

講演音声に対する印象評定尺度の作成

籠宮 隆之（国語研）山住 賢司・榎 洋一（都立大 / 国語研）前川 喜久雄（国語研）

1 はじめに

国立国語研究所・通信総合研究所・東京工業大学では、1999年度より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 [1] によって『日本語話し言葉コーパス』 [2] を作成し、話し言葉に関する様々な研究を行ってきた。その一環として、これまでに、収録時に収録作業者が評定した予備的な印象評定値を分析し、印象評定に関わる講演音声の特徴について報告してきた [3],[4]。また、幾つかの音声的・言語的変異が、「自発性」「明瞭さ」などの印象評定値と関連していることを報告してきた [5],[6]。

ただし、これらの研究で用いられた印象評定データベースには、1) 1つの音声につき1名のみによる評定データであり、評定者の世代差・年齢差なども考慮していない。2) 講演時間の中程で評定したデータであり、時間推移に伴う印象の変化を表していない。—— などの問題点があった。

そこで、『日本語話し言葉コーパス』の一部のデータに対し、1講演につき複数箇所の印象を多数の一般の人により評定させたデータベースを作成することにした。これに伴い、従来の評定項目も見直し、新たな音声評定に関する心理測定尺度を作成した。本発表では、この心理測定尺度の詳細および作成過程を報告する。また、実際に収集した印象評定データに関する簡単な分析も報告する。

2 印象評定尺度の作成

2.1 尺度構成の目的

質問紙などによりある物事や現象に対する印象などの心理的な値を安定して測定するためには、一つの対象に対して一つの質問をするのではなく幾つかのまとまった質問を行い、その合計点をもって対象の得点とする心理尺度法が用いられる。今回の印象データ収集においても、心理尺度法を用いることにする。これまでも音声を与える印象に対する評定尺度は幾つか構成されてきた [7] が、これらは専ら単語・単文の朗読音声を対象としており、今回対象とするようなある程度の長さをもった自発音声に対する尺度ではなかった。

そこで、今回新たに自発音声に対する印象を評定するための尺度を構成することにした。尺度を構成する際には、先行研究によって音声的・言語的変異と関連が見られた「発話スタイル」「発話速度感」「明瞭さ」「自発性」などの印象をより安定して測定できるように設計した。

2.2 評定語の収集

音声や音楽、ニュースなどに対する印象評定を行っている研究発表・論文などから、用いられている評定語を収集した。また、実際に『日本語話し言葉コーパス』に収録されている講演音声を都立大の学生8名に聴取させ、各講演に対する印象を自由記述させた。この自由記述文より音声の印象を表す形容語を抽出した。収集した評定語のうち、同義語や了解性の低い語を削除した。

このようにして得られた評定語を、「発話スタイル」「発話速度感」「明瞭さ」「自発性」などの概念に対応するように仮説的に分類した。この際、できるだけ対語形式になるようにした。収集した語に対語がないものは、筆者らが協議して付け加えた。

以上の手続きにより、計52の評定語対を作成した。

2.3 尺度構成

上記の52評定語対は仮説的に分類したものであり、それぞれの概念を安定して評定できるとは限らない。そこで、52評定語対を実際に用いて評定実験を行い、安定した尺度を構成する。

2.3.1 第1回評定実験

第1回評定実験は、以下の手続きで行った。

評定者 都立大・文教大の学生，および国語研のアルバイト等の男女 50 名（男性 19 名・女性 31 名）。

刺激音声 『日本語話し言葉コーパス』に収録されている講演より，講演者の性別・年代，および発話速度などがバランスするように考慮して選び出した 6 講演。各講演につき，「講演の冒頭」「講演の中盤」「講演の終盤」を 3.1 節で述べる基準で切り出した，計 18 音声。

質問紙 収集した 52 対語からなる質問紙。評定語に対し 7 段階で回答。

手続き 集団法で試行。刺激音声はスピーカーにより呈示。音声聞き終わってから評定するように指示。回答に際して時間的制限は設けず，全員が回答し終わってから次の音声を呈示した。

