

2022 年度
実験計画書

発話スタイルに着目した
会話比較を目的とする尺度の構築

指導教員: 菊池英明

早稲田大学 人間科学部

学籍番号:1J19E058-1

加藤隆聖
Kato Ryusei

目次

1	はじめに	3
1.1	背景	3
1.2	目的	7
1.3	波及効果	7
2	研究計画	8
2.1	研究計画の概要	8
2.2	使用データ	8
2.3	尺度の構築方法	8
2.4	尺度の妥当性	8
	参考文献	9

図目次

1	音声による情報伝達過程の模式図	3
2	パラ言語情報を考慮した音声生成過程の模式図	4
3	The speech chain	4
4	発話向け先ごとの発話の平均 Fo の分布	5
5	快-不快、覚醒-眠気の 2 軸のパラメータで表した図	6
6	親しさ・明瞭さ・社会階層の三尺度でスタイルを捉えた図	6

1 はじめに

1.1 背景

対話場面における情報伝達は複雑である。これは音声には様々な情報が含まれており、会話時には話し手と聴き手の間に高度なやりとりがなされているためである。音声に含まれる情報は、言語情報、パラ言語情報、非言語情報の三つに分類することができる。以下に藤崎による3分類を引用する。言語情報とは、書き言葉によって明示的に表現される情報や文脈から一意かつ容易に推測することが可能である情報のことを指す。パラ言語情報とは、書き言葉に転写すると推測不可能となる情報で、言語情報を補助ないし変容するために話者が意図的に生成する情報のことを指す。特に発話に込められた話者の意図、態度や発話のスタイルなどが該当する。非言語情報とは、話者の年齢、性別、個人性、身体ないし感情の状態などの要因に関わる情報のことを指す。これらの要因は、発話の言語的・パラ言語的内容とは直接に関係せず、話者が意図的に制御することも一般には不可能である [1]。これら三つの情報を伝達するときに、話し手と聴き手の音声伝達の模式図を図1に示す。図1では、聴き手においてパラ言語情報と非言語情報が必ずしも正確に知覚されるとは限らないため、破線の矢印で示されている。パラ言語情報の影響は発話全体にわたって顕在化することもあるが、むしろ多くは発話の冒頭や末尾において局所的に顕在化している [2]。パラ言語情報を分析するためには、発話全体を対象とすることが有効であることが示されている。

このように、音声伝える情報について研究が行われてきたが、パラ言語情報や非言語情報がどのような仕組みで生成され、伝達されているのかについて全てを明らかにすることはできていない。

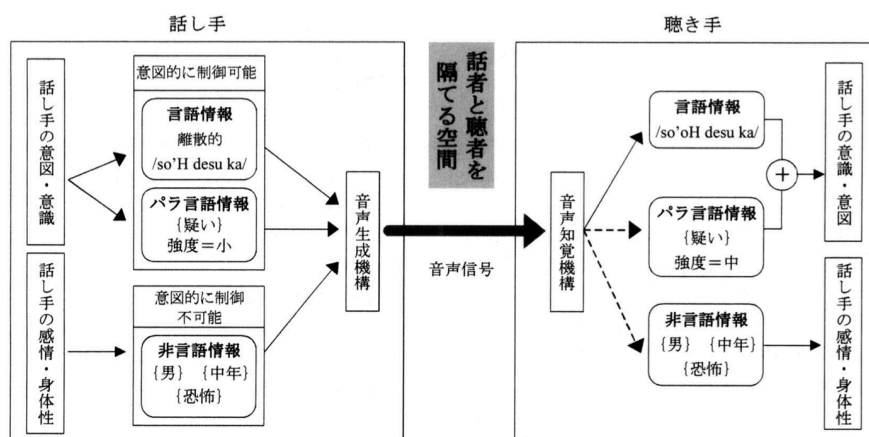


図1: 音声による情報伝達過程の模式図。[4]より抜粋。話し手が聞き手へ言葉発した時の情報の伝達方法を模式的に表している。話し手発した情報の中でパラ言語情報と非言語情報が必ずしも正確には伝わらないことが示されている。

