

2nd Electronic edition

Sophia University
Working Papers
in
Phonetics

Vol.27 2021
Phonetics Laboratory
Sophia University

序

松井さなえ

今年も例年のように、Sophia University Working Papers in Phonetics 2021 が出版され、大変嬉しく思います。この論文集は上智大学の音声学研究室に関わる方々の研究成果をまとめたもので、その内容は音声学、音韻論、言語障害学、応用言語学などの分野にまたがります。

本年の論文集は、Dysarthriaを有する方への指導の仕方を検討するために知覚実験を行なった研究、東京方言の韻律に関する研究、室長雑記から成っています。

刊行に際し、藤本雅子先生、田嶋圭一先生、吉田夏也先生、事務の蓑島さんに心より感謝申し上げます。

未だ、感染症の影響を強く受けている状況ではありますが、この論文集が上智大学音声学研究室に関わる方々をはじめとする多くの方々に読んでいただけることを願い、序文とさせていただきます。

2021年4月

目次

序

松井 さなえ

1. A Study of Voices in Terms of Audibility

Kaori Matsumoto 1

2. 東京方言における枝分かれ構造と韻律の検討

澤田 恵里奈 9

室長雑記

先進国から退きつつあるこの国で音声学を研究することについて: 「選択と集中」の過ち

北原 真冬 26

A Study of Voices in Terms of Audibility

Kaori Matsumoto

Mejiro University

E-mail: k.matsumoto@mejiro.ac.jp

1. Introduction

Verbal Communications are remarkably vital activities which affect the quality of our daily lives. Yet, our verbal conversations are not always conducted in quiet situations. In fact, conversations often occur in noisy surroundings such as outside, in a hospital or a cafeteria, thus, people sometimes feel that it is hard to understand conversations. For instance, it could be hard to listen to a conversation in a train, in a cafeteria, or while you are surrounded by the noise from the dishwasher and other noisy daily chores. Additionally, according to the research conducted by professor Seki in 2009, background noises have significant effects on comprehension of dialogue for elderly people. ¹⁾A sound effect on television could interfere with correctly comprehending what was said. Aforementioned, various noises surrounding us affect our hearing of dialogue. Hence, it is important to talk with a clear enunciation even under noisy conditions. Therefore, to analyze what types of enunciation are easier to understand, we have conducted the following audiological examination.

2. Methods

2.1. Participants

There were 10 participants who were in their 20s, who scored normal results on a standardized pure-tone audiometry evaluation. One male was 21 years old; nine females were 20 or 21 years old, with the average age of 20.9 years old.

2.2. Procedures

2.2.1. Creation of Audiological Documents

2.2.1.1. Selection of Audiological Materials

The words which were selected for the audiological examination were chosen from the two lists that are categorized as level 6 or above, and between level 5 and 6 according to the word familiarity level. We referred to the word familiarity levels that were researched by Dr. Amano at the NTT Communication Science Laboratory. ²⁾ Moreover, we selected a total of 20 words (10 words from each word familiarity level) which are nouns with four-mora. Also, we chose the words considering that the words' initial sounds are from a wide variety of Japanese phonemes. Furthermore, we decided to use the carrier phrase "This is read as xxx," by inserting the target words in the sentence for the examination audio.

2.2.1.2. Recording Method

Before the recording, the audiological documents were shown to the speakers to confirm that the speakers understood all of the words in the document. There were four speakers who were in their 20s (A, B, C, and D). Then, we recorded each speaker saying 40 sentences with the carrier phrase "This is read as xxx," with each of the 20 selected target words inserted twice. We did not specify the speed of speaking. The recording was

conducted in a quiet room which was separated from any surrounding noise (less than 50 dB). We connected the condenser microphone (RØDEMT) to the PCM recorder (TASCAM DR-40), and used it as the recording device. The microphone was placed at a 45-degree angle diagonally towards the face, and the distance from the mouth was 20 cm. In addition, the sampling frequency was set as 44.1 kHz, and the quantization level was set as 16 bits.

2.2.1.3. Material Creation for the Audiological Examination

First, we randomly selected twenty different sentences from each speaker that were not mispronounced during the recording session. Next, the audio of the selected 20 sentences was converted into mono audio using the audio software Wavesurfer1.8.8p5. Then, we adjusted the loudness to the same levels, and we overlaid multi-talker babble noise from the CD used for hearing aid compatibility testing. The level of overlaying with the multi-talker babble noise was set to the level that creates the same loudness balance (0 dB) with auditory stimuli. The presentation order of the audio was carefully determined to avoid subsequent audio by the same speaker or the same sentence. Also, we took into consideration the word familiarity levels in full measure, so that the words from the same level are not subsequently placed. The audio for the audiological examination consisted of four speakers reading 20 sentences each, for a total of 80 sentences.

2.2.2. Audiological Examination

The ten participants were divided into groups of three or four. The audiological examination was conducted in a quiet room which was blocked from any surrounding noise. The audio stimuli were presented next to the ears of each participant with the optimal level (approximately 70 dB A) using the audio speaker (Creative INSPIRE T10). Then each participant wrote down exactly what they heard. Additionally, the definition of a 'correct answer' was set as the following; if the participant wrote the target word accurately, then we documented it as a 'correct answer.'

