

音声コミュニケーションにおける 表現力の測定の研究

Studies on Expressiveness Measurement
in Speech Communication

2013年12月

早稲田大学大学院 理工学研究科

情報・ネットワーク専攻
ヒューマンインターフェース研究

宮島 崇浩

Takahiro MIYAJIMA

目次

第1章 序論	1
1.1 背景	1
1.2 コミュニケーションメディアの定義	3
1.3 コミュニケーションメディア多様化の社会的影響	5
1.4 本研究の目的	6
1.5 本論文の構成	8
第2章 コミュニケーションメディアの種類と評価	11
2.1 コミュニケーションメディアの歴史	11
2.1.1 音声言語の誕生以前	11
2.1.2 音声言語の誕生	11
2.1.3 遠隔コミュニケーションメディアの登場と発展	12
2.1.4 電話の発明	13
2.1.5 インターネット黎明期	14
2.1.6 ブロードバンド回線の普及	14
2.1.7 近年から現在まで	15
2.2 コミュニケーションメディアの性質の分類	17
2.3 近年使われるコミュニケーションメディア	27
2.3.1 音声中心のコミュニケーションメディア	27
2.3.2 文字中心のコミュニケーションメディア	28
2.3.3 複合型コミュニケーションメディア	31
2.4 既存メディアで指摘される問題点	35
2.5 コミュニケーションメディアの評価方法	37
2.5.1 既存の方法	37
2.5.2 本研究における提案	37
2.6 本章のまとめ	39
第3章 コミュニケーションのモデル化	41
3.1 先行研究におけるコミュニケーションモデル	41
3.1.1 Shannon のコミュニケーションモデル	41
3.1.2 Brunswikian Lens Model	42
3.1.3 Brunswikian Lens Model の Scherer による解釈	42

3.2	本研究におけるコミュニケーションモデル	44
3.3	本章のまとめ	46
第4章	音声の表現力の測定方法の検討	47
4.1	関連研究	47
4.2	音声資料の分類	47
4.3	データ収集の方法	50
4.3.1	方針	50
4.3.2	台本のフォーマットの設計	51
4.3.3	台本のインスタンスの生成方法	52
4.3.4	演技者の選定方法	52
4.3.5	発話内容の選定方法	53
4.3.6	演技音声の生成モデル	54
4.4	本章のまとめ	55
第5章	音声収録実験と評価	57
5.1	パイロット版音声の収録と評価	57
5.1.1	台本インスタンスの作成および演技者・発話内容の選定	57
5.1.2	音声収録	58
5.1.3	心理的特徴の分析	59
5.1.4	音響的特徴の分析	62
5.1.5	考察	64
5.2	本実験版音声の収録と評価	65
5.2.1	台本インスタンスの作成および演技者・発話内容の選定	65
5.2.2	音声収録	65
5.2.3	分析用データの選定	66
5.2.4	心理的特徴の分析	67
5.2.5	音響的特徴の分析	72
5.2.6	考察	75
5.3	拡張版音声の収録と評価	77
5.3.1	台本インスタンスの作成および演技者・発話内容の選定	77
5.3.2	音声収録環境と手順	81
5.3.3	音響的特徴の分析	82
5.3.4	考察	83
5.4	議論	85
5.5	本章のまとめ	86
第6章	結論	87
	謝辞	89

参考文献.....	91
研究発表論文	97

図目次

図 1-1 本研究におけるメディアおよびコミュニケーションの分類	4
図 1-2 本研究の流れ	7
図 2-1 コミュニケーションメディアの歴史	16
図 3-1 Shannon の提案したコミュニケーションモデル	43
図 3-2 Scherer による Brunswikian Lens Model の再解釈に対する本研究 における簡易表現	43
図 3-3 本研究で想定するコミュニケーションモデル（メッセージ生成・伝 達モデル）	45
図 4-1 提案方法の概念図	50
図 4-2 本研究の演技音声の生成モデル	53
図 5-1 SEN（パイロット版）と Typical の心理量の主成分得点の比較 ..	61
図 5-2 SEN（パイロット版）と Typical の自然性の印象評価値の比較 ..	62
図 5-3 SEN（パイロット版）と Typical の音響特徴量の主成分得点の比較	63
図 5-4 心理特徴の各主成分の比較（パターン 1）	69
図 5-5 「怒り」の印象評定値の分布の比較	71
図 5-6 「喜び」の印象評定値の分布の比較	71
図 5-7 各特徴量による散布図の比較（パターン 1, 各 160 音声）	73
図 5-8 「怒り」の音響特徴量の分布の比較（パターン 2）	74
図 5-9 「喜び」の音響特徴量の分布の比較（パターン 2）	74
図 5-10 SEN（拡張版）の音響的特徴の第一, 第二主成分の密度プロット (各 200 個)	84

表目次

表 4-1 本研究における音声資料の分類	49
表 4-2 台本のフォーマットとインスタンスの例	51
表 5-1 提案手法を用いて収録した音声一覧	58
表 5-2 Typical における基本感情語一覧	59
表 5-3 Typical の印象評価値に対する主成分分析結果の固有ベクトル	61
表 5-4 SEN と Typical の音響特徴量に対する主成分分析結果の固有ベクト ル	63
表 5-5 本実験版音声の台本の「話し手」の一覧	66
表 5-6 本収録の台本の「発話時の背景」の一覧	66
表 5-7 印象評定値の主成分分析結果（寄与率・固有ベクトル）	69
表 5-8 図 5-4 における各クラスタ内の主成分得点の標準偏差値	70
表 5-9 音声数 160 の時の各特徴量の標準偏差値の比較	73
表 5-10 SEN（拡張版）における発話内容の詳細	79
表 5-11 変更前後の「発話時の背景」の例	80
表 5-12 インスタンスに付与した属性一覧	80

第1章 序論

1.1 背景

情報伝達媒体であるメディアは、人類の発展とともに進化・多様化を続けている。そのなかでも、人間の双方向コミュニケーションを支援するメディア（以下、コミュニケーションメディア）は、20世紀終盤のインターネットの登場、それに続くブロードバンド環境の普及を契機として、急激な発展を遂げている。

電話は、日常で頻繁に使われるコミュニケーションメディアの一つであるが、その登場は1876年であり、今から130年以上前になる。その間、電話の形態は、固定電話、公衆電話、（電話の機能だけを搭載した）携帯電話、スマートフォンと技術革新に伴い発展・多様化しているが、電話による音声通信が我々の日常生活下で果たす役割は普遍的であった。つまり、我々の生活において「遠隔地にいる人間と音声で実時間対話をしたい」という要求は常に顕在し、電話はそれを実現できる唯一のメディアであった。ところが、インターネットの登場は、電話を含む既存のコミュニケーションメディアの在り方に強い影響を与えた。端的に言えば、電話でなければ果たせなかつた役割が、インターネット上の様々なサービスで代替することが可能となった。たとえば、緊急性の高い連絡や比較的の礼儀を重視した連絡においては電話が用いられ、緊急性がさほど高くないものや情報量が多いものは電子メールが用いられる。プライベートにおいては、趣味の話を複数人で楽しむ場合はBBS、自分が知っている情報を発信しそれに対するコメントが欲しい場合はホームページ、Weblog、SNSなどが用いられる。このように、かつて電話が担っていた役割が現在では複数のメディアに分散するようになった。

ここで、各コミュニケーションメディアの役割を決定づける要因はどのようなものであるのかという疑問が生ずる。我々は無意識的に、あるいは社会的枠組みの中での要請にしたがって、利用すべきコミュニケーションメディアを選択している。前述のように、電話と電子メールの使い分けは、緊急性によって判断される場合があるということを経験則で理解できる。そして、同期性の高い対話が可能か否かが電話と電子メールの使い分けの要因であると考えることができる。同様に、多くの言葉の伝達を要する場合に電子メールが好まれるのは、言語情報の揮発性が低いからであろうと考えられる。つまり、コミュニケーションメディアの選択は、物理的性質によって決定づけられている場合が多いという仮説を立てることができる。

一方で、コミュニケーションメディアごとに、人間の心理に及ぼす影響も異なってい

る。たとえば、同一人物との対話でも、電子メールによる対話と音声による対話では、相手の印象が異なる場合がある。この原因として考えられるのは、電子メールの送り手が、作成したメッセージが相手にどのような印象を与えるかをよく理解できていないという点である。電子メールは音声と異なり伝達可能な非言語的情報が乏しいため、言語情報や非言語的情報を丁寧に制御する必要があるが、一般的にそれが上手く作用していないケースが多いと考えられる。同期性や揮発性などの物理的性質と異なり、コミュニケーションメディアの心理的性質は可視化しにくく、明らかになっていない点も多い。

以上のように、コミュニケーションメディアを開発・利用する際には、物理的・心理的性質を適切かつ正確に評価することが重要であると考えられる。それにより、利用者はコミュニケーションメディアの性質を正しく把握し、利用の是非、あるいは利用中の行動をより適切に判断することができると期待できる。しかし現状では、物理的性質については、利用者にとって明確に可視化されているわけではなく、心理的性質に至っては可視化の方法そのものが難しい課題である。

1.2 コミュニケーションメディアの定義

本研究では、メディアに関する用語の意味を次のように定める。

メディア 一般的なメディアの定義に従い、情報を記録・保管・伝達する媒体を指す。ただし、マクルーハンは、メディアは人間の能力の拡張と定義しており[1]、本論文でもその視点を参考にして、幅広い視野をもってメディアか否かを判断する。すなわち、一般的にメディアとして扱われる音声・文字・映像・画像のみならず、5感を通じて何らかの情報を伝えることができる媒体はメディアであるとする。近年ではネットワークを介して触覚や嗅覚を伝達する研究もなされており[2][3]、それらもメディアの一種と捉えることができる。メディアの種類としては様々なものが知られているが、本論文では、メディアの下位分類として記録メディア、単方向コミュニケーションメディア、双方向コミュニケーションメディア、準双方向コミュニケーションメディアを定義する。ただし、メディアの分類は様式だけではなく使い方によって決定するため、あるメディアが单一の分類に属するとは限らない。

記録メディア メディアのうち、情報の記録や保管を目的とするものを指す。コンピュータにおける記憶装置や、IC レコーダーなどがこれに該当する。電子媒体以外では、紙も記録媒体と見做せる。古代の壁画なども記録メディアの一種である。

単方向コミュニケーションメディア メディアのうち、情報発信を目的とするものを指す。新聞やテレビなどのマスメディアがこれに当たるほか、書籍や映画なども情報発信という観点では単方向コミュニケーションメディアである。また、Weblog やホームページなども、コメントなどを求めない限り単方向のコミュニケーションメディアである。

双方向コミュニケーションメディア メディアのうち、人間の双方向（相互）コミュニケーションを支援するものを指す。電話や電子メール、あるいは手紙など、双向あるいは複数人間での情報のやりとりを前提としたメディアが含まれる。古くは、狼煙もコミュニケーションメディアの一種である。

準双方向コミュニケーションメディア メディアのうち、单方向と双方向の中間に位置するものを指す。古くは、ラジオ番組などで、リスナーからの葉書を紹介する行為が該当する。Weblog の場合、コメントなどを求めれば双方向のコミュニケーションが成立しうるので、準双方向コミュニケーションメディアであると見做せる。また、近年次々と登場している SNS (Social Networking Service) もこの分類に属する。また、テレビ

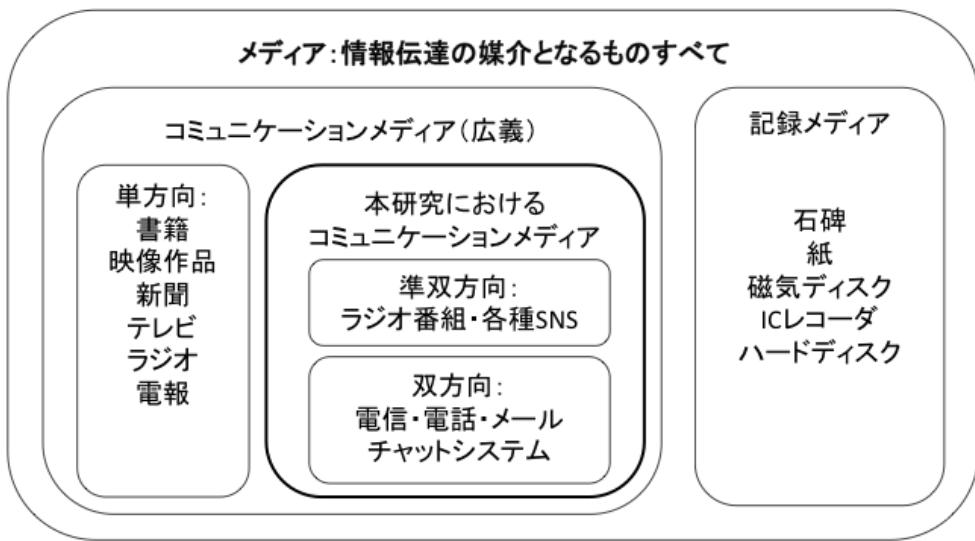


図 1-1 本研究におけるメディアおよびコミュニケーションの分類

でも、デジタル放送が普及するにつれ、視聴者側からもメッセージを送ることが出来るようになるなど、これも準方向コミュニケーションメディアである。また、番組によっては、インターネット上のメディアと連動することで準双方向コミュニケーションを実現している場合がある。以上の分類を整理すると図 1-1 のようになる。

以降、双方向および準双方向コミュニケーションメディアを指して**コミュニケーションメディア**、文脈によっては単に**メディア**と記す。ただし、本研究では、インターネットの通信回線のように、それ自体はメッセージを生成する要素がない媒体に関しては、コミュニケーションメディアには含めない。

また、インターネット上のコミュニケーションメディアを用いたコミュニケーションであることを強調する際には**オンライン（コミュニケーション）**、遠隔コミュニケーションではない場合は強調する際には**オフライン（コミュニケーション）**と記す。この定義では、電話はこのいずれにも属さない。

1.3 コミュニケーションメディア多様化の社会的影響

かつて電話が発明された際、社会に悪影響を及ぼすことが懸念された。たとえば、遠距離での通話が可能であるということから、アリバイ工作などに用いられる可能性が危惧された[4]。しかし、実際には社会問題に発展することはなく、便利な道具として広く社会に受け入れられ、現在まで利用されている。

現在では、インターネットの発展、そしてSNSの登場などによるコミュニケーションメディアの多様化は、我々の社会に様々な影響を及ぼしている。たとえば、コミュニケーションの齟齬やフレーミングと呼ばれる中傷行為が横行している[5]。特に最近では、TwitterなどのSNS上で「炎上」と呼ばれる現象が多発しており、それによって社会的立場を失うなど、深刻な結果となるケースも少なくない[6]。本論文では、このような現象は、コミュニケーションメディアの特性に対する利用者の認識と実態にずれがあることから引き起こされていると仮定する。それが起こる原因のひとつとして、コミュニケーション形式の複雑化が考えられる。かつてから使われているコミュニケーションメディア、たとえば電話や手紙では、その構造は明快であった。電話は高い同期性をもって音声による対話を実現し、手紙の場合は同期性は無いが文字による情報をシンプルに相手に伝えるものであることを、利用者は理解して使用することができた。しかし、現在のインターネットを介したコミュニケーションメディアを利用者が直感的に利用することは、その構造の複雑さを鑑みると、危険である。

また、これらの問題の延長で、オンラインでの問題が連鎖的に生じうる[7]。近年では、このような社会問題に対して、オンラインコミュニケーションが人間の心理に及ぼす影響についての研究が取り組まれている[8][9]。

1.4 本研究の目的

コミュニケーション齟齬について、次のような仮定を考える：電子メールのような比較的シンプルな枠組みのメディアでも、手紙と異なりコミュニケーションに齟齬が生じやすい。手紙も電子メールも、文字情報を伝達するメディアである点は共通しているが、決定的に異なるのは同期性とメッセージ送信の手軽さである。電子メールは、言語情報のみを伝えるものであるのに、受け手は同期性が高い対話、すなわち音声対話や対面対話と同様の非言語的情報の豊かさを無意識に求めている可能性がある。送り手は、言語情報のみしか伝わらないので慎重にメッセージを作成する必要があるのに、送信の手軽さから受け手がどう受け止めるかを考慮していない可能性がある。

この仮定で示したような問題の発生を抑制する手段として、利用者のメディア・リテラシーを向上させることが考えられる。そのために、本研究では、コミュニケーションメディア上でやりとりするメッセージの心理的・物理的な「表現力」を可視化することを試みる。この理由として、表現力はコミュニケーションの品質に直結する重要な要素であるにもかかわらず、これまでにそれを測定し可視化する試みが物理的にも心理的にも行われてきておらず、実際、様々なコミュニケーションメディアがどのような表現力を持っているか不明瞭である点が挙げられる。コミュニケーションメディアの表現力を可視化する手段を提供することで、開発者は生じうる問題点を事前に検討することができ、利用者はメッセージ作成の際に注意深くなるなど、開発者と利用者感は共通の評価軸を持つことが出来る。

本研究では、表現力を可視化する手段として、そのメディア上で出現する多様なメッセージを収集してそれを定量化する方法を提案する。理想的にはそのメディアにおけるあらゆるメッセージを網羅して分析すれば良いが、現実的には限定的なデータからそれを推測しなければならない。そこで、独自のコミュニケーションのモデルを仮定して、モデル上で制約条件を設ける。モデルの検討にあたっては、コミュニケーションメディアの特性や、既存のコミュニケーションモデルを参考にする。そして、多様なコミュニケーションメディアのなかでも、人間にとってもっとも代表的なコミュニケーション手段のひとつである音声を表現力の計測対象とする。具体的には、表現豊かな演技音声の収集する手続きを提案し、それに従って音声の収録実験を行う。そして、統計的な分析を通して、本手法による演技音声の表現力の測定の是非について議論する。

本研究の目的を達するための流れを図 1-2 に示す。

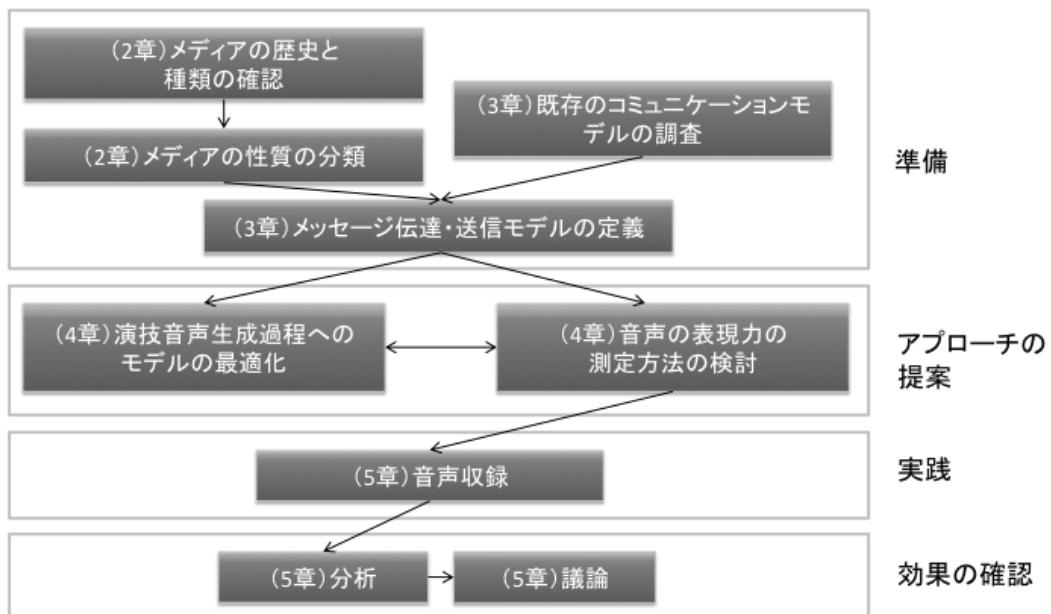


図 1-2 本研究の流れ

1.5 本論文の構成

本論文の構成は以下のとおりである。

第2章では、コミュニケーションメディアの歴史をたどり、それに基づいてコミュニケーションメディアごとの特性を整理する。その後、評価方法に関する先行研究を俯瞰する。本研究では、コミュニケーションメディアの評価方法は、物理的評価と心理的評価に大別されるものとして扱う。物理的評価の方法は既知であるものが多い。たとえば、伝送効率の評価のように、純粋な物理的特性を対象とする場合もあれば、メッセージの揮発性や即時性など、コミュニケーションメディアの物理的な形式が特徴づける高度な特性を対象とする場合もある。いずれも、新しいコミュニケーションメディアが開発される際に、あるいは提供された際にしばしば論じられる要素である。一方、心理的評価は、たとえば遠隔コミュニケーションにおける心理的距離の計測や、非同期コミュニケーションがもたらす利用者の心理への影響、あるいはバーチャル・リアリティにおける感性評価などが先行研究において散見されるものの、実際のコミュニケーションメディアの利用の場面で論じられることは稀である。したがって、物理的・心理的の両側面から先行研究を俯瞰したのち、メディアの評価軸の一環として表現力の評価を実施することの新規性と妥当性について述べる。さらに、この章では、現代における主要なコミュニケーションメディアの特性についても触れる。具体的には、電話などの基本的なコミュニケーションメディアから、ビデオチャットやテキストチャット、電子掲示板などの初步的なオンラインメディア、さらには種々のSNSや動画共有サービスなどの近年の新しいオンラインサービスについて触れる。その過程で、多様化するメディアの中でも音声が基本的かつ重要なものとして位置づけられることを明確にし、本研究で音声を中心に取り扱うことの妥当性を述べる。なお、音声を選んだ理由を簡潔に整理すると次の通りである：音声による対話は(1)単一のモダリティで構造が比較的単純であること、(2)それにもかかわらず離隔なくコミュニケーションが可能であること、(3)代表的な人間のコミュニケーション手段であること、そして(4)コミュニケーションメディアに関する工学的研究の中でも歴史ある分野であること。

第3章では、音声メディアの表現力の測定の前提として、汎用的なコミュニケーションモデルを検討する。表現力を測定するためには、メディア上で生じうるあらゆるメッセージのパターンを網羅する必要があると考えられる。しかし、多くのメディアにおいて、メッセージを符号化した際のパターンは膨大で、その全てを扱うことは難しい。したがって本研究では、特定のコミュニケーションモデルを仮定し、それを制約条件として、限定的な範囲におけるメッセージの分析をおこなう。検討に際しては、まずコミュニケーションモデルに関する関連研究を俯瞰したのち、そのなかで特に重要だと考えられるモデルである、Shannonが提案した情報のコミュニケーションモデルと

Brunswik が提案した視覚モデル（Brunswikian Lens Model）を示す。そして、これらのモデル上での一連の概念、すなわち、送り手によりメッセージが生成されメディア上を伝播し、利き手がメッセージを受け取り解釈するまでの過程において、メッセージの表現に影響すると思われる具体的要因を、第 2 章で示したコミュニケーションメディアの性質を考慮しながら明らかにする。その後、それを参考にして、本研究独自のコミュニケーションモデルを検討する。

第 4 章では、第 3 章で示したコミュニケーションモデルに基づきながら、音声メディアの表現力を測定するための準備として多様なメッセージを収集するための具体的手続きについて検討する。音声の場合、演技音声の収集によってこれを実現することが効率的であると考えられる。この際、演技に関する先行研究や著書を参考に、音声で演技をする際に表現に影響を与えると思われる重要な情報を選定し、コミュニケーションモデルを演技音声に最適化させた。そして、これらの情報を整理して、「台本」として演技者に呈示し、音声表現を得ることにした。また、いかに工夫して音声表現を収集するとしても、演技によって多様な音声表現を偏りなく獲得できると一概に考えるのは危険が伴う。したがって本章では、音声表現の分類を検討したうえで、収集する音声の偏りについても考察する。

第 5 章では、第 4 章で述べた手続きおよび「台本」を用い、実際に実施した音声表現の収録および統計的手法により測定した表現力の詳細について述べる。収録作業は 3 回に分かれており、収録した音声表現はおよそ 9,000 個に達した。また、比較分析のために、台本の代わりに典型的な感情語を呈示してそれを再現してもらった音声も収録した。発話の内容は「ああ、そうですか」「宜しくお願ひします」「いたい」「むずかしい」など全部で 8 通り、発話者は述べ 14 人（男性 4 名、女性 10 名）である。これらに対して、印象評定実験により心理量の定量化、および音響特徴量を用いた物理量の定量化を施して、それらに対する統計的分析を実施し、多様な音声表現を得ることができたのかを確認しながら、本手法による音声の表現力の測定方法の妥当性について議論する。

最後に、第 6 章では、本論文を総括するとともに今後の課題を述べる。

第2章 コミュニケーションメディアの種類と評価

本章では、現在人間のコミュニケーションに用いられるメディアを列挙し、その性質について考察する。古いものは世界的な観点から情報を整理するが、インターネット登場以降はおもに国内の動向を中心に取り扱う。

2.1 コミュニケーションメディアの歴史

コミュニケーションメディアの歴史は非常に古い。本節では、現代までの人のコミュニケーションメディアを振り返り、どのような進化の過程を経てきたのかを整理する。それにより、性質面でどのような要請があり、進化を遂げてきたのかを考察する。

2.1.1 音声言語の誕生以前

イルカや象は、超音波によって驚異的な遠距離のコミュニケーションを可能としていることを有名である。この例が示しているように、すべての生物にとって、音や身振りは根本的なコミュニケーションメディアである。ヒトがかつてこれらをコミュニケーションの道具としていた可能性は大いにあり得るが、ヒトが音声言語を獲得する以前の時代のコミュニケーションに関わる文化や能力についてはあまり明らかではない。