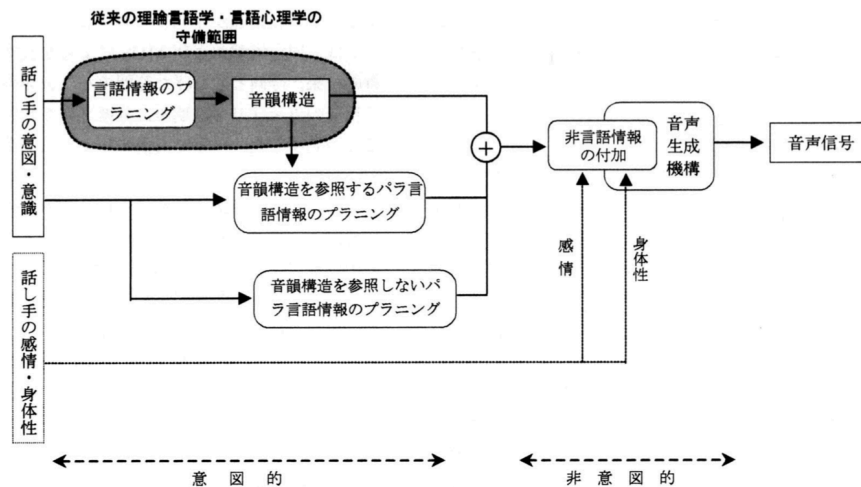


図 2: パラ言語情報を考慮した音声生成過程の模式図。[2] より抜粋。話し手が音声を発する時に、言語情報、パラ言語情報、非言語情報がどのように組み合わせて生成されているのかについて模式的に表されている。この画像から感情と身体性を含む非言語情報の付加は非意図的に生成されていることが示されている。

パラ言語情報を考慮した音声生成過程の模式図を図 2 に示す。図 2 によると、発話する際の音声に含まれる情報には意図的なものと非意図的なものが存在する。さらに、意図的なものに含まれるパラ言語情報には音韻構造を参照するものと参照しないものに分けられている。このように音声のみに対してアプローチしている研究はなされているが、発話者同士の関係や発話者のスタイルに着目はされていない。

話した言葉が話し手の脳から聞き手の脳へ伝わる過程の模式図を図 3 に示す。図 3 では、話し手から発せられた音声と聴き手が音声を知覚するまでの過程がつながりを持っていることを示している。よって、話し手の意図と聴き手が知覚した内容には齟齬が生じることがある。

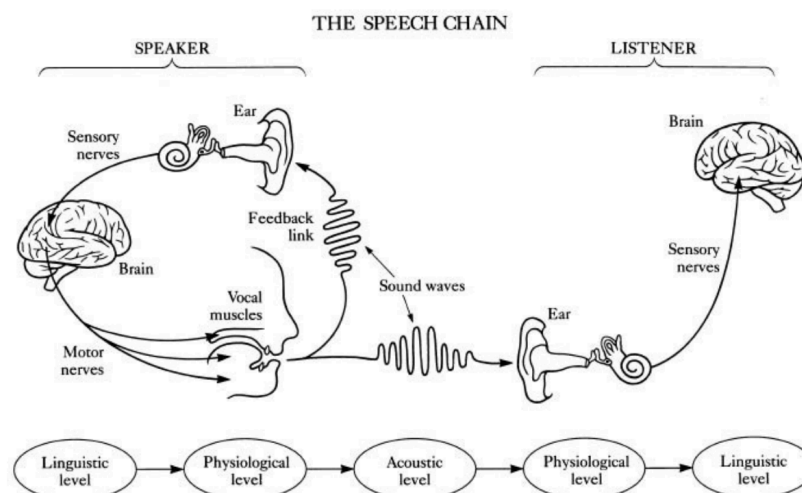


図 3: The speech chain。[?] より抜粋。話した言葉が話し手の脳から聞き手の脳へ伝わる過程が表されている。

Wertsch によると、自分の音声は他者の声から切り離されて存在するのではなく、発話が向けられる他者の

声に基づいて作られるのである。また、音声は他者との関係において成り立つのであり、音声には状況依存性がある [3]。その状況とは、話し手と聴き手が物理的環境を共有することである [4]。このように発話は他者からの声に反応して発せられるものであるため、日常に現れる微細な音声の変化を分析するためには、対話音声进行分析することが妥当である。

