  $\alpha \approx 4.060$  程度の変調を付与しても、少なくとも主観報告ベースでは検知率が 80.35% 程度にとどまる水準であると解釈できる。故に、先行研究のような緩徐な立ち上げに限定せず、より早い段階で変調を導入できる可能性がある。変調開始を前倒しできれば、感情誘導効果の発現タイミングを早められる可能性があり、さらに変調を大きくすることで感情誘導を強められる可能性もあるため、変調の大きさ  $\alpha$  と感情誘導量の関係、および検知(違和感)とのトレードオフを定量的に検討することが重要である。

## 6. おわりに

本稿では、聴覚フィードバックを用いた感情誘導(AAF)の実用化に向け、話者が変調を検知しない範囲である「感情的音声変調の弁別閾」を定量的に測定し、その個人差要因を検討した。

Happy 変調を用いた朗読課題における 2down-1up staircase 法による測定の結果、弁別閾は個人間および条件間で広い分布を示し、標準偏差も大きく、一律の基準を変調強度として設定することが困難であることが示唆された。また、線形混合効果モデルを用いた解析の結果、音楽経験や自己音声への曝露頻度、自己関連指標といった事前に想定した個人差要因は、本実験の範囲内では弁別閾の変動を説明する有意な因子とはならなかった。加えて、無変調の自己音声を変調音声と誤認する被験者群が存在するなど、自己音声知覚の複雑さも観察された。

これらの知見は、万人に共通する「気づかれぬ程度の効果的な変調強度」の単純なモデル化が容易ではないことを示唆すると同時に、実システムにおいては個人ごとの閾値に合わせた適応的なキャリブレーションが必要不可欠であることを裏付けるものである。今後は、サンプルサイズの拡大や骨導音提示の影響、変調の時間的ダイナミクスの

操作を含めた多角的な検討を進め、より精緻な知覚モデルを構築することで、ユーザのメンタルヘルスケアやコミュニケーション支援に資するロバストな感情誘導システムの実現を目指す。

## 6.1 本研究のまとめ

本研究で得られた知見を以下に要約する。

- Aucouturier らの DAVID に基づき、遅延約 18ms のリアルタイム AAF システムを構築し、happy 変調の強度を変調率  $\alpha$  の 1 パラメータで制御可能な実験系を整備した。
- 大学生 16 名の朗読課題において、2down-1up staircase 法 (ステップ比 0.5488) とキャッチ試行 ( $\alpha = 0$  を 10%) により、感情的音声変調の弁別閾を測定した。
- 弁別閾は平均  $4.060 \pm 2.486$ 、中央値 3.470 であり、個人間および条件間で広い分布を示した。このばらつきは、一律の変調強度による設計の一般化が難しいことを示唆する。
- 音楽経験、自己音声曝露頻度、自己関連指標など複数の個人差要因を収集して検討したが、候補因子は lasso により選択されず、主要因子 (朗読文カテゴリ、セッション順序、朗読平均 F0) も弁別閾に対する有意な効果を示さなかった。
- 無変調音声に対する false alarm が極端に高い被験者 (self-loss 群) が一定数存在したが、収集した個人差要因では当該群の特徴を説明できなかった。

以上より、実運用を想定した AAF による感情誘導では、個人ごとの弁別閾に合わせた適応的キャリブレーションを前提に設計する必要がある。

## 6.2 今後の展望

本研究ではサンプル数の制約により、正則化と回帰を同時に行うのではなく、lasso による変数選択と線形混合効果モデルによる推定を分離した二段階アプローチ [30] を採用した。十分なサンプル数が確保できれば、正則化付き線形混合効果モデルを一括して推定することにより、より高精度な予測モデルを構築できると考えられる。

また、無変調と変調を取り違えるケースが一定数存在していることが示されたが、このような事象が起こりやすい実験設計 (例えば変調強度の立ち上がり時間や呈示順序) や、そのような傾向を持つ群の因子分析なども今後の検討課題である。

自己音声の気導提示に基づく感情的変調の弁別閾を検討したが、提示様式の違いによる影響も今後の重要な論点である。骨導による音声提示は、自他声弁別タスク (SOVD) の成績を向上させることが報告されており [41]、骨導提示を用いた場合の弁別閾や個人差要因を検証することが望まれる。

さらに、AAF による感情誘導 [1] で用いられた変調は時間的に段階的な操作を含んでいるのに対し、本研究では発話開始時点から一定強度の変調を付与する実験設計とした。このような発話開始直後から一定強度で付与される変調条件においても感情誘導が生じるかどうかは、現時点では十分に確認されていない。今後は、変調の時間的ダイナミクスを系統的に操作し、時間的に立ち上がりをもつ段階的変調と、発話開始から一定強度の変調とにおける感情誘導効果および弁別閾を比較検討する必要がある。さらに、利用者に気づかれにくい形で変調強度を変化させる手法や、使用者に変調後の音声を「本来の自分の声」と誤認させるための操作条件を明らかにすることも、実用システムの設計において重要な検討課題である。