2.2.3. Analysis Method

2.2.3.1. The Differences of the Number of Correct Answers According to the Word Familiarity Levels

We conducted a statistical analysis utilizing a chi-square test to study the discrepancy of the number of correct answers and incorrect answers according to the different word familiarity levels. The statistic software Js-STAR version 9.7.6j was used. The significance level was set as less than 5%.

2.2.3.2. The Differences of the Number of Correct Answers According to the Speakers

We conducted a statistical analysis utilizing a chi-square test to study the discrepancy of the number of correct answers and incorrect answers according to the different speakers. The statistic software Js-STAR version 9.7.6j was used. The significance level was set as less than 5%.

2.2.3.3. Acoustic Differences of the voices

We conducted acoustic analyzations about the speaking features; such as the speed as well as the pronunciation of consonants of the two speakers who gained the largest number of correct answers and who had the least. For the acoustic analyzation, Wavesurfer1.8.8p5 was used. The audio segmentation was conducted based on the information from Waveform, Spectrogram, Pitch Contour, and Power Plot as referred from the studies by Dr. Yoshida (2015)⁴⁾ and Dr. Kawahara(2018)⁵⁾.

2.2.3.3.1. Measurement System of Word Lengths (Speaking Speed)

The beginning of each word was placed where the characteristic waves or elements were observed on Waveform or Spectrogram for the consonants, and where the sign of Pitch

Contour or Power Plot started. The ending of each word was placed where the audio became inaudible, where the first and the second formants became invisible on Spectrogram, and where Pitch Contour was paused. Furthermore, to determine the speaking speed, the number of morae that were spoken in one second was calculated (4 ÷ word length).

2.2.3.3.2. The Duration of the Consonants Which are the First Sound of the Words

The same measurements were taken to determine the beginning of each consonant at the beginning of each word. The ending of each consonant was set where the waves for vowels were not yet formed before the starting point of the sign by Pitch Contour. In the cases where the words that the first sound included plosives; the data was excluded because it is impossible to determine the precise beginning point. Also, the words which started with vowels were also excluded from the analysis data. Hence, out of the 20 words, a total of 14 words, except the six words which had plosives or vowels as their first sound were analyzed.

2.2.3.3.3. Spectrogram Analysis of the Consonants Which are the First Sound of the Words

We analyzed the visual features of the sound by analyzing the wideband spectrograms.

3. Results

3.1. The Number of Correct and Incorrect Answers According to the Word Familiarity Levels

Table 1 shows the examination’s result of the number of correct and incorrect answers between the different word familiarity levels.

There were 215 correct answers and 185 incorrect answers among the words in level 6 or above. For the words that are between level 5 and 6, there were 174 correct answers and 226 incorrect answers. We conducted a chi-square test to analyze the discrepancies of the number of correct answers and determined that the words in the level 6 and above had a significantly larger number of correct answers compared to the words between level 5 and 6 (p<0.5). Also, the words between level 5 and 6 had a significantly larger number of incorrect answers (p<0.5).

Table 1. The Number of Correct and Incorrect Answers in the Different Word Familiarity Levels

	Level 6 or above	Between level 5 and 6
Correct	215▲* ¹	174▽* ²
Incorrect	185▽	226▲

3.2. The Difference of the Number of Correct Answers According to the Speakers

Table 2 shows the examination’s result of the difference of the number between correct and incorrect answers for each speaker.

The number of correct answers was 73 for Speaker A, 92 for Speaker B, 113 for Speaker C, and 111 for Speaker D. The number of incorrect answers were 127 for Speaker A, 108 for Speaker B, 87 for Speaker C, and 89 for Speaker D. We conducted a chi-square test to analyze the difference of the number of the correct and incorrect answers between each speaker. As a result, we determined that the number of correct answers for Speaker C and Speaker D were significantly larger (p<0.5). On the other hand, the number of incorrect answers for Speaker A was significantly larger (p<.01). In addition, the correct answer rates were 36.5% for Speaker A, 46.0% for Speaker B, 56.5% for Speaker C, and 55.5% for Speaker D.

Table 2. The Number of Correct and Incorrect Answers for Each Speaker

	A	B	C	D
Correct	73▽	92	113▲	111▲
Incorrect	127▲	108	87▽	89▽
The correct answer rate	36.5	46.0	56.5	55.5

3.3. The Audiological Difference of the Voices

Table 3 shows the result of the audiological analysis of consonants and speaking speeds between Speaker A who had the least number of correct answers and Speaker C who had the largest number of correct answers.

3.3.1. Word Length Comparisons (Speaking Speed)

The average word length of Speaker A was 0.465 seconds (0.412~0.592). The number of morae that were spoken in one second was 8.609. The average word length of Speaker C was 0.558 seconds (0.491~0.694). The number of morae that were spoken in one second was 7.168. Speaking speed of Speaker C was slower than Speaker A's speaking speed.

3.3.2. The Duration of the Consonants Which are the First Sound of the Words

Regarding the duration of the consonants which were the first sound of the words, Speaker A's average duration was 0.052 seconds (0.007~0.142). Speaker C's average duration was 0.074 seconds (0.019~0.126). The result shows that Speaker C has a slightly longer duration of the consonants which are the first sound of the words, compared to Speaker A's duration.

3.3.3. The Spectrogram Analysis of the Consonants Which are the First Sound of the Words

Charts 1 to 4 show the information about the words “湾曲” (*Wan-kyo-ku*) and “社会科” (*Sha-kai-ka*) spoken by Speakers A and C that were collected from Waveform, Spectrogram, Pitch Contour, and Power Plot.

