

# 自発音声コーパスにおける 講演の音声特徴と印象との関係について

Analyzing the relation between impression ratings and speech properties  
in the *Corpus of Spontaneous Japanese*

山住賢司<sup>†</sup>      籠宮隆之<sup>††</sup>      槇洋一<sup>†</sup>      前川喜久雄<sup>††</sup>  
Kenji YAMASUMI   Takayuki KAGOMIYA   Yohichi MAKI   Kikuo MAEKAWA

(<sup>†</sup> 東京都立大学人文科学研究科   <sup>††</sup> 国立国語研究所研究開発部門第二領域)

<sup>†</sup> Graduate School of Humanities, Tokyo Metropolitan University

<sup>††</sup> Dept. Language Research, The National Institute for Japanese Language

**Abstract** We analyzed the relationship between impression rating and phonetic property of monologue using *the Corpus of spontaneous Japanese* (CSJ). The data consists of 892 monologue of about 10 minutes long uttered by speakers of various generations (20's to 60's). The impression ratings included two types of ratings: (1) selection from a set of evaluation keywords and (2) five-scale rating of phonetic properties like subjective speaking rate and speech style. The phonetic properties like objectively measured speaking rate and pause ratio were derived from the speech data in CSJ. Using six phonetic properties and five impression ratings as variables, a covariance structure analysis was conducted to explain not only the relations between impressions and phonetic properties but also the relations among impressions. Most striking finding was that subjective speaking rate affects the Fair-Poor evaluation of speech. Viewed as a whole, the faster the perceived rate, the higher the evaluation. Note, however, when the rate is greater than 9.0 mora/sec, the relation was reversed. Extremely fast speaking rate lowered the evaluation.

## 1 はじめに

我々是对話場面などにおいて日常的に話し相手の言葉に対してその意味を受け取ると同時に、何らかの印象も抱いている。例えば、話し方が速い又は遅いとか、上手い又は下手だと感じたりするであろう。そういった印象を評定するのに際して、発話のどのような音声の特徴が影響するかについての報告はすでに行ってきた<sup>1) 2)</sup>。籠宮ら(2001)では講演の評価にはどのような講演音声の特徴が影響するかについて検討しており<sup>1)</sup>、さらに籠宮ら(2002)においては、発話速度の知覚に大きな役割を果たす講演音声特徴の特定を行っている<sup>2)</sup>。

しかしこれらの先行研究においては、特定の1つの印象と音声的特徴との関係をみるにとどまっており、全体として種々の印象がどう関係しており、それらと音声的特徴がどのように結びついているのかは検討されていない。また、人がどのような発話に対してどのような印象を受けるのかを明らかにする上で、音声特徴のあるパラメーターが異なる複数の印象に対して、それぞれどのような影響を及ぼすのかを知ること、重要である。

そこで今回の研究では上記の点について自発的発話の音声的特徴とその印象評定に対する因果モデルを作成し検討を行った。

## 2 方法

### 2.1 音声データ

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「話し言葉の言語的・パラ言語的構造の解明に基づく『話し言葉工学』の構築」の一環として作成中の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>』に収録された音声データの一部を用いた。

今回分析に使用したのは、国立国語研究所で企画・収録された模擬講演の音声データである。模擬講演とは、いくつかの大まかなテーマ(例えば「人生をふりかえって印象に残っていること」など)を与え、そのテーマに従って講演者が具体的な内容を自身で考えて、10～15分程度講演してもらうものであり、20代～60代までの男女の計 892 講演のデータを使用した。聴衆は収録スタッフを含めた 3～5 名ほどであった。

### 2.2 講演に対する印象評定

音声データの収録時に収録スタッフが講演に対する印象評定を行った。評定者は講演ごとに異なり、1講演につき1名の評定者が評定をした。計8名の評定者が印象評定を行ったが、特定の1人の評定者による評定が半数以上を占めている。