2.1.2 音声言語の誕生

ヒトが音声を使うようになったのは、4～5万年前からという説がある[10]。ヒトが使うコミュニケーションメディアの存在について、物的根拠によって明らかになっているのはこの時代からであろう。この説は、解剖学的に音声を発することが可能と考えられる時期であり、ここから言語情報を文字として記録できるようになるまで数万年を要したと言われる。言語情報の記録の歴史としては諸説あ

るが、メソポタミア文明のシュメール期（紀元前 9000 年頃～紀元前 2100 年頃）において粘土板に刻まれた絵文字の記録が最も古いと考えられる[11]。その後、絵文字が抽象化され、楔型文字[12]やヒエログリフ[13]が使われるようになり、この頃には言語の特徴である文法などの性質が備わっている。楔型文字は粘土板、ヒエログリフはパピルス紙に記録され、これらは記録だけではなく言語の伝達としても用いられていた可能性がある。その後、紀元前 1500 年ごろに系統的な漢字が[14]、紀元前 800 年頃に現在用いられている形式のアルファベットが誕生したと考えられ[15]、現在に至っている。このように、音声の誕生以降、文字とそれを記録する媒体が相互に影響を及ぼしながら発展してきた。

さて、以上のような音声言語誕生直後の時期において、コミュニケーションメディアとしての性質をより強く持つ文化としては、古来より情報伝達のために用いられてきた言伝が挙げられる。その起源は明らかではないが、音声の誕生・進化とともに、言伝は自然と用いられるようになったと考えられる。

2.1.3 遠隔コミュニケーションメディアの登場と発展

文字を記録できるようになったことで、ヒトは言語情報を遠隔地に運ぶ手段を獲得した。言語情報を用いた遠隔コミュニケーションメディアの原型の例として、中国の漢の時代で用いられた木札の書欄（木欄）が挙げられる[16]。メソポタミア文明の時代における粘土板やパピルス紙における文字の記録と異なるのは、送り先が特定されていることである。音声メディアが言伝から技術的に発展するのは、電話の登場まで長らく時間がかかったが、文字メディアの場合では、書欄に様々なバリエーションが生まれたことで、送信形態や信頼性、機密性などの点において早期から多様な進歩が見られた。

より高速な情報伝達を可能とする遠隔コミュニケーション手段として、紀元前 220 年頃、秦の始皇帝の時代には存在していたと伝えられる狼煙（烽火）が挙げられる。これは、戦において敵の攻撃を知らせるために考案されたものである。通信間隔は 4～5km であり、時速 100km 程度で情報を高速に伝達したと言われる[17]。日本には弥生～飛鳥時代に導入され、江戸時代には相場情報の伝達に使われていた（後に旗振り信号にとって代わられた）。近年になり新幹線との速度の比較がなされ、ほぼ同程度の速さで伝達可能であることが分かっている[17]。

そのころ、ヨーロッパでは、1790 年頃のフランスで、電信の語源である腕木式通信システム（発明当時は *telegraph*、現在では *semaphore*）がシャッップにより発明された[18]。これは、稼働する腕木を動かして暗号が通信できるシステムであり、狼煙と比べて送信できる情報量が豊かである。10km 間隔で配置が可能で、760km を 10 分という驚異的な速度で、かつ安定した通信が可能であったという。

夜間や荒天時には使えないという欠点が存在したが、やはり軍事用通信として、フランス革命からナポレオンの時代を中心に活躍した。

1830年ごろになると、有線式の電信が実用化された[19]。様々な形式のものが開発されたが、特に、低コストで長距離の通信を可能としたモールス式電信は、モールス信号とともに、世界規模での電信の発展に大きく寄与した。このころ、通信技術の革新的概念として、メッセージの送信コストが考慮されるようになった。1分間に40語という制約の中で、いかに短く、効率的にメッセージを伝えるかという課題に対して、様々な略号が考案された[20]。さらに、その一種として、メッセージに署名を添えることや、アルファベットの大文字が緊急を表すなど、現在でも用いられている習慣が形成された。さらに、特筆すべき現象として、オンライン・コミュニティが形成された点が挙げられる。電信の技師は特殊な能力を必要としたために、技術者によるコミュニティが形成される原因になった。典型的な現象として、技師の中での女性の割合が多く、恋愛関係なども数多く発生したと伝えられる[20]。また、コミュニティ内で様々な規範や、打ち方の癖による個人の特定など、現在のインターネット環境に通ずる概念が数多く生まれた。

2.1.4 電話の発明

1876年、ベルによって、電話すなわち電気を用いた音声通話（音の伝達）が実現した[21]。コミュニケーションメディアの歴史上、言伝に次ぐはじめての音声による通信方法であり、コミュニケーションメディア全体で考えても革新的な出来事であった。しかし、電話の登場の最初期では、必ずしも電話による通信が有益であるとは断言できない出来事が多発した。企業での利用時において、電話による通信の多くがゴシップ程度の会話に用いられていたり、礼儀作法の悪さが目立ったりするなど、問題が数多く指摘された。1920年頃になって礼儀作法が習慣化した電話はようやく社会性を得るに至り、企業以外での利用が通信会社によって推進されるようになった。このとき、宣伝文句として、電話は手紙よりも「親しく」会話できる点が強調されていた[22]。一方、1964年にマクルーハンは、電話の宣伝文句で頻繁に用いられる「phoney（音声の）」という言葉が、身体性の欠如を意味していることを指摘している[23]。すなわち、電話の登場の時期、コミュニケーションメディアの心理的性質・物理的性質がコミュニケーションの質に与える影響についての観点が生まれたという見方ができる。

一方、1900年頃に発明されたラジオは、1920年頃から1940年頃にかけて大きく普及し[24]、1900年ごろから1940年ごろにかけて、多くの技術者によってテレビ受像機の技術開発がなされた[25]。テレビの普及には1960年代まで時間がかかったが、電話・ラジオ・テレビによって、20世紀後半は電子メディア時代の

幕開けとなった。

電話は、1990年台の中盤から大きな革新が続き、ポケットベルやPHS、携帯電話、スマートフォンなど、現在でも多機能化しながら発展を続けている。

2.1.5 インターネット黎明期

インターネットの起源は、1967年のARPANET計画であり[26]、世界で初のパケット通信ネットワークである。当初は政府および学術機関を中心に用いられていたARPANETの技術は、1982年の共通プロトコルであるTCP/IPの標準化、1989年の商標ISP(Internet Service Provider)誕生などを通して、広く民間に普及していった。ISPでは、最初期からダイヤルアップ方式に対応し、専用線が無い環境でも、電話回線の利用によってインターネットへの接続が可能となった。この80年台末から90年台の半ばまでをインターネットの（民間利用の）黎明期と定義するが、この黎明期の日本国内においては、電子メールサービスやニュース配信サービスが、パソコン通信と呼ばれるプロバイダが提供するネットワーク上で、利用者同士のコミュニケーションに利用されていた。この頃は、生活水準に対して高価であったパーソナルコンピュータの所有者や、その中でも前述のパソコン通信などに興味を持った一部の利用者で、趣味などの情報交換に利用されることが殆どであった。また、ここでも、電信や電話の登場時と同様に、様々な規範が生まれた。

1995年のNTTによるISDNの正式サービスの開始を契機に利用者は増大し、情報発信を目的としたホームページの開設、電子掲示板の利用が大きく活性化した。また、チャット(テキストチャット)と呼ばれる文字データによる同期生の高い通信も頻繁に行われるようになった。

また、1990年代後期には、携帯電話の通信網がインターネットと相互接続され、携帯電話から電子メールの送受信が可能となるなど、従来インターネットを利用しない大多数の人間もその利便性に触れる機会が増加していった。

2.1.6 ブロードバンド回線の普及

2000年代に入ると、日本政府はe-Japan戦略を策定した[27]。その政策下において、ADSL、CATV、FTTHなどのブロードバンド回線が低価格化し、インターネット環境が爆発的に普及した。それと同時に、大規模のデータを用いたサービスがインターネット上で利用可能となった。たとえば、インターネットでのデータ通信を利用した音声通話(音声チャット)や、動画による通信対話(ビデオチャット)が実用化されるようになった。音声チャットは、電話の代替手段とし

て用いられるようになったが、単なる代替ではなく、テキストチャットやファイル送信を組み合わせてマルチメディア的な利用が可能であること、複数人の同時通話を気軽に可能したことなどから、電話とは一線を画す位置づけとなった。また、ビデオチャットは、当初はビジネスの場で、遠隔での会議での利用を想定されることが多かったが、友人同士・家族同士で簡単に利用できるようになつた。その一方で、直接対面しての場合とは異なる感覚での対話となることが分かり、現在でもその違和感について研究対象となることが多い[28] [29]。また、従来のホームページとは異なり、気軽に個人のWebサイトを構築し情報発信できるWeblogや、Youtube[30]をはじめとした動画・コメント投稿・閲覧サービスも世界的な流行となつた。その他、オンライン上で楽しむゲームも次々と登場し、ゲームでの記録をWeblogやYoutubeで発信するなどの文化も生まれ、これらのサービスが相互に関係しながらその利用は拡大していった。

また、ブロードバンド回線は、通信量の増大だけではなく、副次的にインターネットへの常時接続環境をもたらした。利用者はインターネットに「接続する」という操作を無意識的に行うか、あるいは操作そのものが不要になり、よりオンラインでの活動を活性化させる契機となつた。

2.1.7 近年から現在まで

日本ではe-Japan戦略に続きICT政策大綱・u-Japan戦略[31]を打ち出し、情報通信技術分野での政策をさらに推し進めた。その結果、2000年代中盤～後半にかけて、モバイル環境も含めたインターネットの普及率が激増した。総務省による調査では2007年におけるモバイルインターネットの普及率は65%であったのに対して、2010年には90%に達したことが報告されている[32]。このように、インターネットがごく日常的に利用されるようになるにつれ、多くの利用者はごく自然にオンラインでの情報発信を行い、他の人間との交流を楽しむようになった。

このころになると、交流の場として、Weblog、電子掲示板、チャットアプリケーション等の既存のサービスと並行でSNS（Social Networking Service）が広く利用されるようになった。SNSの起源は1990年代中盤に遡る。初期のSNSは、テキストチャット機能や友人リスト機能、プロフィール機能、友人検索機能などを備えることによって、オンラインでの社会的繋がりを支援することを可能とした。2000年中盤にはこれらの基本的要素に加え、Weblogに類似した日記機能が搭載されるなどSNSのサービス内容は多様化しており、爆発的流行が生ずる要因となつた。

2010年頃より、3Gデータ通信など、新たな通信規格の登場によって無線通信

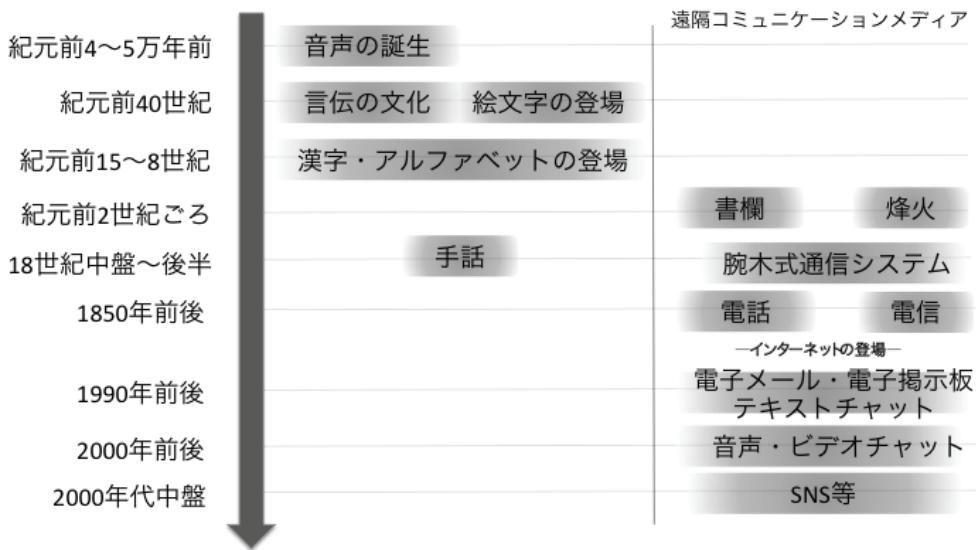


図 2-1 コミュニケーションメディアの歴史

網が高速化し、携帯電話によるインターネットアクセスの利便性がより高まった。加えて、携帯電話は多機能化してパソコンと同様のサービスが享受できるようになり、SNS が携帯電話の標準的機能と見做せるようになった。これらの環境の進化によって、利用者が望むタイミングで、場所を問わず、タイムラグなく様々な形式の情報を発信できるようになり、コミュニケーションメディア多様化の時代が本格的に到来したと言えよう。

本節で示してきたコミュニケーションメディアの歴史についての概観は図 2-1 のようになる。

2.2 コミュニケーションメディアの性質の分類

本節では、2.1 節で述べたコミュニケーションメディアの進化、特に近年使用されているメディアのうち音声関係のメディアと文字関係のメディアを中心に適宜参照しながら、それらが有する物理的な様式に起因する性質について、本研究における定義を列挙する。そして、これらの性質によって、コミュニケーションメディア利用時に人間が受ける心理的影響も変化すると仮定する。本節以降、コミュニケーションメディアによって伝達される情報を**メッセージ**と統一して呼称する。また、送り手が発した情報ではないが、メディアの仕様によって副次的に発せられる情報を**補助的メッセージ**と称する。メッセージのやりとりを**対話**と称する。

以下にそれらの詳細について列挙する。なお、この分類においては、たとえば実時間性や同期性の定義において、既存の概念は考慮せず本研究独自の観点を持って行う。

① 実時間性

メッセージの発信と受信が、メディアの仕様上同時に行われることが強制される場合、実時間性を有すると定義し、この性質を有する対話を**実時間対話**と称する。音声による対話は、電話回線などのネットワークを介した場合は回線の種類や距離によってタイムラグが生じるが、実質的には実時間対話と見做せる。一方でテキストによる対話は、メディアの仕様によってはかなり円滑に対話することが可能ではあるが、遅延が生じた場合でも多くの場合メッセージが相手に伝わるため、実時間対話とは見做さない。

たとえば、実時間性を有するコミュニケーションメディアは、電話や音声チャット、ビデオチャットなどの音声を用いたメディアの他に、手話による対話、身振りや表情を用いた非言語的なメッセージのやりとりなどが該当する。

逆に、音声メディアであっても、録音した音声を配信してコメントを受け取るといった形式のコミュニケーションの場合は、実時間性は備えない。

② 同期性

メディアによって実現可能なメッセージの送受信の同期の程度を示す指標と定義する。もっとも同期性が高い状態は、**話者交代が円滑に行われ、かつ単一の話題が扱われる**場合であるとする。実時間性が強制的な概念を指していることに対して、こちらは可能か否かのみを指す。実時間対話のメッセージの送受信の同期性が最も高くなるが、テキストチャットの場合でも、話者が試みさえすれば、同

期性の高い対話が可能である。特に、対話相手がキーボード入力をしているか否か・メッセージを受信したか否かなどの補助的メッセージは、**アウェアネス**として機能し、話者が同期的な対話をを行っているかどうかの感覚に左右すると考えられる。キーボードを打鍵する際の振動を相手に伝えることでアウェアネスを向上させ、話者交代を円滑化させる試みも存在する[33]。

本研究では、原始的な電子掲示板とテキストチャットの最大の違いは同期性であると見做す（原始的なテキストチャットは、電子掲示板のようにWebサイト上に設置され、パスワードなどのアクセス制限の設定を行わなければ、全世界にその内容は公開されていて、電子掲示板ときわめて類似していた）。電子掲示板は、メッセージの受け手が能動的にメッセージを取得するための動作を必要とするような枠組みであることに対し、テキストチャットは受動的にメッセージを受信することができ、特別な動作は必要としない。電子掲示板もテキストチャットも、オンラインでのテキストコミュニケーションであるにもかかわらず、その役割・位置付けによって意図的に同期性の仕様が異なるように設計されていると考えられる。テキストコミュニケーションを実現するメディアとして近年ではTwitter[34]がチャットのような役割を果たしているが、同期性の観点では、電子掲示板とテキストチャットの中間に位置する。

また、特に近年特徴的な変化を遂げているのが電子メールで、かつては同期性の観点では電子掲示板とテキストチャットの中間点に位置していたものが、スマートフォンの登場などによる電子メールのインターフェースの変化により、テキストチャットのようなイメージで対話が可能となっている。ところが、話者の片方がスマートフォン、もう片方がいわゆるガラパゴス型の携帯電話を使用している場合、スマートフォン利用者からはテキストチャットと認識される一方、携帯電話からは旧来型の電子メールと認識される。この認識のずれにより、コミュニケーションに齟齬が生じる可能性が考えられる。この現象は、電子メールというメディアでも、端末の仕様によって同期性の性質が大きく変わることを示している。

③ 即時性

メッセージの送信開始からその完了まで、あるいは受信開始からその完了までのタイムラグの有無に関する指標であると定義する。音声は、その性質上、メッセージの送信あるいは受信を完了するには発話の完了を待たねばならず、発話の長さのぶんだけタイムラグが生ずる。それに対し、テキストチャットは、作成したテキストを送信すれば、メディアの仕様の範囲内であれば、どんなに長いメッセージも瞬時に送ることができる。そして、メッセージの受け手は、自由にメッセージの読み方を選ぶことができる。大量のテキストの情報を受け取っても、い

わゆる斜め読みによって要点をすぐさま探し出し、効率よく情報を処理することも可能となる。これは電子媒体のみならず、手紙なども同様である。

テキストチャットは、即時性は高いが実時間性を備えておらず、音声対話に比べて情報伝達量の時系列の分布がピークであると言える。一方で音声対話は、即時性は高くはないが実時間性を有し、テキストチャットと比較して、「途切れない」対話が可能である。

④ 挿発性

メッセージが生成され、消滅するまでの時間によって挿発性を定義する。音声対話は、録音・録画装置などを使わない限り挿発性が極めて高く、即座にメッセージは消滅する。電信や腕木式通信システムなども挿発性は音声と同等である。狼煙は、挿発性はあるが、メッセージは即座には消滅しない。テキストによるコミュニケーションメディアは挿発性が全般的に低いが、その中でも差異がある。テキストに限らないが、電子媒体の場合はコピーが容易く、万が一インターネットに流出してしまえば、その削除はかなり難しいことになる。しかし、コピーさえ行わなければ、メディアによっては比較的短時間に履歴が消滅する場合もある。能動的な操作を行えば、完全に削除することも一瞬で可能である。簡易的なテキストメッセージ交換のアプリケーションであれば、履歴自体が残らない場合もある。これに対し、手紙などの紙媒体は、複製も削除も困難ではないが、電子媒体ほど手軽ではない。これらのことより、テキストメディアは、挿発性の観点では非常に複雑な構造をしていると考えることができる。

⑤ 再利用性

メッセージを再利用する際のコストを示す指標と定義する。挿発性とも関連するが、テキストメディア、とくに電子媒体はコピーが容易で、再利用も頻繁に行なわれる。音声・画像メディアの場合、何らかの録音機器が必要であるが、データが大量になると検索が困難になり、場合に応じてアノテーションが必要となるなど、再利用の際にはコストがかかる。

⑥ 編集性

一度発信したメッセージに対して、再編集というかたちで修正を加える事ができるか否かを示す指標と定義する。音声対話や、電信、狼煙など、挿発性の高いものは発信した情報そのものを修正することは不可能であり、訂正の情報を発する必要がある。テキストメディアにおいても、電子メールやチャットアプリケーションでは、一度発した発言を修正することは仕様上不可能となっている。一方、多くのSNSでは、同期性や挿発性が低く、メッセージの再編集が可能である。

⑦ 信頼性

信頼性とは、「メッセージが物理的に確実に相手に届くか」および「メッセージが変化せず相手に届き、保有されるか」を示す指標と定義し、前者を**到達性**、後者を**不变性**と呼称する。

音声による実時間対話では、到達確認を常に行うことになるので到達性は高いが、揮発性の高さから信頼性の保証が難しい。企業のコールセンターでは、対策として録音がなされている。

古来より、不变性の高いメディアとして利用されてきたのはテキストであり、公的権力の介在の有無にかかわらず、契約のために**証文**が用いられてきた。電子媒体におけるテキストの利用では、証文のような拘束力はないが、コミュニケーション齟齬が発生した際の事実確認のために有益である。ただし、編集性が高いテキストメディアについてはこれが当てはまるとは限らない。さらに、電子媒体の場合、たとえば電子メールによるメッセージ送信は、届いてしまえば不变性は高いが、到達確認の難しさや宛先の指定ミス、サーバやネットワークの障害によるメール詰まり・不達の可能性があるなど、仕様上到達性は高いとはいえない。しかし、障害があれば不達が生ずる可能性があるのは手紙などの郵便物でも同様であるから、結局メディアの運用次第で到達性は変動するものであると言える。**伝書鳩**で例えれば、到達性が低いかわりに、同様のメッセージを複数送信することによって信頼性を確保し、軍事的利用に支障が無いように工夫がなされていた。

⑧ 到達確認性

送信したメッセージが、返信メッセージを受け取る以外の方法で、相手に届いているかが確認できる程度を示す指標と定義する。実時間対話でも、対面対話ならば到達確認性があるが、電話などの音声による対話やビデオチャットなどでは、相手からの返信がないかぎりそれを確認することができない。テキストチャットの場合、信頼性が高い環境であれば、即時性も高いことからある程度到達の是非を推測することはできるが、これも確実ではない。

電子メールでは、基本的な仕様（SMTP プロトコル）上では到達確認性は保証されていないので、クライアントアプリケーション上で開封確認の要求および送信が機能として付与されている場合があるが、利用するコミュニティで規則を定めない限り、そのような機能は殆ど使用されない。これは、実時間対話でなければ、到達確認性が低いことが必ずしもコミュニケーション上のデメリットになるとは限らないからである。

最近の SNS では、投稿した記事やコメントを閲覧した利用者が表示される、あるいは、閲覧したことの確認手段が用意されている場合がある。

⑨ 拡散性

同一のメッセージをどれだけの人間が受け取ることができるかの指標であると定義する。一対一の対話がもっとも拡散性が低いが、たとえば電話では、周囲に人間がいれば片方の話者の情報は拡散される。テキストチャットは基本的には閉鎖的な対話空間が用意されるが、再利用性の高さから、意図せずに拡散する可能性がある。対面対話でも電話と同様の拡散性を持つが、拡散する情報は両話者の会話そのものであるという点で異なる。

仕様として、拡散を前提としているメディアも多い。SNSでは、コミュニティを指定して情報を拡散することができる。Twitterでは、設定によっては、メッセージは任意の利用者に拡散する。拡散性の誤解により、コミュニケーションに問題が生ずる場合もある。SNSほどでは無いが、グループ音声対話やビデオチャットシステムでは、グループないで情報が拡散される。

狼煙や腕木式通信システムなどは特殊な例で、メッセージの送受信が一対一の場合でも、メッセージそのものは第三者から閲覧可能であり、強制的に不特定多数に拡散する。

以上より、拡散性は、**意図しない拡散性**と、メディアの仕様上定められた**意図的な拡散性**に分類することができる。

⑩ 同時性

発信されたメッセージを複数人が受信する際、それがどの程度同時に行われるかを示す指標であると定義する。音声対話や狼煙など、実時間性や揮発性が高いメディアでは、同時性は高くなる。テキストメディアでは全般的に同時性は低くなり、SNSや複数人に送信する電子メールでは、同時性は全く保証されない。テキストによるグループチャットで同期性が高い対話が実現されている場合は、同時性は比較的高くなる。紙媒体のテキストメディアの場合、同時性は殆どの場合低くなる。複数人での円滑な対話を実現するためには、同時性の確保が重要であると考えられる。

なお、アナログの掲示板は、同時性の高いテキストメディアであるが、単方向のコミュニケーションメディアであり、本研究の定義におけるコミュニケーションメディアに当てはまらず、概念としてはマスメディアに近い。

⑪ 指向性

メディアの仕様上で、メッセージの宛先が明確に指定されるか否かを示す指標と定義する。電子メールや手紙、電話やテキストチャットなどは、宛先を指定することが対話を開始・進行する際に必要となるため、指向性対話である。

完全な指向性対話以外では、指向性のメッセージとそうでないメッセージが入り乱れることが多い。SNS では、最初のメッセージは不特定多数に発信され、それに対するコメントという形式で返信メッセージが付与される。匿名の電子掲示板である 2ちゃんねる[35]においては、文脈においては一連のつながりがあるものの、宛先が指定されないメッセージの投稿が続く現象がしばしば起こる。一方で、同様の匿名電子掲示板である米国の 4chan[36]は、2ちゃんねると比較して宛先が明確であるメッセージの投稿が非常に多いなど、同じ分類のメディアでも、文化や仕様によって傾向に搖れが存在することが分かる。動画配信サービスのニコニコ動画[37]では、投稿された動画に対してコメントを付与し、利用者はコメントを動的に閲覧しながら動画を楽しむことができるが、この場合でも、動画の投稿者に対するコメントが付与されることもあれば、宛先が不特定多数の視聴者に向かられていると解釈できるコメントも多く発生する。

以上のように、指向性については、近年のコミュニケーションメディアの多様化に伴い、複雑化が進んでいると考えられる。

⑫ 集団性

社会科学で用いられる意味とは異なり、複数人でのコミュニケーションを前提としているか否か、複数人の場合はその規模や形態を示す指標であると定義する。拡散性が高く揮発性が低いテキストメディアでは、集団性を高めることは比較的容易であると考えられるが、音声対話、とくに電話や音声チャットなどの音声のみを用いた対話では、実時間性を持つという性質上、多くの利用者が同時に対話することは困難である。