Speaker C's spectrogram of the word “湾曲” (*Wan-kyo-ku*) (Chart 2) shows a more robust lower formant frequency which is the attribute of the sound [w], compared to Speaker A's (Chart 1). Speaker C's spectrogram of the word “社会科” (*Sha-kai-ka*) (Chart 4) shows a clear attribute of the sound [j]. On the other hand, Speaker A's spectrogram for the word shows the attribute resembling the sound [s] as the energy is concentrated on higher frequency.

Table 3. Speaker A's and Speaker C's Word Lengths and Durations of Consonants

	Aの語長	Cの語長	Aの子音の長さ	Cの子音の長さ
風流	0.439	0.554	0.091	0.124
荷造り	0.592	0.694	0.142	0.094
要約	0.426	0.491	0.011	0.063
卒業	0.546	0.607	0.125	0.126
焼きそば	0.481	0.540	0.021	0.019
ライオン	0.421	0.537	0.022	0.029
内縁	0.442	0.530	0.040	0.050
湾曲	0.412	0.511	0.007	0.063
社会科	0.522	0.587	0.102	0.117
木材	0.460	0.510	0.060	0.102
肉まん	0.451	0.591	0.030	0.080
流失	0.386	0.538	0.016	0.062
山積み	0.496	0.568	0.029	0.050
まな板	0.431	0.554	0.028	0.051
平均	0.465	0.558	0.052	0.074
1秒で発話したモーラ数	8.609	7.168		