印象評定は、①話し方や声に対する印象について31語の評定語から当てはまると思うものにチェックするもの(評定語選択式)、②講演の特徴として「発話速度(速い-遅い)」「講演の自発性(高い-低い)」「発音の明瞭さ(明瞭-不明瞭)」「発話スタイル(あらたまった-くだけた)」について5段階尺度で評定させたもの、の2種類から成る。

①の印象評定に用いた評定語の一覧を表1に示す。これらの評定語は主に話し方について記したものと、声

の特徴について記したものから成っている。評定語の多くは反意関係をなす対(例えば「たどたどしい」-「流暢な」など)であるが、そうでないもの(「きっぱりした」「優しい」など)もある。

なお印象評定は、およそ講演予定時間の半分の時間が経過した時点で行っており、その際、講演の特定の部分に対してでなく、その時点までの講演全体に対しての印象を評定するよう、評定者に指示した。

表1. 評定語選択式の印象評定で用いた評定語

たどたどしい	流暢な	単調な	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	年寄りじみた
元気のある	元気のない	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				

### 2.3 分析方法

#### 2.3.1 評定語選択式の印象評定データの取り扱い

評定語選択式の印象評定については、カテゴリカルなデータ(「あてはまる」か「あてはまらない」かの2値データ)であるため、数量化Ⅲ類を行い軸を抽出して印象評定から見た各講演の心理的距離を求めた。なおチェックされた講演数が35講演以下であった「いらいらした」「裏返った声」「重厚な」「軽薄な」「生意気な」「尊大な」の6語については、分析の対象から除外した。

その結果、第1軸の固有値は0.653と高く、「きっぱりした」「自信のある」「元気のある」といった肯定的表現の語が高い正のカテゴリースコアを取り、一方「落ち着きのない」「たどたどしい」「神経質な」といった否定的表現の語が高い負のカテゴリースコアを取っている(表2)。

表2. 各評定語の数量化Ⅲ類第1軸のカテゴリースコア

自信の無い	落ち着きのない	たどたどしい	緊張した	元気の無い	聞き取りにくい	単調な	小さい声	こもった声	かすれた声	年寄りじみた	低い	高い
-2.933	-2.467	-2.292	-2.115	-1.721	-1.672	-1.534	-1.248	-0.920	-0.814	-0.775	-0.640	-0.273
鼻にかかった	優しい	落ち着いた	リラックスした	聞き取りやすい	表情ゆたかな	大きい声	若々しい	流暢な	元気のある	自信のある	きっぱりした	
-0.131	-0.031	0.244	0.557	0.561	0.609	0.671	0.696	0.703	0.803	0.846	1.004	

このことから第1軸を「肯定的-否定的」評価の軸と解釈しそのサンプルスコアを「講演の評価」に対する尺度値として以後の分析では取り扱った。すなわち数量化Ⅲ類第1軸のサンプルスコアが高い講演ほど肯定的に評価されており、逆にサンプルスコアが低い講演は否定的に評価されていると解釈される。

### 2.3.2 講演音声の特徴量

講演音声からわかる特徴量として以下を音声データの書き起こしテキストから算出した。

1. モーラ数/ポーズ抜き秒 (総モーラ数/発話間のポーズ時間を除いた講演時間)
2. 文節数/ポーズ抜き秒
3. ポーズ比 (発話間のポーズ総時間/総講演時間)
4. ポーズ数/秒
5. 単独の笑い数/秒
6. 笑いながら発話数/秒

なお、5・6における笑いとは、講演者自らが発した笑いのことを指している。

上記の特徴量について因子分析(主成分法・斜交プロマックス回転)を行い3因子を抽出した。因子分析の結果は表 3、4 に示されている。そして、これらの3因子に対して、第1因子からそれぞれ「発話量」「ポーズ」「笑い」の因子と命名し、講演音声の特徴を表す潜在変数として以後の分析で取り扱った。因子間相関からは、これらの因子は互いに独立したものと考えられる。

