

◎籠宮 隆之 (国立国語研究所)

△榎 洋一 (東京都立大学大学院人文科学研究科 / 国立国語研究所)

菊池 英明・前川 喜久雄 (国立国語研究所)

はじめに 自然な音声合成や、円滑な機械と人間とのコミュニケーションのためには、人間がどのような音声に対し、どのような評定を下すのか、知る必要がある。そのためには、大量の音声データに対し、印象評定を行なったデータベースが存在することが望ましい。

以上のことから、国立国語研究所・通信総合研究所・東京工業大学が開放的融合研究「大規模コーパスに基づく『話し言葉工学』の構築」<sup>4)</sup>の一貫として作成している大規模な自発音声コーパスである『日本語話し言葉コーパス』<sup>3)</sup>には、収録された音声に対する印象評定値を付属させる予定である。

現時点では、最終的な評定作業に先立ち、予備的な評定として、講演収録時に収録者による発話の印象評定を行なっている。

本研究は、上記の予備的な印象評定値を分析し、発話の印象に与える要因を探ることを目的とする。

**音声データ** 学会講演と、国立国語研究所で企画した模擬講演をデータとして用いた。両者とも今回用いるのは、2001年度前半までに収録し、書き起こしテキストが用意されているものである。

**学会講演** 1999年度から2000年度にかけて行なわれた、工学系および言語系の学会での講演音声。講演数は、総計341講演。それぞれの講演時間は、工学系では12~20分、言語系では20~25分のものであり、講演者は、工学系の講演では男性の大学院生が多く、言語系の学会では、女性の大学院生が多い。収録場所及び聴衆の数は、講演ごとに大きく異なる。

**模擬講演** 1) 幾つかの大まかなテーマ(例えば、「人生を振り返って楽しかったこと」「人生を振り返って悲しかったこと」など)を与え、そのテーマに則って講演者が内容を考え講演するものと、2) 講演者がテーマを自由に選び講演するものがある。講演数は、総計465講演。講演時間は、10~15分程度で講演するように指示した。講演者は、人材派遣会社を通して募集した20代から60代までの男女、および通訳養成学校の生徒を中心とした20代~30代の男女である。収録場所は、国立国語研究所の録音スタジオ、および都内のオフィスビルの一室である。聴衆は、収録スタッフも含め、3~5人程度である。また、収録スタッフには、講演者をできるだけリラックスさせるよう、指示している<sup>3)</sup>。

**印象評定** 音声データ収録時に、収録スタッフに印象評定を行なわせた。1つの講演につき、1人の評定者が評定を行なった。評定者は講演ごとに異なり、20名が評定を行なった。

評定者には、提示した評定語のうち講演に当てはまるものをチェックさせた。

評定語には、表1の31語を使用した。これらは、話し方に対するものと、声に対するものとに大別される。また、対語になっているものを中心にしたが、そうでないものも含めた。

表1: 提示した評定語

たどたどしい	流暢な	車調な	表情豊かな	自信のある	自信のない
落ち着いた	落ち着きのない	いらいらした	緊張した	リラックスした	
大きい声	小さい声	かすれた声	裏返った声	こもった声	
重厚な	軽薄な	若々しい	年寄じみた	元気のある	元気がない
聞き取りやすい	聞き取りにくい	生意気な	尊大な	鼻にかかった	
高い	低い	きっぱりした	優しい		

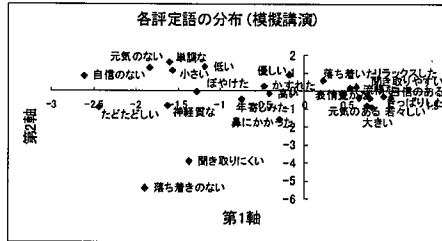
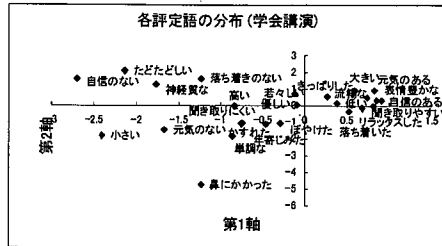


図1: 各評定語の数量化III類の結果による布置。上段は学会講演、下段は模擬講演。第1軸と第2軸を用いて布置。X軸が第1軸、Y軸が第2軸。

**分析** 以上の講演音声データと印象評定の結果を用いて、どのような音声に対し、どのような印象を受けるかを分析する。

それに先立ち、学会講演と模擬講演では、講演の行なわれる場面、話題や講演者の性質が大きく異なるので、両者の違いが印象評定の傾向に差を与えているかを分析した。

また、評定語がどのような特性を示しているかを分析した。

**評定語の分類および評定語間の関係** 林の数量化III類を用いて、各印象評定語に対する反応傾向を分析した。なお、ここでは「あてはまる」とされた講演が5以下であった5評定語は、分析から除外した。

今回は、軸の固有値が学会講演・模擬講演ともに0.49以上と、特に高かった第1軸(固有値:学会講演0.660, 模擬講演0.663)と第2軸(固有値:学会講演0.493, 模擬講演0.517)を用いて分析する。図1に、第1軸と第2軸のカテゴリー数量による各評定語の布置を示す。