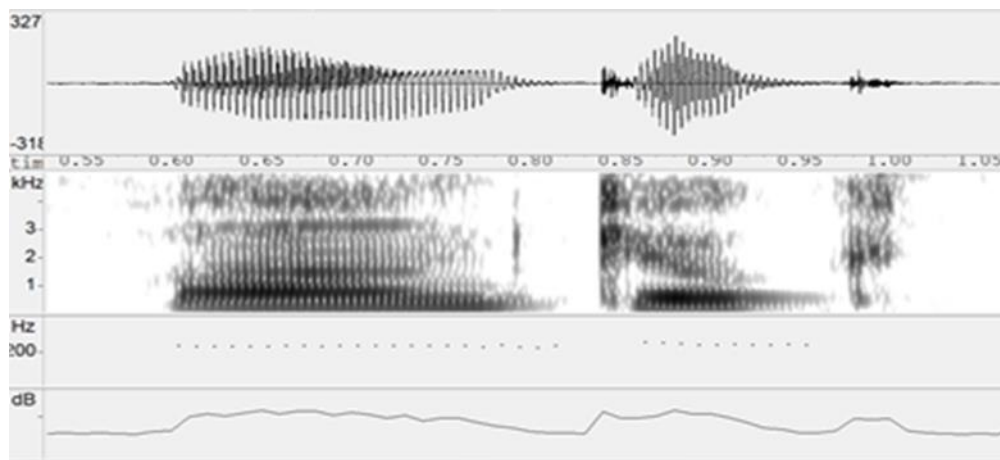


Chart 1. The word “湾曲” (*Wan-kyo-ku*) by Speaker A

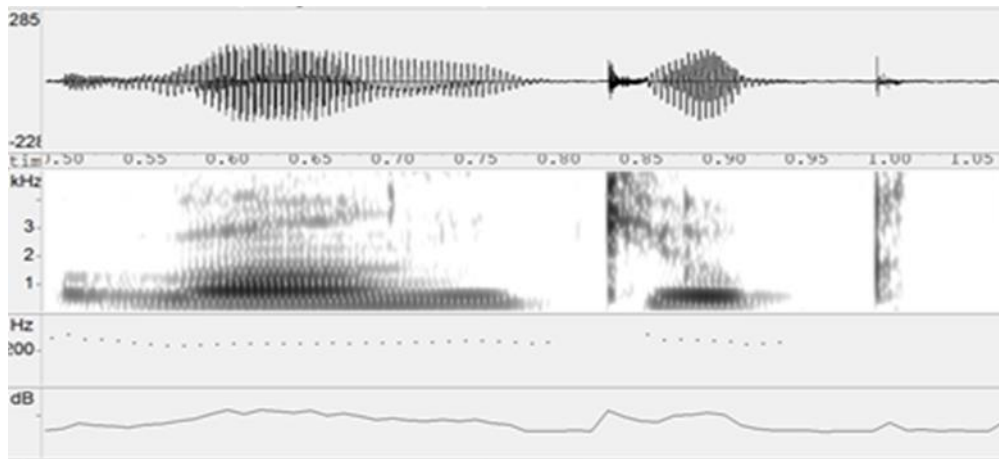


Chart 2. The word “湾曲” (*Wan-kyo-ku*) by Speaker C

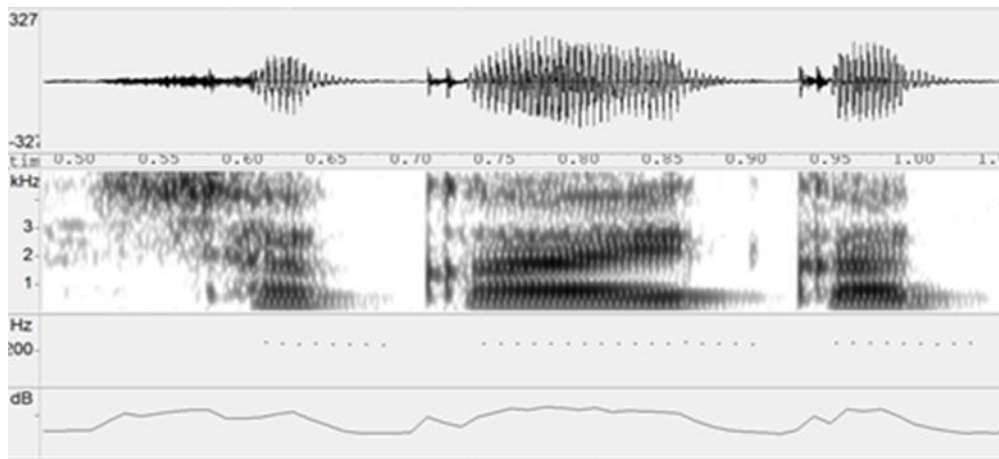


Chart 3. The word “社会科” (*Sha-kai-ka*) by Speaker A

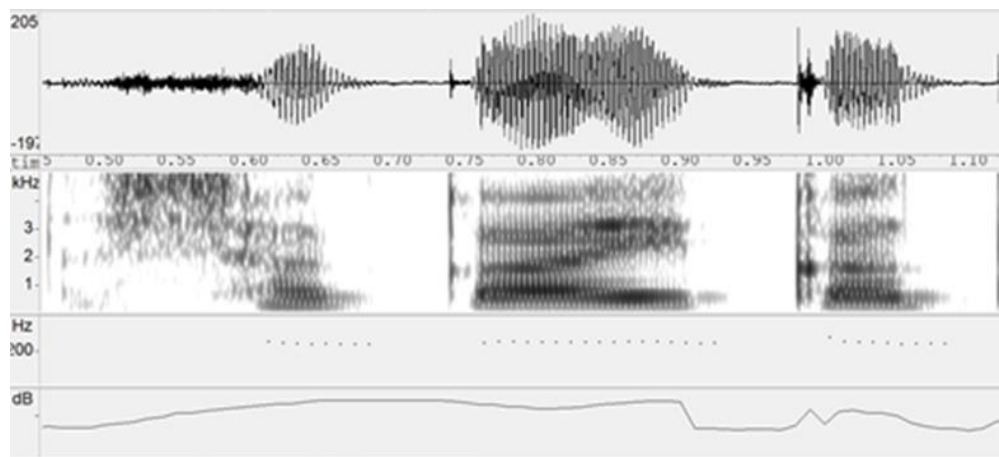


Chart 4. The word “社会科” (*Sha-kai-ka*) by Speaker C

4. Discussion

4.1. Familiarity

This time, we compared difference between the correct and incorrect answers by phonic word familiarity for the purpose of clarifying the voice that it is easy to hear. The result has revealed that the number of words with the familiarity of over 6 was larger than that of those with the familiarity of over 5 and less than 6. The word hearing experiment with background noise superposed for adults in twenties which was performed by Seki (2009) revealed that the correct answer rate of words with high familiarity word was significantly high ¹⁾. Moreover, Sakamoto et al. (2004) reported results of a hearing experiment for which words' familiarity and presentation sound pressure were changed without noise. The experimental result for the influence of familiarity on word intelligibility revealed difference between the words with high and low intelligibility ⁶⁾. The results obtained this time support the above-mentioned report. Therefore, in the case that the person could not understand the sentence at once or in the case of an environment with noise, we infer that it would be better to use familiar words. Further, we presume that consideration has to be given so that words with high familiarity can be used in the conversation as much as possible. Furthermore, Sakamoto et al. (2004) reported that in the case that presentation sound pressure level is low and acquisition of audio information is difficult, it is suggested that collating with mental lexicon be performed more strongly than in the case that presentation sound pressure level is high and acquisition of audio information is easy ⁶⁾. It was supposed that the influence by familiarity would be enhanced in the case that sound pressure was small due to dysphonia or dysarthria.

4.2. Speaking speed

Speaking speeds of the target word of Speaker A, whose correct answer rate was significantly low, and of Speaker C with significantly high correct answer rate were compared. The result showed that the speaking speed of C was slower than A, indicating that slower speech makes a person hear the words more easily. Therefore, there are effects of the speaking speed at the word level as a factor for the sound that is easy to hear. A study by Kawano (1998) reported that the length of 1 mora in Japanese is average 145ms ⁷⁾. The result obtained this time showed that the average number of seconds per mora was 0.116 seconds (116ms) in A and 0.140 seconds (140ms) in C. Therefore, we presume that C, which is close to the average value, was easily heard and it was more difficult for A, which was quicker than the average, to be heard.

4.3. Speech analysis of consonant

A past study reported that longer consonant was identified more easily in hearing by the one with hearing impairment ⁸⁾. Our result showed that the utterer who clearly utters consonants for a long time was heard more easily. Further, a study on speech discrimination of a patient with partial sensorineural hearing loss presented the followings. Voiceless consonant queues except /s/ are contained in low middle frequency bands less than 1,000Hz. In addition, the study reported that information on high frequency band was required for identification of voiced plosive (explosive and frictional sound) and nasal sound in Japanese ⁹⁾. In this way, characteristics of consonants are used for discrimination of sound, and the sounds to which the frequency characteristics of the consonant are reflected is heard more easily. Therefore, more precise utterance of consonants is an important factor for the easiness of hearing by a hearer.