本研究では happy 変調に焦点を当てたが、実際の応用では多様な感情状態の制御が想定される。今後は、sad などの感情カテゴリに対応した変調条件に対しても同様の弁別閾測定と個人差分析を行い、感情次元全体にわたる設計指針を構築する必要がある。

セッション順序 (弁別への慣れ) の影響について有意な効果は認められなかったものの、短期的な学習効果が存在する可能性は否定できない。AAF システムを長期的に使用した場合の弁別閾や感情誘導効果の変化についても、今後の検討が望まれる。

**謝辞** 本研究の一部は、科研費 23K18474, 25K21221, JST 創発的研究支援事業 JP23KJ0828, JST ムーンショット型研究開発事業 JPMJMS201 の助成を受け実施しました。

## 参考文献

- [1] J.-J. Aucouturier, P. Johansson, L. Hall, S. Satoh, and K. Watanabe, "Covert digital manipulation of vocal emotion alters speakers' emotional states in a congruent direction," *Proceedings of the National Academy of Sciences*, vol. 113, no. 4, pp. 948–953, 2016.
- [2] J. J. Gross and H. Jazaieri, "Emotion, emotion regulation, and psychopathology: An affective science perspective," *Clinical Psychological Science*, vol. 2, no. 4, pp. 387–401, 2014.
- [3] J. Costa, M. F. Jung, M. Czerwinski, F. Guimbretière, T. Le, and T. Choudhury, "Regulating feelings during interpersonal conflicts by changing voice self-perception," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–13.
- [4] L. Goupil, P. Johansson, L. Hall, and J.-J. Aucouturier, "Influence of vocal feedback on emotions provides causal evidence for the self-perception theory," *bioRxiv*, 2019, preprint.
- [5] G. T. Fechner, "Elements of psychophysics, 1860," in *Readings in the History of Psychology*, W. Dennis, Ed. Appleton-Century-Crofts, 1948, pp. 206–213.
- [6] N. E. Scheerer and J. A. Jones, "Detecting our own vocal errors: An event-related study of the thresholds for perceiving and compensating for vocal pitch errors," *Neuropsychologia*, vol. 114, pp. 158–167, 2018.