集団性の境界線が明確であるのは対話の参加者が 3 人以上か否かであり、それ以上になると、どこまでが「双方向コミュニケーション」が成り立っていてグループ対話と見做せるのかの判定は難しくなる。これはメディアの様式によって様々な解釈が可能である。

大規模なコミュニケーションメディアとしては Twitter や Facebook[38]、あるいは前述の匿名掲示板である 2ちゃんねる等は、国内の利用者数はいずれも 1000 万人を超えると推定あるいは発表されている[39][40][41]。しかし、実際にコミュニケーションが行われるのはトピック単位であり、たとえばある投稿に 100 のコメントが付与され、それらのコメント間でコミュニケーションが成立していかなければ、ごく短い対話が 100 回行われただけに過ぎない。これに対し、トピックの中心者が居ない場合は、单一のトピック内で同時多発的にコミュニケーションが行われる可能性がある。このような対話は、匿名掲示板や SNS の機能として実装されたコミュニティの枠組みの中で生じうる。

⑬ 機密性

本研究において、機密性は、メッセージそのものの機密性と、通信経路上の機密性が存在すると考え、特に前者についてを取り扱う。後者は、インターネットの場合では、通信プロトコルが暗号化されている場合を指す。

メッセージそのものの機密性は、メッセージが暗号的意味を有している場合と、メッセージへのアクセスに制約が存在する場合がある。たとえば電信は、解読方法を知らない人間にとっては、メッセージそのものが暗号である。同様に、野球におけるサインなども暗号化されたメッセージである。音声やプレーンテキストは、メッセージそのものは平文であり機密性を有していない。

それに対し、たとえば電子メールは、メッセージは平文であるがその閲覧にはパスワードの入力を要する。SNS も、設定によって指定された利用者以外には情報を閲覧することができないようになることができる。これらはアクセス制限によって機密性を確保している例である。

大規模な電子掲示板においては、制限が存在せず誰でも閲覧することができるようになっている場合が多い。あるいは、そのサービスの会員となることを条件にほぼ制約なくメッセージの閲覧や送信が可能となる場合もある。

⑭ 社会性

本研究における社会性とは、コミュニケーションメディア上で、利用者の社会的情報を明示・操作する枠組みや、社会的規範があることと定義する。

社会的情報の明示としては利用者のプロフィール、社会的情報の操作としてはフレンドリストを例として上げることができ、定義上それらを備えている SNS の社会性は極めて高い。電子メールの場合でも、SNS と比べて社会性の程度は低下するが、メールのアプリケーション上で署名を付すことによるプロフィールの表示や、電話帳の作成が可能である。匿名の電子掲示板の場合では社会性はかなり低くなる。さらに、携帯電話の場合、電話番号そのものがプロフィールとして機能している。

また、社会的規範については、コミュニケーションメディアの物理的性質によって規定される枠組みとは別に、メディアごとに独特のエチケットやマナーが形成されうる。それは、コミュニケーションメディアを提供する側が誘導して形成される場合もあれば、利用者が利用の過程で自主的に形成する場合もある。2.1.4 節で述べたように、電話の場合、それが使われるようになった直後は、その利用に不慣れな人々が様々なコミュニケーション上のトラブルを抱えたことが知られている[4]。このときは、電話の通信会社が先導してエチケットやマナーを普及させ、トラブルが減少し円滑に利用できるようになったことで、電話が広く普及する大きな動機となった。電話におけるマナーとは、現在我々も社会的マナーとし

てよく知っている、「もしもし」という言葉や（これは場合によっては失礼にあたるが），自分の名前を名乗ること，そしてTPOに応じて定型的な挨拶を交わすことなどが該当する。

社会性が比較的低い匿名掲示板であっても，独特的のマナーが存在し，コミュニケーションメディアの円滑な利用のためにはそれらを意識することがきわめて重要になる。

⑯ 受動性

対話中，メッセージを受動的に受信することが強制されているか否かを指す指標であると定義する。音声対話などの実時間対話では必然的に受動性は高くなる。電子メールの場合でも，携帯電話のメールは自動受信設定となっているので受動性は高い。パソコンの電子メールの場合は，受信した場合でもアプリケーションを立ちあげなければメッセージを閲覧するにはいたらず，受動性は低い。

⑰ 個人性

メッセージあるいはメディアに，特定の個人によるメッセージであると第三者から判別できる情報が含まれる度合いと定義する。音声対話においては，声質や口調が個�性となる。テキストコミュニケーションの場合でも，文章の作り方の癖や，絵文字や顔文字の使い方，句読点や改行の挿入の傾向などから個�性が表出する。さらに，電報のように非言語的情報が少ないメッセージ上でも個�性は現れる。2.1.3節で示したように，電報の打ち方の癖で個人がある程度特定可能であったことが知られている。

⑱ 匿名性

ここでいう匿名とは，利用者間で匿名性が確保されていることと定義する。したがって，匿名掲示板の場合でも，通信記録をたどれば，大抵の場合には利用者の個人情報が判明するが，それについてはここでは考慮しない。

匿名性そのものの定義は，様々なものがあるが，Pfitzmannらによって情報工学の観点から厳密に規定されたものがあり[43]，その強弱によってレベル分けがなされている。以下，強いものから列挙する。

Unobservability :

メッセージの送受信者やコミュニケーションの存在が観察できないこと

Unlinkability :

複数のメッセージ間で同一人物であることが特定できないこと

Pseudonymity :

メッセージの送受信者が識別子などの仮名を与えられた状態であること

これらの定義に従えば、本研究においては、利用者間の匿名性を扱うことから、**Unobservability** に該当するものは存在しないので考慮しない。

「2ちゃんねる」など、国内の匿名掲示板においては識別子が付与されている場合が多く、**Pseudonymity** が担保されている。Twitter などの多くの SNS では仮名を名乗ることが許可されており、これも **Pseudonymity** を満たしているように思われるが、社会性が強いため、実質的には匿名には該当しないと見做せる。これに対し、**Unlinkability** を特徴としたコミュニケーションメディアも存在する。国内では、匿名の日記サービスである「はてな匿名ダイアリー[43]」が有名であるほか、米国発祥のニュース雑談サービスである「スラッシュドット[44]」が該当する（スラッシュドットは匿名か否かを任意に選択できる）。米国の匿名掲示板である 4chan でも、**Unlinkability** を満たす仕様を採用している。

なお、匿名性は個人性と相反する概念ではない。個人性が表出したテキストであっても、メッセージに識別子のみが付与されていなければ同一人物であることを特定することができず、**Unlinkability** は担保されるからである。

⑯ 自己観察性

コミュニケーションの過程で自らのメッセージを客観的に観察できることと定義する。実時間対話ではこの性質を満たさない。テキストメディアの多くはこの性質を満たしていて、自らのメッセージを含めた対話の内容を観察することが可能である。

⑰ モバイル性

場所を問わずコミュニケーションが可能かどうかを示す指標であると定義する。電話においては、固定電話の場合モバイル性は無いが、携帯電話はモバイル性を有する。インターネット上のメディアにおいては、かつては多くのサービスがモバイル性を有していなかったが、近年では、モバイルでのインターネット環境が充実し、場所を問わず利用できるようになっている。

⑲ ハンズフリー性

そのメディアを利用している際に、物理的に他の操作が可能かどうかを示す指標であると定義する。音声対話ではハンズフリー性は比較的高く、特にヘッドセットを用いた場合は高くなる。ボイスチャットでは、ヘッドセットやパソコン搭載のマイクやスピーカーを利用する事が前提となる場合が殆どのため、ハンズフリー性はきわめて高い。逆に、テキストメディアは、メッセージを作成してい

る瞬間は、ハンズフリー性は低い（携帯電話のメールのように、片手のみで操作する場合もある）。

以上の性質は独立した概念ではなく、関係性が深いものも存在するが、メディアの違いを特徴づけるものを列挙している。また、これらの性質は並列の関係ではなく、階層の関係である場合も考えられるが、ここでは考慮しない。

2.3 近年使われるコミュニケーションメディア

本節では、近年広く利用されているコミュニケーションメディアのうち、オンラインコミュニケーションにかかるものを中心に列挙し、2.2節で述べた各性質の観点から、コミュニケーションにおける特徴を考察する。ただし、近年におけるメディアの関係性は階層的であることに留意する必要がある。

2.3.1 音声中心のコミュニケーションメディア

携帯電話 音声メディアで現在最も広く利用されているのは、携帯電話であろう。近年では高機能化した携帯電話を指してスマートフォンと呼ぶ場合があるが、それについては2.3.3節において触れる。国内の携帯電話の普及率は、平成24年度末の時点での人口の98%に達している[45]。旧来型の携帯電話の特徴は、固定電話の特徴に加えて高いモバイル性を有することであり、音声通信だけではなく電子メールによるテキストの通信、Webブラウジング機能を経由したSNSの利用などにモバイル性を与えており、利用者にとって料金面・操作面でのコストが高かったため普及しなかったが、特定のキャリア・特定の端末でグループ通話も実現していた（集団性）。また、社会性の観点では、携帯電話は、通常特定の個人に帰属する性質を有するため、その個人であることを示す材料と成り得る。ただし、契約や解約を行う機会が頻繁に通信キャリアにより設けられているため、社会性も変化が生じやすい。

固定電話 固定電話も、携帯電話と並んで最も利用されている音声メディアであるが、近年は、家庭用・事務用・公衆電話すべてにおいて、その利用者数・契約者数は激減している[47]。固定電話は、特定の個人に属さない場合もあり、本研究で定義する社会性は携帯電話より低い。ただし、契約や解約は住居情報と強く結びつきがあるため、一般的には信頼性の高い情報として取り扱われる。また、端末にFAX機能があれば、文字通信も可能である。一方、固定電話単体でのグループ通話のサービスは存在しない。

インターネット電話（IP電話） インターネット回線上で音声対話を実現するメディアである。端末ごとに電話番号を持つため、固定電話と類似した社会性を備える。インターネットの回線を流用するという仕様上、固定電話にくらべ比較的低くなる通話品質（音質）の違いを除けば、固定電話と同様の使い勝手があり、固定電話や携帯電話との相互通信も可能である。さらに、回線が高速化するにつ

れ、グループ通話のサービスも提供されている。なお、本研究では考慮しないが、通信経路上での機密性が高く、固定電話や携帯電話で行われるような方法で盗聴されにくいという利点がある。

ボイスチャット ブロードバンドの普及によって音声のような大量データ通信が可能となり、2000年代に入り急激に普及した。サービスの種類や設定によっては電話番号への発信が可能だが、基本的には特定のサービスに登録した利用者同士での通話に限定されており、コミュニティを形成しやすい。電話番号の代わりに利用者IDを指定して通話を開始するため、利用者リストはおのずと作成される。また、アプリケーション上で各種プロフィールを設定することができる場合が多く、社会性は高い。ボイスチャットの機能のみを備えるアプリケーション・サービスは殆どなく、多くの場合でテキストチャットとの併用、ビデオチャットへの切り替えなどが可能である。端末としてパソコンを利用することからハンズフリー性が高く、他のメディアを操作しながら利用することも多い。その例として、オンラインゲームでのボイスチャットの利用が挙げられる。

2.3.2 文字中心のコミュニケーションメディア

電子メール 電子メールは、インターネットが普及する以前の黎明期から使われてきた。その本質は宛先に対して題目と本文を送信することである、近年では、Web上でメールを操作できる環境（Webメール）や、携帯電話との機能統合が進み、利用形態は変質しつつある。初期のインターネットにおける電子メールは、同期性は低く、受動性も低かったが、携帯電話への電子メール機能の搭載によってモバイル性が向上し、それにともなって同期性・受動性ともに大きく向上している。また、メーリングリストなどの機能によって、集団性を確保することができる。そのような使い勝手の良さから、電子メールは長い間コミュニティ内の情報共有手段として用いられてきたが、編集性がないことや、到達確認性が基本的には存在しないなどの不便さも存在するため、近年ではコミュニティ内の情報共有にSNSが利用されるケースが増加傾向にある。

テキストチャット テキストチャットも電子メールと同時期から利用されている。電子メールとの違いは同期性の高さが確保された設計が施されている点である。きわめて初期のテキストチャットは、単純なテキストメッセージの交換の機能を備えるのみであった（UNIXのターミナルにおけるtalkコマンドがそれに近い）が、比較的初期から仕様が大きく分かれている。すなわち、IRC（Internet Relay Chat）[48]に代表されるような、ルーム入室型（参加者全員で会話する型）のも

のと、ICQ[49]を起源とする、フレンドリストを生成したのちにそこから対話の相手を選択するインスタントメッセンジャー（IM）型のものである。ルーム入室型のテキストチャットは、1990年代から2000年代初頭にかけてWebサイト上にCGIを用いて設置されることが多かったが、次第にIM型のアプリケーションの利用が増加していった。IMは、フレンドリストやプロフィール表示など、社会性が高く、匿名性もPseudonymityを満たす程度に抑えられているため、ルーム入室型に比べてトラブルが生じにくい。また、機密性も相対的に高い。さらに現在では、ボイスチャットアプリケーションや、SNSの多くにIM機能が搭載されていることから、ルーム入室型のテキストチャットの利用者は減少していると考えられる。多くのIMにおいて、対話相手の入力状態の表示機能が搭載されており、同期性の高い円滑な対話を実現する工夫がなされている。最近ではメッセージの開封確認機能なども搭載されたものが増えてきている。一方で電子メール同様に、殆どのもので編集性は存在しない。

電子掲示板 電子掲示板も電子メールやテキストチャット同様古い歴史を持つメディアであり、多様な規模・種類のものがある。インターネット普及期には、個人が自らのホームページにCGIを用いて設置することが流行し、個人間での交流のために役立っていた。この場合、個人サイトに設置されるという背景から、ある程度の社会性を備えていると見做せる。一方で、パスワードを設定しない場合など機密性が低い際には、匿名の利用者によるメッセージや、スパムメッセージが書き込まれる場合もある。掲示板は、雑談の場として利用されるほか、何らかの主題が存在する場合も多い。組織によって運営される大規模の掲示板では、多様な話題を同時に扱うための工夫が必要となり、スレッドフロートと呼ばれる、同一の話題がツリー形式でまとまって表示される（スレッド）方法がよく利用される。非常に大規模な掲示板では、多数の掲示板がグループごとに整理され、それぞれの掲示板でスレッドフロート型が使われているなど、階層構造がある。その程度の規模掲示板になると匿名性はかなり高く、UnlinkabilityないしはPseudonymityの状態であることが多い。一方で、大学や企業などWebシステム上の機能としても掲示板は活用され、その場合、匿名性は存在せず社会性が高い状態で運用される。なお、匿名性の高い掲示板では指向性が非常に複雑である。たとえば、単一のスレッド内では、特定のトピックにかかわる不特定多数による対話が進行していくが、メッセージの送信先が明示してある場合とない場合が入り乱れたり、突如同期性の高い対話が発生したり、さらにそれらを多くの第三者が楽しんで閲覧していたりするなど、メディアそのものは単純な構造であるにもかかわらず、複雑な対話様式を観察することができる。

Weblog 国内ではブログ（blog）と呼ばれることが多い。個人や組織が様々な記録を公開するためのサービスであり、基本的機能として情報を Web ブラウザ上から簡単に更新できることから、国内では 2000 年代に入り大きく普及した。配布・販売されているツールを用いれば個人の Web サイトにサービスを構築することも可能だが、企業によって提供されるブログサービスを利用する場合は、利用者登録するだけで簡単に利用を開始することができる。コミュニケーションメディアとしては、コメント機能を標準で搭載しており、記事を読んだ利用者がコメントを付け、それに対して記事の作者がさらにコメントを返すなどの比較的短いコミュニケーションが発生しうる。ブログは、同期性や即時性、到達確認性は低いが、編集性が非常に高い。社会性や匿名性は利用者ごとに揺れが大きく、実名で運用するケースも非常に多い。また、指向性に関して、掲示板同様に非常に複雑な構造をしている。記事は不特定多数に公開されるだけではなく、明暗問わず特定の人物に向けて発信されることもある。コメント欄が掲示板におけるスレッドのような機能を果たすときもある。

マイクロブログ マイクロブログというカテゴリは 2006 年にサービスが開始された Twitter[34]がさきがけである。国内では 2008 年頃から大きく普及した。Twitter の主旨は、個人の置かれている状況をどこでも記録できること、および緩やかな繋がりによって形成されたネットワーク上の複数の知人の記録を单一のストリーム上で閲覧できることである。仕様上、投稿するメッセージは 140 文字という短文に限定されるため、利用者はブログと比較して気軽に利用することができる。コミュニケーションの形式としては、数人から数十万人まで、利用者が閲覧したいメッセージの発信者を自由に選択できることから、集団性は非常に高い。利用者は対話の機会を常に与えられており、基本的には各利用者は非同期かつ関連のないトピックに関するメッセージを不定期に発信するが、興味がある話題がストリーム上に発生した場合、そのメッセージに反応して対話を始めることができる。その場合、同期性の高い対話を実現することも可能である。Twitter の主体は 140 文字のテキストであるから、分類はテキスト中心のメディアであるが、Twitter で扱うことのできるメッセージはテキストだけではなく拡張的機能として画像や動画など、様々なものがある。また、API の公開により、他のコミュニケーションメディアとの連携が次々と行われている。たとえば、Twitter と連携している Web ニュースの記事を読んだ時、その URL と記事に対する自らのコメントを簡単に投稿することができる。それとは反対に、Twitter のユーザーアカウントを用いて自分のコメントを残すことが可能なニュースサイトなどがある。このように、Twitter は、マイクロブログと表現されるような短い情報の投稿のみが特徴として流行したのではなく、それを軸として多様かつ複雑なコミュニケ

ーションを実現しているために面白みが演出され、世界的な普及に至ったと解釈するべきであろう。なお、Twitter の発展以降、マイクロブログに属する様々なサービスが登場したが、その多くは「短い情報を気軽に投稿できる」機能を全面に打ち出したものである。

2.3.3 複合型コミュニケーションメディア

スマートフォン インターネット接続機能やアプリケーション搭載機能を強化したタイプの携帯電話である。普及率が急激に上昇しており、2012年6月の段階では国内の普及率が18.0%であったのに対し、2013年8月の調査では28.2%に達したという報告がある[50]。スマートフォンは、その多機能性から、利用者によって用途が大きく分かれており、音声通信を中心に用いる場合もあれば、データ通信を中心に用いる場合もある。データ通信においては、本節で述べている様々なメディアが利用可能であり、携帯電話同様にそれらに高いモバイル性を与えている。

ビデオチャット・ビデオ会議 ビデオチャットは、音声および映像情報を、インターネットを通じて送受信するメディアであり、ボイスチャットの普及と同時期に、やはりブロードバンドの普及によって利用されるようになった。ボイスチャットやIM型テキストチャットの付属的機能、あるいはそれらと並ぶ機能としてクライアントアプリケーションが提供されていることが多い。そのような仕様上、話者間で異なるメディアを用いて対話することが可能である（ボイスチャットとIM型テキストチャットでも同様の現象が起こりうる）。また、対話の人数も、1対1の対話だけではなく、アプリケーションによって差はあるが、複数での対話が可能である。その場合、企業などにおけるビデオ会議に利用されることがある。複数の映像情報を同時に扱うために、専用の機材を利用する場合もある。音声および映像情報をメッセージとして送受信するので、対面対話と同様のコミュニケーションの特徴を多く備えているが、異なる点も多い。たとえば、複数人対話における重要な要素である視線情報をビデオチャット上で扱うことは難しく、様々な技術が開発されている[51]。また、映像情報のデータはサイズが大きいため、しばしば遅延が生じうる。すなわち、実時間対話であるのに同期性が低下する可能性があり、それがコミュニケーションの品質に影響を及ぼすことが指摘されている[52]。

SNS (Social Networking Service) SNSは、社会的繋がりをオンライン上で具現化することを主な目的としたサービスであり、基本的・中心的機能としてプロ

フィール機能やフレンドリスト機能、コミュニティ機能、メッセージ送受信機能などを備える。社会的繋がりをオンラインで見出すという概念の登場は比較的古く、1990年代の中盤には、オンラインサービスを提供する企業はそれを強く意識していた。ポータルサイト運営とホームページ開設サービスが連携することで、個人のホームページ同士をコミュニティという枠でくくり、リンクを生成し、利用者同士の交流の活性化を促進した。現在におけるSNSの代表格のひとつとしてFacebook[38]が挙げられる。前述の各機能が搭載されているほか、画像や映像の挿入、他サイトとの連携、スマートフォン上での操作の充実化などが図られており、マイクロブログのような使用方法が可能である。Twitterと決定的に異なるのは、テキストがサービスの中心に置かれていないこと、強力なコミュニティ機能を備えていること、利用者同士の繋がりの強さ、そして匿名性の低さである。Facebookは2013年11月現在、原則として実名での利用が定められており、オフライン社会における人間関係を軸としていることが窺える。また、SNSの多くは利用者同士が相互に許可したうえでフレンドリストに加わるようになっており、一方通行の繋がりが可能であるTwitterとは大きく異なる。また、Facebook以外にも、各国独自のものや世界的に展開しているものを含め、数多くのSNSが利用されているが、Myspace[53]やiKnow[54]のように、音楽や語学学習など特定のテーマを扱うSNSも存在する。最近では、かつてはSNSではなかったものが、コミュニティ機能を搭載してSNSへと変化したサービスも多い。また、企業や大学などの組織で、かつて掲示板に変わって、コミュニケーションを促進するために独自のSNSが導入されていることも多く、SNSがオンラインコンテンツの中心となっている。

動画投稿サイト 2005年に動画共有サービスであるYoutube[30]が登場したこと、不特定多数の人物が動画を投稿し、それを一般利用者が閲覧して楽しむという文化がオンラインで定着した。Youtubeの特徴は投稿された動画にコメントが付与できることであり、動画のページに掲示板のスレッドが添付されているような形式でコミュニケーションが発生する。非会員でも動画を閲覧することができるが、会員になることで投稿やコメント付与、動画リストの作成や公開などが可能になる。したがって、Youtubeの会員同士ではSNSのような社会性が発生する。日本発祥であるニコニコ動画[37]は、もともとはYoutubeの動画にコメントを付与するサービスとして2006年に登場したが、その後、動画の投稿も可能になった。ニコニコ動画のコメントは、動画の画面上を時系列に沿って流れる形で表示される。このことにより、本来は異なる時間軸に付与されたコメントを、あたかも視聴者が動画を視聴している際に実時間で付与されたかのような感覚で閲覧し楽しむことができる。たとえば、本来は数年のタイムラグがあり、全く同期

していないコメント同士が、閲覧時はまるで同期しているかのように機能するなど、コミュニケーションの形態としては非常に独自性が高い。コメント同士の関係を時系列で整理すると、通常のコミュニケーションメディアではありえない構造をしている場合がある。このコメント形式は他の様々な動画共有サイトに参考にされている。また、ニコニコ動画もYoutube同様に、次第に機能が拡大し、動画や画像、さらには動画の実時間配信などを中心としたSNSとして機能している。

動画の実時間配信 インターネットを通じて映像を実時間配信できるサービスである。ライブビデオストリーミングとも呼ぶ。2007年にサービスを開始したUSTREAM[55]が先駆け的存在で、国内ではニコニコ動画が運営するニコニコ生放送[56]の知名度が高い。配信者がおもに映像でメッセージを発信するのに対して、閲覧者はテキストコメントを付与して、相互にコミュニケーションを行う。USTREAMのコメントは、映像画面とは別枠で表示され、Twitterと連携することもできる。ニコニコ生放送の場合は、前述のニコニコ動画と同様の形式でコメントを閲覧することができる。コミュニケーションの様式としては、高い同期性で配信者と閲覧者がメッセージを交換できることから、各種SNSやニコニコ動画と比較すると比較的単純な構造をしている。最近では、携帯電話ネットワーク回線の高速化やスマートフォンの登場により、スマートフォンで記録した映像を配信することなどができるようになっている。

オンラインゲーム 1990年代の末期より、インターネットを通じて他の利用者と通信しながら楽しむことのできるオンラインゲームが登場し始めた。初期は、ゲームの進行に必要なデータのみを交換して楽しむサービスが多くたが、次第に、ゲーム内にテキストチャット機能が実装されるようになり、アバターを用いるゲームでは、それを操作して様々なコミュニケーションを行うことができるようになった。この場合、一般的なテキストチャットと異なり、互いに同一の仮想空間でアバターを操作しながら会話するので、ほとんどの場合で同期性は非常に高くなる。その他、ゲームと平行で他のテキストチャットアプリケーションや、ボイスチャットアプリケーションを利用して、より円滑なコミュニケーションを図る文化が生まれている。最近では、家庭用ゲーム機や携帯型ゲーム機にも通信機能が搭載され、無線LAN環境の充実化にともなって、気軽にオンラインゲームが楽しめるようになっている。携帯型ゲーム機などでは、ハードウェアの仕様上、テキストの入力がやや難しいため、必要に応じてあらかじめ登録しておいた定型文を駆使しながら同期性の高い対話を進める必要がある。この場合個人性が通常のテキストチャットと比較して低下するが、プレイヤーが自分の個性を演出するために、定型文に独自性を与えるなどの工夫がみられる。

その他 本節でこれまでに列挙したコミュニケーションメディアは、現在インターネットで利用することができるものの一部にすぎない。たとえば、ブログサービスの一種である Tumblr[57]は、Twitter のようなゆるやかなフレンド機能を備えており、フレンドが投稿した記事を引用し自分の記事のストリーム上に表示することができる。Tumblr で扱うことのできる記事は、テキストのほか、音声、動画、画像、他の Web サイトの引用など様々であり、極端な例では情報の転送だけでコミュニケーションが成立する非常に独自性の高いサービスとなっている。また、嗅覚や触覚などの 5 感を活用したコミュニケーションの研究も行われております[2][3]、バーチャル・リアリティによるオンラインコミュニケーションの発展に期待が高まっている。