4.4. Speech instruction by speech therapist

The results have revealed that words with high familiarity are heard more easily, and that speaking speed that is close to the average speed rather than fast speaking speed and more precise utterance of consonants would allow the hearer to hear the words easily. These results can be incorporated into the

points for instruction of dysarthria. The movement of tongue and oral cavity in a person with dysarthria is affected, which often makes their articulation unclear. Therefore, of course, training is performed for improvement of the movement of tongue and oral cavity, although it is not necessarily improved. For such a case, the following clues were considered for instruction. Firstly, it is presumed that it is effective to instruct the patients to use easy words in conversations. Secondly, it is already known well yet it is thought, after all, to be important to instruct them to say individual words slowly. Thirdly, there is a possibility that it is effective to instruct them to speak words clearly for a slightly longer time in articulatory training, so that the characteristics of consonants are transmitted well.

5. The limitations of this study and future task

Since the experimental sounds in this study were created from utterance of normal young persons, experiments with sounds of dysarthria persons will be needed in the future. Further, it is necessary to examine elements other than speaking speed and the first consonant of the word as for characteristics of sounds that are easily heard and those that are not.

Acknowledgments

I sincerely thank all the people for their understanding and cooperation.

References

- Amano, S., & Kondo, T. (2003). *Glossary characteristics of Japanese the 1st period (Vol. 1 - Vol. 6)*. Sanseido.
- Kawahara, S. (2018). *Visual phonetics*. Sanseido, Tokyo, 140-142, 156-158.
- Kawano, M. (1998). Psycholinguistic discussion of mora, syllable and rhythm. *Sound study*, 2, 16-24.
- Kodera, K., Horiuchi, M. (1995). Speech discrimination of patient with partial sensorineural hearing loss. *Audiology Japan*, 38, 298-304.
- Sakamoto, S., & Amano, S. (2004). Influence of familiarity on mora identification in word intelligibility. *Acoustical Society of Japan journal*, 60, 351-357.
- Seki, H. (2009). Study on sounds that elderly people easily hear (special feature: design of sound in reliable society). *Japan Acoustical Materials Association "Sound technique"*, 38, 11-14.
- Yasu, K., Arai, T. (2012). Influence of deterioration of hearing characteristics in elderly persons on identification of silent frictional sound and explosion sound. *Acoustical Society of Japan journal*, 68, 501-512.
- Yonemoto, K., & Tateishi T. (1989). CD(TY-89) for hearing aid adaptation assessment and single syllable articulation and sound pressure of the 57-S vocabulary list. *Audiology Japan*, 32, 429-430.
- Yoshida, T. (2015). *Introduction to acoustics of speech therapist*. Kaibundo, Tokyo, 3, 144-151.

Notes

-
- 1 significantly larger
2 significantly smaller

東京方言における枝分かれ構造と韻律の検討

澤田恵里奈

上智大学大学院 言語科学研究科 言語学専攻

E-mail: e-sawada-e5f@eagle.sophia.ac.jp

1. 導入

1.1. 先行研究

日本語のイントネーションに関しては1980年代初頭から広く研究されるようになった (Igarashi, 2018)。イントネーションはその韻律特徴と機能の間に恣意的な関係を持つものの、多くの言語で共通した関係性が見られる。例えば、高い F0 や上昇調であれば継続性を示すのに対して、低い F0 や下降調では終端性を示すといった対比が多くの言語で認められている。しかし、その逆の結びつきを示す言語もあるため、完全に非恣意的と呼ぶことはできない。Igarashi (2018) によると、日本語におけるイントネーションを最も特徴付けているのは基本周波数である。しかし、韻律特徴は F0 曲線の変化だけではない。Ladd (2008) はイントネーションを「"後語彙的"なあるいは文レベルの語用論的意味を伝えるための超分節的な音声特性の言語学的な体系的使用」と定義している。この超分節的な音声特性とは、音の強さ、長さなども含まれている。

韻律特徴が重要となる一つの例として、統辞的曖昧性のある文の弁別があげられる。窪菌 (1988) は以下のような左枝分かれ文と右枝分かれ文では、F0 曲線が異なっていることを指摘した (Venditti, 1994 参照)。

左枝分かれ文 (LB) : [[ao'yama-ni a'ru] daigaku] a university in Aoyama

右枝分かれ文 (RB) : [ao'yama-no [a'ru daigaku]] a certain university in Aoyama

左枝分かれ文は句頭上昇からダウンステップと呼ばれる階段のような F0 の下降が現れるのに対し、右枝分かれ文は二つ目のピーク (peak2) が一つ目のピーク (peak1) の高さとは変わらない。窪菌はこの違いに対して、右枝分かれ部分の最も左の要素に対して、F0 を上げる操作 (= "metrical boost") を行っていると提案した。

一方、Pierrehumbert & Beckman (1988) のモデルでは、左枝分かれでは単一の韻律句であったのに対し、右枝分かれの場合は韻律句が2つになることでダウンステップのリセットが行われているとした¹。Venditti (1994) の研究では、この二つの提案の比較を目的とした音声産出と知覚の実験を行った。窪菌の指摘したように、図 1 は peak2 の高さが左枝分かれ文と右枝分かれ文で異なることを示している。