表 3. 因子負荷量行列(因子パターン)

	Factor 1	Factor 2	Factor 3	共通性
モーラ数	0.945	-0.083	-0.062	0.894
文節数	0.940	0.061	0.076	0.897
単独の笑い数	-0.087	0.866	0.061	0.743
笑いながら発話数	0.066	0.839	-0.076	0.732
ポーズ数	0.204	0.044	0.853	0.754
ポーズ比	-0.205	-0.059	0.821	0.741

表 4. 因子間相関

	Factor 1	Factor 2	Factor 3
Factor 1	1.000		
Factor 2	0.080	1.000	
Factor 3	-0.039	-0.064	1.000

### 2.3.3 共分散構造分析

以上、講演音声に対する6つの特徴量と、講演の評価に関する尺度値(数量化Ⅲ類第1軸スコア)、発話速度・講演の自発性・発音の明瞭さ・発話スタイルの5段階評定値を用い、「講演音声の特徴(因子分析により抽出された3因子)」と「講演の評価・発話速度・講演の自発性・発音の明瞭さ・発話スタイル」との関係について因果モデルを構築し、SAS(Ver. 8.2)のCALISプロシジャを用いて共分散構造分析を行った。

5段階評定の尺度については、「発話速度(速い-遅い)」「講演の自発性(高い-低い)」「発音の明瞭さ(明瞭-不明瞭)」「発話スタイル(あらたまった-くだけた)」それぞれに5~1の尺度値を与え数値化してある。例えば「発話スタイル(あらたまった-くだけた)」については尺度値5が「あらたまった」発話スタイルを示し、尺度値3は「普通」な発話スタイル、尺度値1が「くだけた」発話スタイルを表している。

因果モデルを構築するにあたっては、講演を聞いてから印象評定を行うという時間的順序を考慮し、因子分析により得られた講演音声の特徴をあらわす3つの潜在変数(発話量・ポーズ・笑い)が講演の評価などの印象評定にそれぞれ影響を及ぼすという仮説をたてパス図を作成した。

なお印象評定に関してはモデルの識別性を保証するため、各印象評定の尺度値が1つの潜在変数を構成するという1指標条件<sup>4)</sup>を適用した。これは潜在変数から観測変数へのパスの因果係数を1.0に固定し、同時に誤差変数の分散が0であるという仮定をおくものである。

また各印象評定間でもお互いに影響を与えている可能性を調べるため、各印象評定間に異なる因果パスを設定した複数のモデルを構築し、どのようなパスを含むモデルが最適なものかを比較・検討した。

## 3 結果と考察

共分散構造分析に用いた変数の相関係数行列を表5に示す。

表 5. 分析に用いた全変数間の相関係数行列

	ポーズ比	ポーズ数	モーラ数	文節数	笑い数	笑いながら数
ポーズ比	1.000					
ポーズ数	0.494	1.000				
モーラ数	-0.230	0.022	1.000			
文節数	-0.143	0.129	0.780	1.000		
笑い数	-0.033	-0.032	-0.046	0.082	1.000	
笑いながら数	-0.191	-0.008	0.093	0.207	0.469	1.000
講演の評価	-0.460	-0.124	0.295	0.208	0.058	0.175
発話速度	-0.538	-0.103	0.474	0.367	0.061	0.128
自発性	-0.163	-0.005	0.201	0.279	0.159	0.290
明瞭さ	-0.216	-0.131	-0.006	-0.098	-0.058	-0.013
発話スタイル	0.092	0.030	-0.150	-0.278	-0.266	-0.350

	講演の評価	発話速度	自発性	明瞭さ	発話スタイル
講演の評価	1.000				
発話速度	0.389	1.000			
自発性	0.303	0.234	1.000		
明瞭さ	0.373	0.070	0.054	1.000	
発話スタイル	-0.164	-0.113	-0.230	0.199	1.000