2.4 既存メディアで指摘される問題点

2.3 節で述べたように、近年では非常に多様なコミュニケーションメディアが次々と登場し、人間の文化が大きく発展していることが分かる。一方で、電話が初めて世の中に登場したときのように、問題点も数多く見られる。以下、代表的なものについて列挙する。

フレーミング オンライン上、とくにテキストコミュニケーション上で、人間はしばしば攻撃的になることが知られている。それをフレーミングと呼び、オンラインコミュニケーションが広まった当初からこの現象に関しての研究が心理学の分野ですでにいる。具体的なフレーミング行動として、攻撃的な口調や発話の内容、感情的な言動、誇張表現の使用などが挙げられる。この原因については諸説あるが、匿名性の高い環境が原因の可能性や、揮発性の低いテキストのメッセージが反論を誘発する可能性が指摘されている[5]。また、これらのことから、本来は揮発性が低く情報を保持しやすいため、議論に向いていると思われるテキストメディアは、心理的には議論には向いていないおそれがある。

炎上 オンライン上で生じた何らかのメッセージを対象に、集団で攻撃的行為を行うことが、慣習として炎上と呼ばれる。集団心理やフレーミングの延長線上にある現象だと考えられるが、単純に攻撃的な言動をとるだけではなく、たとえばメッセージから犯罪性が読み取れれば、警察への通報や自治団体への連絡など、それを正すような社会的行動を行うことがある。しかし、しばしばこれらの行為は加熱し、司法の判断を仰がない私的制裁に近い行為が行われる。この過程で、攻撃の対象者は、個人情報をオンラインに流出され、社会生活が困難になることもあり問題視されている。

炎上の原因 炎上は問題視すべき社会現象であるが、炎上の原因の多くは軽率な情報発信である。最近では、未成年の飲酒や、窃盗など、犯罪的行為をSNS上に投稿して炎上するケースが多発している[6]。このような行動がオンライン上で多発する理由は明確ではないが、フレーミングの行動の一環である誇張表現の使用と同様の心理が働いている可能性が考えられる。また、情報の発信者は多くの場合、発信したメッセージが不特定多数に閲覧されると捉えていなかつたと考えられ、コミュニケーションメディアへの不慣れも原因の一つである可能性がある。

情報の流出 オンライン上での軽率な行為のひとつに、犯罪ではないが不用意な情報発信が挙げられる。注意すべき点は、一度送信した情報は編集性が無い場合が多く、一度拡散してしまうと殆どの場合回収することができないことがある。SNSでは、情報公開の範囲の設定をあやまって、プライベートの写真を意図せず全世界に公開するケースや、企業などの組織において部外秘の情報を漏らしてしまうケースが散見される。SNSが広く普及してから10年も経っておらず、その利用に不慣れな利用者が多いことが一因であると考えられるほか、コミュニケーションメディアの形態そのものが複雑化しているため、そもそも利用者がメッセージの流れや状態を把握しきれるものではない可能性が考えられる。

コミュニケーションの齟齬 オンライン上では、しばしばコミュニケーションに齟齬が生ずる。たとえば、近年では端末が多様化し、チャットのようにメールの内容を表示するなど、同期性をかなり高くすることができる電子メールクライアントが登場している。しかし、双方のクライアント間の同期性に差異があった場合、送受信者間で円滑なコミュニケーションが成立しているかどうかの感覚にギャップが生ずる危険がある。たとえば、受信者が旧来型の電子メールクライアントを用いている場合、テキストチャットのような即座な返信は感覚的に要求されないので、メディアが備える同期性に即した程度の時間差で返信をすると考えられるが、テキストチャットに類似した同期性の高い電子メールクライアントを用いている発信者側は、返信までのタイムラグが長く感じられ、過度に反応してしまうおそれがある。また、テキストチャットにおいて、送信したメッセージが相手に到達したかどうかが確認できる際、メッセージを相手が閲覧したにもかかわらず即座な返信がない場合に、心理的不安が生ずる可能性がある。コミュニケーションメディアに対する感覚は個人差があり、かつ、遠隔地にいる対話相手の状況は殆どの場合把握できないので、必ずしも自らが想定したタイミングでメッセージのやりとりが行われるわけではないことを良く理解しておかなければならぬ。

2.5 コミュニケーションメディアの評価方法

2.4 節で述べた諸問題に対応していくには、利用者や開発者が、コミュニケーションメディアが備える性質を正しく把握して利用する必要がある。本節では、それについての既存の方法と、その問題点を踏まえた本研究における提案を示す。

2.5.1 既存の方法

コミュニケーションメディアの利用に際して、利用者は、メディアの知名度や機能、コスト、通信品質、利用可能端末などについて検討して決めるのが通常であると思われる。あるいは、知人から推薦されて「試しに利用する」といった場合もあり得る。知名度の高いSNSを比較した文献や調査では、評価の際に、匿名性や送信可能メッセージの種類や規模、指向性、インターフェース、コメント数によるアクティブ率など、いずれも物理的な特徴が取り上げられている[58][59][60]。

コミュニケーションメディアの開発サイドの評価の事例としては、各種物理的性質のほか、遠隔コミュニケーションにおける心理的距離の計測[61]や、メッセージ遅延がもたらすコミュニケーションがもたらす利用者の心理への影響[52]、あるいはバーチャル・リアリティにおける感性評価[62]などが先行研究にみられる。

2.5.2 本研究における提案

既存の方法では、利用者と開発者が共通して扱えるような汎用的な評価の観点は、コミュニケーションメディアの物理的な仕様に起因するものだけだと考えられる。そこで本研究では、コミュニケーションメディア上におけるメッセージの表現力に着目し、物理的な観点からだけではなく、心理的な観点から可視化することで、利用者・開発者双方が利用できる新しい評価軸の提案を試みる。本研究では、その測定方法について検討し、それを実際に音声メディアに適用して測定方法の妥当性について議論する。

コミュニケーションメディアの表現力を測定する理由として、表現力はコミュニケーションの品質に直結する重要な要素であるにもかかわらず、これまでにそれを測定し可視化する試みが物理的にも心理的にも行われてきておらず、実際、様々なコミュニケーションメディアがどのような表現力を持っているか不明瞭で

ある点が挙げられる。利用者にとっては、これをあらかじめ把握しておくことで、コミュニケーションメディアを利用開始する際の判断材料として活用でき、利用開始後も対話相手のメッセージの表現を正しく解釈する視点を得ることが期待できる。開発者にとっても、コミュニケーションメディアを設計する際の評価軸のひとつとして活用することができる。

また、本研究における表現力測定の対象として音声を選ぶ。本章でこれまでに述べたように、音声は人間にとってもっとも馴染み深く初步的なコミュニケーションメディアのひとつであり、その誕生から現在に至るまでその役割は殆ど変化していない。したがって、次のような特徴を備えていると考えられる：音声による対話は(1)単一のモダリティで構造が比較的単純であること、(2)それにもかかわらず離隔なくコミュニケーションが可能であること、(3)代表的な人間のコミュニケーション手段であること、そして(4)コミュニケーションメディアに関する工学的研究の中でも歴史ある分野であること。

(1)から(3)の理由により、音声が備える表現力が明らかになれば、他のメディアの表現力の測定時に音声の表現力を基準に議論することが期待できる。また、(1)および(4)の理由により、他のメディアよりも表現力の測定の際に参考となる知見が多く、目的の達成が比較的容易であると考えられる。

2.6 本章のまとめ

本章では、コミュニケーションメディアの評価にかかる情報を整理し、それに基づいて、本研究で提案する評価方法とその重要性について述べた。

コミュニケーションメディアの評価にかかる情報として、コミュニケーションメディアの歴史・性質・種類・問題点・既存の評価方法の5つの観点から整理した。1点目の歴史については、人間の文明や科学技術の変化とともにどのような変遷を遂げてきたのかを振り返り、音声言語の誕生からインターネットの普及に至るまで、いかに多くの種類のコミュニケーションメディアが使われてきたのかを述べた。2点目の性質については、コミュニケーションメディアの歴史を振り返った際に登場した多くのコミュニケーションメディアにおいて、物理的な観点からどのような特徴の差があるのかを論じながら、20の項目について列挙した。3点目の種類については、とくに近年頻繁につかわれているオンラインコミュニケーションに関わりが深いメディアを、音声メディア・文字メディア・複合型メディアのカテゴリごとに代表的なものをとりあげ、前述の20の性質について触れながらそれぞれのメディアの詳細について説明することで、インターネット上のコミュニケーションメディアがいかに多様化しているかを確認した。4点目の問題点については、近年とくに問題視されているオンラインコミュニケーション上で発生する事象について列挙し、コミュニケーションメディアの多様化による影響の可能性について述べた。5点目の既存の評価方法については、先行研究や事例における評価の例を示し、利用者や開発者による既存の評価においては、主にコミュニケーションメディアの物理的な仕様に起因する特徴の観点から行われていることを指摘した。

最後に、本研究で提案する評価方法を述べた。本研究では、コミュニケーションメディア上におけるメッセージの表現力に着目し、物理的・心理的な側面から表現力の可視化をおこなうことで、メディアの利用者・開発者双方が共通で扱える新しい評価の観点を与えることを試みる。そして、表現力の可視化の対象として、本研究では人間にとってもっとも基本的なコミュニケーションメディアのひとつである音声を取り扱うことを示した。

第3章 コミュニケーションのモデル化

コミュニケーションメディアの表現力を測定するためには、メディア上で生じるあらゆるメッセージのパターンを網羅する必要があると考えられる。しかし、多くのメディアにおいて、メッセージを符号化した際のパターンは膨大で、その全てを扱うことは難しい。したがって本研究では、特定のコミュニケーションモデルを仮定し、それを制約条件として、限定的な範囲におけるメッセージの分析をおこなうことを考える。本章では、先行研究で提案してきたコミュニケーションモデルを俯瞰し、それを参考にして、本研究の目的を達するために適した汎用的なコミュニケーションモデルを検討する。

3.1 先行研究におけるコミュニケーションモデル

3.1.1 Shannon のコミュニケーションモデル

コミュニケーションに関するモデルは、古くから多くの研究者達によって提案・検討されてきた。情報理論の考案者である Shannon は、コミュニケーションを「情報を伝達する過程」に還元することを提案した[63]。Shannon のコミュニケーションモデルは、ネットワーク技術における伝送効率の最適化が目標である。図 3-1 はそのモデルの概要であり、6 つの要素から構成されている。このモデルにおいては、情報発信源 (Information Source) でメッセージ (Message) が生成され、送信器 (Transmitter) で符号化されたのち、伝送経路 (Channel) を通して受信機 (Receiver) に伝達され、そこで復号されたメッセージを送信先 (Destination) が受け取るというシンプルな構造に加え、伝送経路上でノイズ (Noise Source) による影響を受けて、メッセージが変質しうることが示されている。

ここでいうノイズとは、もともとは通信回線上で生ずる物理的なノイズを示していたが、このモデルを人間のコミュニケーションに当てはめた場合、コミュニケーションメディアが備えている様々な性質が、ノイズとして機能し、メッセージを変質させると見做すことが可能である。

3.1.2 Brunswikian Lens Model

Shannon と同時期に Brunswik が発表した、人間の視覚における認知過程のモデル[64]は、Shannon らが主張した情報伝達過程のほかに、生態学的妥当性が考慮されていることが特徴であり、Shannon らのモデルと比較すると、印象形成などの心理的な作用が意識されている。生態学的妥当性とは、Brunswik が提唱した、「人為的な環境下での実験結果がどの程度実際の環境下で役立つかの程度を示す指標」である。

3.1.3 Brunswikian Lens Model の Scherer による解釈

近年になり、Scherer が Brunswik のモデルにおける生態学的妥当性の重要性を指摘した上で、このモデルの音声コミュニケーションへの適用を提案した[65]。Scherer による Brunswik のモデルの解釈を、簡単のために我々が簡易表現したものが図 3-2 である。発信者 (Criterion) から受信者 (Attribution) までメッセージが伝達される際、メッセージに含まれる情報が符号化 (Encoding) され、遠手がかり (Distal Cue) となる。遠手がかりは、音声においては、音声が人間の口から発せられる瞬間の音響特徴そのものをイメージすれば良い。遠手がかりは、通信媒体を通過して (Transmission)，実際に知覚される情報、すなわち近手がかり (Proximal Cue) となり、受信者に呈示 (Representation) される。

基本的な構造は Shannon のモデルと類似しているが、異なる点は、メッセージが符号化された際に、符号化された情報である遠手がかりが複数発生しうる点が強調されていることと、メッセージが遠手がかりに符号化される際、および近手がかりが受信者へ呈示される際に、メッセージそのものに含まれない、社会的背景や状況もあわせて処理されることが暗に示されている点である。

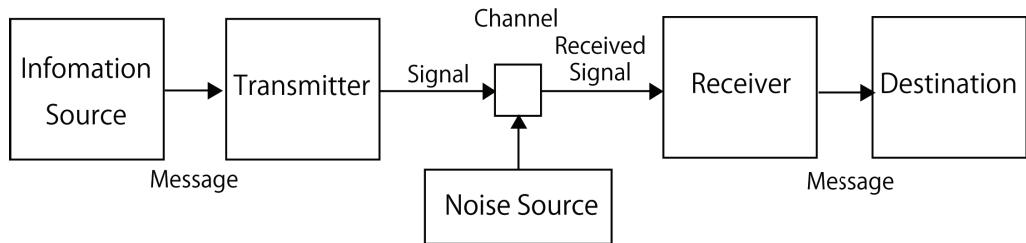


図 3-1 Shannon の提案したコミュニケーションモデル

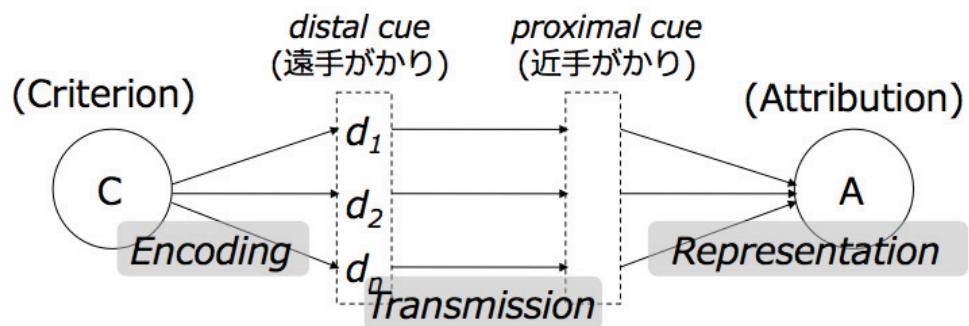


図 3-2 Scherer による Brunswikian Lens Model の再解釈に対する本研究における簡易表現

3.2 本研究におけるコミュニケーションモデル

我々は、Shannon によるコミュニケーションモデルおよび、Scherer による Brunswikian Lens Model の解釈を参考にしながら、2.2 節で述べた各性質を明示したモデルを想定することにした。具体的には、Shannon のモデルから Noise Source の概念を踏襲し、Scherer によるモデルからは生態学的妥当性および符号化および復号の処理、すなわち「メッセージそのものに含まれない、社会的背景や状況もあわせて処理されること」の概念を踏襲した。そして、Noise Source および社会的背景や処理をより具体化するため、2.2 節で述べた各性質をモデル上に配置することにした。

図 3-3 に、本研究で我々が提案するコミュニケーションモデルを示す。このモデルは、大別して、送信者 (Seder)，受信者 (Receiver)，メディアの性質 (Media Property)，そして Brunswikian Lens Model を参考にした Encoding から Decoding までの流れから構成される。Scherer は、メッセージの内容は単に社会的背景や状況に影響されると述べたが、我々のモデルでは、送信者や受信者に属する要素として、「メッセージそのものに含まれない印象を制御する要因」をより具体的に示している。

たとえば、メッセージの送信者は、受信者との関係や、自己自身の社会的背景、置かれている状況などに応じてメッセージを変化させる可能性があるが、受信者の置かれている状況や感情状態などによっても、メッセージの印象に変化が生ずる。加えて、メディアの性質は、伝送経路上でメッセージに影響を与えるほか、送信者・受信者のメッセージの生成・解釈の際にも直接影響を与える。たとえば、機密性が低いメディアを用いるのであれば送信者はメッセージの生成に注意深くなるであろうし、受動性が高いメディアであれば受信者はメッセージの受信を強制されているような心理状態となることなどが考えられる。

このモデルは、どのようにメッセージが生成され、伝送および受信の際に変質するかを示しているので、コミュニケーションのなかでも特にこれらに着目した、メッセージ生成・伝達モデルと呼ぶことができる。

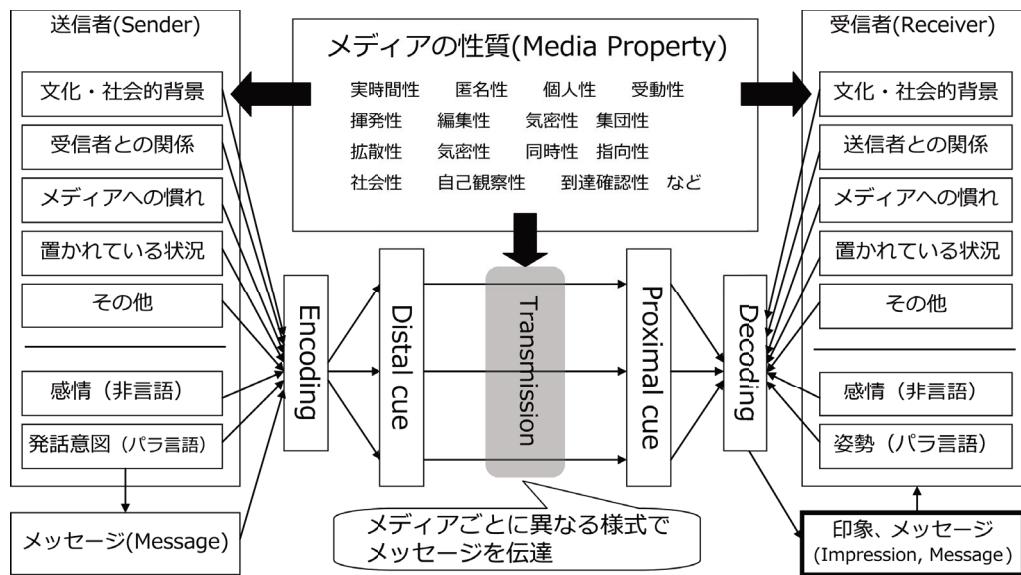


図 3-3 本研究で想定するコミュニケーションモデル（メッセージ生成・伝達モデル）

3.3 本章のまとめ

本章では、コミュニケーションメディアの表現力を測定する際に問題となる、メッセージの符号化パターンの膨大さを緩和するために、メッセージに制約条件を付与するためのコミュニケーションのモデルを検討した。そのために、まず先行研究として Shannon のコミュニケーションモデルと、Brunswikian Lens Model の Scherer による解釈を示し、本研究にとって重要であると思われる概念をそれぞれから参考することにした。そして、それらに加えて、2.2 節で述べたコミュニケーションメディアの性質を考慮しながら、独自のコミュニケーションモデルである、メッセージ生成・伝達モデルを検討した。次章では、この汎用的なコミュニケーションモデルを基盤として、演技音声の表現力測定に最適化したモデルを検討する。

第4章 音声の表現力の測定方法の検討

本章では、第3章で設計したコミュニケーションモデルに基づきながら、実際に音声メディアの表現力を測定するための準備として、音声メディアにおける幅広いメッセージの表現を収集する方法について述べる。具体的には、おもに演技指示に工夫を施すことで、表現豊かな感情音声の獲得を試みる手順について説明する。

また、いかに工夫して音声表現を収集するとしても、演技によって多様な音声表現を偏りなく獲得できると一概に考えるのは危険が伴う。したがって本章では、音声表現の分類を検討したうえで、収集する音声の偏りについても考察する。

4.1 関連研究

表現の豊かさを追求した自発音声のデータベース構築を主題としたものは多く存在する。たとえば、日常生活下あるいは模擬的な生活環境下での自発音声の収録を試みた The JST/CREST ESP (Expressive Speech Processing) Corpus[65] や The FAU Aibo Emotion Corpus[67]が挙げられる。その一方で、演技音声に関しては、表現の豊かさは殆ど追求されてこなかった。音声分野における従来研究で用いられてきた演技音声表現の多くは、演技指示が基本感情語で説明できるような簡易なものを用いている。この方法では、得られる感情表現も典型的なものに偏り、表現豊かな演技音声資料を得ることは難しいと予想される。

4.2 音声資料の分類

演技音声の定義はどのような行為を「演技」とみなすかによって、かなり幅がある。そこで本章では、目的とする音声の収集にあたり、本研究における演技音声・感情音声の定義と分類について明確にする。類似した問題として前川は、自発音声と朗読音声の定義を行っており[68][69]、これらは本章における分類を考慮するうえで重要な概念である。その定義に従えば、朗読とは棒読みとは限らず、たとえば、MULTEXT-J[70]では感情を込めた朗読音声が収録されている。

(ア) 演技音声

従来の音声研究で使われてきた多くの演技音声は、典型的な感情語などを刺激として与えて、発話者が表現を統制した音声を発話したものである。これを「朗読型の演技音声」と呼ぶこととする。他方、たとえば日常生活下において、「店員」「教師」などの役割を与えられて発する音声は、一般に自発音声となり、「自発的演技音声」と言える。

(イ) 表現豊かな音声

Erickson[71]は、表現豊かな音声の研究では、音声に含まれる言語的情報・パラ言語的情報・非言語的情報（それぞれ Fujisaki の定義[72]に従う）すべてが対象になると述べた。本研究では特にパラ言語・非言語的情報に着目したうえで、「表現豊かな音声」の解釈として、個々の発話の観点と多数の発話でみられる変動幅の観点を考える。

- 表現豊かな音声とは「ある発話においてパラ言語・非言語的情報が（豊かに）含まれている」ことを指す（個々の観点）
- あるいは、大量の音声資料を大局的に観察した際に「パラ言語・非言語的情報のバリエーションが幅広い」ことを指す（多数発信の観点）

本研究では、これらにおけるパラ言語・非言語的情報のうち、感情のみを扱うことにして、上記の 1.を指して「感情音声」、2.を指して「多様な感情表現」という。

(ウ) 感情音声

「感情音声」における「感情」という言葉の解釈は幅広く、また一般的に多層であるとされるが、本研究では感情の定義や分類が本題ではないので、ここでは、「情動 (emotion)」の意味である、「特定の事物、事象に触れることによって、身体的、精神的な存在としての自己が脅かされたり、勇気づけられたりしたときに生じる心[73]」を「感情」と定義するに留める。たとえば、喜怒哀楽やそれらの中間層などの心理状態がこれに含まれる。また、従来の感情音声研究では、発話者本人の意識した感情の状態と発話を聞いた第三者の印象として捉えられるものが混同されて扱われている場合が多いという森山らの指摘[74]に従い、本研究では第三者が印象として捉える発信者の感情を単に「感情」として統一して扱う。また、発話を聞いた際に生ずる第三者の感情状態（「生意気な」など）は含まない。

(エ) 多様な感情表現

本研究の目的は、音声の表現力を測定する材料とするため、演技音声の表現の豊かさを可能な限り高めることであるが、その達成は、いかに多様な感情表現を

表 4-1 本研究における音声資料の分類

分類	演技音声(1)	演技音声(2)	非演技音声(1)	非演技音声(2)
自発/朗読	朗読	自発	自発	朗読
感情音声の出現	あり	あり	あり	なし
説明	刺激で統制された音声	日常生活下の役割的音声	日常生活下の自然な発話	朗読における棒読み音声
発生する状況	実験音声・創作物中など	演説や仕事など社会的役割	家族との対話・独り言など	棒読みによる読み上げ音声
コーパスの例	既存の実験用感情音声	特定の職業の発話コーパス	一般的な自発音声コーパス	一般的な朗読音声コーパス

含む演技音声資料を収集できたかを調べることで確認できる。ここで、「多様な感情表現」の要件を「感情音声の持つ音響的・心理的特徴を幅広く把握した表現の集合」と定義する。

(才) 本研究が対象とする音声

以上を踏まえ、自発音声か朗読音声かおよび感情音声が現れ得るか否かという基準で、演技音声と非演技音声を分類すると表 4-1 のようになる。本研究ではまず、表 4-1 の「演技音声(1)」における表現の豊かさを高め、限定的な範囲で多様な音声表現を得ることを目指す。一般的に、従来型の感情語を話者に呈示する方法による演技音声だけでは、得られる感情音声は典型的な感情表現に偏り、自然に発生する多様な感情表現のバリエーションを包含することはできないと考えられる。したがって、次章では、データ収集の方法について、従来型とは異なる方法を検討する。

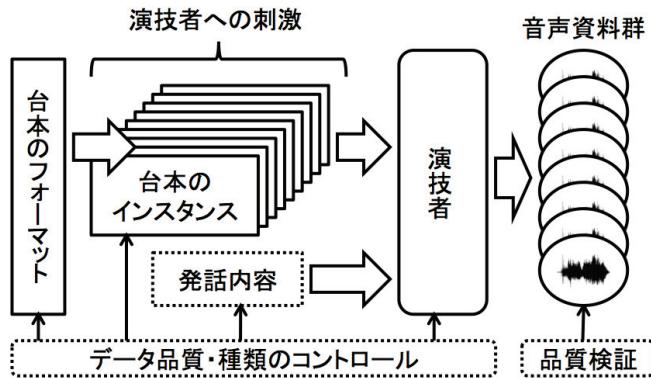


図 4-1 提案方法の概念図

4.3 データ収集の方法

本章では、演技音声を用いて多様な感情表現を得るための具体的手順について述べる。特に、基本感情語あるいは基本感情語を主軸とした演技指示では得られないような、ステレオタイプではない音声表現を得ることができる手続きについて検討する。