上記の 18 音声×50 人の評定データに対し，因子分析（最尤推定法・斜交プロマックス回転）を行い，因子の抽出を行った。この際，因子負荷量が小さいものや複数の因子にまたがって高い因子負荷量を持つ評定語を削除していった。最終的に 41 評定語が残り，5 因子が抽出された。尺度の内の一貫性を示す α 信頼性係数はどの因子も 0.70 を超えており，尺度として安定していた [8]。

抽出された 5 つの因子は以下のように命名し，それぞれを下位尺度として扱う。

第 1 因子 「好きな—嫌いな」「心地よい—不快な」などの評定語対より構成される因子。『好悪』と命名。

第 2 因子 「流暢な—たどたどしい」「上手い—下手な」などの評定語対より構成される因子。『上手さ』と命名。

第 3 因子 「声の大きい—声の小さい」「元氣のある—元氣のない」などの評定語対より構成される因子。『活動性』と命名。

第 4 因子 「速い—遅い」「せわしげな—のんきな」などの評定語対より構成される因子。『速さ感』と命名。

第 5 因子 「礼儀正しい—無礼な」「丁寧な—ぞんざいな」などの評定語対より構成される因子。『スタイル』と命名。

2.3.2 第 2 回評定実験

第 1 回予備評定において 5 因子からなる尺度が得られたが，41 対語と評定項目が多い。そこで，各因子につき因子負荷量の高い 4 語を抽出し，計 20 対語からなる尺度として再構成した。

この 20 項目からなる尺度が 41 項目からなる尺度と 1) 同じ因子構造を持つか。2) 同等の信頼性を示すか —を確認するため，再度評定実験を行った。第 2 回評定の手続きのうち，第 1 回評定と異なるものを以下に示す。

評定者 都立大・文教大の学生，および国語研のアルバイト等の男女 51 名（男性 26 名・女性 25 名）。

質問紙 収集した 20 対語からなる質問紙。各評定語対に対し 7 段階で回答。

上記の 18 音声×51 人の評定データに対し，因子分析（最尤推定法・斜交プロマックス回転）を行い，因子の抽出を行った。表 1，表 2 に，第 2 回評定実験の因子分析結果を示す。その結果，因子の順番の入れ替わりがあったものの，第 1 回評定実験の結果と同様の 5 因子が得られた。また，各評定語対も，第 1 回評定実験と同じ因子に属していた。 α 信頼性係数についても，各因子ともに 0.78 以上と高い値を示しており，20 対語でも尺度として安定していることが分かった [8]。

3 『日本語話し言葉コーパス』コアに対する評定データ収集

以上の手続きにより作成した印象評定尺度を用いて，『日本語話し言葉コーパス』に対する大規模な評定作業を行った。

3.1 対象音声

『日本語話し言葉コーパス』のうち，「コア」[2] と呼ばれるデータセットに対して印象評定を行った。コアを対象とするのは，1) コアには分節音ラベルや韻律ラベル [9] が付与されるため，印象評定結果と音声特徴の比較を行いやすい。2) また，分節音ラベルや韻律ラベルに関わる種々の変異現象の解析に印象評定結果が必要となる。—などの理由による。コアには対話音声や朗読音声も含まれるが，印象評定の対象には，独話の講演（187 講演）のみを用いた。

講演の時間的推移に伴う印象の変化を評定するには，講演を分割する必要がある。また，各講演は短いものでも 8 分以上あり，講演全体に対して印象評定を行なうのは多大な負担がかかる。そこで，以下の方法で各講演から 3 箇所聴取単位を切り出した。

表 1: 第 2 回評定実験における因子分析結果 (因子パターン行列)