### 3.1 モデルの評価

モデルの適合度(GFI・AGFI)・情報量(AIC)に関する指標を参考にし、最終的に採択したモデルのパス図(標準化解)を図1に示す。

パス図において四角で囲まれた変数名は直接測定された観測変数を表し、楕円で囲まれた変数名は直接に観測できない潜在変数を表す。矢印の向きは因果関係の方向を表し、矢印に付記された数値は標準化された因果係数の値を表す。なおパス図中に記された e1-e11 並びに d1-d5 はそれぞれ測定方程式並びに構造方程式における誤差変数を意味する。また e1 と e3、e1 と e4、e2 と e3 の間にはモデルの適合度上昇のため、共分散を仮定している。これらの誤差変数に対応する観測変

数間にはやや低いものの相関関係が存在しており(表5参照)この仮定は妥当なものであると考えられる。

全ての因果係数は5%水準で有意である。モデルの適合度に関する指標の値は GFI が 0.967、AGFI が 0.929 といずれも高く、構成されたモデルはデータの標本共分散行列をよく説明していると判断される。

講演音声の特徴に関する潜在変数から、その潜在変数を構成する各観測変数へのパスの因果係数に関しては、標準化されているにも関わらず一部で 1.00 を超えているものがあるが(ポーズからポーズ比への因果係数)、これは潜在変数を構成する観測変数間の内部相関の高さから生じた多重共線のためだと考えられる。この場合、因子分析の結果自体は良好であったことから、因果係数は不安定となっているが構成概念そのものは安定していると解釈して差し支えないであろう。すなわち、潜在変数から観測変数へのパスの因果係数の大きさについて解釈することは出来ないが、因果モデルを構成する潜在変数としては安定した構造を持っていると見なすことが出来る。

### 3.2 モデルの解釈

本モデルからは、発話量が多くポーズが少ないものを速い発話と評定していることが示されている。また発話量・笑いが多く発話スタイルがぐだけており、ポーズの少ないものを自発性が高い講演と評定することも示され

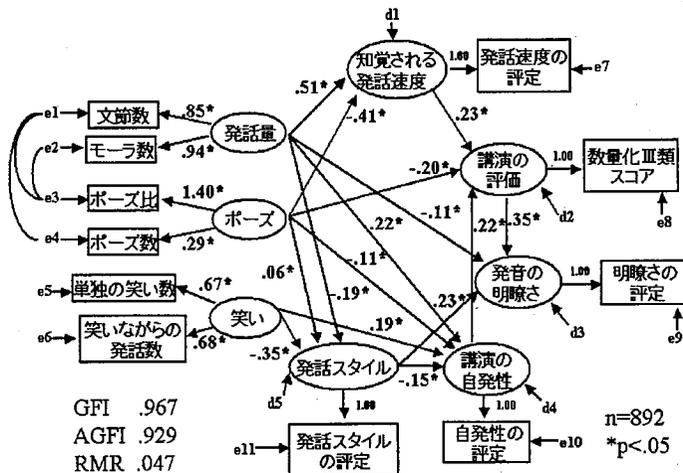


図1. 因果モデルのパス図

ている。講演の評価には知覚される発話速度・講演の自発性・ポーズといった要因が寄与している。発音の明瞭さでは講演の評価が肯定的で、あらたまった発話スタイルで、発話量の少ないものを明瞭な発話と評価している。

先行研究<sup>1)2)</sup>で確認された、講演の評価並びに知覚される発話速度と、講演音声の特徴との関係は、今回のモデルでも有意なパスとして示されている。すなわち、講演の評価に関してはポーズ比が大きな役割を果たしており、発話速度の知覚にはポーズ比・モーラ数/秒が共に大きな影響を与えているという結果と、本モデルで示されている因果関係は矛盾していない。