- [7] R. A. Lester-Smith, A. Daliri, N. Enos, D. Abur, A. A. Lupiani, S. Letcher, and C. E. Stepp, "The relation of articulatory and vocal auditory-motor control in typical speakers," *Journal of Speech, Language, and Hearing Research*, vol. 63, no. 11, pp. 3628–3642, 2020.
- [8] G. R. Iannotti, P. Orepic, D. Brunet, T. Koenig, S. Alcoba-Banqueri, D. F. A. Garin, K. Schaller, O. Blanke, and C. M. Michel, "EEG spatiotemporal patterns underlying self-other voice discrimination," *Cerebral Cortex*, vol. 32, no. 9, pp. 1978–1992, 2022.
- [9] L. Kishon-Rabin, O. Amir, Y. Vexler, and Y. Zaltz, "Pitch discrimination: are professional musicians better than non-musicians?" *Journal of Basic and Clinical Physiology and Pharmacology*, vol. 12, no. 2 Suppl, pp. 125–143, 2001.
- [10] J. Thomas, J. Jogia, M. Barbato, and R. Bentall, "Me, not-me: Voice note use predicts self-voice recognition and liking," *Computers in Human Behavior Reports*, vol. 15, p. 100446, 2024.
- [11] R. Alemi, A. Lehmann, and M. L. D. Deroche, "Adaptation to pitch-altered feedback is independent of one's own voice pitch sensitivity," *Scientific Reports*, vol. 10, no. 1, p. 16860, 2020.
- [12] A. M. Kring, D. A. Smith, and J. M. Neale, "Individual differences in dispositional expressiveness: Development and validation of the emotional expressivity scale," *Journal of Personality and Social Psychology*, vol. 66, no. 5, pp. 934–949, 1994.
- [13] N. Braun, S. Debener, N. Spychala, E. Bongartz, P. Sörös, H. H. O. Müller, and A. Philipsen, "The senses of agency and ownership: A review," *Frontiers in Psychology*, vol. 9, p. 535, 2018.
- [14] R. Ohata, T. Asai, S. Imaizumi, and H. Imamizu, "I hear my voice; therefore i spoke: The sense of agency over speech is enhanced by hearing one's own voice," *Psychological Science*, vol. 33, no. 8, pp. 1226–1239, 2022.
- [15] M. Schechtman, "The narrative self," in *The Oxford Handbook of the Self*, S. Gallagher, Ed. Oxford University Press, 2011, pp. 394–416.
- [16] 徳永侑子 and 堀内孝, "邦訳版自己概念の明確性尺度の作成および信頼性・妥当性の検討," *パーソナリティ研究*, vol. 20, no. 3, pp. 193–203, 2012.
- [17] R. D. Melara and L. E. Marks, "Interaction among auditory dimensions: timbre, pitch, and loudness," *Perception & Psychophysics*, vol. 48, no. 2, pp. 169–178, 1990.
- [18] K. Naruse, S. Takamichi, T. Tanikawa, S. Yoshida, T. Narumi, and M. Hirose, "Estimating confidence in voices using crowdsourcing for alleviating tension with altered auditory feedback," in *Asian HCI '19: Proceedings of Asian CHI Symposium 2019: Emerging HCI Research Collection*, 5 2020, pp. 15–22.
- [19] N. E. Tomassi, M. E. Castro, L. T. Sund, M. E. Díaz-Cádiz, D. P. Buckley, and C. E. Stepp, "Effects of sidetone amplification on vocal function during telecommunication," *Journal of Voice*, vol. 37, no. 4, pp. 553–560, July 2023.
- [20] A. Stuart, J. Kalinowski, M. P. Rastatter, and K. Lynch, "Effect of delayed auditory feedback on normal speakers at two speech rates," *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 2237–2241, May 2002.
- [21] 梶原智之, "WRIME: 主観と客観の感情強度を付与した日本語データセット," *自然言語処理*, vol. 28, no. 3, pp. 907–912, 9 2021.
- [22] L. Cosmides, "Invariances in the acoustic expression of emotion during speech," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 9, no. 6, pp. 864–881, 1983.
- [23] H. Levitt, "Transformed up-down methods in psychoacoustics," *The Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 467–477, 1971.
- [24] M. A. García-Pérez, "Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties," *Vision Research*, vol. 38, no. 12, pp. 1861–1881, 1998.
- [25] M. Lin, S. Soi-Kawase, R. Narita-Ohtaki, M. Itoh, and Y. Kim, "Reliability and validity of a self-report emotional expressivity measure: The Japanese version of the Berkeley Expressivity Questionnaire," *Japan Journal of Nursing Science*, vol. 13, no. 1, pp. 196–201, 2016.
- [26] H. Yanagida, Y. Ijima, and N. Tawara, "Effect of personal traits on impressions of one's own recorded voice," Aug. 2024, preprint, not peer-reviewed. [Online]. Available: <https://ssrn.com/abstract=4928104>
- [27] T. Asai, N. Kanayama, S. Imaizumi, S. Koyama, and S. Kaganoi, "Development of embodied sense of self scale (ESSS): Exploring everyday experiences induced by anomalous self-representation," *Frontiers in Psychology*, vol. 7, p. 1005, 2016. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01005/full>
- [28] M. Morise, F. Yokomori, and K. Ozawa, "WORLD: A vocoder-based high-quality speech synthesis system for real-time applications," *IEICE Transactions on Information and Systems*, vol. E99.D, no. 7, pp. 1877–1884, 2016.
- [29] N. M. Laird and J. H. Ware, "Random-effects models for longitudinal data," *Biometrics*, vol. 38, no. 4, pp. 963–974, 1982. [Online]. Available: <https://www.jstor.org/stable/2529876>
- [30] J. Schelldorfer, P. Bühlmann, and S. van de Geer, "Estimation for high-dimensional linear mixed-effects models using  $\ell_1$ -penalization," *Scandinavian Journal of Statistics*, vol. 38, no. 2, pp. 197–214, 2011.
- [31] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available: <https://jmlr.org/papers/v12/pedregosa11a.html>
- [33] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in *Proceedings of the 9th Python in Science Conference (SciPy 2010)*, 2010, pp. 57–61. [Online]. Available: <https://conference.scipy.org/proceedings/scipy2010/seabold.html>
- [34] C. C. Wier, W. Jesteadt, and D. M. Green, "Frequency discrimination as a function of frequency and sensation level," *The Journal of the Acoustical Society of America*, vol. 61, no. 1, pp. 178–184, 1977.
- [35] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho,

