

OGI 多言語電話音声コーパスにおける日本語自然発話音声の分析

荒井隆行 (上智大・理工), Natasha Warner (Max Planck Institute / Univ. of Arizona),
Steven Greenberg (International Computer Science Institute)

1. はじめに

OGI 多言語電話音声コーパス (OGI Multi-Language Telephone Speech Corpus) [1] は実環境に対する自動言語識別のために OGI (Oregon Graduate Institute of Science and Technology) で収集された自然発話音声のコーパスで、このコーパスは自動言語識別の研究のみならず音声研究全般で幅広く用いられている。自然発話音声では朗読音声や実験室環境で録音された音声などと違い、様々な音声現象が観測され [2, 3]、自然発話音声に対する観測や分析を行うことは実環境における音声処理に対する重要な知見を与えると同時に言語学的な理論の裏付けにも貢献するものと考えられる。そこで本稿ではいくつかの項目に焦点を絞り、日本語自然発話音声の特性について報告する。

2. 分析

OGI 多言語電話音声コーパスでは複数の話者による自然発話音声が含まれており、その中でも話者自身が選んだ話題に対する 1 分間の発話のうち、頭から約 50 秒を切り出したものに関し 30 名分を分析対象とした [4]。各発話に対し、波形・スペクトログラムの観察ならびに聴取によって、音素ラベルを付与した。

2.1 母音

2.1.1 母音の無声化

日本語の場合、母音の無声化は /C[-voice] V[+high] C[+voice]/ という環境で観測されることが一般に言われている [5]。例えば、「でした」/deshita/ や「ました」/mashita/ の /i/ がそうである。文末の「です」/desu/ や「ます」/masu/ の /u/ も多くの場合、無声化する [5]。

後続子音が有声音であっても (/C[-voice] V[+high] C[+voice]/) 無声化が観測されることがあり、本コーパスでもしばしば観測された。例えば、「五月の」/gogatsu no/ の /u/ や、「話します」/hanashimasu/ の /i/、「来る」/kuru/、「する」/suru/、「趣味」/shumi/、「すごく」/sugoku/ の最初の /u/ などがある。

本コーパスの自然発話音声においては、さらに典型的な無声環境以外でも無声化が多く見られた。例えば、/C[+voice] V[+high] C[-voice]/ (例: 「自転車」/jitensha/ の /i/) や /C[+voice] V[+high] C[+voice]/ (例: 「はじめ」/hajime/ の /i/) などで

ある。

本コーパスで観測された無声化の頻度を表 1 に示す。非高母音であっても無声化することも文献に報告されているが [5]、本コーパスでもいくつか観測された。

2.2 母音の出現頻度と持続時間

母音の出現頻度と持続時間を表 2 と表 3 に示す。長母音と短母音の平均持続時間の比は 117.0 ms / 76.4 ms = 1.53 となった。これらの数値は、文献で報告されているフォーマルな音声に対しての比よりも小さい値になっている。

2.3 子音

2.3.1 発話の多様性

自然発話音声では一般的に話速が速く閉鎖子音の調音が完全ではなくなり、口腔内で完全な閉鎖が作られずに接近音化する spirantization と呼ばれる現象がしばしば観測される。この現象は英語でもしばしば観測される [6]。本コーパス内においても、有声閉鎖音 /b, d, g/ ではしばしばこの現象が観測された (例えば、「大学」/daigaku/ の /g/、「おばさん」/obasan/ の /b/、「～方です」/... kata desu/ の /d/ など)。

本コーパス内における /g/ の異音の割合を調べたところ、クリアな破裂が観測された [g] は /g/ 全体の 20.4%、明瞭な破裂がなく摩擦のみ観測された [x] は 72.2%、軟口蓋鼻音 [ŋ] は 7.4% であった。

日本語の /r/ は弾音 (flap) と言われているが、実際には多様性があり自然発話音声ではそれが顕著である。本コーパス内における /r/ の異音の割合を調べたところ、母音に挟まれた /r/ のうち弾音 [r] は /r/ 全体の 82.3%、弱い弾音 [r̥] は 6.4%、/r/ 自身がほとんど観測されないものが 4.7% あった。また、弾音 [r] の直前が子音であったケースが /r/ 全体の 1.7% あった。さらに、破裂音 [d] として観測されたものは /r/ 全体の 4.9% (母音に挟まれたケースが 2.8%、無音に後続するケースが 2.1%) 存在した。その他、側音化したケースもあった。

表 1. 母音の無声化頻度

		無声化	非無声化	無声化率
高母音	無声化環境	279	10	96.5%
	非無声化環境	157	1021	13.3%
非高母音		43	3856	1.1%

* Analysis of Spontaneous Japanese in OGI Multi-Language Telephone-Speech Corpus

By Takayuki Arai (Sophia Univ., Tokyo, Japan), Natasha Warner (Max Planck Institute for Psycholinguistics, the Netherlands / Univ. of Arizona, Department of Linguistics, Tucson, Arizona, U.S.A.),

