

# B 6

## 講演の印象評定に寄与する発話要因

籠宮 隆之 (国立国語研究所)

槇 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)

前川 喜久雄 (国立国語研究所)

### 1 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『話し言葉工学』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』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。

### 2 データ

#### 2.1 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。

##### 2.1.1 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を収録した。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分のもので殆どである。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。

##### 2.1.2 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 — とがある。

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。詳しくは参考文献[4]、[3]を参照。

#### 2.2 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、以下の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。

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・優しい 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・年寄りみだ・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
---

### 3 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。

#### 3.1 学会講演と模擬講演の印象評定からの特性と評定語の分類

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。

ここでは、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軸と第2軸を用いて分析する。表1-1、1-2に各評定語の固有値を示す。また、図1-1、1-2に模擬講演の第1軸と第2軸の値による各評定語の散布図、及び各講演の散布図を示す。同様に、図1-3、1-4に学会講演の散布図を示す。

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の傾向に差を与えていないと解釈できる。

#### 3.2 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労力を要する。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

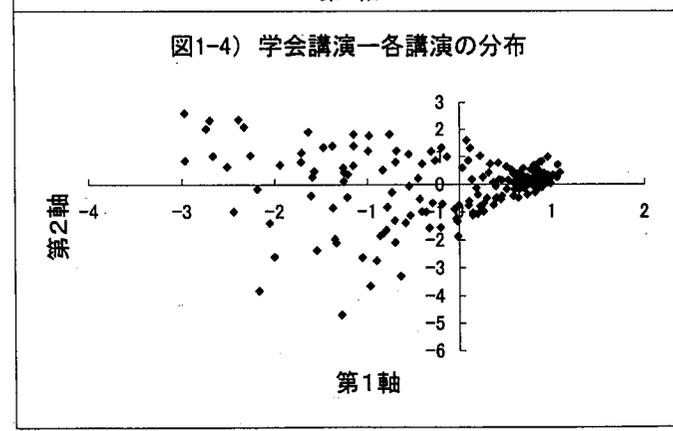
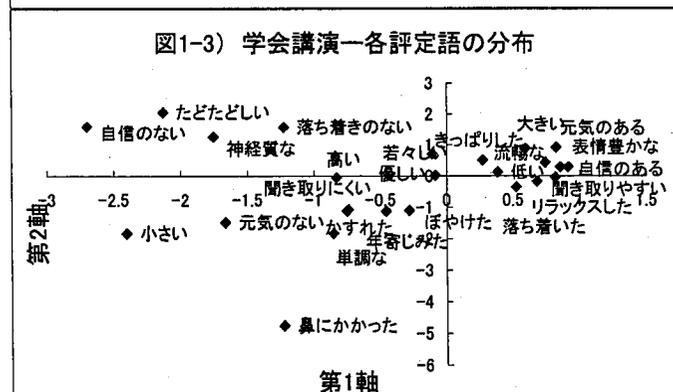
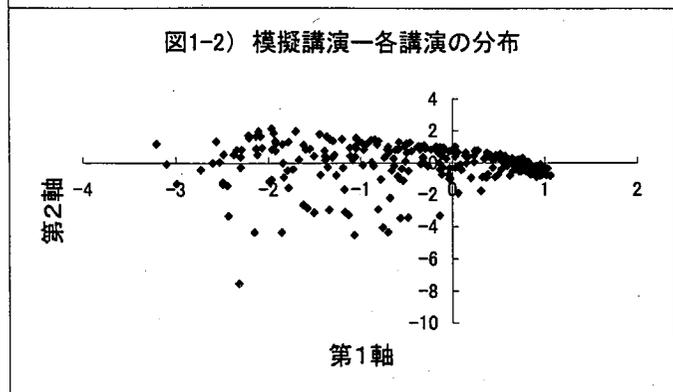
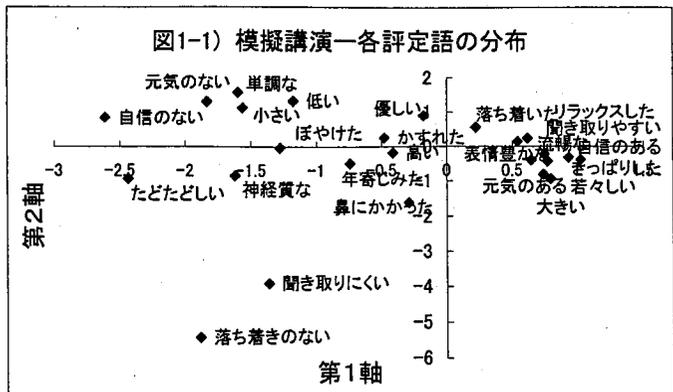