加えて本モデルでは講演の評価には知覚される発話速度が影響するということが確認できた。このことは講演音声の特徴と印象評定の関係を単独に眺めているだけでは明らかに出来なかったものである。本研究のような自発音声をういたものではないが、音声合成された比較的短い発話の速度を変化させ、推測される発話者の性格特性の変化を検討した内田(2000)の研究もある。その結果は「外向性、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<sup>5)</sup>」といったもので、推測される性格特性も発話から受ける印象の一種と考えれば、本研究の結果とも一致するものである。

講演の評価には、上記の知覚される発話速度以外にも、講演の自発性も影響しており、発話スタイルは間接的ながら講演の評価に影響していることもわかる。講演音声の特徴だけでなく、それにより形成されたある印象が他の印象にも影響を与えているという複雑な図式が明らかになっている。現実場面において発話から受ける印象というものは、もちろん1つだけということではなく、同時に又は継時的に複数の印象を抱くことだろう。そうした現実には則って考えるなら、単独の印象について焦点を絞った研究のみならず、複数の印象間の関係にも着目した研究を進めていく必要があると考えられる。今回のモデルは発話の印象に関わる現実の事象全てを網羅したものではないにせよ、印象間の複雑な関係の一端を明らかにしている。

### 3.3 発話速度別のモデルの検討

本モデルでは知覚される発話速度から講演の評価へのパスの因果係数は正の値となっており、つまり発話速度が速いほど講演の評価が高くなるという関係が示されている。しかし発話速度が速い方が良いといっても、それには限度が存在するであろう。すなわち発話速度が一定の速さを超えると講演の評価がそれ以上高くない、もしくは低下するのではないかと予測される。

前述の内田(2000)の結果においても性格特性の評価のピークは「やや速い」発話速度であった。このようなことから印象評定においても同様にモードが変化するポイントが存在している可能性がある。

そこで知覚される発話速度に影響を与える観測変数のうち「モーラ数/ポーズ抜き秒」「ポーズ比」について、散布図を考慮しながらいくつかのポイントでデータを二分割し、講演の評価との相関係数の変化を調べた(表6)。

表6. データ分割による相関係数の変化

上段:モーラ数/ポーズ抜き秒による分割

下段:ポーズ比による分割

左列に分割ポイントの値を示し、そのポイントでデータを分割した際の、講演の評価との相関係数をそれぞれ示してある

分割ポイント(モーラ/s)	ポイント前相関係数	ポイント後相関係数
6.0	0.369	0.195
7.0	0.037	0.067
8.0	0.237	-0.222
8.5	0.310	-0.368
9.0	0.303	-0.523
10.0	0.265	-0.418
11.0	0.228	0.968

分割ポイント(ポーズ比)	ポイント前相関係数	ポイント後相関係数
0.15	-0.171	-0.437
0.20	-0.155	-0.378
0.25	-0.207	-0.307
0.30	-0.284	-0.248
0.35	-0.364	-0.265
0.40	-0.417	-0.465
0.45	-0.420	-0.651

その結果、「ポーズ比」に関してはデータを二分割することで相関係数が大きく異なるポイントはなかったが、「モーラ数/ポーズ抜き秒」においては、9.0 モーラ/ポーズ抜き秒でデータを二分割した際、最も劇的に相関係数の値が変化することがわかった。そこで全データ

を 9.0 モーラ/ポーズ抜き秒を境界として二分割し、図1のモデルに対して、それぞれのデータをあてはめて、因果係数の変化を比較した(図2, 3)。

図1のモデルと比べてパスのいくつかは有意でなくなっており、特に図3のモデルにおいてその傾向は著しい。これは図3のモデルではデータ数が 65 と極端に減少していることの影響があらわれていると考えられる。しかし図2・3のモデルともに知覚される発話速度から講演の評価に対するパスは有意なまま残っている。そして、その因果係数の値は見事に逆転しており、図2のモデルでは知覚される発話速度から講演の評価へのパスの因果係数は 0.32 であるのに対して、図3のモデルでは -0.35 となっている。すなわち 9.0 モーラ/ポーズ抜き秒までのデータでは講演の評価は発話速度の上昇にともない高くなっているが、9.0 モーラ/ポーズ抜き秒を超えるデータでは逆に講演の評価は低下する。