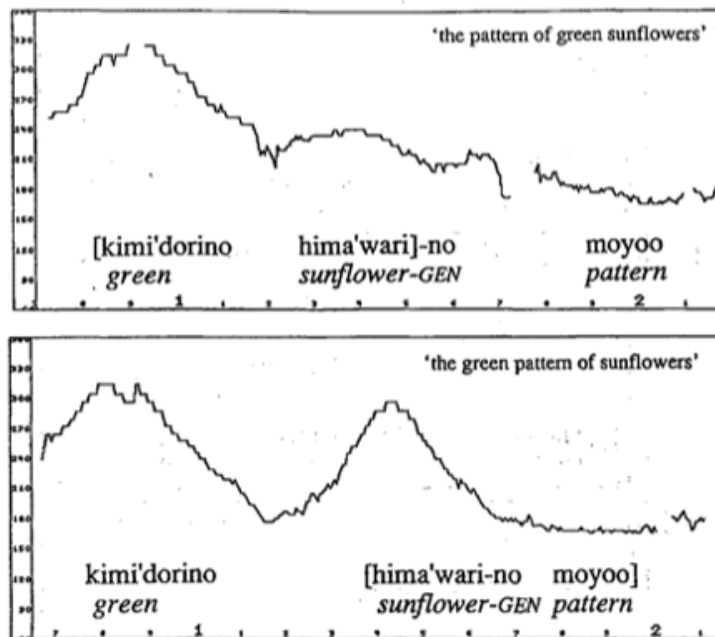


図 1:Venditti (1994) で確認された左枝分かれ文 (上) と右枝分かれ文 (下)

Venditti は先行研究で用いられたような形式の名詞句だけでなく、関係節を含む曖昧文も刺激として使用している。結果として、発話者 3 名のうち 1 名は、先行する語のアクセントの有無によって peak2 に差が表れず、ダウンステップがリセットされるモデルを支持する結果となった。しかし、残りの 2 名は先行語のアクセントの有無によって、peak2 の F0 に差が表れたため、ダウンステップの特徴が表れていた。さらに、関係節を含む右枝分かれ文の読み上げでは、F0 の変化ではなく peak1 と peak2 の間にポーズを入れることで弁別を行っていることが指摘されている。Venditti (1994) は話者によって統辞的曖昧性を解決する方略が異なっていると述べている。

さらに、広瀬 (2006) では単語のアクセント核の有無によって枝分かれ構造を韻律に反映する難しさが異なることを指摘した。広瀬 (2006) はアクセント核の F0 とポーズ長、母音長について、AP1 (Venditti の peak1 に対応) の語アクセントが有核・無核の場合で比較検討した。はじめの 2 つの文節が有核語である場合は、ダウンステップが起こる。Venditti (1994) での指摘と異なり、広瀬 (2006) の場合 AP1 から AP2 は LB と RB 条件どちらも F0 が下降傾向であった。しかし、下降幅が RB より LB の方が有意に大きいことも指摘している。また、AP2 (Venditti の peak2 に対応) から AP3 にかけて、LB 条件では F0 が有意に上昇することも指摘している。ポーズについては、LB 条件について AP2 の後により長いポーズが置かれていた。一方で、AP1 が無核の場合は、AP1 から AP2 でダウンステップが起こらない。F0 の AP1-AP2 にかけての下降傾向は両枝分かれ

条件において確認され、下降量において有意な違いがないことがわかった。ポーズ長については、LB 条件において、AP2 後のポーズがやや長い傾向が確認された。また、AP1 と AP2 の末尾の母音長については有意差が認められた。結果として、ダウンステップが発生しない場合に F0 以外の時間長の手がかりにおいて区別されるといった傾向はあまり見られなかったと報告している。

知覚実験においても、日本語の枝分かれ構造と対応する韻律に関して検討されてきた。例えば、Azuma & Tsukuma (1990) は近畿方言を用いた知覚実験を行い、F0 ピークを上昇させると RB 構造を支持する解釈が増えたと報告している。また、彼らは F0 とポーズのうち F0 の方が弁別においてより重要であるということを示した。Hirose (2020) では、視覚-単語パラダイムを使用して、枝分かれ構造が曖昧な文の音声処理中の視線計測を行った。結果として、AP2 のアクセント核が上昇している場合も、AP1-AP2 へ通常のダウンステップが起こっている場合も、初めは LB に一致する絵の方を見つめていた。言い換えると、AP2 が上昇すると、音の変化から即時的に対象を変えるのではなく、音声処理の終わりに RB に一致する刺激を注視する傾向にあることを報告している。

これらの研究が示す通り、話者は枝分かれ構造を様々な韻律的情報を用いて区別している。特に先行研究で指摘されてきたのは、AP2 のアクセント核の F0 の違いであろう。また、AP 間のポーズや AP 末の母音長についても検討が行われていた。さらに、筆者が昨年度行ったパイロット実験では新たな F0 の手がかりがある可能性が示された。3 名の東京方言話者のうち 1 名について、AP 末音調が一部の発話において体系的に上昇するパターンが確認された。その話者については、統語境界と隣接する AP のうち、先行する AP の末尾（助詞など）がアクセント核よりも F0 が上昇傾向であった²。