石本によると、発話向け先との関係の違いによって日常対話の声の高さが変化することがわかっている。また、会話場面の同席者している参与者との関係によって韻律的多様性が生じていることがわかっている [5]。発話向け先との関係の違いや同席している図を 4 に示す。図 4 によると、家族に向けた発話は一般的に低い Fo で発声され、丁寧な発話ほど Fo が高くなることが読み取れる。つまり、対話相手によって発話スタイルが変わるため、パラ言語情報の表出の仕方が変わるのである。よって、パラ言語情報が発話者のみによって発せられるのではなく、発話相手によってパラ言語情報の表出の仕方が変わるため、音声のみに注目してはいけなくと考えられる。音声のみに注目するのではなく、対話者の発話のスタイルや対話場面についても考慮すべきであると考えられる。

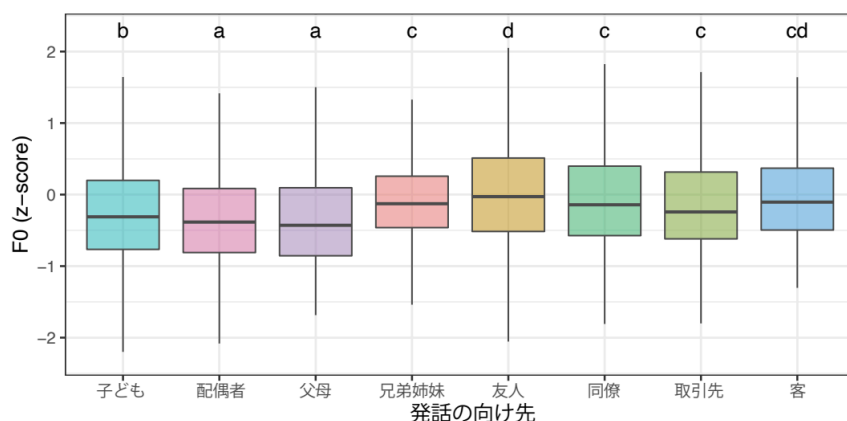


図 4: 発話向け先ごとの発話の平均 Fo の分布。[5] より抜粋。

従来までは、対話場面が限定されたコーパスしかなかったため、日常的かつ自然な音声の分析を行うことはできなかった。例えば、オンラインゲームを行いながら話しているコーパスについての分析 [6] や講演を行なっている音声のコーパスについての分析 [7] である。これらの研究には、パラ言語情報の表出としての表現が限定的になってしまうという問題がある。音響特徴量は感情ラベルに比べコーパスの特性の違いの影響を受け易いことがわかっている [8]。よって、従来の対話音声研究では、パラ言語情報が収録したコーパスの特性の影響を受けやすいため、パラ言語情報を普遍的に明らかにするには至らなかった。しかし、国立国語研究所では 2016 年より、日常生活の中で自然に生じる多様な場面・話者の会話 200 時間を対象とする『日本語日常会話コーパス』(CEJC) の構築され 2022 年に公開された [9]。多様な関係性における対話を収録した日常対話コーパスによって、コーパスの特性に限定されことなく対話音声进行分析できるようになった。また、従来までの研究は、異性と同性・親密さの違いなどの限られた関係の分析にとどまっている。そのため、多様な関係性を考慮した包括的な分析を行うことができていない。そこで、日常的かつ自然な音声に対して多様な話者の関係性を考慮した研究が必要であると考えられる。

音声进行分析する方法としては、横軸に快-不快、縦軸に覚醒-眠気の 2 軸をとりこれらのパラメータによって感情を捉える方法 [10] や基本 8 感情 (acceptance, anger, anticipation, disgust, joy, fear, sadness,

surprise)[11] として捉える方法があるが、日常に現れる微細な発話をこのような方法で捉えることは難しい。そこで、日常的な会話を包括的に捉えることのできるような、親しさ・明瞭さ・社会階層 [12] という三尺度で捉えた方法が有効であると考えられる。