第1軸は、模擬講演・学会講演ともに、原点の右側に「自信がある」「聞き取りやすい」「元気のある」「流暢な」などの好感を抱く印象評定語が集まり、原点から離れた左端には「自信のない」「たどたどしい」「落ち着きのない」などの否定的な評定語が集まっている。このことから、第1軸は、「肯定的—否定的」を示していると解釈できる。

第2軸は、模擬講演では、原点から上に「低い」「元気がない」「落ち着いた」「単調な」などが集まり、原点から下に「落ち着きのない」「大きい」「元気のある」などの評定語が集まっている。一方、学会講演では、一部の評定語を除いて模擬講演とは逆の傾向を示している。これらの評定語は、動きの多少を示す語と考えられるので、第2軸は「活動性」を示していると解釈できる。

また、学会講演と模擬講演とは、第2軸が逆転しているものの、同じような傾向を示すことから、両者の違いが印象評定

\* The factors for impression test on Corpus of Spontaneous Japanese : KAGOMIYA Takayuki(NIJLA), MAKI Yohichi(Tokyo Metropolitan Univ./ NIJLA), KIKUCHI Hideaki, MAEKAWA Kikuo(NIJLA)

表 2: 評定語と講演特徴との関係。'○'は、「評定語があてはまる」とされた講演の講演特徴の数値が「評定語があてはまらない」とされた講演の講演特徴の数値に対してt検定の結果5%水準以上で有意に大きいことを示す。'×'は、同様に有意に小さいことを示す。上段は学会講演、下段は模擬講演。評定語は、数量化III類の分析結果で得られた第1軸のカテゴリの順に並んでいる。なお、評定語は短縮表記した。

学会講演	自信ない	たどたどし	神経質	元気がない	単調	優しい	落着いた	聞き取り易	リラックス	表情豊か	聞き取	元気ある	流暢	自信ある
第1軸カテゴリ-数量	-2.702	-2.134	-1.759	-1.670	-0.858	0.088	0.267	0.520	0.680	0.740	0.816	0.821	0.853	0.912
モーラ数=発話速度	×	×	×	×	×	×				○		○	○	○
ポーズ抜き発話速度	×				×							○	○	○
ポーズの比率	○	○	○		○									×
発話速度のゆれ											○			
フィルター数													○	
言い淀み数														○
単独の笑い数					×			×				×		
発話中の笑い数	×				×			○						
模擬講演	自信ない	たどたどし	神経質	元気がない	単調	優しい	落着いた	聞き取り易	リラックス	表情豊か	聞き取	元気ある	流暢	自信ある
第1軸カテゴリ-数量	-2.615	-2.438	-1.837	-1.624	-1.603	-0.182	0.221	0.541	0.614	0.644	0.738	0.743	0.929	1.018
モーラ数=発話速度	×	×	×	×	×	×			○	○	○	○	○	○
ポーズ抜き発話速度	×	×	×	×	×	×			○	○	○	○	○	○
ポーズの比率	○	○	○	○	○	○		×	×	×	×	×	×	×
発話速度のゆれ								×						
フィルター数									○			○		
言い淀み数														×
単独の笑い数			×		×	×	×	×		○	○		×	×
発話中の笑い数			×		×	×	×	×	○	○	○	○	×	×

の傾向に差を与えていないと解釈できる。

い評定語ではポーズの比率が小さい。つまり、ポーズの比率の大きい講演は否定的な評定をされる。

印象評定と音声との関係 どのような音声特徴をもつ講演に対し、どのような印象評定が行なわれているかを分析した。

分析方法は、1) それぞれの評定語に対し、評定語があてはまる」とされた講演と、評定語があてはまらない」とされた講演とに、講演を分類し、2) 両者の間に、幾つかの音声特徴に差が見られるかを分析する — という方法をとった。なお、「あてはまる」とされた講演が20以下であった17評定語は、今回の分析から省いた。

印象評定に影響を及ぼす音声の特徴量は、ピッチの変動やスペクトル特性など、さまざまな値が考えられる。しかし、これらの値を各講演について調べるには多大な努力を要する。そこで、今回は書き起こしテキストから導き出せる以下の値を用いた。ここで示す講演時間には、発話間のポーズも含まれている。また、フィルターや言い淀みなどの認定に関しては、テキストの書き起こし基準<sup>1)</sup>に従った。