1.2. 本研究の目的と仮説

上述の通り、枝分かれ構造と対応した韻律については様々な検討が行われてきた。本研究の目的は、東京方言話者は統語論の枝分かれ構造を韻律にどのように反映するのか、その中でもどのような韻律的特徴が特に重要であるかを明らかにすることにある。また、Venditti (1994) で指摘されていたような、読み上げる文のもつ特徴がどれほど韻律に反映されているのかについても検討する。先行研究に基づき、以下の仮説を設けた。

- 仮説1. LB 読みに比べ、RB 読みの際に 2 つ目のアクセント核の F0 がより高くなる。
- 仮説2. LB 読みに比べ、RB 読みの際に 1 つ目の要素の句末の F0 がより高くなる。
- 仮説3. LB 読みに比べ、読みの際に 1 つ目と 2 つ目の要素の間のポーズがより長くなる。
- 仮説4. ポーズ長や句末音調に比べ、アクセント核の F0 がより重要である。

仮説 1-3 は、統語境界に位置する韻律的特徴が強調されるという前提に基づいている。例えば、3つの文節からなる句の RB 構造でははじめの2つの文節の間に統語的区切りがある。本研究ではこの区切りを強調するような F0 および時間長での操作が行われると考える。つまり、これらの仮説はその境界を強調するために、2つ目の文節の F0 を高めたり、直前の文節末の F0 を高めたり、比較的長いポーズを入れることが予想される。

また、仮説 4 は先行研究や筆者の行ったパイロット実験に基づき設けられたものである。前述の通り、韻律は話者や発話によって変化するものであるが、2つ目の要素のアクセント核の F0 が LB と RB の間で異なっていることは多くの研究で支持されてきた (e.g. Kubozono, ; Igarashi, 2010)。そこから、2つ目の要素のアクセント核の F0 が今回の研究で用いる変数の中で最も安定した指標であると考えられる。

2. 手法

2.1. 参加者

本研究には東京都内の大学に所属する 20 歳から 23 歳の大学生 10 名が参加した (男性 = 3, 女性 = 7)。全ての参加者は東京方言の母語話者である。

2.2. 実験文

実験に使用した文はターゲット 7 文、フィラーが 3 文の合計 10 文である。表 1 にターゲット文を示す。ターゲット文の一部は先行研究で用いられていたものを使用している³。便宜上、それぞれターゲット文の中に使用されている名詞から文 ID をつけている。文構造は 3 文節からなる名詞句 (NP) が 5 文と関係節が含まれる 2 文 (RC) がある。関係節文の文節数はコントロールされていない。

また、語アクセントについても無核・有核どちらも含まれている。文節数や語アクセントのばらつきは、文そのものが持つ特徴によって韻律にどのような影響をもたらすか検討をするためである。それぞれの実験文は 2 つの枝分かれの条件があるため、カウンターバランスをとり、2 つのリストに分けられた。また、実験では枝分かれ構造を示唆するヒント文を用いた。表 2 にそれぞれのターゲット文に対応するヒント文を示す。

表 1：実験に使用したターゲット文。下線は無核語を示す⁴

ID	Type	Num. of AP	Sentence
<i>police</i>	RC	6	警官は <u>自転車に</u> <u>乗って</u> <u>逃げる</u> <u>泥棒を</u> 追いかけた
<i>cap</i>	NP	3	水色の <u>ブタの</u> <u>帽子</u>
<i>scarf</i>	RC	4	去年 <u>編んだ</u> <u>マフラーが</u> <u>盗まれた</u>
<i>hanky</i>	NP	3	<u>桃色の</u> <u>バラの</u> <u>ハンカチ</u>
<i>umbrella</i>	NP	3	青い <u>ネコの</u> <u>傘</u>
<i>bark</i>	NP	3	大きな <u>犬の</u> <u>鳴き声</u>
<i>cd</i>	NP	3	青い <u>CDの</u> <u>ケース</u>

表 2：ターゲット文に対応したヒント文

ID	Sentence	RB	LB
<i>police</i>	警官は自転車に乗って逃げる泥棒を追いかけた	自転車に乗っているのは泥棒だ	自転車に乗っているのは警官だ
<i>cap</i>	水色のブタの帽子	水色なのは帽子だ	水色なのはブタだ
<i>scarf</i>	去年編んだマフラーが盗まれた	昔編んだマフラーを去年盗まれた	去年マフラーを編んだが、その後盗まれた
<i>hanky</i>	桃色のバラのハンカチ	桃色なのはハンカチだ	桃色なのはバラだ
<i>umbrella</i>	青いネコの傘	青いのは傘だ	青いのはネコだ
<i>bark</i>	大きなイヌの鳴き声	大きいのは鳴き声だ	大きいのはイヌの体だ
<i>cd</i>	青い CD のケース	青いのは CD のケースだ	青いのは CD だ

2.3. 手順

密閉空間での COVID-19 のリスク等を考慮し、今回の実験はリモートで行われた。参加者は自室および静かな空間で Zoom を介して実験者から教示や刺激の呈示を受けた。実験は 2 つの日程に分かれており、1 日程に 1 つの実験文リストを用いた。実験に移る前に、テスト録音を行った。録音状態の確認をし、参加者が正しく録音機材などを用いているかを確認した⁵。また、実験文を読む前に 2 文、実験と全く同じ手続きで仮録音を行い、参加者が手順などについて質

問する時間を事前に設けた。図 2 は実験中の画面の遷移を示す。参加者ははじめに共有された画面に表示された実験文を黙読する (①)。その後、画面上に示された 2 つの枝分かれ構造に基づいた解釈を確認し (②)、本人のタイミングで声に出して実験文を読み上げた (③)。参加者は実験者に合図を送り、画面の表示を実験者が適宜変更した。参加者は全ての文を 2 回繰り返し読んだ。