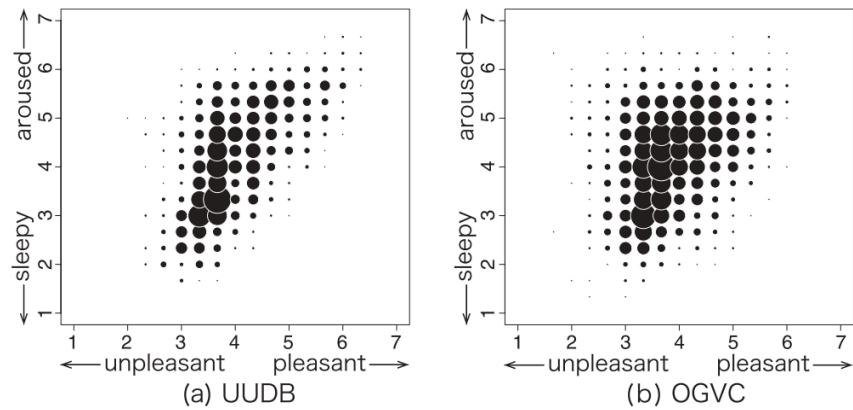


図 5: (a) は「4コマまんがの並び替え課題」を与えられた会話、(b) はオンラインゲームをしながら行った会話である [10] より抜粋。

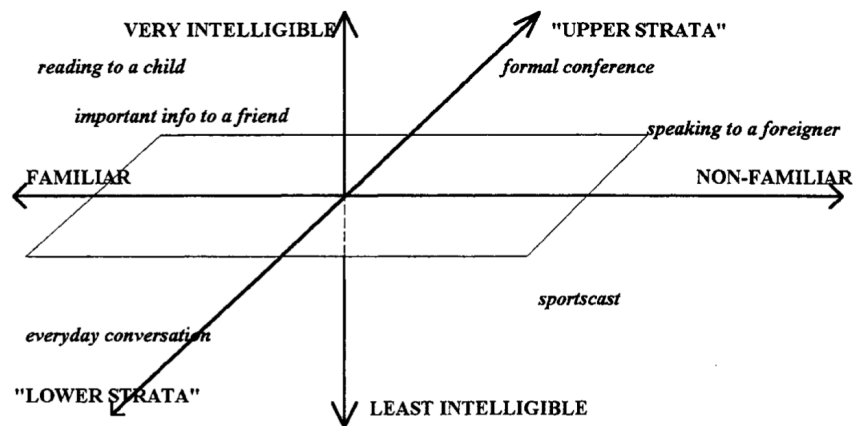


図 6: 親しさ・明瞭さ・社会階層の三尺度で speaking style を捉えた図 [12] より抜粋。

1.2 目的

本研究では、日常的かつ自然な対話音声进行分析するための尺度を構築することを目的とする。日常会話に現れる同一話者の発話や異なる話者の発話を比較することができるような尺度を構築する。この目的を達成するために、次の三つのリサーチクエスチョンを設定した。

1. 日常会話を比較するための尺度は何なのか
2. 見つけた尺度は妥当であるのか
3. 日常に現れる発話スタイルを包括的に表すことができているのか

1.3 波及効果

パラ言語情報を構成する要素を明らかにすることで、パラ言語情報を分析する際の一つの指標として確立することができる。また、パラ言語情報の表出方法の仕組みを明らかにすることにもつながるであろう。さらに、スタイルの違いに応じたパラ言語情報の表出方法がどのようなものであるかを明らかにできる。その結果、場面に適した音声を使うことで、教育やトレーニングを行う際にも、どのような音声が指導に適切であるか知ることができる。それに加え、パラ言語情報についての知見を応用することで、音声合成、音声変換の多様性を実現できるであろう。

2 研究計画

2.1 研究計画の概要

日常に現れる会話を分析するためには、様々な年齢、場面、状況、対話相手との関係性などを比較する必要がある。しかしこれらの要素が複雑に絡み合ったものを分析する尺度がないと言える。そのため、複雑に絡み合った要素を包含するような尺度が必要である。

2.2 使用データ

使用データは、多種多様な関係性の会話が 200 時間収録された『日本語日常会話コーパス』(CEJC)を用いる。様々な年齢、場面、状況、対話相手との関係性などを比較する必要があり、『日本語日常会話コーパス』はこれらの要素を多く含んでいるからである。データの数はいくつあるか、その中でも二人で行われた会話 280 データを用いる。日常会話コーパスは協力者 40 名に機材を貸し出し、日常生活で自然に生じる会話を記録している。協力者の年齢は 20 代から 60 代以上までの男女と幅広い年齢層の会話が収録されている。収録方法には個人密着法と特定場面法の 2 種類がある。個人密着法は協力者に機材をつけてもらい会話を収録するというものである。特定場面法は、個人密着法のみでは不足してしまうような会話を補うことを目的としているため、会議や会合、中高生の雑談打ち合わせなどの場面に絞って収録している。データとしては、会話の形式・場所・活動内容・話者間の関係などのメタ情報まで収録されている。