		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
好悪	好きな—嫌いな	0.906	0.028	0.075	-0.067	0.010
	心地よい—不快な	0.872	0.096	0.000	-0.067	0.013
	感じの良い—感じの悪い	0.835	-0.097	-0.026	0.067	0.148
	親しみやすい—親しみにくい	0.749	-0.049	-0.115	0.138	-0.109
上手さ	流暢な—たどたどしい	-0.055	0.859	0.042	0.045	0.012
	話し慣れた—話し慣れていない	-0.064	0.846	-0.017	0.039	-0.027
	なめらかな—しどろもどろな	0.014	0.818	0.000	-0.032	0.005
	上手い—下手な	0.116	0.813	-0.009	0.039	-0.008
速さ感	速い—遅い	0.047	0.096	0.952	-0.027	-0.003
	スピード感のある—ゆったりした	0.025	0.079	0.929	0.008	-0.035
	せわしげな—のんきな	-0.178	-0.025	0.794	0.020	0.088
	落ち着きのない—落ち着きのある	-0.110	-0.389	0.571	0.118	-0.069
活動性	声の大きい—声の小さい	-0.107	-0.041	-0.123	0.826	0.064
	力強い—弱々しい	-0.092	0.121	0.031	0.795	-0.008
	元気のある—元気がない	0.237	-0.064	0.049	0.783	-0.073
	積極的な—消極的な	0.067	0.116	0.155	0.656	0.029
スタイル	礼儀正しい—無礼な	0.093	-0.004	-0.009	0.000	0.795
	まじめな—ふまじめな	-0.215	-0.001	0.112	0.066	0.743
	丁寧な—ぞんざいな	0.129	-0.001	-0.140	0.072	0.635
	上品な—下品な	0.171	-0.008	0.008	-0.133	0.573
α 信頼性係数		0.910	0.903	0.904	0.865	0.780

表 2: 因子間相関

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Factor 1 (好悪)	1.000				
Factor 2 (上手さ)	0.317	1.000			
Factor 3 (速さ感)	-0.447	0.068	1.000		
Factor 4 (活動性)	0.118	0.283	0.413	1.000	
Factor 5 (スタイル)	0.356	0.216	-0.187	-0.012	1.000

- 講演を、「講演の冒頭」「講演の中盤」「講演の終盤」に分ける。
- これらの中から意味上のまとまりを持ち、文末表現で終わる箇所を選ぶ。
- この際、200ms 以上のポーズまたは明確な文末表現で区切られた発話単位 [2] で切る。
- 「これから○○というタイトルで発表します」などのようなメタ的表現を述べている箇所は外す。
- 10 秒以上のポーズを含む箇所は対象としない。
- 1 つの聴取単位が 1 分前後の長さになるようにする。

3.2 評定者

人材派遣会社を通じて募集した、20 代男性・20 代女性・50 代男性・50 代女性それぞれ 5 名ずつ、計 20 名。過去に音声学・言語学・心理学に関する仕事に従事したことの無いものを募集した。

3.3 評定項目

尺度構成により作成した印象評定尺度 20 対語。これに単項の項目として、以下の評定語対を追加。

聞き取りやすい—聞き取りにくい、その場で考えて話している—原稿を読み上げている
 あらたまった—くだけた、きまじめな—奔放な、きちんとした—くつろいだ
 甘えた—そっけない、愛想がない—愛想がよい

また、話し方に関する項目とは別に、和田さゆりにより開発された日本語 Big Five 尺度 [10] の短縮版 [11],[12] を用いて、講演者の性格推定を行わせた。表 3 に日本語 Big Five 尺度の項目を掲げる。

3.4 刺激呈示・回答方法

刺激呈示には計算機を用いた。評定者は計算機のサウンドボードよりヘッドホンを通じて音声を聴取した。刺激音声はランダムイズして呈示した。

回答も計算機上のプログラムにより行った。刺激音声終了するとともに評定作業を行なうようにした。評定語もランダムに呈示するようにした。ただし、尺度構成して作成した項目は対語形式に

表 3: Big Five 尺度短縮版

外向性	話し好き, 無口な*, 陽気な, 外向的な
情緒不安定性	悩みがち, 不安になりやすい, 心配症, 気苦労の多い
経験への開放性	独創的な, 進歩的, 洞察力のある, 想像力に富んだ
誠実性 (勤勉性)	いい加減な, ルーズな*, 怠惰な*, 計画性のある
調和性 (協調性)	温和な, 寛大な, 親切な, 協力的な

(*は逆転項目)

なっているが, Big Five 尺度は対語形式になっていない。そこで両者が混在しないようにし, 両者が切り替わる際に評定者に注意を促す警告を表示するようにした。