- “The mismatch negativity (MMN) in basic research of central auditory processing: a review,” *Clinical Neurophysiology*, vol. 118, no. 12, pp. 2544–2590, 2007.
- [36] A. B. Watson and D. G. Pelli, “Quest: A bayesian adaptive psychometric method,” *Perception & Psychophysics*, vol. 33, pp. 113–120, 1983.
- [37] N. Ball, M. Wisniewski, B. Simpson, and E. I. Mercado, “The impacts of training on change deafness and build-up in a flicker task,” *PLOS ONE*, vol. 17, no. 11, p. e0276157, 2022.
- [38] P. Orepic, H.-D. Park, G. Rognini, N. Faivre, and O. Blanke, “Breathing affects self-other voice discrimination in a bodily state associated with somatic passivity,” *Psychophysiology*, vol. 59, no. 7, p. e14016, 2022.
- [39] C. Rosa, M. Lassonde, C. Pinard, J. P. Keenan, and P. Belin, “Investigations of hemispheric specialization of self-voice recognition,” *Brain and Cognition*, vol. 68, no. 2, pp. 204–214, 2008, iSSN 0278-2626.
- [40] M. Candini, E. Zamagni, A. Nuzzo, F. Ruotolo, T. Iachini, and F. Frassinetti, “Who is speaking? implicit and explicit self and other voice recognition,” *Brain and Cognition*, vol. 92, pp. 112–117, 2014.
- [41] P. Orepic, O. A. Kannape, N. Faivre, and O. Blanke, “Bone conduction facilitates self-other voice discrimination,” *Royal Society Open Science*, vol. 10, no. 2, p. 221561, 2023.

## 付 録

### A.1 アンケート

アンケート項目のうち、実験前の気分・感情はベースラインとして少なくとも実験前に測定する必要がある。一方で、感想や判別の精度に対する自信、自分の声を「見失った」と感じた度合いは、実験の経験を踏まえて評価されるべき指標であるため、実験後に収集する必要がある。さらに、感情発露傾向などの感情特性に関する質問を実験前に多数提示すると、被験者が「本実験では感情的な変調を扱っている」と察知し、変調検出や弁別行動に影響を与える可能性がある。加えて、質問数が多くなると被験者の負担も増大する。これらを考慮し、本研究ではアンケートを二つに分け、実験前・実験後にそれぞれ実施した。

#### A.1.1 実験前アンケート

実験前に実施したアンケートの質問項目を表 A-1 に示す。これらの質問について、実験開始前に Google Form を用いて回答を求めた。

本研究の被験者には聴覚異常を有する者が存在しなかったため、対応する質問項目の結果は解析には使用しなかった。また、学んだ楽器に関する質問については、被験者ごとに回答内容が多様であり、サンプル数の制約からカテゴリ化が困難であったため、本研究の解析には含めなかった。

#### A.1.2 実験後アンケート

実験後に実施したアンケートの質問項目を表 A-2 に示す。これらの質問について、実験終了後に Google Form を用いて回答を求めた。

音声加工内容の推定に関する質問および実験全体に対する自由記述は、探索的情報として収集したものであり、本研究の定量的解析には使用していない。なお、音声加工内容の推定においては、実際には存在しない「声が低くなる変調」に言及する回答も一部に認められた。しかし、当該質問は自由記述形式であり、被験者全員に特定の変調内容への言及を促す設計ではなかったため、これらの記述のみを根拠として被験者を self-loss 群に分類することは妥当ではないと判断した。

### A.2 実験詳細

#### A.2.1 機材・前準備

本研究では、朗読中の自己音声をリアルタイムに加工し、ヘッドフォンから気導提示した。ヘッドフォンは SONY MDR-MV1、マイクはコンデンサマイク NEU-MANN TLM103 mt を用いた。

本番開始前に音量の較正と練習を実施した。較正では、

表 A-1: 実験前アンケートの質問項目と対応する因子・尺度

Demographics
年齢
性別
Emotion
幸福である。
うれしい。
心地よい。
楽しい/面白い。
イライラしている。
不愉快だ。
怒り/敵意を感じる。
気持ちが沈んでいる/憂うつである。
何となく心配だ/不安だ。
Music experience
音楽経験 (楽器・合唱の継続的練習) の有無。
Music genre
経験のある音楽ジャンル (クラシック・それ以外 (現代音楽))
Music years
演奏経験の年数。
Self-concept clarity (SCC)
私は自分の性格の側面に矛盾を感じることはめったにない (1-5)。
自分自身に関する信念は頻繁に変わる (1-5)。
自分自身に関する信念はしばしば互いに矛盾する (1-5)。
これまでの自分がどのような人間であったか確信がない (1-5)。
日ごろの自分が本当の自分ではないと感じる (1-5)。
性格の記述は日によって異なる (1-5)。
自分が誰であり何者であるかを明確に自覚している (1-5)。
別の日には異なる自己意見を持つ (1-5)。
自分が何をしたいかわからず決心が困難 (1-5)。
本当の自分について考えることに多くの時間を費やす (1-5)。
本当の自分について誰にも話さない (1-5)。
自分より他者のほうがよくわかっていると感じる (1-5)。
Excluded from analysis
これまでに聴覚に異常 (耳の病気や難聴など) を経験したことはありませんか。
(はいの場合) よろしければ具体的な症状や診断名を記入してください。
学んだ主要な楽器 (または歌唱)。
他に学んだ楽器 (または歌唱)。