- モーラ数 (1分あたり) = 発話速度 —— 講演の総モーラ数 / 講演時間
- ポーズ抜き発話速度 —— 講演の総モーラ数 / (講演時間 - 発話間のポーズ総時間)
- ポーズの比率 —— 発話間のポーズ総時間 / 講演時間
- 発話速度のゆれ —— 各講演内での、個々の発話単位の発話速度の分散
- フィルター数 (1分あたり) —— 講演の総フィルター数 / 講演時間
- 言い淀み数 (1分あたり) —— 講演の総言い淀み数 / 講演時間
- 単独の笑い数 (1分あたり) —— 講演の総単独の笑い数 / 講演時間
- 笑いながらの発話の数 (1分あたり) —— 講演の総笑いながらの発話数 / 講演時間

表2に、以上の分析の結果を示す。'○'は、「評定語があてはまる」とされた講演の講演特徴の数値が「評定語があてはまらない」とされた講演の講演特徴の数値に対してt検定の結果5%水準以上で有意に大きいことを示す。'×'は、同様に有意に小さいことを示す。評定語は、数量化III類の分析結果で得られた第1軸のカテゴリ-数量の順に並べた。また、上段は学会講演、下段は模擬講演である。

結果と考察 表2をみると、「発話速度」「ポーズ」「フィルター」「言い淀み」「笑い」と、全ての講演特徴が、講演の印象評定に関係していることが分かる。

ここでは、特に顕著に見られる傾向を列挙し、考察する。

1. 学会講演・模擬講演ともに、第1軸のカテゴリ-数量が大きい評定語では発話速度が速く、カテゴリ-数量が小さいものでは発話速度が遅い。つまり、発話速度の速い講演は肯定的な評定をされ、発話速度の遅い講演は否定的な評定をされる。
2. 学会講演ともに第1軸のカテゴリ-数量が小さい評定語では、ポーズの比率が大きい。また、模擬講演では、「流暢な」「きっぱりとした」を除いてカテゴリ-数量が大きい

発話速度に関しては、内田は比較的短い発話の発話スピードから発話者の性格特性を推測させた実験を行ない、「外交性、開放性、及び誠実性は、やや速い発話に評価のピークがあり、発話速度の低下と共に低下した」と述べている<sup>5)</sup>。今回の結果からでは、内田の挙げる性格特性までは分からない。しかし、肯定的な印象が、発話速度の上昇とともに高くなったという点で、内田の研究と同様の結果を示している。

ただし、ポーズを除いた講演時間から求めた発話速度では、評定と発話速度とのあいだで、それほど顕著な関係は見られなかった。一方、ポーズの比率が多い講演は否定的な評定をされた。つまり、講演においては単純な発話速度のみで肯定的・否定的な印象を得ているわけではないことを示している。

これは、講演とは聴衆に情報を伝えるものであること、および今回の評定は評定者も広い意味での聴衆であることが関係していると考えられる。つまり、発話速度が速い講演やポーズの少ない講演では単位時間あたりの情報伝達量が多くなり、発話速度が遅い講演やポーズの多い講演では単位時間あたりの情報伝達量が少くなる。聴衆は講演者からより多くの情報を得られると感じれば肯定的な印象評定をし、より少ない情報しか得られないと感じれば否定的な印象評定をしている可能性がある。

まとめと今後の課題 講演の印象評定と講演の音声特徴との関係を調べた。その結果、発話速度およびポーズの比率が講演の印象評定に大きな役割を果たしていることが分かった。

今後は、今回扱わなかった数量化III類の第2軸以降のカテゴリ-数量と講演特徴との関連も調べたい。

また、評定語の追加や絞り込みなどを行なって、項目を整理すると共に、評定者を固定するなどして、より信頼のおける評定データベースを作成したい。

更に、現在作業を進めている分節音ラベリングや韻律ラベリングなど<sup>2)</sup>の結果を生かし、今回より詳細な音声特徴と印象評定の関係を検討したい。

### 参考文献

- 1) 小磯, 土屋, 間瀬, 斉藤, 籠宮, 前川. 「日本語話し言葉コーパス」における書き起こしの方法とその基準について. 日本語学, 9, 2001.
- 2) 菊池, 前川, 籠宮, 竹内. 自発音声に対する J.ToBI ラベリングの問題点検討. 日本音響学会 2001 年春季研究発表会講演論文集, 2001.
- 3) 前川, 籠宮, 小磯, 小椋, 菊池. 日本語話し言葉コーパスの設計. 音声研究, 4(2), 2000.
- 4) 古井, 前川, 井佐原. 科学技術振興調整費開放的融合研究推進制度 — 大規模コーパスに基づく『話し言葉工学』の構築 —. 日本音響学会誌, 56(11), 2000.
- 5) 内田. 音声の発話速度の制御がピッチ感及び話者の性格印象に与える影響. 日本音響学会誌, 56(6), 2000.