尺度構成で得られた印象評定尺度, Big Five 尺度ともに 7 段階で評定させた。

3.5 その他

評定作業は, 一日あたり 75~90 聴取単位を行った。長時間・複数日に渡って収集作業をするために, 休憩から作業を再開する時に安定した評定が行えない可能性がある。そこで, 作業再開時などに適宜「コア」以外の講演からダミー音声を配置した。また, 評定者が同一の音声に対して一貫した印象評定が行えるかを確認するために, 「コア」以外の講演から日をおいて 2 回ずつ呈示するものを含めた。

4 分析

以上の手続きにより, 収集した印象評定データの分析を行う。ただし, 今回は評定の一貫性確認のために用いた音声の相関係数が 0.4 以上と比較的良好であった 8 名のデータを用いる。

4.1 評定データの扱い

各項目につき 7 段階で評定させた結果に対し, 1~7 点の得点を与えた。単項の項目としたものは, この得点をその聴取単位に対する項目の得点とする。尺度構成により得られた尺度, 及び Big Five 尺度はそれぞれの下位尺度が 4 項目ずつより構成されるので, 下位尺度に属する項目の得点合計を下位尺度の得点とする。よって各下位尺度につき 4~28 点の得点が得られる。今回の分析では, 上記の得点の評定者 8 名の平均を, 各聴取単位に対する得点として扱う。

4.2 講演特徴との関係

どのような講演の特徴が印象評定に寄与しているかを分析する。今回の分析では, 書き起こしテキスト [13] より得られる以下の項目を講演特徴として扱う。

- 「モーラ/秒」 —— 聴取単位中の総モーラ数 / 聴取単位の発話単位の総計
- 「モーラ/秒のゆれ」 —— 「モーラ/秒」の分散
- 「文節/秒」 —— 聴取単位中の文節数の総計 / 聴取単位の時間
- 「ポーズ比」 —— 聴取単位の発話単位の総計 / 聴取単位の時間
- 「ポーズ/秒」 —— 聴取単位中のポーズ数 / 聴取単位の時間
- 「フィラー/秒」 —— 聴取単位中のフィラーの総計 / 聴取単位の時間
- 「語断片/秒」 —— 聴取単位中の語断片の総計 / 聴取単位の時間
- 「笑いながら/秒」 —— 聴取単位中の笑いながらの発話の総計 / 聴取単位の時間
- 「笑い/秒」 —— 聴取単位中の単独の笑いの総計 / 聴取単位の時間

これらの講演特徴を説明変数, 印象評定結果得られたそれぞれの下位尺度を目的変数とした重回帰分析を行い, モデルを構築し, AIC (赤池情報量基準) や多重共線性の問題を考慮しながら最適なモデルを選択した。この結果, 「上手さ」「速さ感」「活動性」に関して比較的良好な結果が得られた(表 4)。

この結果から, 1) 「活動性」および「上手さ」には「ポーズ比」が大きな役割を果たしている。2) 「速さ感」には「モーラ/秒」と「ポーズ比」が大きな役割を果たしている。—— ことが分かる。これは, 講演の収録時に収集した印象評定の分析 [3],[14],[15] と同様の結果を示している。

表 4: 各下位尺度を目的変数とした重回帰分析

「上手さ」	標準誤差	標準化偏回帰係数	寄与率
ポーズ比	1.211	-0.611	0.390
言い淀み/秒	2.501	-0.280	0.064
モーラ/秒	0.101	0.141	0.045
自由度調整済み決定係数 0.496, $p < 0.001$			
「速さ感」	標準誤差	標準化偏回帰係数	寄与率
モーラ/秒	0.091	0.578	0.412
ポーズ比	1.095	-0.380	0.223
笑い/秒	5.073	-0.104	0.016
言い淀み/秒	2.287	0.072	0.009
自由度調整済み決定係数 0.657, $p < 0.001$			
「活動性」	標準誤差	標準化偏回帰係数	寄与率
ポーズ比	1.261	-0.620	0.335
フィラー/秒	0.823	-0.319	0.029
文節/秒	0.481	0.214	0.023
言い淀み/秒	2.763	-0.129	0.015
自由度調整済み決定係数 0.398, $p < 0.001$			