4.3.1 方針

具体的手順の全体を概念図で示せば図 4-1 のようになる。この中で、音声資料の品質確保のほか、収集効率や評価を考慮して、各部分すなわち、発話内容の生成、および演技者への発話内容・感情刺激の指定については、次の基本方針を設ける。

- 基本感情のステレオタイプに属さない、「多様な感情表現」を促進するような複数の情報を組み合わせて刺激を設計する
- 効率性・拡張性の確保のため、発話内容に依存しない刺激となるように設計する
- 演技者と発話内容の各組み合わせにつき、多様性を判断するのに十分な音声数を収録する

これらの方針を満たすことのできる具体的なデータ収集の方法として、以下、「台本のフォーマット」の設計、「台本のインスタンス」の作成、および「演技者」「発話内容」の選定について述べる。

表 4-2 台本のフォーマットとインスタンスの例

台本フォーマット		インスタンス	台本の一例
話し手と 聞き手 共通の情報	発話時の場所・ 状況	コント, ニュース, 学校(ドラマ), など バラエティ番組, 調理場, 病院(ドラマ), など	学生寮のホール (ドラマ)
	話し手と 聞き手の関係	学校の友人, 芸人の先後輩, リポーターと一般人, 生徒と教師, 父と娘, 兄弟, など	同じ学校, 寮に住む親しい友人達
	発話時の背景	組織の過ちの許しを静かに請うように 自分のプライバシーを指摘されてごまかすように, 目に涙を浮かべ淡々と親の仇に対し恨みを語るように, など	周りの騒ぎっぷりを, 呆れてみると
話し手が持つ 聞き手の情報	年齢/性別	10代未満~60代の男女	17~18歳, 男
	職業・続柄	高校生, MC, CA, アイドル, 芸人, アナウンサー, 教師, 医者, など	高校生
	人物像	冷徹で欲深い, 男勝りの活発な, 明るく騒がしい, 天然, 重症で危篤, 持論をじっくりもったベテラン, など	騒々しく, 騒いでいる最中
話し手自身の 情報	年齢/性別	10代後半~70代の女性	17歳, 女
	職業・続柄	高校生, 女優, アナウンサー, 主婦, 料亭の女将, コメンテーター, 知識人, など	高校生
	人物像	体育会系, お調子者, 天然, 清楚, ツンとして気の強そう, 活発で男勝り, おっとり, ベテランで落ち着いた, など	男のフリをした女 冷静だが, 恋愛に鈍感

4.3.2 台本のフォーマットの設計

台本のフォーマットは、演技者に呈示する刺激の内容およびそれを与える方法などを整理したものである。たとえば、話し手の年齢や職業がそれに該当し、「話し手の年齢」であれば「10代」「20代」「60代」などが変数の中身(=インスタンス)に成り得る。

つまり、台本のフォーマットの設計には、演技者が感情音声を表現する際に、変化を生じさせるであろう要因を適切に選定しなければならない。

4.3.2.1 演技方法に関する先行研究

心理学の分野の先行研究や一般著書の中で、演技の質を高めるための重要な情報に関する記述が散見される。

平田[75]は、「発話は常に他者との関係において行われる」点を指摘しているほか、「相手との距離、部屋の大きさ、外からの雑音などと、様々な要素から無意識のうちに影響を受けている」とも述べている。また、後安ら[76]は、演劇の現場での演出家の指示を分析し、指示は抽象的指示と具体的指示に分類でき、ある劇の稽古においては93%が具体的指示であったことを述べた。具体的指示の内容は、発話のタイミングやセリフ回し、姿勢・表情・知覚などに分類されるとしている。このように、演技の際に利用される情報には多様な種類・階層がある。

一方、安藤[77]は、演技は脚本解釈と演技計画、そして演技遂行段階に分かれているとし、演技の熟達者ほど演技計画が脚本解釈を上回ると述べている。すな

わち、熟達者は一つの脚本解釈から多くの演技計画、つまり、どのように表現するかを生み出すのである。本研究においても、あまり細かな指示からは、多様な発話パターンが得られにくくなると考えた。そして平田[75]が述べた「他者との関係性」を重視し、話し手・聞き手の情報、そして両者の関係性、さらに我々が無意識に影響を受けている様々な要素を「発話時の場所・状況」「発話時の背景」と整理して刺激のフォーマットの項目として記述することにした。

4.3.2.2 台本のフォーマットの具体例

具体的に本実験で設定した、台本のフォーマットの例を表 4-2 に示す。ここで、表現の多様性を演出出来る程度に、単純とならないようにこれらの項目を決定する一方で、効率性を損なわない程度の必要最低限の項目となるよう留意する。また、項目の過不足などについては、収録した音声の分析を通して検討することができる。

4.3.3 台本のインスタンスの生成方法

台本のインスタンス（以下、単に「台本」と記す場合はインスタンスを指す）は、方針に従い、多様性判断に十分な数を作成するが、実際の収録にあたっては、何度かに分けて精度を確認しながら次第に数を増やしていくことにした。

生成方法については様々な手順が考えられる。たとえば、完全に創作でも良いし、現実世界や創作物から様々なエピソードを抽出するという手順もある。理想的には、多様な感情表現を生み出せるよう調整したインスタンスを大量に創作するのが望ましいが、創作によって品質の高いインスタンスを作るためには、専門的知識・技術が必要となる。そこで、本研究では、次に示すように段階的にインスタンスの生成方法を変化させることにした：(1)現実世界や創作物からのエピソード抽出によるインスタンス、(2)既存のエピソードにおける各項目の組み合わせ、(3)インスタンスの完全な新規作成

なお、あくまでも多数のインスタンスの効果を大局的に観察したときの多様性の向上が第一の目的であるため、あるインスタンスに対する再現可能性および安定性は考慮しない（Krippendorff による定義[78]を参考にし、本研究における安定性は「同一演技者が複数回演技した際の表現の類似性」、再現可能性は「複数の演技者が演技した際の表現の類似性」を指すこととする）。

4.3.4 演技者の選定方法

まず、我々は演技者を次のように簡単に分類する。

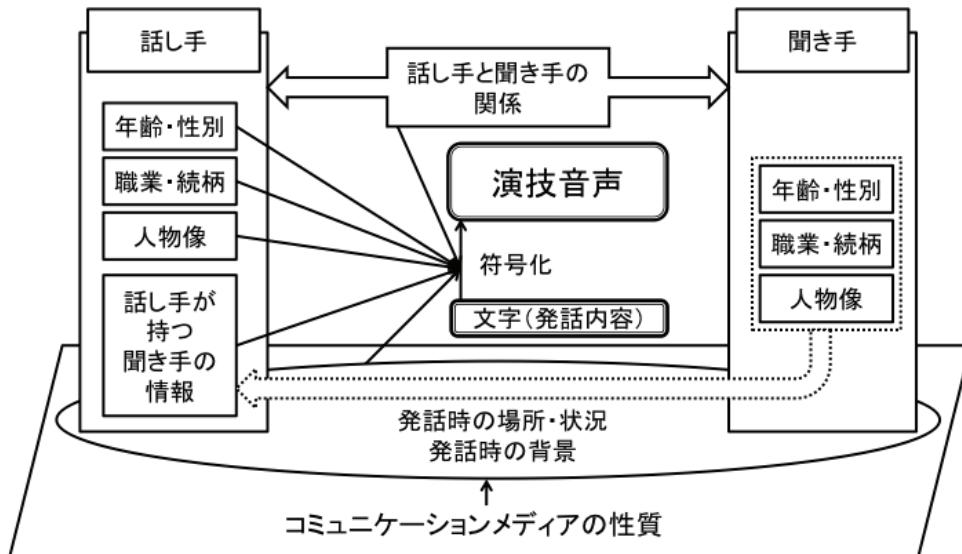


図 4-2 本研究の演技音声の生成モデル

声優：音声のみでの演技経験があるもの

俳優：音声を含めた多様なモダリティで演技経験があるもの

一般人：演技経験がほとんどないもの

これらの分類において、いずれかが絶対的に本収録方法に対して最適だとは言い切ることは難しい。表現力の幅に違いはあるかもしれないが、ある演技者の演技音声の表現内容が他の演者の感情表現を包含する関係になるとは限らないからである。また、声優も俳優も、より詳細な分類方法がある。たとえば、声優の場合「ナレーター中心」「アニメーション中心」など、俳優の場合「映像中心」「舞台中心」などで違いがあることが予想されるが、本研究の初期段階ではそのような分類方法による差異を研究対象としない。ただし、実際の選定の際には、これらの分類およびキャリア年数、年齢、性別を考慮する。

4.3.5 発話内容の選定方法

本手法において演技者は、インスタンスを刺激として、決められた発話内容による表現をおこなうが、4.3.1 節で述べたように、本手法では、インスタンス中には発話内容が限定されるような情報は組み込まないことを前提としている。ただし、演技や分析の難しさを緩和するため、以下の制約条件を設けることにした。

- 文意が中立的 (neutral) である：表現の多様性が文意に影響を受けないようにするため
- 有核のアクセント句を含む：韻律の上昇下降部分を活用した文章への表現の

付与を可能とするため

- 日常で頻繁に使用される発話である：インスタンスと発話内容の組み合わせの段階での不自然性を緩和するため

なお、前川・北川によるパラ言語情報付与の実験[79]でも、発話内容として14種類の有核のアクセント句を含む文章を採用している。

4.3.6 演技音声の生成モデル

これまで述べてきた演技に関する要素を用いて、3.2節で示したコミュニケーションモデルを演技音声の生成過程に最適化させると図4-2のようになる。人間は様々な情報（周囲の状況、相手の様子、自分の心的状態等）を言語情報に乗せて符号化（音声を生成）すると仮定している。また、コミュニケーションメディアの性質は、発話時の場所・状況や、発話時の背景における設定を通して演技者の表現に影響を与える。

本研究はこのモデルに基づきながら、多様な感情表現を抽出することが目的となる。

4.4 本章のまとめ

本章では、音声メディアの表現力を測定するための準備として多様なメッセージを収集するための具体的な手順について検討した。

まず、音声資料の分類をおこない、本研究で対象とする音声メッセージは、演技音声の一部とすることを示した。そして、表現力の測定に十分役立つような表現豊かな演技音声を得るために工夫として、本研究独自の「台本のフォーマット」を提案した。台本のフォーマットの設計にあたっては、演劇に関する先行研究や著書を参考にしながら、音声で演技をする際に表現に影響を与えると思われる重要な情報を選定した。その情報に基づき、第3章で示したコミュニケーションモデルを音声に最適化させた。さらに、表現豊かな音声表現の獲得のための、台本のインスタンスの生成方法、発話内容および発話者の選定に関する方針を定めた。

第5章 音声収録実験と評価

本章では、第4章で述べた枠組みにしたがって収録した音声（以下、『千通りの表現』の意を表す『SEN』と表記）と、比較用として用意した従来用いられてきた感情語を表示して収録した音声（以下、『Typical』と表記）を実際に収録し、それらの比較を通して表現豊かな演技音声ができたか、そして表現力の測定に貢献することができるものかを論ずる。

収録は、段階的に行われ、パイロット版と本収録版、そして拡張版が存在する。それらの概要は表 5-1 の通りである。

5.1 パイロット版音声の収録と評価

5.1.1 台本インスタンスの作成および演技者・発話内容の選定

まず、表 4-2 で定めたフォーマットおよび 4.3.3 節で示した方針にしたがって、台本のインスタンスを生成することにした。具体的には、多様な音声が得られやすいと考えられるテレビ番組から、様々な場面を抽出することにした。抽出対象として、国内の放送局をひとつ選び、24 時間の録画を行った。放送局の選定条件は、固い雰囲気のニュース番組からバラエティまで、幅広いプログラムを用意していることである。そして、24 時間の録画データの中から、十分に多様なインスタンスが生成できるよう考慮しながら、主観的に 300 個のインスタンスを作成した。表 4-2 における「台本の一例」は、実際に作成したインスタンスから選んだものである。「発話時の場所・状況」および「発話時の背景」は、音声による対面対話、テレビ上での情報発信、および中継による遠隔対話から構成されている。

次に、演技者を選定した。パイロット版で選定した演技者は女性 1 名で、演技者 A とする。演技者 A は、声優・舞台両方のキャリアで 17 年という長いキャリア（収録当時）を有し、特にプロフェッショナルの声優として活動しているため、多様な演技音声を表現するために十分な能力があると判断した。

発話内容は「ああ、そうですか」とした。これは、4.3.5 節で述べた条件を満たすと判断したほか、「そうですか」の箇所が先行研究[79]で用いられているため、将来の比較分析を可能にするためである。

表 5-1 提案手法を用いて収録した音声一覧

収録種別	インスタンス作成方法 / 発話内容	台本数 (音声数) / 人数
パイロット版	TV 番組から主観的に作成 「ああ、そうですか」	304(100) 1名
本収録版	パイロット版の「話し手」および「発話時の背景」の組み合わせ 「ああ、そうですか」	280(1400) 5名
拡張版	全て新規で書き起こし 「いたい」「からい」「すばらしい」など 6 通り	1200(7200) 6名

5.1.2 音声収録

5.1.2.1 収録環境と手順

収録環境は、早稲田大学内のマルチメディアスタジオに、パソコンおよび USB オーディオデバイス(Roland UA-25)を設置した。マイクはヘッドセットタイプ(audio-technica AT810F)を採用した。サンプリング周波数は 44,100Hz, ビットレートは 16bit に設定した。

収録に際しては、300 個のインスタンスから、発話者の設定が小学生となっているものなど、演技者 A の演技に無理が生ずる可能性があるものを除き、さらに、十分に内容が異なるものを 100 個主観的にピックアップした。そして、1 枚につき 1 台本が記された用紙を実験者が順次表示して、それにしたがって演技をしてもらった。台本に順序関係は存在しないので、ここではインスタンス作成時に割り当てた ID 順に表示をおこなった。このとき、実験者の判断で、台本とあまりにもかけ離れている表現であると判断した場合、または、発話者がやり直しを希望した場合は再収録を実施した。

5.1.2.2 比較用音声の収録

SEN の特徴を明らかにするための比較用の音声として、典型的な感情語を演技者に表示して音声による感情表現を収録した(Typical)。従来研究の多くでこのような音声が用いられているため[65]、それらに対する比較という観点で、Typical では最低限の要素である典型的な感情語のみを表示するに留めることとした。典型的なものとして扱われる感情語のセットは、研究目的によって根拠が異なり、多数存在する。本研究では、Ortony and Turner[80]を参考に、14 種類の基本感情語のセットを用意し収録を実施した（表 5-2）。感情語は重複を含め延べ 79 種類で、これに無感情を含めて延べ 80 種類とした。

演技者は SEN と同様に A の 1 名で、発話内容および収録環境も同様である。

表 5-2 Typical における基本感情語一覧
 (各文献については Ortony and Turner[80]を参照)

提唱者	感情語
◎ Arnold (1960)	怒り, 反感, 勇敢, 落胆, 欲望, 絶望, 恐怖, 嫌悪, 希望, 愛, 悲しみ
◎ Ekman, et al. (1982)	怒り, 嫌悪, 恐怖, 喜び, 悲しみ, 驚き
◎ Frijda (1968)	欲望, 幸福, 興味, 驚き, 怪しみ, 後悔
Gray (1982)	激怒, 恐怖, 不安, 喜び
◎ Izard (1971)	怒り, 軽蔑, 嫌悪, 恐怖, 自責, 興味, 耾辱, 驚き
James (1884)	恐怖, 悲嘆, 愛, 激怒
◎ McDougall (1926)	怒り, 嫌悪, 高揚, 恐怖, 服従, 親愛, 怪しみ
Mowrer (1960)	苦痛, 喜び
Oatlay, et al. (1987)	怒り, 嫌悪, 不安, 幸福, 悲しみ
◎ Pandsepp (1982)	期待, 恐怖, 激怒, 混乱
◎ Plutchik (1980)	受諾, 怒り, 興味, 嫌悪, 喜び, 恐怖, 悲しみ, 驚き
Tomkins (1984)	怒り, 興味, 軽蔑, 嫌悪, 悩み, 恐怖, 喜び, 耿辱, 驚き
Watson (1930)	恐怖, 愛, 激怒
Weiner, et al. (1984)	幸福, 悲しみ

収録に際しては、多様な感情表現を得ることを重点的に考慮してデザインされた SEN の収録手法に対して、典型的な感情語を用いる方法の範囲で、SEN とできるだけ情況を近くして比較ができるような方法を検討した。具体的には、感情語のセットごとに刺激として呈示し、「セット内で出来るかぎり幅広い表現を試みてください」という指示をおこなうことで、できるだけ演技者に意識をさせずに多様な感情表現が引き出せるようにした。これにより、同じ「怒り」という感情語でも、演技者が意図すること無く異なる演技表現が得られる可能性があると考えた。

80 語の収録を、休憩を挟んで 2 回実施し、160 個のデータを得た。

5.1.3 心理的特徴の分析

本章では「多様な感情表現」を満たす要件のうち、心理的特徴について検証する。まず、印象評定の手順と結果について述べたのち、評定結果の分析をおこなう。

5.1.3.1 評価尺度

感情音声の多様性の検証にあたって、感情音声の印象評定に適した尺度を決める必要がある。森山ら[74]は、感情音声に含まれる客観的感情（心理量）と感情表現パラメータ（物理量）の相互変換を検討する過程で、従来研究で利用された46の感情語の中から感情音声を効率的に印象評定することが出来る9つの感情語を抽出した。これを用いれば、少ない評価項目で多様な音声表現の評価ができるほか、評価者の疲労を軽減することができる。その感情表現語は、「怒り、喜び、皮肉、恐れ、悲しい、驚き、こび、穏やか、おかしい」の9語から成る。本実験もこの9語を評価軸とする。

5.1.3.2 印象評価実験

評価者は20代の大学生、男女12名である。評価軸の9語を用いて、「全然あてはまらない(1)～どちらでもない(4)～非常にあてはまる(7)」の7段階のラベルを設定し、評価尺度とした。また、評価対象の音声が自然な表現であったかどうかを評価するため、自然性の評価として、「非常に不自然(1)～どちらでもない(4)～非常に自然(7)」の7段階の評価項目を追加した。

評価対象は、SENとTypicalからそれぞれ50語抽出した。SENは完全無作為に50語選び、Typicalは、表5-2の一覧を見る限り類似した感情語が多く含まれることから、完全無作為では感情表現に偏りが生ずる可能性があるので、感情語のセット単位で選択することにした。本実験で実際に選んだ感情語のセットは、表5-2の「◎」の印が付いた7セットである。

評価はWebを用いたアンケート方式で実施し、SENとTypicalの合計100の音声を被験者ごとにランダムに呈示した。評価は、1人あたり3時間程度の時間を要した。疲労による評価のぶれを防ぐため、評価者は任意のタイミングで休憩をとって良いことを伝えた。

5.1.3.3 感情表現語の印象評価値の主成分分析

各音声に付与された10の印象評価値に対して、12人の評価者の平均値を求め、最終的な評価値とした。そして、心理評価で一般的に用いられる軸（たとえば「感情値」や「覚醒」など）上での考察をおこなうため、この評価値を用いて主成分分析を実施した。

ここで、Typicalが従来型の演技音声として、それに対する相対的なSENの位置づけを確認するために、Typicalの50音声に対して主成分分析を実施した。表5-3は、分析の結果得た固有ベクトルであり、この結果より、第一主成分は感情値（Valence、寄与率52.2%）、第二主成分は覚醒（Arousal、寄与率30.1%）と解釈した。これを用いて、SENとTypical各50音声の主成分得点を求めた。図5-1

表 5-3 Typical の印象評価値に対する主成分分析結果の固有ベクトル

感情語	第 1 主成分	第 2 主成分
怒り	-0.57	0.264
喜び	0.402	0.353
皮肉	-0.382	0.329
恐れ	0.025	-0.384
悲しみ	-0.091	-0.696
驚き	0.372	-0.029
こび	0.379	0.214
穏やか	0.269	0.005

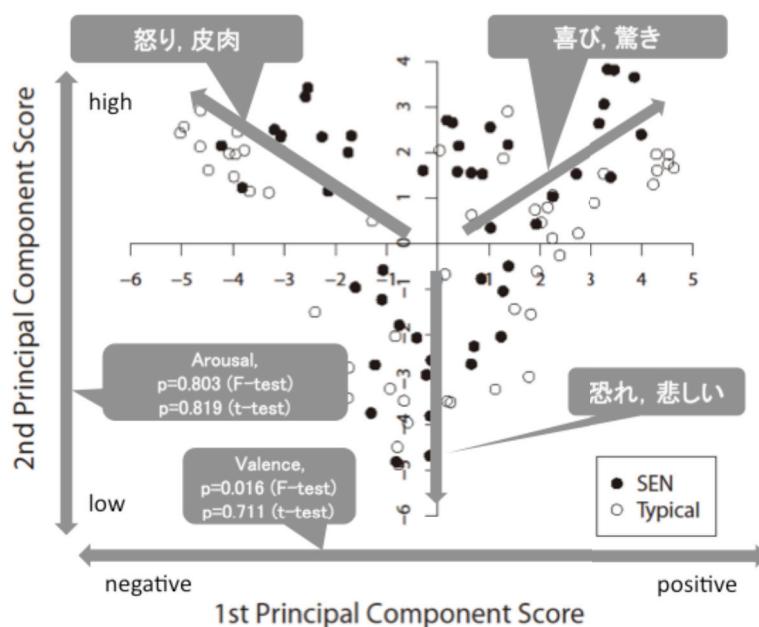


図 5-1 SEN (パイロット版) と Typical の心理量の主成分得点の比較

は第一主成分得点と第二主成分得点の散布図である。散布図は、中心から大きく分けて 3 方向に分散し、図中に感情語で示したような散らばりの傾向が見られた。わずかに、SEN のほうが内側に分散しているような傾向が確認されたので、平均値の検定および分散の検定を行ったところ、第一主成分では $p=0.711(t\text{-test})$, $p=0.016(F\text{-test})$ となり、分散値が有意に異なることが分かった。第二主成分では、

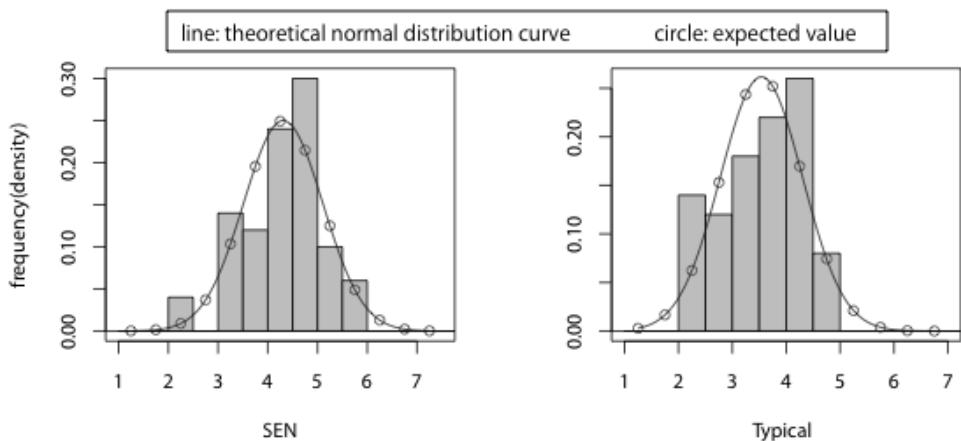


図 5-2 SEN (パイロット版) と Typical の自然性の印象評価値の比較

$p=0.819$ (t-test), $p=0.803$ (F-test)となり、いずれも有意差は確認されなかった。このことより、SEN は、Typical と比較して、感情値（ポジティブ—ネガティブ）の軸で、強度の低い感情表現を得られやすいことが分かった。

5.1.3.4 自然性の評価値の比較

図 5-2 は、自然性の評価値のヒストグラムおよび正規分布曲線である。ヒストグラムの値は 0.5 間隔で 1.0 から 7.0 までを範囲とした。比較のため、平均値の差の検定を実施したところ、 $p=0.063$ となり、5%水準で有意傾向であった。

5.1.4 音響的特徴の分析

音響的分析の対象は、印象評定を実施したものと同じデータ、SEN と Typical 各 50 個とした。音響特徴量として、各音声の発話全体に対する基本周波数 (F_0 、平均／標準偏差／最大値／最小値／レンジ)、パワー(平均／標準偏差／最大値)、そして発話長を求めた。そして、SEN と Typical を混合した 100 音声に対して主成分分析を施した。表 5-4 は、その結果得た固有ベクトルであり、第一主成分はパワーの大きさおよび F_0 の高さ（寄与率 60.8%）、第二主成分は発話内の F_0 の変動（寄与率 14.3%）を主に示す成分であると解釈した。図 5-3 は、SEN および Typical の主成分得点の散布図である。第一・第二主成分軸上で、SEN と Typical の平均値の差の検定を実施したところ、第一主成分では $p=0.037$ 、第二主成分では $p=0.0013$ となり、ともに有意であった。SEN は、Typical に比べて発話内の F_0 変動が大きく、パワーや F_0 の最大値の値が低くなる傾向が確認された。

表 5-4 SEN と Typical の音響特徴量に対する主成分分析結果の固有ベクトル

音響特徴量	第 1 主成分	第 2 主成分
基本周波数 (平均)	-0.38	-0.22
基本周波数 (標準偏差)	-0.33	0.52
基本周波数 (最大値)	-0.39	0.13
基本周波数 (最小値)	-0.25	-0.57
基本周波数 (レンジ)	-0.33	0.51
パワー (平均)	-0.38	-0.15
パワー (標準偏差)	-0.37	-0.17
パワー (最大値)	-0.37	-0.11
発話長	0.05	0.11