所定のマイク位置 (6 cm) において、マイク入力とヘッドフォン出力の音圧が等しくなるよう事前に調整した。これにより、自己音声フィードバックの提示レベルを安定化させた。

また、練習時に「本番では、演技をせずに、文の場面を想像しながら自然な状態で文を朗読してください」と教示し、マイクとの距離を一定に保つように口頭で指示した。

#### A.2.2 システム実装

音声変調は、AAF による感情誘導を扱った先行研究 [1] が配布する Max 用パッチ DAVID をベースとして用いた。元パッチに含まれる変調アルゴリズムおよび関連する外部オブジェクトは基本的にそのまま利用し、本研究では主として発話検出部と実験制御部を拡張した。具体的には、(i) 発話区間検出を外部 (Python) 側で安定化させ、(ii) 検出

表 A.2: 実験後アンケートの質問項目と対応する因子・尺度

Question item
<i>SVE</i>
1 週間に、自分の声を含む動画を SNS に投稿する頻度.
1 週間に、自分の声を録音する頻度.
<i>Voice self-loss</i>
実験中に自分の声をどれだけ見失ったか (1-7).
<i>Discrimination confidence</i>
判別の精度に対する自信 (1-7).
<i>EES</i>
私は自分の感情を他人にみせない (1-5).
強い感情でも外に表さない (1-5).
他人は私を感情的だと思っている (1-5).
他人は私の感情を読み取れる (1-5).
自分の気持ちを内に秘める (1-5).
感情に気づかれにくい (1-5).
自分の感情を表すほうだ (1-5).
感情的ではないと思われる (1-5).
感情を知られるのが嫌だ (1-5).
感情を隠すことができない (1-5).
感情表現が豊かではない (1-5).
無感動な人だと思われる (1-5).
人前で泣くことができる (1-5).
感情的でも周囲にみせない (1-5).
自分は感情表現が豊かだと思う (1-5).
他人の解釈と自分の感情は異なる (1-5).
気持ちを内に閉じ込める (1-5).
<i>Ownership</i>
服が重く感じることもある (1-5).
体がロボットのように感じる (1-5).
考えが他人のもののように感じる (1-5).
服の感触が気になる (1-5).
自分の姿に気づかない (1-5).
身体が冷えていることに気づく (1-5).
存在が現実でないように感じる (1-5).
身体が軽く感じる (1-5).
自分を遠くから観察している感覚 (1-5).
<i>Narrative</i>
自分の性格がわからなくなる (1-5).
性格は状況で変わると感じる (1-5).
無関係な考えが入り込む (1-5).
頭の中の音楽が止まらない (1-5).
過去と現在の自分が別人のように感じる (1-5).
性格を誤解されていると感じる (1-5).
記憶が抜け落ちる期間がある (1-5).
思い出し笑いをしてしまう (1-5).
<i>Agency</i>
物を失くしても気づかない (1-5).
物をよく落とす (1-5).
物や人にぶつかる (1-5).
声が大きすぎると言われる (1-5).
霊的なものに気づかないかもしれないと感じる (1-5).
手探りで物を探すのが難しい (1-5).
言おうとしたことを忘れる (1-5).
自分はずぐったがりだと思う (1-5).
<i>Free comment</i>
今回の実験における音声加工内容の推定.
実験についての自由記述.

結果と変調条件を Max/MSP 側へ送信して試行制御と同期させる構成とした。

音声入出力には `sounddevice`<sup>\*4</sup> を用い、サンプリング周波数は 48 kHz とした。OS は Windows とし、オーディオバックエンドには WASAPI を使用した。システム遅延が約 18 ms となるようバッファ設定を行い、朗読中の音声リアルタイムで処理されてヘッドフォンヘッドバックされるよう構成した。

検出結果（発話開始・終了など）と変調条件（変調率  $\alpha$  など）は、OSC プロトコルを介して Max/MSP 上の DAVID パッチへ送信した。これにより、外部（Python）側での検出と Max/MSP 側の音声変調処理をリアルタイムに連携させ、実験制御を実現した。