表1-1) 模擬講演に関する数量化Ⅲ類から得られた各評定語の固有値

	第1軸	第2軸
たどたどしい	-2.438	-0.919
流暢な	0.743	-0.204
単調な	-1.603	1.580
表情豊かな	0.644	-0.378
自信のある	0.929	-0.309
自信のない	-2.615	0.849
優しい	-0.182	0.904
落ち着いた	0.221	0.561
落ち着きのない	-1.890	-5.410
神経質な	-1.624	-0.821
リラックスした	0.614	0.236
大きい	0.793	-0.930
小さい	-1.565	1.135
かすれた	-0.483	0.268
ぼやけた	-1.280	-0.025
若々しい	0.768	-0.424
年寄じみた	-0.744	-0.474
元気のある	0.738	-0.801
元気のない	-1.837	1.303
聞き取りやすい	0.541	0.147
聞き取りにくい	-1.369	-3.912
鼻にかかった	-0.295	-1.601
高い	-0.417	-0.167
低い	-1.181	1.331
きっぱりした	1.018	-0.365

表1-2) 学会講演に関する数量化Ⅲ類から得られた各評定語の固有値

	第1軸	第2軸
たどたどしい	-2.134	2.060
流暢な	0.853	0.295
単調な	-0.858	-1.827
表情豊かな	0.740	0.453
自信のある	0.912	0.303
自信のない	-2.702	1.596
優しい	-0.088	0.027
落ち着いた	0.520	-0.340
落ち着きのない	-1.230	1.585
神経質な	-1.759	1.269
リラックスした	0.680	-0.160
大きい	0.590	0.894
小さい	-2.407	-1.820
かすれた	-0.755	-1.099
ぼやけた	-0.286	-1.083
若々しい	-0.113	0.710
年寄じみた	-0.460	-1.114
元気のある	0.821	0.926
元気のない	-1.670	-1.478
聞き取りやすい	0.816	-0.037
聞き取りにくい	-0.745	-1.080
鼻にかかった	-1.231	-4.747
高い	-0.834	-0.048
低い	0.381	0.132
きっぱりした	0.267	0.526

表2-1～2-14) 学会講演・模擬講演における評定語ごとの各発話特徴の平均値および検定の結果

表2-1 「たどたどしい」

	学会講演		模擬講演	
	×	○	×	○
データの個数	262	26	473	45
モーラ数(/1分)	398.56	359.57 **	371.32	301.56 **
発話速度のゆれ	3.88	3.85	3.90	3.90
フィラー数(/1分)	12.80	11.85	10.27	8.97
言いよどみ数(/1分)	2.12	2.51	1.74	1.74
単独の笑い数(/1分)	0.02	0.02	0.16	0.12
発話中の笑い数(/1分)	0.08	0.05	0.54	0.37

表2-2 「流暢な」

	学会講演		模擬講演	
	×	○	×	○
データの個数	216	72	374	144
モーラ数(/1分)	384.08	427.90 **	355.17	391.49 **
発話速度のゆれ	3.88	3.88	3.90	3.89
フィラー数(/1分)	12.19	14.28 **	9.96	10.67
言いよどみ数(/1分)	2.12	2.24	1.68	1.89
単独の笑い数(/1分)	0.02	0.02	0.14	0.20
発話中の笑い数(/1分)	0.08	0.09	0.46	0.71 **

表2-3 「単調な」

	学会講演		模擬講演	
	×	○	×	○
データの個数	224	64	466	52
モーラ数(/1分)	401.56	372.15 **	369.02	331.56 **
発話速度のゆれ	3.88	3.86	3.90	3.90
フィラー数(/1分)	12.94	11.90	10.27	9.20
言いよどみ数(/1分)	2.19	2.03	1.75	1.63
単独の笑い数(/1分)	0.02	0.01 **	0.17	0.02 **
発話中の笑い数(/1分)	0.10	0.02 **	0.57	0.17 **

表2-4 「表情豊かな」

	学会講演		模擬講演	
	×	○	×	○
データの個数	257	31	353	165
モーラ数(/1分)	392.01	420.15 **	359.30	378.01 **
発話速度のゆれ	3.88	3.85	3.91	3.88
フィラー数(/1分)	12.81	11.87	10.20	10.07
言いよどみ数(/1分)	2.18	1.93	1.68	1.87
単独の笑い数(/1分)	0.02	0.02	0.09	0.30 **
発話中の笑い数(/1分)	0.07	0.12	0.36	0.89 **

表2-5 「自信のある」

	学会講演		模擬講演	
	×	○	×	○
データの個数	219	69	424	94
モーラ数(/1分)	387.56	418.76 **	355.81	407.88 **
発話速度のゆれ	3.88	3.88	3.90	3.88
フィラー数(/1分)	12.46	13.51	9.96	11.05
言いよどみ数(/1分)	2.11	2.28	1.74	1.74
単独の笑い数(/1分)	0.02	0.02	0.17	0.07 **
発話中の笑い数(/1分)	0.07	0.11	0.54	0.49