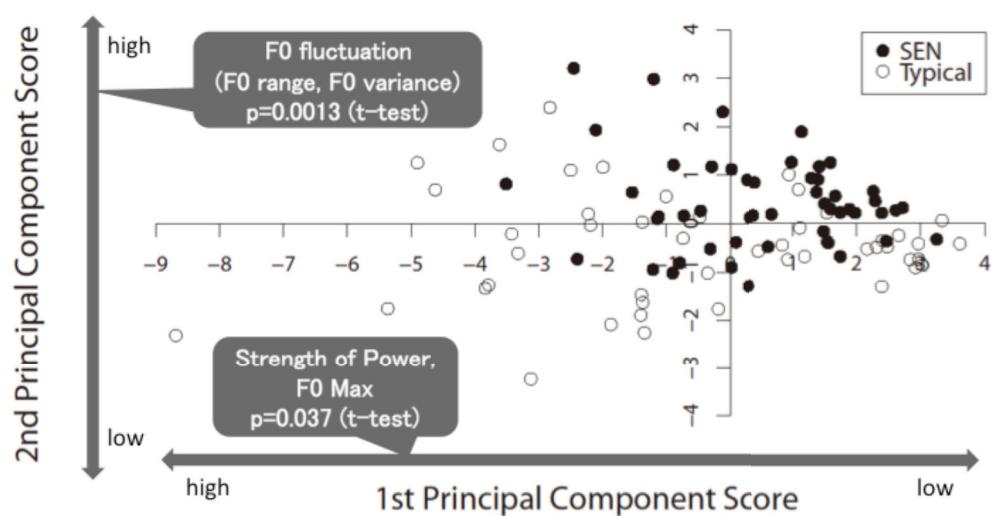


図 5-3 SEN (パイロット版) と Typical の音響特徴量の主成分得点の比較

5.1.5 考察

心理特徴量および音響特徴量の主成分分析の結果を整理すると、SEN は Typical よりも感情の強度が低い表現が多く得られ、その要因として、パワー最大値や F_0 の最大値の低さが考えられる結果となった。一方で、SEN は発話内での F_0 の変動値が大きく、それを活かした表現がなされている可能性が示唆された。また、自然性に関するSEN のほうが高くなる傾向が見られた。

しかし、これらの結果は話者 A の 1 名による表現の結果であり、台本も主観的に作成したものである。従って、次の収録では、話者および台本の拡充を行うことにした。

5.2 本実験版音声の収録と評価

パイロット版で、本手法の有効性がある程度確認できたため、つぎに、本実験版として、話者と台本数を拡大し、パイロット版で確認されたような傾向についてあらためて検証する。

5.2.1 台本インスタンスの作成および演技者・発話内容の選定

台本のフォーマットは、パイロット版同様表 4-2 の形式を採用し、インスタンスは、パイロット版で用いたインスタンスの項目を組み合わせて作成した。具体的には、できるだけ音声表現が多様となるように、「話し手」のパターン 10 通り（表 5-5）および「発話時の背景」のパターン 28 通り（表 5-6）を選択し、組み合わせて 280 個を用意した。一部の「発話時の背景」については、試験的に、感情の強度を 3 段階指定することにした。インスタンスのその他の項目は、パイロット版で用いた台本で用いられていたもののなかから違和感がないものを適当に選んだ。

発話内容は「ああ、そうですか」であり、パイロット版と同様である。

演技者は、パイロット版の話者である A に加え、B～E の 4 名を加えた計 5 名であるが、同一のインスタンスを用いることができるよう、全員女性に統一した。演技者 B～E は劇団員として舞台上での演技のキャリア（5 年～10 年程度）が主である。本実験ではまず、A～C の 3 名分の音声を分析対象とする。

5.2.2 音声収録

5.2.2.1 収録環境・手順

収録環境は、5.1.2.1 節で示したパイロット版の環境と同様である。

収録手順については、パイロット版の方法に対して工夫を施した。パイロット版の収録では、音量がレベルオーバーしてしまったり、小さすぎたりしてしまい、音質に問題があった。そのため、その改善策として、音声収録の開始時に、発話者の練習と収録機器の音圧設定の調整を兼ねて、「平静」「遠くの相手に対して話しかけるように」「ささやきかけるように」の 3 条件を用紙で呈示し、「ああ、そうですか」を発話してもらった。その後、収録の全体像を把握してもらうため、全ての台本を印刷した用紙を配布し、ひと通り目を通してもらった。

表 5-5 本実験版音声の台本の「話し手」の一覧

年齢	職業・続柄	人物像
10 代前半	中学生	大人びてしっかりした子
10 代後半	高校生	冷静だが、恋愛に対して鈍感
20 代	芸人	物静かだが、我が強い
	CA 志望 の学生	自分の感情を押し殺して、 中に溜め込んでしまうタイプ
	アナウンサー	清楚をうりにしている
	料理人	お調子もので体育会系のノリ
	アイドル	明るく活発だが天然
30 代	看護士	仕事を淡々とこなす人
40 代	主婦	自己中心的で自分の意見を貫く
	タレント	自由人で飘々としている

表 5-6 本収録の台本の「発話時の背景」の一覧

感情語	強度	発話時の背景
呆れる	普通	周りの騒ぎっぷりを呆れて見るように
恥	普通	憧れの人を前に、自分の恥ずかしいをことを友人に言われ、それを慌てて止めるように
激昂	普通	2人の友人が喧嘩している様子を見て、失望し、激昂するように
拒絶	普通	相手がしつこく喋ってくることに対し、語気を強めて、遮るように
恐怖	普通	雰囲気の恐ろしい大先輩の事を恐怖するように
傷つける	普通	鼻についた女の子に対し、優しいふりをして、傷つけるように
嫌味	普通	鼻を高くし、あからさまな不満をこめて、嫌味っぽく
無関心	普通	相手の空回りの元気に対し、無関心な様子で語りかけるように
疑い	普通	相手に紹介されたモノの信頼性を大げさに疑うように
安心	普通	久しぶりの秋晴れにほっとするように
怒り	普通・強・弱	日々の努力が報われない自分への怒りを吐露するように
恐れ	普通・強・弱	自分の身の回りで殺人事件が起こり、まだ犯人が逮捕されていないという恐怖を伝えるように
嫌悪	普通・強・弱	セクハラを繰り返す人に対する嫌悪感を相手に伝えるように
悲しみ	普通・強・弱	大好きな恋人との別れを悲しむように
喜び	普通・強・弱	自分の作った料理に対し、お世辞でもおいしいと言ってくれた事に対し、喜ぶように
驚き	普通・強・弱	新商品のお菓子(普段目にしてないもの)を食べ、意外にもおいしく感じビックリするように

収録開始以降の流れは、5.1.2.1 節で示したものと同様である。

5.2.2.2 比較用音声の収録

本実験版についても、比較用音声を用意した。その内容はパイロット版の Typical と同様である。演技者は、B および C の 2 名であり、発話内容および収録環境も同様である。話者 A の Typical に関しては、パイロット版を収録した際のものと同じものを用いた。

80 語の収録を、休憩を挟んで 2 回実施し、1 人あたり 160 個のデータを得た。

5.2.3 分析用データの選定

SEN は 5 名分で 1400 音声、Typical は 3 名分で 480 音声収録しているが、そ

これら全てに対する印象評定は実験規模が非常に大きく実施が困難である。したがって、分析対象の音声を特定の規則にしたがって選別し、それらに母集団を代表させる。絞り込むにあたり、できるだけ多角的な分析ができるよう、話者を A～C の 3 名に限定したうえで次の 2 通りの方法で選別を行う：

パターン 1：SEN と Typical からそれぞれ同数を無作為あるいは作為性の低い方法で抽出

パターン 2：SEN と Typical の刺激において、「怒り」「喜び」など、感情語によって区分可能なものを抽出

つぎに、それぞれの抽出方法について詳述する。

5.2.3.1 パターン 1：無作為による抽出

SEN の A～C の 3 話者の各 280 発話から各 50 発話、Typical の A～C の 3 話者の各 160 発話から各 50 発話を、無作為あるいはそれに準ずる方法で抽出した。

SEN の抽出は、完全無作為抽出を行う。A～C の 3 話者でインスタンスの内容は同一である。

Typical の抽出は、パイロット版の分析時と同様、表 5-2 の「◎」の印が付いた 7 セットの各 50 発話である。

ただし、音響的特徴の分析に際しては、上述の各 50 発話に加え、各 110 語 (SEN は無作為) を加えた合計 160 音声を対象としている。

5.2.3.2 パターン 2：感情語をキーとした抽出

SEN と Typical から、感情表現のカテゴリが同じと思われる表現を抽出する。具体的には、「怒り」および「喜び」の表現をそれぞれ抽出する。

SEN は、表 5-6 における、「感情語」が「怒り」および「喜び」に属するものを抽出した。各カテゴリにおいて、1 話者分あたり 30 音声がそれらに該当する。

Typical は、表 5-2 のすべての感情語セットのなかから、「怒り」および「喜び」に属すると思われるものを抽出した。「怒り」は 11 音声の収録 2 回分で合計 22 個、「喜び」は 8 音声の収録 2 回分で合計 16 個が該当する。

5.2.4 心理的特徴の分析

本章では「多様な感情表現」を満たす要件のうち、心理的特徴について検証する。まず、印象評定の手順と結果について述べたのち、評定結果の分析をおこなう。

5.2.4.1 評価尺度

パイロット版と同様に、森山らによる感情表現語 9 語を用いる。

5.2.4.2 評価の流れ

20 代の男女学生各 8 名、合計 16 名に評価を依頼した。評価はアンケートによって実施し、5 章で選んだ、SEN と Typical 合わせて 583 個の音声（重複を除いた個数、のべ 684 個、欠損データ 1 個あり）をランダムに呈示した。

感情表現語については、それぞれの評価語に対して、「全く当てはまらない(1)～どちらでもない(4)～非常に当てはまる(7)」の 7 段階を設定した。評価結果は間隔尺度として扱うことにし、12 人の平均値を求め、それを音声に対する最終的な評価値とした。

なお、パイロット版同様、Web 上でアンケートを実施した。パイロット版と比較して評価の規模が大きいため、1 人あたり 8~12 時間程度の時間を要した。評価の精度が低下しないよう、休憩をよく挟むことを強調して指示した。

5.2.4.3 パターン 1 の分析

まず、パターン 1 の評価結果の比較を行う。印象評価値に対して主成分分析を行い、主成分軸上の比較において、両者での違いを確認した。

表 5-7 に、第 3 主成分までの寄与率と固有ベクトルを示す。累積寄与率は、第 2 主成分までで 74.3% と高い値を示した。固有ベクトルを確認すると、第 1 主成分は、正の方向にポジティブさ、負の方向にネガティブさが現れていることが見て取れる。すなわち、感情価 (valence) の軸であると言える。第 2・第 3 主成分は、一般的に覚醒 (arousal) と知られている概念が二つの成分に分かれて現れていると考えられる。これらの傾向は、パイロット版とほぼ同様であった。

図 5-4 は、パターン 1 における話者 A から C における第 1、第 2 主成分のカーネル密度プロットである。カーネル関数として Gaussian カーネルを選び、バンド幅は Silverman[81] の手法に基づき決定した（本稿における以降のカーネル密度分布はすべて同様の設定である）。話者 A の第 1 主成分、および話者 C の第 2 主成分で、Typical の密度が低い中央付近を、SEN が埋めている。すなわち、SEN では、個人差があるものの感情価や覚醒の強度の強いものから弱いものまで万遍なく得られることがわかる。

表 5-7 印象評定値の主成分分析結果（寄与率・固有ベクトル）

Words (寄与率)	PC1	PC2	PC3
	0.47	0.27	0.10
怒り	-0.43	0.23	-0.16
喜び	0.42	0.22	0.01
皮肉	-0.37	0.34	-0.06
恐れ	0.02	-0.55	-0.31
悲しみ	-0.16	-0.54	0.01
驚き	0.33	-0.13	-0.63
こび	0.42	0.11	-0.07
穏やか	0.35	-0.11	0.61
おかしい	0.24	0.38	-0.29

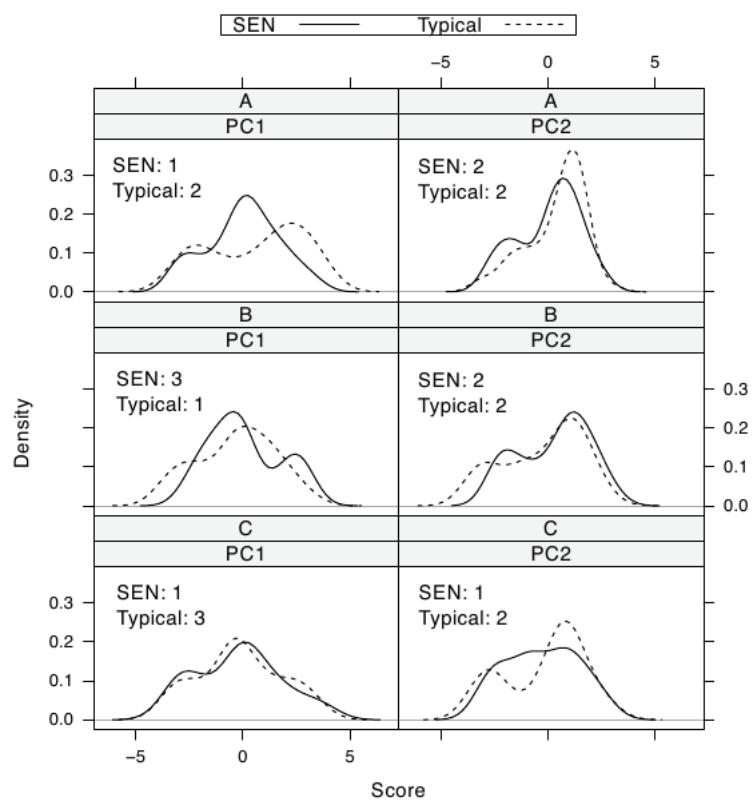


図 5-4 心理特徴の各主成分の比較（パターン 1）

表 5-8 図 5-4 における各クラスタ内の主成分得点の標準偏差値
(括弧内はクラスタ数)

話者	主成分	SEN		Typical	
		全体	クラスタ平均	全体	クラスタ平均
A	PC1	1.701	1.701(1)	2.222	0.957(2)
	PC2	1.456	0.652(2)	1.378	0.899(2)
B	PC1	1.676	0.480(3)	1.829	1.829(1)
	PC2	1.658	0.761(2)	1.858	0.883(2)
C	PC1	2.000	2.000(1)	1.957	0.626(3)
	PC2	1.698	1.698(1)	1.857	0.745(2)

この傾向を確認するため、各分布に対して、混合正規分布のクラスタ数の自動推定を行った。クラスタリングは EM アルゴリズムを用い、最もベイズ情報量規準値が小さいモデルを最適と見做した。ここで、クラスタ数が多いほど表現の分布に偏りが生じていると仮定する。推定の結果を、図 5-4 上に数字で示す。話者 A の第 1 主成分、話者 C の第 2 主成分において、SEN の母集団は 1 つの正規分布、Typical の母集団は 2 つの正規分布からなることが分かり、前述で視覚的に確認した傾向を裏付けている。その他に関しても、話者 B の第 1 主成分以外は、SEN と Typical のクラスタ数は同等か、SEN のほうがクラスタ数が少ないことが分かった。

また、主成分得点の標準偏差を求め、各クラスタがどの程度幅広い表現の音声が含まれているかを確認した。表 5-8 は、話者および主成分ごとの、データ全体の標準偏差値および各クラスタにおける標準偏差値の平均値である。SEN は Typical に比べて全体的に標準偏差値が低く、それぞれの軸で強度の低い音声表現が多いことが分かる。また、Typical は標準偏差値が比較的高い一方、クラスタ数も多く、SEN と比べて表現にやや偏りがあると解釈できる。たとえば、話者 A の第一主成分および話者 C の第二主成分では、全体では Typical の標準偏差値が大きいが、クラスタごとに見ると、SEN は強度が低い表現を中心とした幅広い一つの分布、Typical は強度が高い表現を中心とした比較的狭い二つの分布からなることが分かる。これらの結果は、前述した図 5-4 で視覚的に確認できる傾向を裏付けるものである。

5.2.4.4 パターン 2 の分析

図 5-5 および図 5-6 は、パターン 2 における「怒り」および「喜び」の評価値

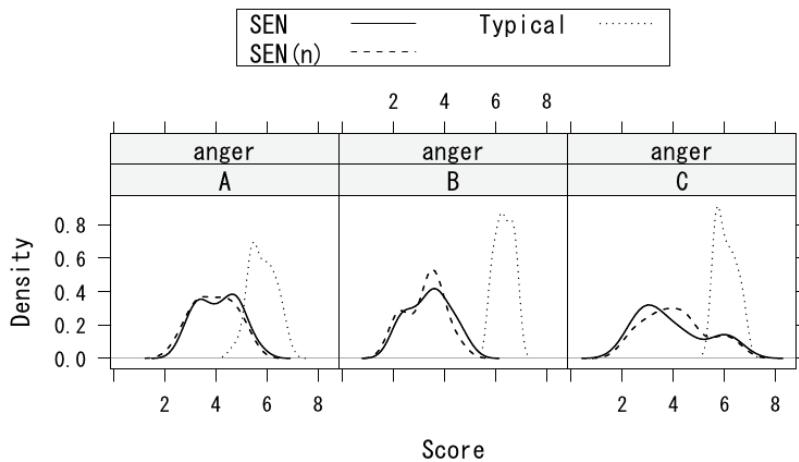


図 5-5 「怒り」の印象評定値の分布の比較

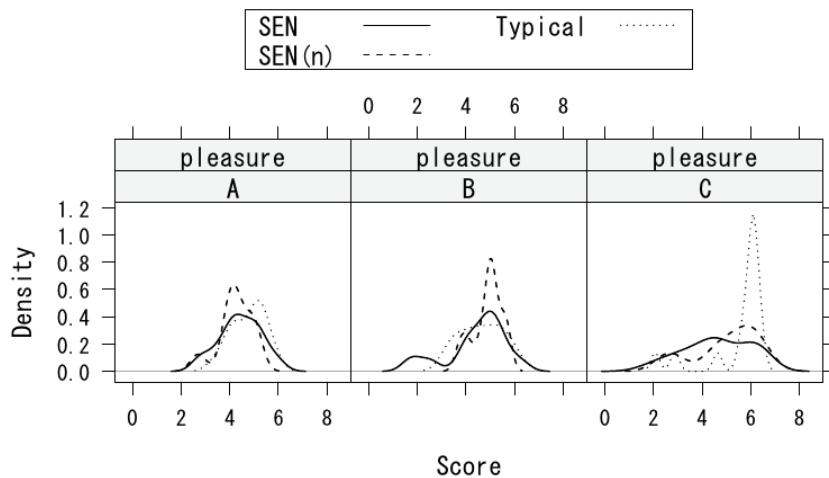


図 5-6 「喜び」の印象評定値の分布の比較

の実測値のカーネル密度プロットの比較である。SENに関しては、「怒り」「喜び」の強度が3パターン（各10音声）あり、そのままではTypicalと対等な比較ができないため、強度が「普通（表5-6参照）」の10音声のみの密度プロットをSEN(n)として示してある。「怒り」の場合は話者3名分において、「喜び」の場合は話者Cにおいて、Typicalが強度の高い部分に極端に偏っているのに対して、SENは幅広い感情強度の分布を示していることが分かる。また、これらのSENの傾向は、SEN(n)でも大差なく見られている。

Typicalの分布とSEN及びSEN(n)の分布の違いを確認するため、各話者の2感情について、それぞれ分布の適合度の検定、分散の検定、及び平均値の検定を

実施した。ここで、説明変数はデータの種類 (Typical, SEN, SEN(n)), 目的変数は話者ごとの「怒り」および「喜び」の印象評価値である。

コルモゴロフ・スミルノフ検定の結果、話者 A の anger, 話者 B の anger, 話者 C の anger で、Typical と SEN, Typical と SEN(n) の分布が有意に異なっていた（それぞれ、 $p < 0.001$, ボンフェローニの補正済み）。SEN と SEN(n), 及び全ての話者の pleasure では有意差はなかった。

また、F 検定の結果、話者 B の anger, 話者 C の anger で、SEN と SEN(n) の分散は有意に Typical の分散よりも大きかった (B(anger)について, SEN-Typical: $F=5.65$, $P < 0.001$, Typical-SEN(n): $F=0.22$, $P < 0.01$ 。C(anger)について, SEN-Typical: $F=11.62$, $P < 0.001$, Typical-SEN(n): $F=0.10$, $P < 0.001$ 。ボンフェローニの補正済み)。SEN と SEN(n), 話者 A の anger, 及び全ての話者の pleasure では有意差はなかった（ただし話者 A の anger では Typical と SEN が有意傾向）。

平均値の差の検定（等分散では分散分析、異分散ではウェルチの法）の結果、話者 A の anger, 話者 B の anger, 話者 C の anger で、Typical と SEN, Typical と SEN(n) の平均値の差が有意に異なっていた（それぞれ、 $p < 0.001$ ）。SEN と SEN(n), 及び全ての話者の pleasure では有意差はなかった（ただし話者 C の pleasure では Typical と SEN, Typical と SEN(n) の差が有意傾向）。

以上の結果は、「怒り」において Typical の分布形状が、SEN 及び SEN (n) の分布の形状とは異なっており、この異なりは SEN 及び SEN (n) の分散の広さによることを示している。一方「喜び」においては、一部の平均値の差が有意であったことを除き、Typical と SEN, SEN (n) の分布の差は見られなかった。このように、「怒り」「喜び」とともに分布の形状を裏付ける結果が得られた。

5.2.4.5 心理的特徴のまとめ

SEN と Typical による心理的特徴の比較を大規模に行った結果、あらためて、4.2 節で示した「多様な感情表現」の要件を満たすために有意義な結果が得られた。パターン 1 の分析においては、感情音声で一般的だとされる主成分である感情価や覚醒の軸において、SEN は強度の強いものから弱いものまで万遍なく得ることができ、パターン 2 の分析においても、「喜び」と「怒り」共に、SEN は幅広い感情強度の音声が得ることができた。

5.2.5 音響的特徴の分析

パイロット版同様に、「多様な感情表現」を満たす要件のうち、音響的特徴に関して検証する。ただし、特微量の選択に関しては様々な観点がある。本節では、

表 5-9 音声数 160 の時の各特徴量の標準偏差値の比較

	SEN	Typical
データ数	160	160
F_0 標準偏差	話者 A	0.89
	話者 B	0.90
	話者 C	1.01
発話長	話者 A	0.70
	話者 B	0.67
	話者 C	1.13

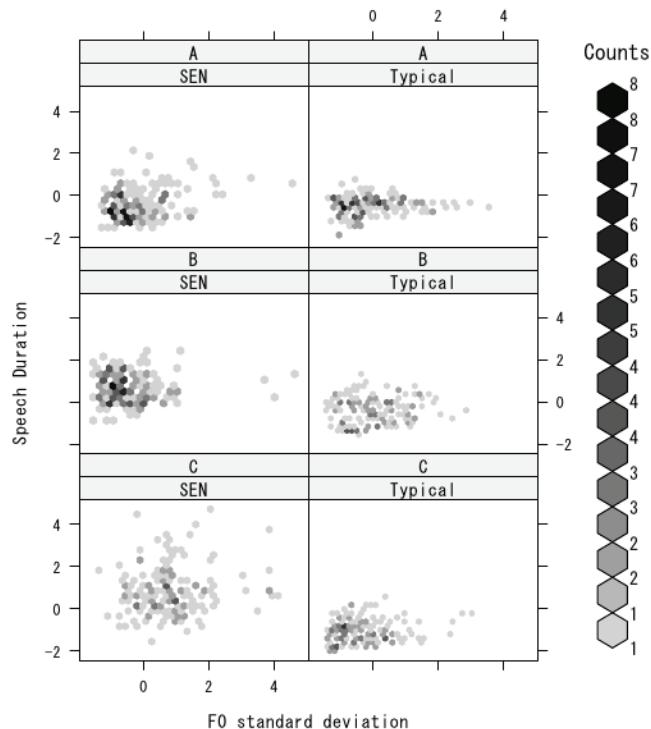


図 5-7 各特徴量による散布図の比較（パターン 1, 各 160 音声）

Erickson[71]が整理した、先行研究で感情研究に重要とされる様々な特徴量とともに検討した結果、 F_0 の変動量を表す特徴として発話単位における F_0 の標準偏差、および発話長に着目することにした。 F_0 の標準偏差を扱うのは、 F_0 の最大値やレンジなどと比べて F_0 の誤抽出の影響を受けにくい特徴量と考えたからである。また、SEN と Typical すべてを対象に、各特徴量間のスケールを合わせるための正規化を施す。