### A.2.3 発話区間検出 (VAD) の詳細

元の DAVID パッチに含まれる発話検出は本研究の環境では十分に機能しなかったため、発話区間検出には Webrtc の GMM ベース VAD を、Python バインディングである `webrtcvad`<sup>\*5</sup> を通じて使用した。入力音声を 10 ms フレームに分割し、各フレームについて VAD の二値判定 (speech/non-speech) を得た。

フレーム単位の判定は揺らぎが大きいため、60 ms の移動窓に含まれる speech フレーム比 (speech ratio) を算出し、これをもとに発話開始・終了を決定した。さらに誤検出低減のため、開始判定と終了判定で異なる閾値を用いるヒステリシス制御、および無音継続時間の閾値（無音が一定時間以上続いた場合に終了とみなす）を導入した。この処理により、短い無音や瞬間的ノイズによる不要な開始・終了判定を抑制した。

## A.3 階段法集計における最終反転数 $R$ の妥当性

### A.3.1 $R$ 決定手順

本研究では、staircase の安定期に対応すると考えられる最後の  $R$  回のステップ方向反転点における変調率  $\alpha$  を抽出し、それらの平均値を当該条件における弁別閾と定義した。具体的には、 $R = 3, 4, 5, 6, 7, 8$  の各値について弁別閾を算出し、分布の安定性と分散の大きさを比較したうえで、最終的に採用する反転回数  $R_{\text{choice}}$  を決定した。

さらに、選択した  $R_{\text{choice}}$  が staircase の安定期を適切に捉えているかを確認するため、各条件について最終  $R_{\text{choice}}$  回の反転点を時間順に並べ、その出現順序に  $1 \sim R_{\text{choice}}$  の通し番号を付与した。この通し番号を説明変数、対応する変調率  $\alpha$  を目的変数として単回帰分析を行い、各条件ごとに傾きの推定値を得た。続いて、得られた傾きの推定値が 0 から有意に異なるかどうかを一標本  $t$  検定により評価

\*4 <https://python-sounddevice.readthedocs.io/en/0.5.3/>

\*5 <https://pypi.org/project/webrtcvad/>

表 A.3: 反転回数  $R$  ごとの弁別閾の記述統計量. 平均値の最小は  $R = 6$  であり, 中央値の最小は  $R = 7$  である. なお, 有効なサンプル数  $N$  は  $R$  が増えるごとに単調減少する.

$R$	Mean	Median	$N$
3	4.599 ± 2.898	3.716	45
4	4.330 ± 2.683	3.631	43
5	4.225 ± 2.651	3.610	41
6	4.060 ± 2.486	3.470	40
7	4.135 ± 2.443	3.466	39
8	4.192 ± 2.503	3.516	35

し, 最終  $R_{\text{choice}}$  回の反転区間において変動率が十分に収束しているかどうかを検証した.

あわせて, 以下の条件を満たすデータは信頼性を欠くと判断し, 解析から除外した.

- 反転回数が  $R_{\text{choice}}$  未満である条件
- 弁別閾の  $Z$  スコアの絶対値が 3 以上である外れ値
- キャッチ試行に対する False Alarm 率が 90 % 以上

### A.3.2 結果

反転数  $R$  を変更した際の弁別閾の分布を図 A.1 に, それぞれの  $R$  における平均値・標準偏差・中央値・有効条件数を表 A.3 に示す. 弁別閾の平均値は  $R = 3$  で 4.599 と最も大きく,  $R = 4, 5, 6$  と増やすにつれて 4.330, 4.225, 4.060 と単調に減少した.  $R = 7, 8$  では平均値がそれぞれ 4.135, 4.192 とわずかに増加するものの,  $R = 6$  を含め 4.0 台前半に収まっており, 大きな変化は見られない. 標準偏差も  $R = 3$  の 2.898 から  $R = 6 \sim 8$  では 2.4 ~ 2.5 の範囲に縮小しており,  $R = 6$  以降で弁別閾分布が相対的に安定していることが示唆される. 一方で, 有効条件数  $N$  は  $R = 3$  の 45 から  $R = 8$  の 35 まで単調に減少しており,  $R$  を大きくしすぎるとサンプルサイズの低下が問題となる.

これらを踏まえ, 本研究では分布の安定性とサンプルサイズのバランスが良い  $R = 6$  を最終反転数  $R_{\text{choice}}$  として採用した. 選択した  $R_{\text{choice}} = 6$  について, 最終 6 回の反転点に対する線形回帰を行ったところ, 傾きの平均値は 0.0280 ( $\pm 0.1895$ ) であった. この傾きが 0 から有意に異なるかどうかを一標本  $t$  検定により評価した結果,  $t(39) = 0.935$ ,  $p = 0.356 > 0.05$  となり, 有意な差は認められなかった. したがって, 最終 6 回の反転区間において弁別閾は十分に収束していることが示唆された. よって本文での解析には  $R = 6$  に基づく弁別閾を用いた. なお,  $R = 6$  の弁別閾に対して残りの除外基準 ( $Z$  スコアの絶対値が 3 以上・FA 率が 90 % 以上) を適用した結果, 有効データは 40 条件中 34 条件となった.