表2-6 「自信のない」

	学会講演		模擬講演	
	×	○	×	○
データの個数	265	23	491	27
モーラ数(/1分)	398.14	359.33 **	368.27	310.62 **
発話速度のゆれ	3.88	3.89	3.90	3.91
フィラー数(/1分)	12.88	10.82	10.27	8.10
言いよどみ数(/1分)	2.16	2.08	1.76	1.26 **
単独の笑い数(/1分)	0.02	0.00	0.15	0.17
発話中の笑い数(/1分)	0.08	0.02 **	0.54	0.32 **

表2-7 「優しい」

	学会講演		模擬講演	
	×	○	×	○
データの個数	158	34	417	101
モーラ数(/1分)	400.11	391.42 **	367.22	357.17 **
発話速度のゆれ	3.86	3.83	3.91	3.87
フィラー数(/1分)	12.93	10.63	10.29	9.62
言いよどみ数(/1分)	2.15	1.82	1.82	1.39
単独の笑い数(/1分)	0.01	0.02	0.17	0.09 **
発話中の笑い数(/1分)	0.05	0.07 **	0.57	0.34 **

表2-8 「落ち着いた」

	学会講演		模擬講演	
	×	○	×	○
データの個数	150	138	251	267
モーラ数(/1分)	395.73	394.29	366.57	364.03
発話速度のゆれ	3.85	3.91	3.89	3.90
フィラー数(/1分)	12.64	12.79	10.31	10.02
言いよどみ数(/1分)	2.18	2.12	1.97	1.52 **
単独の笑い数(/1分)	0.01	0.02	0.20	0.11 **
発話中の笑い数(/1分)	0.08	0.07	0.62	0.44 **

表2-9 「神経質な」

	学会講演		模擬講演	
	×	○	×	○
データの個数	230	58	424	94
モーラ数(/1分)	398.48	381.39 **	372.04	334.69 **
発話速度のゆれ	3.88	3.88	3.91	3.87
フィラー数(/1分)	12.65	12.96	10.36	9.25
言いよどみ数(/1分)	2.08	2.42	1.71	1.88
単独の笑い数(/1分)	0.02	0.02	0.16	0.15
発話中の笑い数(/1分)	0.08	0.09	0.54	0.46

表2-10 「リラックスした」

	学会講演		模擬講演	
	×	○	×	○
データの個数	220	68	355	163
モーラ数(/1分)	391.91	405.15	359.99	376.74 **
発話速度のゆれ	3.87	3.90	3.91	3.88
フィラー数(/1分)	12.70	12.76	9.80	10.94 **
言いよどみ数(/1分)	2.10	2.31	1.75	1.71
単独の笑い数(/1分)	0.02	0.02	0.14	0.18
発話中の笑い数(/1分)	0.08	0.09	0.43	0.74 **

表2-11 「元気のある」

	学会講演		模擬講演	
	×	○	×	○
データの個数	265	23	440	78
モーラ数(/1分)	392.51	424.14 **	361.11	388.71 **
発話速度のゆれ	3.88	3.87	3.90	3.88
フィラー数(/1分)	12.63	13.62	10.18	10.07
言いよどみ数(/1分)	2.14	2.28	1.76	1.59
単独の笑い数(/1分)	0.02	0.00 **	0.14	0.24 **
発話中の笑い数(/1分)	0.08	0.07	0.49	0.73 **

表2-12 「元気のない」

	学会講演		模擬講演	
	×	○	×	○
データの個数	268	20	480	38
モーラ数(/1分)	398.48	348.94 **	367.85	332.58 **
発話速度のゆれ	3.88	3.90	3.90	3.89
フィラー数(/1分)	12.73	12.43	10.24	9.10
言いよどみ数(/1分)	2.16	2.05	1.76	1.47
単独の笑い数(/1分)	0.02	0.01	0.17	0.02 **
発話中の笑い数(/1分)	0.08	0.02 **	0.56	0.16 **

表2-13 「聞き取りやすい」

	学会講演		模擬講演	
	×	○	×	○
データの個数	200	88	275	243
モーラ数(/1分)	391.40	403.29	361.26	369.79
発話速度のゆれ	3.87	3.91 **	3.93	3.86 **
フィラー数(/1分)	12.83	12.45	10.49	9.79
言いよどみ数(/1分)	2.20	2.03	1.84	1.62 **
単独の笑い数(/1分)	0.02	0.02	0.18	0.12 **
発話中の笑い数(/1分)	0.07	0.11	0.54	0.51