2.3 尺度の構築方法

言語情報から Speaking style を推定するモデル [13] を日常会話コーパスに用い分布を確かめる。分布の仕方や明らかに異なるような発話が近くに付置されていないかを確かめる。言語情報のみでは不十分だった場合には音声情報を使ったモデルを構築し、尺度を構築する。さらに、親しさ・明瞭さ・社会階層 [12] の三尺度がだとなかった場合には、音響的特徴量のみで比較する方法やインターネット上から日本語の口調を表す語を集めた口調 [14] を元に日常会話を比較する尺度を構築していく。集められた口調を元に七件法の質問を行い、質問結果から印紙分析を行い日常会話音声評価のための尺度構築を行う。さらに、尺度の妥当性をはかるために二回目の印象評定を行う。因子において因子付加量の高い項目を選抜し、尺度として再構築する。

2.4 尺度の妥当性

尺度の一貫性をはかる G-P 分析や I-T 相関分析で妥当性をはかる。

参考文献

- [1] Fujisaki, Hiroya. "Prosody, models, and spontaneous speech." *Computing prosody*. Springer, New York, NY, 1997. 27-42.
- [2] 前川喜久雄, and 北川智利. "音声はパラ言語情報をいかに伝えるか." *認知科学* 9.1 (2002): 46-66. 17.1: 109-118.
- [3] Wertsch, James V. *Voices of the mind: Sociocultural approach to mediated action*. Harvard University Press, 1993. pp. 46 - 63 (ジェームス・ワーチ著, 田島信元ほか訳 (2004) 心の声, 福村出版, 東京, pp. 67 - 89)
- [4] 前川喜久雄 (1998). 音声学. 『岩波岩波講座言語の科学 第2巻, 音声』, 1-52, 岩波書店.
- [5] 石本祐一. "日本語日常会話コーパスから見える会話場面と声の高さの関係性." *言語資源活用ワークショップ発表論文集= Proceedings of Language Resources Workshop*. No. 5. 国立国語研究所, 2020.
- [6] 有本泰子, et al. "感情音声のコーパス構築と音響的特徴の分析-MMORPG における音声チャットを利用した対話中に表れた感情の識別." *情報処理学会研究報告音楽情報科学 (MUS)* 2008.12 (2008-MUS-074) (2008): 133-138.
- [7] 前川喜久雄. "『日本語話し言葉コーパス』を用いた自発音声の分析." *研究報告 人文科学とコンピュータ (CH)* 2011.4 (2011): 1-4.
- [8] 永岡篤, 森大毅, and 有本泰子. "感情音声コーパス共通化のための新たな感情ラベル推定における既存感情ラベル併用の効果." *日本音響学会誌* 73.11 (2017): 682-693.
- [9] 小磯花絵, 菊池英明, and 山田高明. "『日本語日常会話コーパス』への韻律ラベリング-ラベリングの設計と日常会話の韻律の特徴." *人工知能学会研究会資料 言語・音声理解と対話処理研究会*. 一般社団法人 人工知能学会, 2020.
- [10] 永岡篤, 森大毅, and 有本泰子. "感情音声コーパス共通化のための新たな感情ラベル推定における既存感情ラベル併用の効果." *日本音響学会誌* 73.11 (2017): 682-693.
- [11] Plutchik, Robert. "Emotion." *A psychoevolutionary synthesis* (1980).
- [12] Eskenazi, Maxine. "Trends in speaking styles research." *Third European Conference on Speech Communication and Technology*. 1993.
- [13] 沈睿, and 菊池英明. "音声言語コーパスにおける speaking style の自動推定—転記テキストに着目して—." *自然言語処理* 21.3 (2014): 445-464.
- [14] 郡史郎. "日本語の「口調」にはどんな種類があるか (i 特集, i イントネーション研究の現代的課題)." *音声研究* 10.3 (2006): 52-68.