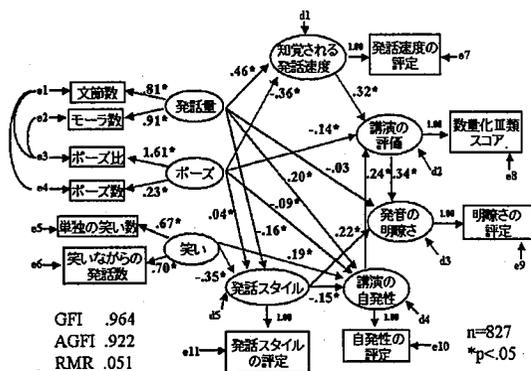


図2. 9.0(モーラポーズ抜き秒)未満のデータによる因果モデルのパス図

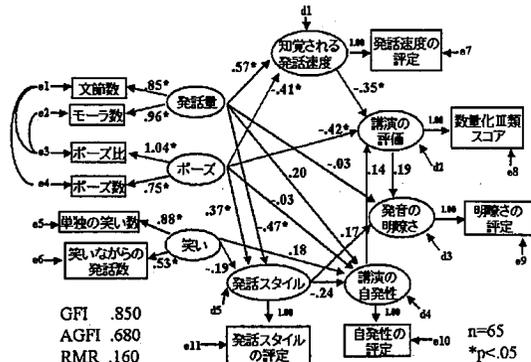


図3. 9.0(モーラポーズ抜き秒)以上のデータによる因果モデルのパス図

予測されたように、ある発話速度までで講演の評価の上昇は頭打ちとなり、それ以降は逆に発話速度が速くなるとむしろ評価が悪くなるポイントが存在することが、図2・3のモデルによって示された。

## 4 まとめ

本研究では「日本語話し言葉コーパス」に収録されている模擬講演のデータを用い、講演音声の特徴と印象評定との関係について共分散構造分析を用いて検討した。音声特徴と印象評定との関係は先行研究と一致するものであったが、さらに印象評定間の関係性についても明らかになった。

特に知覚される発話速度と講演の評価との関係は、線形的なものでなく、発話速度によって関係性が逆転するような性質のものであることもわかった。

今回は講演音声の特徴として6つの特徴量のみ用いたが、現在『日本語話し言葉コーパス』作成において進められている韻律ラベリング<sup>6)</sup>が完了することで韻律情報や分節音情報なども利用できるようになる。今後は、それらの変数を組み込んだ上で、さらなるモデルの洗練を目指していきたい。

## 参考文献

- 1) 籠宮隆之, 榎洋一, 菊池英明, 前川喜久雄. 自発音声コーパスにおける印象評定とその要因. 日本音響学会 2001 年秋季研究発表会講演論文集, 2001.
- 2) 籠宮隆之, 山住賢司, 榎洋一, 前川喜久雄. 発話速度の違いが印象評定に及ぼす影響. 日本音響学会 2002 年春季研究発表会講演論文集, 2002.
- 3) 前川喜久雄, 籠宮隆之, 小磯花絵, 小椋秀樹, 菊池英明. 日本語話し言葉コーパスの設計. 音声研究, 4(2), 2000.
- 4) 豊田秀樹. SASによる共分散構造分析. 東京大学出版会, 1992.
- 5) 内田照久. 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.
- 6) 前川喜久雄, 菊池英明, 五十嵐陽介. X-JToBI: 自発音声の韻律ラベリングスキーム. 電子情報通信学会技術報告 SP2001, 2001.

謝辞 共分散構造分析を用いたデータの分析にあたり大学入試センターの内田照久氏に有益なコメントをいただきました。ここに深く感謝いたします。