表2-14 「きつぱりした」

	学会講演		模擬講演	
	×	○	×	○
データの個数	170	21	483	34
モーラ数(/1分)	397.41	407.75	363.16	394.23 **
発話速度のゆれ	3.86	3.83	3.90	3.90
フィラー数(/1分)	12.70	10.97	10.15	9.88
言いよどみ数(/1分)	2.13	1.83	1.77	1.33 **
単独の笑い数(/1分)	0.01	0.00 **	0.16	0.05 **
発話中の笑い数(/1分)	0.05	0.04	0.54	0.29 **

各表2-1における○は「あてはまる」、×は「あてはまらない」を示す。  
 \*\*はT検定において5%水準以上で有意差が認められたことを示す。  
 「」内の単語は評定語を示す。

含まれている。また、フィラ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に従った。書き起こし方法や、フィラーや言い淀みの認定の詳細については、参考文献[1]を参照。

- 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
- 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
- フィラー数 (1分あたり) —— 講演の総フィラー数 / 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 2-1 から表 2-14 に、各評定語ごとに発話の特徴の平均値、および t 検定の結果を示す。模擬講演と学会講演とに分けて示した。

## 4 結果

以上の分析から、以下の結果が得られた。

「たどたどしい」と評定された講演 学会講演においては、そうでない講演よりも発話速度が遅かった。模擬講演においてもまた同様の傾向を示した。

「流暢な」と評定された講演 学会講演においては、そうでない講演よりも発話速度が速く、フィラーが多かった。模擬講演においては、発話速度が速く、笑いの数が多かった。

「単調な」と評定された講演 学会講演においては、そうでない講演よりも発話速度が遅く、笑いの数が少なかった。模擬講演においてもまた同様の傾向を示した。

「表情豊かな」と評定された講演 学会講演においては、そうでない講演よりも発話速度が速かった。模擬講演においては、発話速度が速く、笑いの数が多かった。

「自信のある」と評定された講演 学会講演においては、そうでない講演よりも発話速度が速かった。模擬講演においては、発話速度が速く、笑いの数が少なかった。

「自信のない」と評定された講演 学会講演においては、そうでない講演よりも発話速度が遅く、笑いの数が少なかった。模擬講演においては、発話速度が遅く、言い淀みが少なく、笑いの数が少なかった。

「優しい」と評定された講演 学会講演においては、そうでない講演よりも発話速度が遅く、笑いの数が多かった。模擬講演においても、同様の傾向を示した。

「落ち着いた」と評定された講演 学会講演においては、どの項目でも違いがみられなかったが、模擬講演においては、言い淀みが少なく、笑いが少なかった。

「神経質な」と評定された講演 学会講演においては、そうでない講演よりも発話速度が遅かった。模擬講演においても、同様の傾向を示した。

「リラックスした」と評定された講演 学会講演においては、どの項目でも違いがみられなかったが、模擬講演においては、発話速度は速く、フィラーが多く、笑いが多かった。

「元気のある」と評定された講演 学会講演においては、そうでない講演よりも発話速度が速く、笑いが多かった。模擬講演においても、同様の傾向を示した。

「元気のない」と評定された発話 学会講演においては、そうでない発話よりも発話速度が遅く、笑いが少なかった。模擬講演においても、同様の傾向を示した。

「聞き取りやすい」と評定された講演 学会講演においては、そうでない講演よりも発話速度のゆれが大きかった。模擬講演においては、発話速度のゆれが大きく、言い淀みが多く、笑いの数が多かった。

## 5 考察

以上の結果から、「フィラー」「言い淀み」「発話速度」「笑い」と、全ての講演特徴が、講演の印象評定に寄与していることが分かった。

ここでは、特に発話速度について述べたい。

上記の結果から、発話速度が、多くの評定語と関係すること分かった。発話速度と評定語の特性の観点から結果をまとめると、以下のようになる。

- 肯定的な評定を表わすと考えられる、「流暢な」「表情豊かな」「自信のある」「リラックスした」「元気のある」「聞き取りやすい」という評定を与えられた講演では、発話速度が速い。
- 否定的な評定を表わすと考えられる、「たどたどしい」「単調な」「神経質な」「元気がない」という評定を与えられた講演では、発話速度が遅い。

参考文献[5]において、内田は「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。今回の結果からでは、内田の挙げる性格特性までは分らない。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、試みに数量化III類の第1軸の値に対する発話速度の相関を計算してみたところ、模擬講演が相関係数0.348、学会講演が相関係数0.252と、相関は見られなかった。この理由については、今後検討したい。

## 6 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、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、評定者を固定して、更に信頼のおける評定データを作成したい。

また、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（参考文献[2]参照）の結果を生かし、今回よりも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.
- [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, 菊池, 小椋. モノログを対象とした自発音声コーパス —その設計について—. 第14回日本音声学会全国大会予稿集, 2000.
- [4] 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.
- [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.