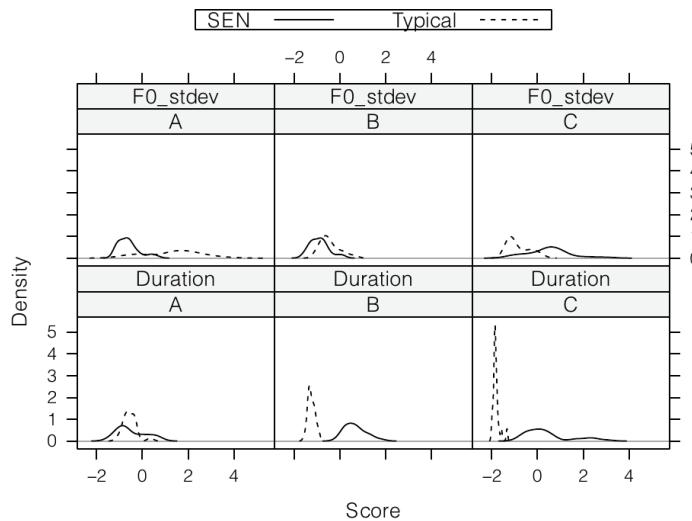


図 5-8 「怒り」の音響特徴量の分布の比較（パターン 2）

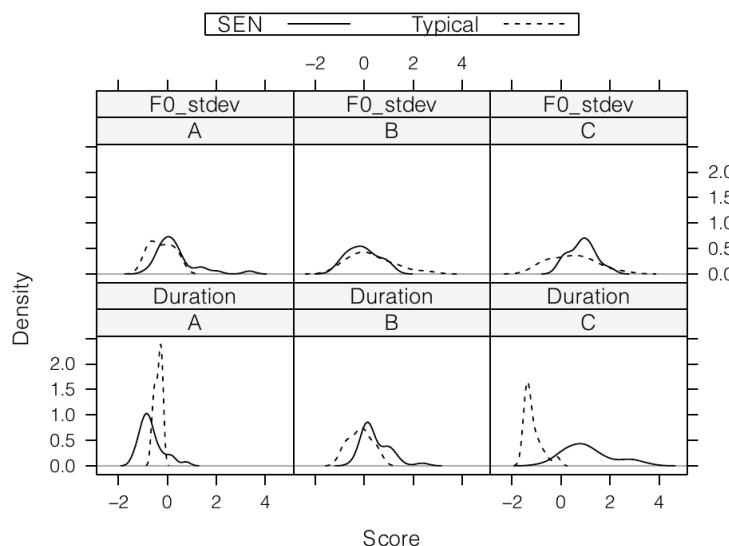


図 5-9 「喜び」の音響特徴量の分布の比較（パターン 2）

5.2.5.1 パターン 1 の分析

心理的特徴の分析のパターン 1においては、印象評定を行う都合上、各 50 音声を分析対象としたが、ここではパターン 1 で無作為抽出した 50 音声に、さらにそれぞれ 110 音声を加え、Typical で利用できる最大の数である 160 音声を対象とする（SEN は無作為抽出）。

まず、表 5-9 は、SEN および Typical の各 160 音声の、「 F_0 の標準偏差」お

より「発話長」の標準偏差の比較である。話者 A の F_0 の標準偏差を除き、全体的に Typical に比べて SEN の分布が幅広いことがわかる。

次に、図 5-7 は各 160 音声の F_0 の標準偏差および発話長を、Carr et al.[82] が提案した六角形を用いた散布図で示したものである。ここで、同一話者間の SEN と Typical の分布の差異について検定を実施したところ、話者 B・C の両方の特徴量で SEN と Typical 間の差異が有意であることが分かった。話者 A の発話長についても、コロモゴロフ・スミルノフ検定で Typical は $p=0.35$ となり正規性があるのに対し、SEN は $p=0.02$ となり正規性がなく、分布の形状は異なることが分かった。これらの結果は、SEN は Typical と異なる音響的表現が表れやすいことを示していると考えられる。

5.2.5.2 パターン 2 の分析

図 5-8 および図 5-9 は、「怒り」および「喜び」のラベルが付与されたデータにおける、各音響特徴量のカーネル密度プロットである。「怒り」「喜び」とともに発話長の分布は SEN は広がっており、 F_0 の標準偏差に関しても、「怒り」の話者 C、「喜び」の話者 A において SEN は広がっている。

等分散および平均値の検定を実施したところ、いずれにおいても SEN と Typical 間の分布は不等分散あるいは平均値が有意であり、分布の形状が異なることが示された。

5.2.5.3 音響的特徴のまとめ

F_0 の標準偏差および発話長を対象に分析を行った。いずれの話者においても SEN は Typical と異なる分布を示すことが分かったが、特に発話長において、SEN は Typical よりも幅広い値を得られていることが分かった。4.2 節で示した「多様な感情表現」の要件を満たすために有意義な結果があらためて得られた。

5.2.6 考察

パイロット版で確認できた提案方法の効果を、話者および台本を大幅に確認し、分析の手続きを見直しながら再確認を行った。話者ごとに差が大きいが、それを前提として整理すると次のようになる：

- 「喜び」「怒り」の感情表現の幅を音響的・心理的に大きく拡張させる
- 感情価や覚醒の軸においても、典型的な感情語を刺激とした音声にあらわれにくく强度の低い表現を得ることができる
- 音響的多様性に関しては、SEN は発話長が Typical に比べて値の分布が幅広く、 F_0 標準偏差も部分的に同様の傾向である

以上の結果より、適切な演技者に我々が設計した演技指示である台本を呈示すれば、4.2 節で示した要件を備えた「多様な感情表現」を獲得できることが示された。

一方で、台本の効果の確認方法について、検討すべき課題が存在する。今回の実験では、台本の項目（「発話時の背景」「聞き手」「話し手」など）すべての効果を同時に確認している。したがって、特定の項目のみの効果を確認できるような実験を行うことで、Typicalに対するSENの効果がより明確になることが期待できる。また、音響的特徴と心理的特徴の相互関係を調査することで、今回の実験で確認できた音響的多様性は、純粋に感情表現の多様性と強く関連しているのか、あるいは、「聞き手」「話し手」などの話者性による表現の違いが関連しているのかを明らかにすることができる。

5.3 拡張版音声の収録と評価

SEN（パイロット版）およびSEN（本実験版）の心理的・音響的分析によって、本手法の適用による自然性の向上および心理的・音響的多様性向上の可能性が示唆された。とくに、5.2.5 節で示した、SEN（本実験版）の音響的特徴において話者ごとの違いが顕著に確認されたことは、本研究における要点だと考えられる。我々は、次の課題として、演技者のカテゴリや発話内容の違いによる多様性の広がりの差異について調べることにした。この目的のため、SEN（拡張版）では、発話者（演技者）および発話内容を精密に選定することにし、さらに、表現豊かな発話を促すため、台本の改善も試みた。

5.3.1 台本インスタンスの作成および演技者・発話内容の選定

5.3.1.1 発話者カテゴリの再考察

これまでの収録では、声優や舞台俳優としてのキャリアを持つ人物を主観的に選定してきた。しかし、SEN（本実験版）までの収録した音声を聴取し考察するなかで、同じ声による演技のプロフェッショナルでも表現方法の違いはかなり大きく、本手法における音声表現の多様性に大きく左右することが分かつてき。そこで、いくつかの仮説を考えた。

仮説1：声のみによる表現と複数モダリティによる表現では表現方法が異なる

仮説2：聴衆者をどのように意識して演技するかで表現方法が異なる

これらの仮説に基づいて、我々は声を用いたプロフェッショナルの演技者のカテゴリ（発話者カテゴリ）として、次の区分を定義した：

- ・声優：声のみを用いた表現者、観客は意識しない
- ・舞台系俳優：複数モダリティを用いた表現者、観客を意識する
- ・映像系俳優：複数モダリティを用いた表現者、観客を意識しない

さらに、3 カテゴリに関して、次の仮説を考えた。

仮説3：声の表現の多様性は、映像系俳優、舞台系俳優、声優の順に高くなる

仮説4：声の表現の自然性は、声優、舞台系俳優、映像系俳優の順に高くなる

本章では、仮説 3 における音響的多様性の確認をおこなう。声優は、声のみで多様な表現を求められるため、最も多様性が高いと想像できる。また、舞台系俳優と映像系俳優は、観客に対して演技することを意識するかどうかで、表現の度合いの大きさに差が出る可能性がある。たとえば、舞台系俳優は観客を意識した演技をすることに対して、映像系俳優は観客を意識せず、日常生活に近い演技を行っていると考えられる。なお、観客の存在を意識して演技する点の重要性は、先行研究[77]で述べられている。

以上のように、キャリアに基づく演技手法の質の違いにより、本手法による同一の手続きに従った演技でも、表現の広がりの差異を生み出すことが期待できる。また、その差異は、本研究の目的である、表現の多様性の獲得に強く関連するであろう。

さらに、男女による表現の違いも検討する。これまでの収録では主に女性を中心収録を実施してきたが、基本周波数の特性の違いを考慮すると、女性の音響的な表現力が豊かである可能性が高い。これを検証するため、SEN（拡張版）では、各発話者カテゴリについて、男女を別カテゴリとして扱い、計 6 カテゴリの話者を選定することにした。

表 5-10 SEN（拡張版）における発話内容の詳細

セット	発話内容	極性	音声単語 親密度	アクセント 型	末尾の音素列
セット1 3モーラ 形容詞	うまい	Positive	6.312	2	/ai/
	からい	Neutral	6.250		
	いたい	Negative	6.312		
セット2 5モーラ 形容詞	すばらしい	Positive	6.344	4	/asii/
	いそがしい	Neutral	6.156		
	むずかしい	Negative	6.031		

5.3.1.2 オーディションの実施

演技者の選定は、オーディション形式を採用することにした。手続きとして、12通りの台本サンプルを用意し、発話を録音してもらったうえ返送してもらうよう依頼した。台本サンプルの形式は、SEN（拡張版）で用いた台本と同一であり、発話内容は「ああ」と「そうですか」の2種類（各6通り）を用意した。「ああ」を選んだ理由は、極めて短い発話でも表現力が豊かであるかどうかを確認するためであり、「そうですか」を選んだ理由は、既存のSENのデータと比較しながら多様性を演出できる人物かどうか検討するためである。

依頼に際しては、声優・舞台系俳優・映像系俳優というカテゴリライズを良く理解でき、かつ各カテゴリに対して人脈を持つ人物を仲介人として起用した。仲介人経由で、声優は特定の事務所、舞台系俳優は特定の劇団、映像系俳優は特定の映画監督に依頼し、本研究の主旨を伝えたうえで、適正があると判断したもらった人物を各複数名選定してもらった。

最終的には、サンプル音声を聴き比べ、筆者らの主観で表現力の多様性が同程度であると思われる人物を、各カテゴリから男女各一名選定した。

5.3.1.3 発話内容のデザインの拡張

発話内容は、SEN（本実験版）までは中立的な内容であることを条件としてきた（「ああ、そうですか」）が、5.2.4節で述べたSEN（パイロット版）の心理的特徴の傾向の考察から、提案手法は感情価の強度の意識的なコントロールが重要であると考えられるため、発話内容の（感情価の）極性の違いが多様性に及ぼす影響の検証、さらに発話内容の極性毎による音響パラメータと心理パラメータの差異の検証が必要であると考えた。

また、発話の長さによっても多様性に違いが生まれる可能性がある。理想としては、どのような長さでも同様の多様性が確保できることが望ましいが、これま

表 5-11 変更前後の「発話時の背景」の例

SEN (パイロット版)	まだ見ぬヒーローをカッコいいと妄想し、憧れるように
SEN (本実験版)	
SEN (拡張版)	学校から帰る電車の中。A は、手帳とにらめっこしながら今後数週間は一日も空かずに予定が埋まっていることを改めて確認する。それを見た瞬間、先を考えると過酷・・・と思いながら出た一言。

表 5-12 インスタンスに付与した属性一覧

属性	内容	個数
話し手の年代	10,20,30-40,50-60 代	各 25 通り（計 100 通り）
聞き手の年代	上記なし（独り言）	話し手と同じ年代の場合：10 通り 異なる場合：4 通り、独り言：3 通り
聞き手の性別	同姓、異性	聞き手の各年代で半数ずつ（5 通り、2 通り）

での収録の過程で、複数の演技者より、短すぎる発話はバリエーションのある演技は難しいという意見が得られているため、これも合わせて検証できるような発話内容を選択することにした。

これらの観点にしたがって、発話内容の極性および長さを複数用意することにした。極性は、先行研究[83][84]で提案された極性評価が記された辞書に従い、Positive/Negative/Neutral をそれぞれ選ぶことにした。また、発話の長さに関しては、モーラ数およびアクセント型を揃えることにした。さらに、全ての発話において、BPM などの重要な韻律情報の比較ができるよう、末尾の音韻が統一されるようにした。その他、音声単語親密度[85]や品詞の統一なども考慮した。

以上の条件に従い、表 5-10 に示した 6 種類の発話を収録することにした。

5.3.1.4 台本の改善

① 日常性と非日常性

SEN (パイロット版) では、台本のフォーマットに従い、インスタンスを TV 番組からランダムに抽出するという手法を取った。また、SEN (本実験版) では SEN (パイロット版) から抽出した項目の組み合わせで台本のインスタンスを拡張した。

SEN (パイロット版) および SEN (本実験版) での問題点のひとつとして、Typical に比べて自然性は向上したと考えられるものの、やはり非現実的な場面、

すなわちアニメーション等の創作物やバラエティ番組で聞くことができるような、やや偏った表現が多いと思われる点であった。我々の目的は、そのような非現実的な場面での表現も、我々が日常生活下で聴取できるような表現も全て収集できる手続きを確立することにある。

そこで、台本のインスタンス生成の際、意識的に「日常的な」インスタンス、「非日常的な」インスタンスの両方を生成することにした。具体的には、一つの発話につき、日常的な場面を意識した台本を 100 本、非日常的な場面を意識した台本を 100 本用意することにした。非日常的な場面とは、SEN（パイロット版）で用いたような、テレビ番組でのワンシーンの再現であり、日常的な場面とは我々が日常生活下で遭遇するワンシーンの再現である。これらを今回は完全に創作で作成した。これらのストーリーは、パイロット版および本実験版で利用された、対面対話、情報発信の発話、中継による遠隔対話に加え、聞き手が存在しない「独り言」を加えている。また、日常性や非日常性をより強調するため、次節で示すように発話時の背景の長文化を施した。台本の原案は発話内容 1 種につき各 200 本の合計 1200 本であるが、発話者の性別による違和感の無いよう改変し合計 2400 本を用意した。

② 台本フォーマット「発話時の背景」の長文化

SEN（拡張版）の台本フォーマットは SEN（パイロット版）と同様であるが、発話時の背景を長文化し、ショートストーリー仕立てにした。これは、これまでの収録で、発話時の背景の情報が少ないために強引に表現を広げようとし、不自然な音声表現を招く可能性があるという指摘を受けてのことである。変更前後の文章の例を表 5-11 に示す。長文化することで、より自然な表現、また、発話者カテゴリ毎に異なる多様な表現を導くことが目的である。

③ インスタンスの各項目への属性の事前付与

SEN（拡張版）では、インスタンス作成は完全ランダムではなく、日常・非日常の各 100 本について、話し手の年代・聞き手の年代および性別の組み合わせ（属性）を創作前に指定し、音声データへのクエリとして利用できるようにした。詳細を表 5-12 に示す。

5.3.2 音声収録環境と手順

国立国語研究所内のマルチメディアスタジオに、パソコンおよび USB オーディオデバイス(Roland UA-25)を設置した。マイクはヘッドセットタイプ(SHURE WH20XMR)を採用した。サンプリング周波数は 44,100Hz、ビットレートは 16bit

に設定した。また、バックアップとしてリニア PCM レコーダ(SONY PCM-M10)による録音を同時に行つた。さらに、将来的な検証のため、デジタルビデオカメラによる撮影を行つた。

収録は、約 1 ヶ月にわたり断続的に実施した。1 名あたり 1200 発話の収録を 4 回に分け、1 回あたりの時間は発話者の任意で休憩を挟みながら 3 時間程度であった。刺激の呈示はスタジオ内に設置したディスプレイを用いて行つた。呈示順は全員ランダムであったが、日常的な台本から始めたほうが演技しやすいとのコメントを受け、そのようにコントロールを行つた。以上の手続きで、合計 7200 音声の収録を完了した。

5.3.3 音響的特徴の分析

SEN (拡張版) の 7200 個の音声について、発話ごとの F_0 (平均・標準偏差・最大・最小・レンジ) および発話長関連 (発話長・平均モーラ長・合計ポーズ長) の特徴量を算出し、主成分分析を試みた。第一主成分は基本周波数関連、第二主成分は発話長関連の成分が強く含まれている。第二主成分まで累積寄与率は 72% であった。図 5-10 に、3 モーラ、5 モーラの単語ごと、極性ごと、そして話者ごとの計 36 通りの音響特徴量の密度プロット (各 200 個) を示す。

5.3.3.1 発話長ごとの分布の傾向

全体の傾向を確認する限り、3 モーラ発話と比較して 5 モーラ発話は分布が狭い傾向が見られた。仮説では、5 モーラ発話のほうが表現が豊かになると想えていたが、今回の主成分分析で用いた特徴量で判断する限りでは、3 モーラ発話のほうが表現が豊かであるような結果が得られた。これらの特徴量は、発話全体に対して計算したものとも単純な算出方法で求めたものだったので、短い発話ではそのような単純な特徴量で表現の多様性が演出された可能性がある。モーラ長による表現力の違いをより精密に検証するためには、BPM など局所的な音響特徴量の確認が必要である。

5.3.3.2 極性ごとの分布の傾向

すべての話者に関して、極性ごとに大きな差異は見られなかった。本手法が必ずしもこの傾向を必ず保証することではないが、少なくとも今回用意した台本では、極性に左右されず、多様な表現を生み出すことが出来たと考えられる。

5.3.3.3 話者カテゴリごとおよび性別による分布の傾向

発話者カテゴリおよび性別による表現力の比較であるが、これも仮説と反し、

声優と舞台系俳優については男性のほうが音響的多様性が高いことが確認された。特に女性声優については、収録中の表現の主観的印象は非常に幅広かったので、筆者らにとって意外な結果であった。一因として、この分析では用いていない音声パワーを良くコントロールして多様な表現を再現していたので、このような結果になったのだと推測される。全体を通しては、分布の違いは個人ごとの差が大きく、話者カテゴリや性別の仮定を明確に裏付ける結果を得るには至らなかった。

5.3.4 考察

今回起用した6名の話者が、各発話者カテゴリを代表する特徴を持つとは限らないが、全員が発話内容6種類のいずれもほぼ同様の分布を示したことは興味深い結果であり、この分布は各演技者の表現様式が可視化されたものではないかと推測できる。少なくとも、音響的多様性において、発話内容による影響は比較的小さいと考えられる。

また、男性映像俳優（MAM）の分布がいずれの発話においても一番小さい分布を示しているが、台本に対する整合性は他の話者カテゴリと同程度だと思われた。映像俳優は、非常に微妙な声の表現の際で台本の状況を再現しており、興味深いデータが得られた。

VAF	: 女性声優	SAF	: 女性舞台系俳優	MAF	: 女性映像系俳優
VAM	: 男性声優	SAM	: 男性舞台系俳優	MAM	: 男性映像系俳優

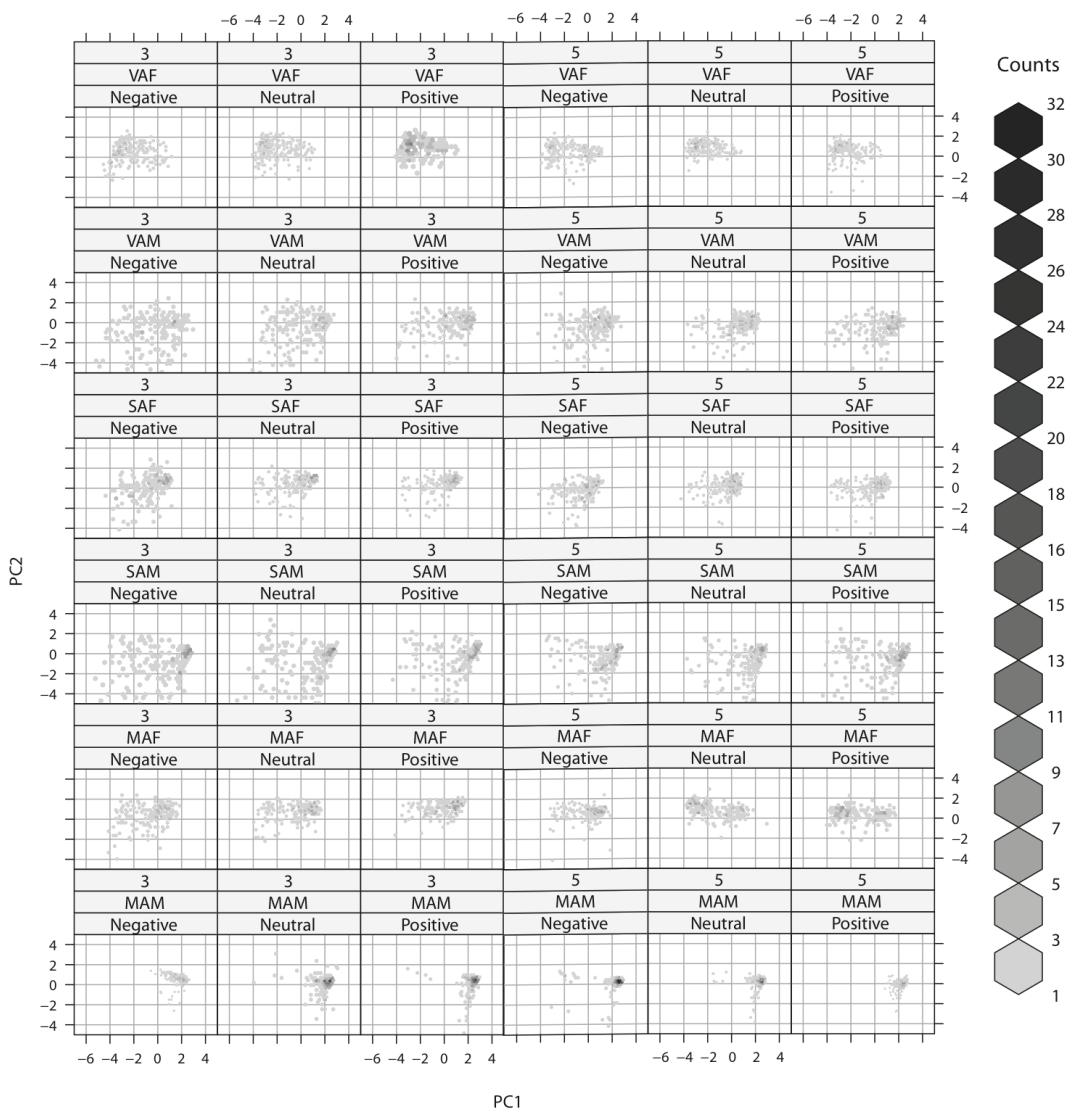


図 5-10 SEN(拡張版)の音響的特徴の第一、第二主成分の密度プロット(各200個)

5.4 議論

本節では、3回の音声収録実験の結果を踏まえて、本研究の目的である音声の表現力の測定を、提案手法によって実現することができたかを論じる。

本研究では、音声の表現力の測定を行うために、制約条件を設けたうえで多様な音声表現（表現豊かな音声）を収集する方法を検討した。多様な音声表現を得られたかどうかの判断基準を設定することは困難であることから、パイロット版および本収録版では、典型的な感情語から生成された音声データを比較対象として、比較対象に対してどのように多様であったかを相対的に調べる方法を採用した。拡張版の段階では、典型的な感情語との比較は十分であると判断したので、複数話者・話者カテゴリ間での比較を行い、話者間で表現方法が全く異なることを確認した。

これらの結果を踏まえて、演技音声に限定したうえで、音声の表現力の測定の可能性について考察する。提案手法では、拡張版の段階で、一人あたり1200個（1発話あたり200個）の台本を用意したが、図5-10で示したように、同一話者の異なる発話内容の音響的分布が相当似通うことが分かった。従って、台本の規模に関しては、最大でも200個程度あれば、その話者の表現力を測定するのに十分であるという判断ができる。現状では、話者あるいは話者カテゴリで表現方法が大きく異なることがほぼ判明しているため、台本数がどこまで削減できるかを検討したうえで、話者数をできる限り増やすことで、演技音声における表現力を測定することができるようになることが期待できる。ただし、拡張版の段階では心理的評価は実施していないため、心理的評価軸の見直しも含めて、大規模な心理実験を通してこの考察を再確認する必要がある。

また、本手法を適用するかぎり、あくまでも演技音声を扱うことになるため、あらゆる音声表現を得られるという断定はできない。しかし、これから心理実験で確認する必要があるが、収録の過程で実験者が受けた印象では、拡張版における男性舞台俳優の演技は自発音声に近く、今後多様な話者を拡大していくなかで自発音声に近い音声表現を得られる可能性は存在すると考えられる。

5.5 本章のまとめ

本章では、第4章で定めた手続きに則って、3回の収録実験を行い、それぞれの結果を述べたのち、音声の表現力の測定の可能性について論じた。

まず、パイロット版の収録では、1名の話者と主観的に作成した台本を用いた小規模な収録を行い、心理的・音響的な豊かさ、および自然性の観点で、少なくとも既存の演技音声より高品質な音声資料が得られることを確認した。

パイロット版での分析結果を踏まえて、2回目の収録を本実験と位置づけて、3名の話者、ある程度パターン化された台本を用いて、規模を拡大した実験を行った。その結果、提案手法による音声表現は、やはり典型的な感情語と比較すると性質の異なる表現が得られることが分かったが、パイロット版で確認できた傾向が必ずしも全話者で確認できたわけではなく、個人差が課題であることが分かった。

最後に、3回目の収録として、話者および発話内容をさらに拡大して、総音声数はパイロット版の72倍という規模で、音響的な側面からの多様性の確認を行った。5.4節で論じたように、音響的分析の結果から判断する限りでは、台本の数は十分であると推測でき、特に、発話内容が異なっても、音響的な表現の分布は同一話者でほぼ同様となることから、個人の演技の表現力という観点では測定が実現できていると考えられる。ただし、心理特徴の計測も含めて追加の実験をおこない、さらなる検証が必要である。