表 A.4: 因子間相関

Factor A	Factor B	Cross-factor corr.
Emotion expression	Narrative	0.702
Self voice exposure	Gender	-0.674
Narrative	Agency	0.666
Music experience	Genre category	0.604
Music experience	Music experience	0.595
Voice self-loss	Discrimination confidence	-0.588
Positive emotion	Negative emotion	-0.528
Music experience	Genre: classic vs other	0.522
FA rate	Agency	-0.511

## A.4 因子間の交互作用

採取した因子同士の交互作用を確認するため, 各因子について被験者ごとに平均値を算出し, 16 名ぶんのデータに対して相関係数を計算した. 朗読 F0 平均 および False alarm 率についてはセッションごとに平均をとったうえで, 被験者単位で集約した. 得られた相関係数の絶対値が特に大きい組み合わせを表 A.4 に示す. 最大でも相関係数はおよそ 0.7 程度であり, 極端に強い相関を示す組み合わせは見られなかった. このことから, 本研究で扱った因子間に顕著な多重共線性が存在する可能性は低いと考えられる. 一般に被験者が自然に朗読した際, ポジティブな刺激文である程, 朗読 F0 平均が高くなる [22], しかし図 A.2 に示されるように強い相関は持たない. なお, 文の感情はカテゴリ因子であり, 朗読 F0 平均は連続因子のため, 相関を算出してはいない.

## A.5 使用した刺激文

実際に使用した刺激文を, 表 A.5 (happy 条件), 表 A.6 (neutral 条件), 表 A.7 (sad 条件) に示す.

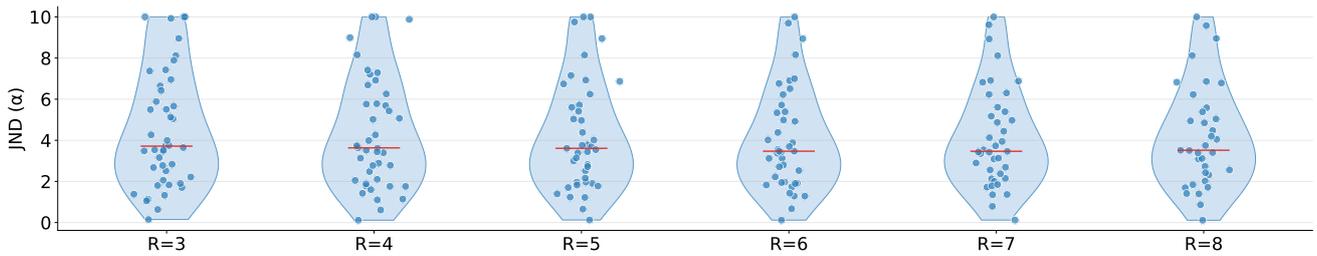


図 A-1: 弁別閾分布 ( $R = 3 \sim 8$  の比較).  $R$  が大きくなるにつれて分布の広がり小さくなり、中央値 (赤線) も低下する傾向がみられる。

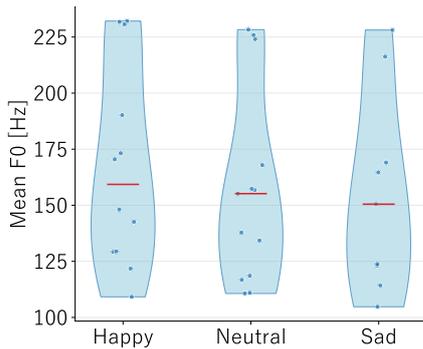


図 A-2: 朗読平均 F0 の感情カテゴリ別分布. 上から happy, neutral, sad の順に朗読 F0 平均の中央値が少しだけ高い傾向にあることが示される。

表 A-5: 使用した刺激文 (happy 条件)