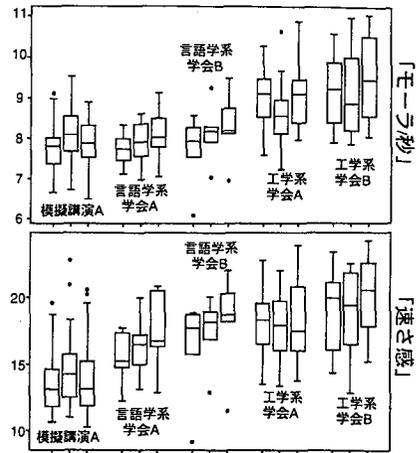


図 1: 講演種別の「モーラ/秒」と「速さ感」の対比

「速さ感」に関しては、講演の時間的推移に伴う発話速度の違いも反映されている。図 1 に、講演種別に、「モーラ/秒」および「速さ感」の得点を掲げる。各講演種別に左より「講演冒頭」「講演中程」「講演終盤」の順に並べる。この図から、講演の時間的推移に伴う「モーラ/秒」の違いに「速さ感」が対応していることが分かる。

「スタイル」に関しては重回帰分析で良好な結果が出なかったが、講演種別で差が見られた。図 2 に、講演種別の「スタイル」の得点を掲げる。模擬講演と学会では、模擬講演の方が「スタイル」の得点が低い。これは、模擬講演の収録の際に講演者をできるだけリラックスさせるようにしたことを反映していると考えられる。特に模擬講演 D の得点が低いのは、模擬講演 D の講演者は学生・大学院生が中心であり、また模擬講演の収録スタッフも 20 代の者が中心のため、よりリラックスして収録が行われたためであると考えられる。

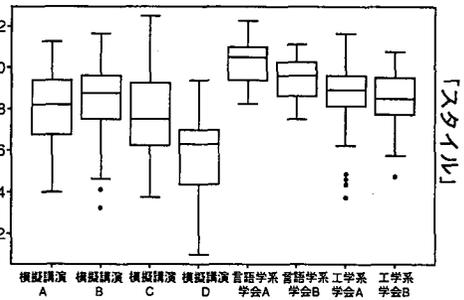


図 2: 講演種別による「スタイル」得点の変化

4.3 各印象項目間の関係

講演の収録時に評定した印象評定の分析では、講演の特徴と印象評定項目の間の他に、それぞれの印象評定項目間でもさまざまな因果関係が見られた [4]。今回収集した各尺度および Big Five 項目間の相関係数を表 5 に掲げる。「好悪」と「上手さ」「外向性」「協調性」との間や、「上手さ」と「外向性」「情緒不安定性」「経験への開放性」との間などに、強い相関関係が見られる。

また、先行研究 [4] では、「発話速度感」が速くなれば「高評価」になるが、ある限界を超えると「低評価」になることを明らかにした。今回のデータにおける「速さ感」とその他の尺度との関係を示す散布図を図 3 に、「速さ感」と Big Five 尺度との関係を示す散布図を図 4 に掲げる。各図ともに x 軸が「速さ感」であり、y 軸がそれぞれの下位尺度である。「速さ感」に対するそれ以外の下位尺度に対する回帰を 1 次式から 3 次式まで計算し、それぞれの回帰式に対して AIC を計算したところ、2 次式が最も AIC が低かった。よって、図 3 には 1 次式と 2 次式による回帰式を掲げた。

どの下位尺度も「速さ感」が中間 (16) 付近までは上昇（「情緒不安定性」は下降）するが、それを境に上昇が止まる、もしくは下降している。これは先行研究 [4] の結果、および内田の研究 [12] に類似した結果である。