2.4 子音の出現頻度と持続時間

子音の出現頻度と持続時間を表4と表5に示す。これらの表では、/N/は撥音、その他の音素は主だった異音に分類している。長子音と短子音の平均持続時間の比は $142.4 \text{ ms} / 76.9 \text{ ms} = 1.85$ となった。これらの数値は、母音同様、文献で報告されているフォーマルな音声に対しての比よりも小さい値になっている。

3. おわりに

OGI多言語電話音声コーパスを対象に、日本語の自然発話音声に対して各音素の多様性や出現頻度・持続時間などについて見てきた。今後、より自然な音声を対象とした研究を行うには引き続き自然発話音声に対する調査を進める必要がある。

参考文献

- [1] <http://www ldc.upenn.edu/Catalog/LDC94S17.html>
- [2] Arai, T. 1999. A case study of spontaneous speech in Japanese. In *Proc. Int'l Cong. on Phonetic Sciences*, pp. 615-618.
- [3] Arai, T. and Warner, N. 1999. Word Level Timing in Spontaneous Japanese Speech. In *Proc. Int'l Cong. on Phonetic Sciences*, pp. 1055-1058.
- [4] Muthusamy, Y.K., Cole, R.A. and Oshika, B.T. 1992. The OGI Multi-Language Telephone Speech Corpus. In *Proc. Int'l Conf. on Spoken Language Processing*, pp. 895-898.
- [5] Vance, T.J. 1987. *An Introduction to Japanese Phonology*. State University of New York Press, Albany.
- [6] Greenberg, S. 1997. The Switchboard Transcription Project, Technical Report, 1996. *Johns Hopkins CSLP Workshop on Innovative Techniques for Large Vocabulary Continuous Speech Recognition*, Baltimore, MD.

表2. Frequency of occurrence and duration (in ms) for short vowels.

Segment	N	Mean	20%	50%	80%
/a/	1855	82.3	53.0	73.3	103.0
/e/	848	85.7	48.8	71.0	113.4
/i/	1022	67.5	41.9	59.0	86.2
/o/	1196	75.4	47.0	66.0	94.3
/u/	447	56.8	33.0	48.1	71.2
total	5368	76.4	46.7	66.1	97.1

表3. Frequency of occurrence and duration (in ms) for long vowels.

Segment	N	Mean	20%	50%	80%
/a:/	31	122.4	94.0	110.0	136.7
/e:/	44	120.5	93.0	109.9	145.2
/i:/	58	123.4	90.0	122.0	148.8
/o:/	188	116.4	82.0	110.0	145.0
/u:/	86	110.3	67.5	101.0	137.8
total	407	117.0	83.0	110.0	145.0

表4. Frequency of occurrence and duration (in ms) for short consonants.

Segment	N	Mean	20%	50%	80%
/p/	32	74.1	40.3	68.0	106.2
/t/	640	76.5	56.0	76.4	96.3
/k/	816	83.1	61.0	80.0	104.9
/b/	115	55.9	36.0	52.9	73.4
/d/	401	44.6	27.0	43.0	61.0
/g/	240	40.3	26.7	37.0	51.0
[s]	429	105.5	69.0	92.6	133.4
[ʃ]	320	97.1	72.0	92.0	120.0
[h]	146	69.4	46.9	66.0	84.5
[ɸ]	32	86.2	63.2	84.0	101.0
[ts]	77	108.7	90.0	108.3	130.0
[tʃ]	105	102.7	78.9	100.8	122.1
[(d)z]	51	61.8	40.4	62.0	75.1
[(d)ʒ]	106	69.1	48.0	67.0	92.0
/r/	495	29.3	20.4	27.2	37.0
/w/	197	47.8	29.3	45.0	61.6
/y/	221	46.5	29.0	41.7	60.3
/m/	521	61.2	46.0	61.4	75.6
/n/	635	52.8	35.1	49.3	68.0
/N/	312	77.3	55.0	72.5	96.0
total	5891	67.2	37.0	63.0	92.0

表5. Frequency of occurrence and duration (in ms) for long consonants.

Segment	N	Mean	20%	50%	80%
/pp/	9	139.6	118.0	140.0	156.5
/tt/	127	148.7	115.0	142.0	174.0
/kk/	38	172.2	123.9	153.0	211.3
[ss]	1	114.0	*****	114.0	*****
[ʃʃ]	9	165.9	130.0	144.0	177.0
/mm/	5	92.2	75.0	86.8	115.0
/nn/	64	113.8	78.2	102.0	136.0
total	253	142.4	96.0	135.0	174.0