Stimulus sentence
ホットカーベットに寝転がるのが気持ちよくて最高。
大好きな友達とご飯を食べるのが、本当に幸せ。
何にも縛られない久しぶりの休日。何をしようかな？
中学生の頃読んでた漫画の作者と偶然話せた。テンション上がった。
友達と数年ぶりに連絡取れた。嫌われてなくてよかった。
修論発表一応合格もらえた。頑張ったかいがあった。
今日もいい夢が見れた。最高の気分になった。
高校野球、母校が初戦突破。おめでとうございます。
いやー、本当に日本酒飲み放題幸せだった。
明日は休み。もらった写真をゆっくり鑑賞するんだ。
今日は高校からのお友達の結婚式。晴れ姿楽しみだ。
今日もよく笑った。心から笑える人と一緒にいると人生楽しいな。
オンラインで飲み会ができる時代。最高に楽しかった。
この連休は愛する家族とたくさん一緒に過ごせて楽しかった。
ダイエットの成果に気づかれた。たくさん褒められて嬉しい。
知名度やお金はないけど、人に恵まれてるからとっても幸せ。
初めての給料日。いろいろな支払いで消えるけど、それでも嬉しい。
部活の先輩みんないい人ばかりで本当に恵まれてるな。
すぐに明日遊べる友達が見つかる。これって幸せだな。
お寿司食べすぎておなか痛い。幸せものだな私は。
今日はとても天気が良いです。快晴。気分が良いですね。
笑顔が素敵って、それだけで人として魅力的だよな。
来月は有休を使って夏休みをとるぞ。どこにお出かけしようかな。
サイン入りのプログラム抽選にあたりました。ありがとうございます。
初めてブログの感想をもらえた。とっても嬉しい。

表 A-6: 使用した刺激文 (neutral 条件)

Stimulus sentence
娘、この人の顔がかっこいいって言ってる。
今日夢見たんだけど、内容忘れちゃった。
買い出しから帰るまで、夜遊びに出かける人を見送った。
明後日は不動産屋にいてこない。雨が降らないといいけど。
土日の電車は快速にしようかな。
好きな曲とカラオケで歌う曲って別だと思うな。
6時半に起きたけど、二度寝したから今起きた。
これまで身内が出した本しか買ってこなかったな。
聲 (こえ) の形はうちの息子にはまだ難しいみたい。
東急の日吉駅は仕事で毎日乗り降りしてた。
娘は雨がやんでからの下校になったらしく、さっき帰ってきた。
ドタキャンに怒るのは、相手が 100% 来ると思ってるから。
買うか迷うってことは、必要ないものだよな。
「結婚しないと結構暇」っていう意見も、分かる気がする。
需要があるか分からない場合は、小さく始めるのが鉄則かも。
身長 170cm って、男性ならそこまで高くないのかな。
教授が休みだし、研究室いってもやることないよな。
今すぐ寝てもいいけど、小腹がすいてきたな。
大体の雨は降っていないものと考えて傘を差さずに歩いてる。
ディーラーニングそのものにはあまり興味はないけど下準備として学んでおくか。
夢の続きが見たいがためにもう一度寝たい。
いろいろと足りないものはあるが、差し当たりコンセントが足りない。
自分にとってはいらないものでも他の人から見たら欲しいものもあるよね。
20 歳になって俺も自立しなきゃと思いつつも、仕送りを増やしてもらった。
ライバルとの差別化をしようと思って、需要のないことをやってしまうのは逆効果。

表 A.7: 使用した刺激文 (sad 条件)

Stimulus sentence
ああもうしんどいしんどい。2025 年は大失恋の年だ。 靴を履こうとしただけでぎっくり腰になって泣ける。 胃の潰瘍（かいよう）が酷くて、ご飯が食べられない。 帰省するとき毎回会うおじいちゃん、ガンだった。 人の死に直面するのは辛いな。 明日も無駄なサービス残業が待っているのかな。 昨日は娘と色々あって、凹んで号泣しちゃった。 明日着るつもりのアウターに油汚れがついちゃった。 せっかくのバレンタインなのに風邪ひいて家で寝込んでた。 夫がストレスで子供にあたってしまい、もうだめ。 なんかいろいろあって涙出てきた。電車の中で。 機種変更したら、友達との思い出が全部消えちゃった。 かつての親友が苦手になっちゃった。こうして友人が減っていく。 人間関係に嫌気がさしてきた。山にでもこもりたい。 あまり言いふらしたくないものではないけど、今日親友の命日。 顔や声についていじられるたび笑って返してたけど、今日はただ苦しい。 悲しすぎて泣けてきちゃう。家族がそばにいない寂しさ。 精神が不安定だ。仕事辞めたい。もう人と関わりたくない。 会社辞めたいなあ。毎日頭痛がする。上司に合うのが嫌だ。 頻繁に起こる痙攣（けいれん）もどうにもできず、最期（さいご）は レタス 1 口食べて亡くなっちゃった。 今の部活メンバーで毎日バカできなくなるって考えると、なんだか 寂しい。 やりがいのない長時間労働はつらい。また環境変えるしかないのか なあ。 自分に余裕がないから心が狭くなって、思いやりを忘れてしまう。 歳を取ると大切な子供と過ごした日々とかも思い出せなくなるのか な。