今後は話者を拡大していくことで、個人性や演技音声か否かという枠を超えて、音声メディアの表現力そのものの測定が可能となることが期待できる。

第6章 結論

本論文では、コミュニケーションメディアが多様化する現代において、メッセージの表現力を可視化することの重要性を指摘したうえで、人間の代表的なコミュニケーション手段である、音声における表現力の可視化の手法を検討し、その妥当性を確認した。

まず、第2章では、コミュニケーションメディアの評価にかかる情報として、コミュニケーションメディアの歴史・性質・種類・問題点・既存の評価方法の5つの観点から整理した。そして、本研究では、コミュニケーションメディア上におけるメッセージの表現力に着目し、物理的・心理的な側面から表現力の可視化をおこなうことで、メディアの利用者・開発者双方が共通で扱える新しい評価の観点を与える目的を持つことを述べた。そして、人間にとってもっとも基本的なコミュニケーションメディアのひとつである音声を対象に手法の適用を試みる意義を述べた。

第3章では、コミュニケーションメディアの表現力を測定する際に問題となる、メッセージの符号化パターンの膨大さの緩和を目的として、メッセージへの制約条件の付与を可能とするためのコミュニケーションのモデルを検討した。そのため、まず先行研究として Shannon のコミュニケーションモデルと、Brunswikian Lens Model の Scherer による解釈を示し、本研究にとって重要なと思われる概念をそれぞれから参考することにした。そして、それらに加えて、2.2節で述べたコミュニケーションメディアの性質を考慮しながら、独自のコミュニケーションモデルを検討した。

第4章では、音声メディアの表現力を測定するための準備として多様な音声メッセージを収集するための具体的手続きについて検討した。まず、音声資料の分類をおこない、本研究で対象とする音声メッセージは、演技音声の一部とすることを示した。そして、表現力の測定に十分役立つような表現豊かな演技音声を得るために工夫として、本研究独自の「台本のフォーマット」を提案した。台本のフォーマットの設計にあたっては、演劇に関する先行研究や著書を参考にしながら、音声で演技をする際に表現に影響を与えると思われる重要な情報を選定した。その情報に基づき、第3章で示したコミュニケーションモデルを音声に最適化させた。さらに、表現豊かな音声表現の獲得のための、台本のインスタンスの生成方法、発話内容および発話者の選定に関する方針を定めた。

第5章では、第4章で定めた手続きに則って、3回の収録実験を行い、それぞれの結果を述べたのち、音声の表現力の測定の可能性について論じた。パイロット版の収録では、1名の話者と主観的に作成した台本100個を用いた小規模な収録を行い、心理的・音響的な豊かさ、および自然性の観点で、少なくとも既存の演技音声より高品質な音声資料が得られることを確認した。そして、2回目の収録を本実験と位置づけて、3名の話者、ある程度パターン化された台本280個を用いて、規模を拡大した実験を行った。その結果、提案手法による音声表現は、やはり典型的な感情語と比較すると性質の異なる表現が得られることが分かったが、パイロット版で確認できた傾向が必ずしも全話者で確認できたわけではなく、個人差が課題であることが分かった。最後に、3回目の収録として、話者および発話内容をさらに拡大して、台本数は1200個、総音声数はパイロット版の72倍という規模で、音響的な側面からの多様性の確認を行った。5.4節で論じたように、音響的分析の結果から判断する限りでは、台本の数は十分であると推測でき、特に、発話内容が異なっても、音響的な表現の分布は同一話者でほぼ同様となることから、個人の演技の表現力という観点では測定が実現できていると考えられる。したがって、今後は話者を拡大していくことで、個人性や演技音声か否かという枠を超えて、音声メディアの表現力がより精度高く可能となることが期待できることを述べた。

ただし、本研究で示した、音声収集から分析までの手続きには多くの課題も残っている。まず、本稿で示した実験の範囲では、台本の項目ごとの効果を確認していない。したがって、表現の豊かさに貢献する項目・貢献しない項目について明らかにし、提案手法の効果をより明確にする必要がある。また、分析方法についても、心理的・音響的両側面において、特徴量の選択が十分であるとは言えず、多様な心理尺度や音響特徴量を用いた多角的な分析が必要である。

最後に、今後の展望について述べる。本研究の延長線上に、多様なメディアにおける表現力の測定およびメディア間の表現力の比較が存在する。それらの実現可能性についてであるが、たとえばテキストなどのインターネット上のデータを活用できるメディアであれば、本稿で示したように演技のプロフェッショナルを用意する必要はない。一方で、インターネット上の多くのメディアは複雑な構造をしており、コミュニケーションモデルの最適化が難しく、多様なメッセージを網羅することは音声やテキスト中心のモダリティに比べて難易度は高くなる可能性がある。また、メディア間の表現力の比較についてであるが、比較するメディアが明らかなときは、共通で用いることのできる、ある程度汎用的なモデルを検討する方法が考えられる。

本研究で提案した手法が、今後の社会において、安全・安心なインターネット活用を促進するための一助となれば幸いである。

謝辞

私が社会人のおりに博士課程へ進学することを決意して以来、数多くの方々の支えがあり、本研究をまとめることができました。

学部3年時の私の研究室配属から現在までの15年間、常に暖かく厳しくご指導してくださった、指導教官である早稲田大学名誉教授の白井克彦先生に心から感謝申し上げます。とくに、私が修士の学生のころに頂いた「大事なのは結果ではなく、最後までやり遂げることである」というご助言、そして博士入学直後に頂いた「社会問題と向き合う際には、ネガティブな姿勢ではなくポジティブな姿勢で望み解決をはかること」というご指摘は、常に私の行動の指針として心に根付いております。そして、本研究を進めるための実験の計画から遂行、分析、論文の執筆に至るまで細かくご指導してくださった、早稲田大学人間科学学術院の菊池英明先生に感謝申し上げます。

本論文をまとめるにあたり、核心をついた貴重なご助言の数々と、心強い多くの励ましをくださいました、早稲田大学理工学術院の小林哲則先生、ならびに、ご多忙のなか快く副査を引き受けてくださった松山泰男先生に感謝申し上げます。

これまでの多くの学外発表の原稿執筆にあたり、共著者としてご助言くださいました電気通信大学名誉教授の樽松明先生、千葉工業大学工学部の大川茂樹先生、東洋大学総合情報学部の村上真先生に感謝申し上げます。

「研究は楽しく取り組むこと」「体調に負担をかけないこと」と、博士入学直後から現在に至るまで頻繁にご心配くださいり、研究への姿勢について常にご指導くださいました、早稲田大学理工学術院の東山三樹夫先生に感謝申し上げます。

今後の私の進路や活動についてご相談に乗ってくださいました、早稲田大学名誉教授の市川薰先生、早稲田大学グリーンコンピューティングシステム研究機構・知覚情報システム研究所の岩田和彦先生に感謝申し上げます。

本研究で取り扱った実験音声の収集と分析においては、人間科学学術院菊地英明研究室卒業生の佐藤安里さん、福田節君をはじめ、多くの方々のご協力なくして達成はできませんでした。特に、拡張版の収録において、大規模な収録の機会を与えてくださいました国立国語研究所の前川喜久男先生には大変お世話になりました。ここに感謝の意を表します。

7年間の博士課程において主な活動拠点であった早稲田大学理工学術院白井克彦研究室ならび人間科学学術院菊地英明研究室の皆様にも感謝申し上げます。自

井研究室出身の谷口徹博士には、博士入学時に多大なご協力を頂き、本論文の執筆に関してもアドバイスを頂きました。久保陽太郎博士には、研究者としてのストイックなスタンスを背中から学ばせていただきました。菊池研究室出身の宮澤幸希博士には、各種実験の分析において、統計的観点からのアドバイスや、博士論文執筆において多くのアドバイスをいただきました。その他、学生の皆さんには、研究室生活の面で数多くのサポートを頂きました。そのすべては、とてもここに書ききれるものではありません。

同世代の博士課程として、現在は様々な分野で活躍しておられる多くの研究者の皆さんにも、助手業務や研究室間交流などを通じて親しくさせていただき、学位取得のおおきな励みとなりました。有難うございました。

そして最後に、7年もの長い間、私を支えてくれた家族に感謝の意を表します。

参考文献

注：URLへのアクセスは2013年11月15日時点のものである。

- [1] M. マクルーハン著, メディア論 人間の拡張の様相, 粟原裕・河本仲聖訳, みすず書房, 1987.
- [2] 水上陽介, 内田啓治, 澤田秀之, “糸状形状記憶合金の振動を利用した高次知覚生紀による触覚呈示,” 情報処理学会論文誌, vol.48, no.12, pp.3739-3749, 2007.
- [3] 垣内祐一, 石澤正行, 重野寛, 岡田謙一, “感性語を媒介にした香りコミュニケーションモデル,” 情報処理学会論文誌, vol.47, no.12, pp.3414-3422, 2006.
- [4] A.N ジェイソン著, “1.3 節 会わずに話すこと：電話,” インターネットにおける行動と心理, 三浦他訳, 北大路書房, p.15, 2004.
- [5] A.N. ジェイソン著, “3.2 節 フレーミングと反社会的行動,” インターネットにおける行動と心理, 三浦他訳, 北大路書房, p.68, 2004.
- [6] 小林直樹著, ソーシャルメディア炎上事件簿, 日経BP社, 2011.
- [7] A.N. ジェイソン著, “3.3 節 インターネット上の人間関係：すぐ親密になりすぎる?,” インターネットにおける行動と心理, 三浦他訳, 北大路書房, p.15, 2004.
- [8] 安藤玲子, 坂本章, 他, “インターネット使用が人生満足感と社会的効力感に及ぼす影響—情報系専門学校男子学生に対するパネル調査”, パーソナリティ研究, vol.13, no.1, pp.21-23, 2004.
- [9] 藤桂, 吉田富二雄, “インターネット上の行動内容が社会性・攻撃性に及ぼす影響：ウェブログ・オンラインゲームでの検討より,” 社会心理学研究, vol.25, no.2, pp.121-132, 2009.
- [10] J. ダイアモンド著, 人間はどこまでチンパンジーか 人類進化の栄光と驕り一, 長谷川真理子訳, 新曜社, 1993.
- [11] 松島英子, “メソポタミアにおける文字と思考についての一考察,” 法政大学キャリアデザイン学部紀要3, pp.257-275, 2006.
- [12] 原田慶吉著, 楔形文字法の研究, 弘文堂, 1949.
- [13] B.マクダーモット著, 古代エジプト文化とヒエログリフ, 竹田悦子訳, 産調

出版, 2005.

- [14] CBI Online, “漢字の起源と変化” : <http://japanese.cri.cn/chinaabc/chapter14/chapter140507.htm>
- [15] 中川勝昭, “アルファベットの原理,” 北海道大学独文学科研究年報, vol.23, pp.55-72, 1996.
- [16] 大庭脩, “漢墓出土の簡牘 一特に王杖簡牘について—,” 書学書道史研究, vol.1996, no.6, pp.3-15, 1996.
- [17] 津金学校, “烽火” :<http://www.tsugane.jp/meiji/rekisi/sutama/norosi.html>
- [18] 中野明著, 腕木通信 ナポレオンが見たインターネットの夜明け, 朝日選書, 2003.
- [19] 白戸健一郎, 放送メディアにおけるソフト・パワー論の系譜(文献展望), 京都大学生涯教育学・図書館情報学研究, vol.11, pp.25-38, 2012.
- [20] A.N ジェイソン著, “1.3 節 会わずに話すこと：電信,” インターネットにおける行動と心理, 三浦他訳, 北大路書房, p.13, 2004.
- [21] NTT 東日本, “通信偉人伝 電話を発明したベルとグレイ, そしてエジソン,” : http://www.ntt-east.co.jp/business/magazine/network_history/02/
- [22] A.N ジェイソン著, “1.3 節 会わずに話すこと：電話,” インターネットにおける行動と心理, 三浦他訳, 北大路書房, p.16, 2004.
- [23] A.N ジェイソン著, “1.3 節 会わずに話すこと：電話,” インターネットにおける行動と心理, 三浦他訳, 北大路書房, p.17, 2004.
- [24] 日本ラジオ博物館 : <http://www.japanradiomuseum.jp/>
- [25] テレビる毎日, “テレビの歴史年表” : <http://cozalweb.com/ctv/shiryo/rekishii.html>
- [26] JPNIC, “JPNIC アーカイブス インターネット歴史年表” : <https://www.nic.ad.jp/timeline/>
- [27] 首相官邸 Web サイト, “e-Japan 戦略” : http://www.kantei.go.jp/jp/it/network/dai1/1siryou05_2.html
- [28] 小柴等, 加藤直孝, 國藤進, “グループ意思決定におけるアウェアネス：通信環境と GDSS の観点から,” 情報処理学会論文誌, vol.47, no.1, pp.77-86, 2006.
- [29] 坂本大介, 神田崇行, 他, “遠隔存在感メディアとしてのアンドロイド・ロボットの可能性,” 情報処理学会論文誌, vol.48, no.12, pp.3729-3738, 2007.
- [30] Youtube : <http://www.youtube.com/>
- [31] 総務省, “u-Japan 政策” : http://www.soumu.go.jp/menu_seisaku/ict/u-japan
- [32] 総務省, 平成 25 年度版情報通信白書, “インターネットの利用状況” : <http://www.soumu.go.jp/seisaku/ict/u-japan>

- //www.soumu.go.jp/johotsusintokei/whitepaper/ja/h25/html/nc243120.htm
l
- [33] 山田裕子, 平野貴幸, 西本一志, “TangibleChat : 打鍵振動の伝達によるキーボードチャットにおける対話状況アウェアネス伝達の試み,” 情報処理学会論文誌, vol.44, no.5, pp.1392-1403, 2003.
- [34] Twitter : <https://twitter.com/>
- [35] 2ちゃんねる : <http://www.2ch.net/>
- [36] 4chan : <http://www.4chan.org/>
- [37] ニコニコ動画 : <http://www.nicovideo.jp/>
- [38] Facebook : <https://www.facebook.com/>
- [39] インターネットコム, “意外と高い「ウィキペディア」「2ちゃんねる」の職場利用——ネットレイティングス調べ” : <http://japan.internet.com/wmnews/20081224/4.html>
- [40] In the loop, “mixi, Twitter, Facebook, Google+, Linkedin 2013年3月最新ニールセン調査” : <http://media.looops.net/sekine/2013/05/08/neilsen-netview-201303/>
- [41] 総務省, “平成23年度版情報通信白書, ICTにより国民生活はどう変わったか” : <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h23/html/nc213120.html>
- [42] PRIVACY AND DATA SECURITY, “Anon Terminology Paper” : http://dud.inf.tu-dresden.de/Anon_Terminology.shtml
- [43] はてな匿名ダイアリー : <http://anond.hatelabo.jp/>
- [44] スラッシュドット・ジャパン : <http://slashdot.jp/>
- [45] 総務省, “移動体通信（携帯電話・PHS）の年度別人口普及率と契約数の推移” : http://www.soumu.go.jp/soutsu/tokai/tool/tokeisiryo/idoutai_nenbetu.html
- [46] 総務省, “移動体通信（携帯電話・PHS）の年度別人口普及率と契約数の推移” : http://www.soumu.go.jp/soutsu/tokai/tool/tokeisiryo/idoutai_nenbetu.html
- [47] 総務省, “平成24年度版情報通信白書, 固定通信” : <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h24/html/nc245220.html>
- [48] IRC普及委員会 : <http://irc.nahi.to/>
- [49] ICQ : <http://www.icq.com/>
- [50] 日経BPコンサルティング, “「携帯電話・スマートフォン“個人利用”実態調査2013』より” : <http://consult.nikkeibp.co.jp/consult/news/2013/0830sp/>

- [51] 宮里勉, 岸野文郎, 寺島信義, “臨場感通信会議における参加者の対面状況の保持特性の評価,” 電子情報通信学会論文誌, vol.J79-A(2), pp.518-526, 1996.
- [52] 大西 仁, 望月要, “電子通信技術と実験心理学: 映像と音声の伝送遅延の心理的影響を測定する,” 基礎心理学研究, vol.21, no.2, pp.148-152, 2003.
- [53] Myspace : <https://myspace.com/>
- [54] iKnow : <http://iknow.jp/>
- [55] USTERAM : <http://www.ustream.tv/>
- [56] ニコニコ生放送 : <http://live.nicovideo.jp/>
- [57] Tumbler : <https://www.tumblr.com/>
- [58] 折田明子, “SNS に集約する情報: ネットワーキングからライプログへ,” 情報の科学と技術, vol.61, no.2, pp.70-75, 2011.
- [59] 山本仁志, 他, “コミュニケーション構造の推移による大量 SNS サイトの分類,” 日本社会情報学会学会誌, vol.23, no.1, pp.33-43, 2011.
- [60] マイナビニュース, “Google+は Twitter・Facebook と何が違う? どう使い分ける? ” : http://news.mynavi.jp/articles/2013/07/22/google_howto/
- [61] 田村武志, 上西慶明, 佐藤文博, “マルチメディア遠隔教育システムの評価と学習者インターフェースの検討”, 情報処理学会論文誌, vol.34, no.6, pp.1235-1245, 1993.
- [62] 石川智治, 宮原誠, “バーチャルリアリティに重要な体感に基づく心理生理的評価の一提案,” 映像情報メディア, vol.60, no.3, pp.446-448, 2006.
- [63] C.E. Shannon, “A Mathematical Theory of Communication,” The Bell System Technical Journal, vol.27, pp.379–423, 623–656, 1948.
- [64] E. Brunswik, Perception and the Representative Design of Psychological Experiments, University of California Press, Berkley, 1956.
- [65] K.R. Scherer, “Vocal communication of emotion: a review of research paradigms.” Speech Communication, vol.40, pp.227-256, 2003.
- [66] N. Campbell, “Developments in corpus-based speech synthesis: approaching natural conversational speech,” IEICE Trans. Inf. & Syst., vol. E88-D(3), pp.376-383, 2005.
- [67] S. Steidl, Automatic Classification of Emotion-Related User States in Spontaneous Children's Speech. Logos Verlag Berlin, 2009.
- [68] 前川喜久雄, “言語研究における自発音声,” 日本音響学会春季研究発表会公演論文集, pp.425-426, 2001.
- [69] 前川喜久雄, “自発音声とデータベース,” 日本音響学会誌, vol.61, no.9, pp.544-549, 2005.

- [70] S. Kitazawa, T. Kitamura, K. Mochiduki, and T. Itoh, "Preliminary Study of Japanese MULTTEXT: a Prosodic Corpus." Proc. of International Conference on Speech Processing, pp.825-828, 2001.
- [71] D. Erickson, "Expressive speech: production, perception and application to speech synthesis." Acoust. Sci. & Tech., vol.26, no.4, pp.317-325, 2005.
- [72] H. Fujisaki, "Prosody, information, and modeling -With emphasis on tonal features of speech." Speech Prosody 2004, pp.1-10, 2004.
- [73] 伊藤正男, 梅本守, 他, 岩波講座 認知科学 6 情動, 岩波書店, 1994.
- [74] 森山剛, 斎藤英雄, 小沢慎治, "音声における感情表現語と感情表現パラメータの対応付け", 信学論, vol.J82-D-2(4), pp.703-711, 1999.
- [75] 平田オリザ, 演劇入門 (講談社現代新書 1422), 講談社, 1998.
- [76] 後安美紀, 辻田勝吉, "演劇操作におけるシステムダイナミクス," 認知科学, vol.14, no.4, pp.509-531, 2007.
- [77] 安藤花恵, "演技の熟達化-脚本の読み取りから演技計画, 演技遂行まで," 心理学研究, vol.74, no.7, pp.373-379, 2002.
- [78] K. Krippendorf, Content Analysis. Sage Publications, 1980.
- [79] 前川喜久雄, 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, vol.9, no.1, 46-66, 2002.
- [80] A. Ortony and T.J. Turner, "What's the basic about basic emotions?" Psychological Review, vol.97, no.3, pp.315-331, 1990.
- [81] B.W. Silverman, Density Estimation of Statistics and Data Analysis. Chapman & Hall, 1986.
- [82] D. Carr, N. Lewin-Koh, and M. Maechler, Hexbin: hexagonal Binning Routines, R package version 1.26.2, <http://cran.r-project.org/web/packages/hexbin/index.html>
- [83] 小林のぞみ, 乾健太郎, 松本裕治, 他, "意見抽出のための評価表現の収集," 自然言語処理, vol.12, no.2, pp.203-222, 2005.
- [84] 東山昌彦, 乾健太郎, 松本裕治, "述語の選択選好性に着目した名詞評価極性の獲得," 言語処理学会第 14 回年次大会論文集, pp.584-587, 2008.
- [85] 天野成昭, 他(編・著), 日本語の語彙特性, 三省堂, 1999.

研究発表論文

国際会議

○T. Miyajima, H. Kikuchi, K. Shirai, S. Okawa, “Method for Collection of Acted Speech Using Various Situation Scripts,” Proc. 8th Intl. Conf. on Language Resources and Evaluation (LREC2012), pp.1179-1182, May. 2012.

○T. Miyajima, H. Kikuchi, K. Shirai, “Collection and analysis of emotional speech focused on the psychological and acoustical diversity,” Proc. 17th Intl. Congress of Phonetic Sciences (ICPhS2011), pp. 1394-1397, Aug. 2011.

T. Miyajima, T. Fukuda, H. Kikuchi, K. Shirai, “Method for Collection of Diverse Speech for Emotion Research Database,” Proc. Oriental COCOSDA 2010, Paper-17, Nov. 2010.

Y. Ogawa, M. Murakami, J. Katsuyama, T. Miyajima, K. Shirai, “Construction of Decision Model for a System to Start Communicating with a Human Using Hidden Markov Model,” Proc. 9th IEEE/ACIS Intl. Conf. on Computer and Information Science, pp.401-406, Aug. 2010.

T. Miyajima, H. Kikuchi, A. Kurematsu, K. Shirai, “Analysis of Alteration of Impressions During Conversion of Voice to Text,” Proc. 2009 RISP International Workshop on Nonlinear Circuits and Signal Processing, pp.57-60, Mar. 2009.

国内研究会

菊池英明, 宮島崇浩, 沈叡, “多様な音声表現コーパスにおける句末音調のクラスタリング,” 第3回コーパス日本語学ワークショップ予稿集, pp.23-28, 2013年2月.

宮島崇浩, 菊池英明, “多様な話者による演技感情音声の収集と特徴の比較,” 第 2 回コーパス日本語学ワークショップ予稿集, pp.245-254, 2012 年 9 月.

菊池英明, 宮島崇浩, “日本語話し言葉コーパスにおける句末音調のバリエーション,” 第 1 回コーパス日本語学ワークショップ予稿集, pp.351-354, 2012 年 3 月.

端千尋, 宮島崇浩, 村上真, 白井克彦, “能動的情報収集対話システム構築を目的とした対話の活性化,” 人工知能学会研究会資料, SIG-SLUD-A903-9, pp.49-54, 2010 年 2 月.

宮島崇浩, 菊池英明, 横松明, 白井克彦, “印象語から想起した音声情報の特徴量空間の分析,” 人工知能学会研究会資料, SIG-SLUD-A803-12, pp.63-68, 2009 年 3 月.

宮島崇浩, 菊池英明, 横松明, 白井克彦, “印象空間における音声と文字の対応関係の分析,” 人工知能学会研究会資料, SIG-SLUD-A801-2, pp.7-14, 2008 年 7 月.

国内全国大会

中山瑛梨, 宮島崇浩, 菊池英明, “パーソナリティ表現を用いた音声合成用インターフェースの検討,” 日本音響学会春季研究発表会講演論文集, Vol.1-7-3, pp.259-260, 2013 年 9 月.

菊池梨佳子, 宮島崇浩, 菊池英明, “多様な音声表現コーパスを用いた記憶に残りやすい話し方の分析,” 日本音響学会春季研究発表会講演論文集, Vol.1-7-2, pp.257-258, 2013 年 9 月.

赤塚祐斗, 宮島崇浩, 菊池英明, “多様な音声表現コーパスを用いた演技経験と演技感情表現の関係の分析,” 日本音響学会春季研究発表会講演論文集, Vol.1-7-1, pp.255-256, 2013 年 9 月.

青木由希, 宮島崇浩, 菊池英明, 塩見格一, “音声から疲労程度を推定するスマートフォン用アプリケーションの開発,” 第 12 回情報科学技術フォーラム (FIT2013), E-037, pp.269-270, 2013 年 9 月.

真木恵, 宮島崇浩, 菊池英明, “創作物における声優演技音声の声質のステレオタ

イブ,” 日本音響学会秋季研究発表会講演論文集, Vol.2-Q-a3, pp.515-518, 2012 年 9 月.

菊池英明, 宮島崇浩, 前川喜久雄, “表現豊かな音声の収集における多様性の追求,” 日本音響学会秋季研究発表会講演論文集, Vol.1-2-16, pp.263-264, 2012 年 9 月. 宮島崇浩, 福田節, 菊池英明, 白井克彦, “感情音声コーパスのための多様な音声収集手法,” HCG シンポジウム 2010, A1-2, pp.10-17, 2010 年 12 月.

端千尋, 宮島崇浩, 村上真, 白井克彦, “再利用可能情報収集のための構造化データベースの提案,” 電子情報通信学会総合大会講演論文集, D-5-6, p.47, 2009 年 3 月.

鈴木智史, 宮島崇浩, 村上真, 白井克彦, “相手にあわせた発話調整機能のための言語情報からのユーザモデル生成手法の検討,” 電子情報通信学会総合大会講演論文集, D-5-7, p.48, 2009 年 3 月.

宮島崇浩, 菊池英明, 横松明, 白井克彦, “音声対話システム利用時における感情表出誘因性に関する考察,” 第 7 回情報科学技術フォーラム(FIT2008), J-028, pp.463-464, 2008 年 9 月