表 5: 各下位尺度間および Big Five 尺度との相関係数

	Fac. 1	Fac. 2	Fac. 3	Fac. 4	Fac. 5	Big5 1	Big5 2	Big5 3	Big5 4	Big5 5
Factor1 (好悪)	1.000									
Factor2 (上手さ)	0.656	1.000								
Factor3 (速さ感)	-0.194	0.297	1.000							
Factor4 (活動性)	0.444	0.668	0.528	1.000						
Factor5 (スタイル)	0.419	0.411	0.157	0.163	1.000					
Big5 1 (外向性)	0.703	0.671	0.076	0.699	-0.006	1.000				
Big5 2 (情緒不安定性)	-0.509	-0.658	-0.053	-0.615	0.169	-0.809	1.000			
Big5 3 (経験への開放性)	0.489	0.779	0.489	0.728	0.360	0.592	-0.529	1.000		
Big5 4 (勤勉性)	0.225	0.476	0.518	0.386	0.832	0.013	0.110	0.546	1.000	
Big5 5 (協調性)	0.885	0.491	-0.275	0.321	0.363	0.696	-0.417	0.360	0.134	1.000

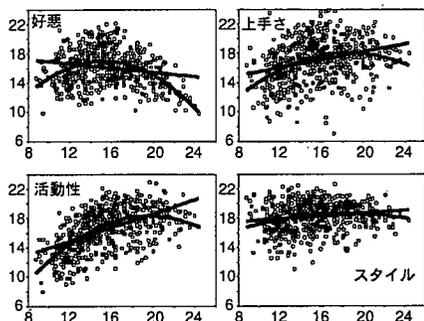


図 3: 「速さ感」(横軸) とその他の下位尺度 (縦軸) との相関

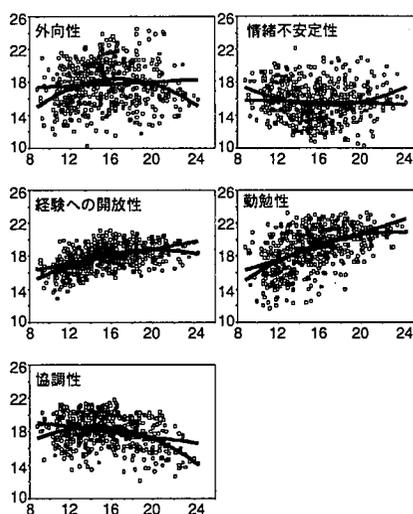


図 4: 「速さ感」(横軸) と Big Five 尺度 (縦軸) との相関

5 まとめと今後の課題

講演音声の特徴を捉える尺度を作成し、実際に『日本語話し言葉コーパス』に付与した。今後は、1) 印象評定項目間の関係について因果関係を探る。2) 分析のパラメータに韻律ラベルを用いる。— など、更に研究を進めたい。

参考文献

- [1] 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).
- [2] 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).
- [3] 籠宮, 榎, 前川: “講演の印象評定に寄与する発話要因”, 第 15 回日本音声学会全国大会予稿集 (2001).
- [4] 山住, 籠宮, 榎, 前川: “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”, 人工知能学会研究会資料 SIG-SLUD-A202 (2002).
- [5] 小磯, 斉藤, 間淵, 前川: “話し言葉における助詞の撥音化現象の実態—『日本語話し言葉コーパス』を用いて—”, 社会言語科学会第 10 回研究大会予稿集 (2002).
- [6] 前川: “話し言葉における長母音の短呼—『日本語話し言葉コーパス』を用いた音声変異の分析—”, 国語学会 2002 年度春季大会要旨集 (2002).
- [7] 木戸, 粕谷: “通常発話の声質に関連した日常語の抽出”, 日本音響学会誌, 55, 6 (1999).
- [8] 山住, 籠宮, 榎, 前川: “講演音声の特徴を捉える評定尺度の構築”, 日本音響学会 2003 年秋季研究発表会講演論文集 (2003).
- [9] 前川, 菊池, 五十嵐: “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 SP2001-106 (2001).
- [10] 和田: “性格特性用語を用いた Big Five 尺度の作成”, 心理学研究, 67, (1996).
- [11] 萩生田, 繁樹: “因子数選択のための新基準の提案”, 心理学発表論文集 (1995).
- [12] 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).
- [13] 小磯, 土屋, 間淵, 斉藤, 籠宮, 菊池, 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”, 日本語科学, 9, (2001).
- [14] 籠宮, 山住, 榎, 前川: “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).
- [15] 榎, 籠宮, 前川: “講演の印象評定に影響を与える発話要因”, 第 8 回社会言語科学会研究大会予稿集 (2001).