図 2：実験中の画面。ヒント文だけでなく、状況を表す画像も条件によって加えられた⁶

音声の録音では、参加者は単一指向性のピンマイク (Hiroumer, HR-MKF-L01) を首元につけ、マイクを iPhone に接続し、iOS アプリのボイスメモを使用した。全ての録音はサンプリング周波数 48kHz の Apple ロスレス形式で録音された。次に、Sound Converter というオーディオ変換アプリケーションを使用して、録音データを wav 形式に変換した。分析では 2 度目の録音のみを対象とし、計 140 発話を使用した。MacBook Pro 上の Praat (ver 6.0.35) を用いてアノテーションを行った。

本研究ではアクセント核の F0 (文の冒頭から P1, P2)、文節末 1 モーラの F0 (F1, F2) ⁷、ポーズ長の 3 つの韻律情報を取得した。以下に、各ターゲット文の F0 を取得した位置を表 3 に示す。この範囲の中の F0 の最大値を分析対象とした。また、アノテーションでのピッチ取得範囲を図 3 に示す。

表 3 : 韻律情報の F0 取得位置

ID	P1	F1	P2	F2
<i>police</i>	keHkan	wa	jitensya	ni
<i>cap</i>	mizuiro	no	buta	no
<i>scarf</i>	kyo'	neN	a'N	da
<i>hanky</i>	momoiro	no	bara	no
<i>umbrella</i>	ao'	i	ne'ko	no
<i>bark</i>	o'Hki	na	inu	no
<i>cd</i>	ao'	i	CD	no

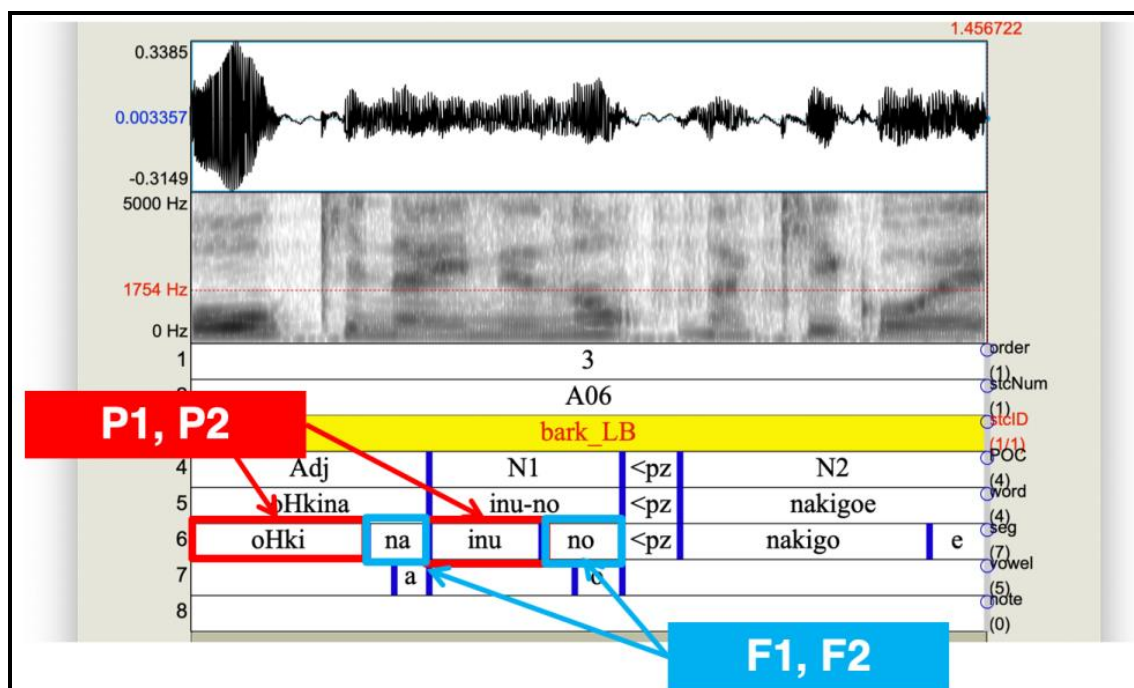


図 3 : F0 ピークの取得範囲例。赤枠で囲まれているのがアクセント核の範囲 (P1, P2)、青枠で囲まれているのが句末の範囲 (F1, F2)。

ポーズ長については F1 と P2 の間のポーズを T1 とし、名詞句文では F2 と P3 の間のポーズを T2 とした。関係節文では、F3 と P4 の間を T2 とした。

3. 結果

140 発話の F0 とポーズ長の情報を分けて分析を行った。次の 2 節では大まかにデータの外観を述べた後、統計検定の結果を示す。また、3.1 節では F0 の 2

変数を含めたロジスティック回帰分析を行った結果を示す。

3.1. F0

P1, P2 の平均と SD を表 4 に示す。P1 においては枝分かれ条件間で平均値はほぼ同じであるが、P2 は RB の方が 5Hz ほど高い。P1, P2 の参加者別の F0 最大値の平均を以下の図 4 に示す。P1 においては、RB が LB に比べてやや低い傾向が見られるが、P2 においては RB が LB よりも高い参加者は 7 名であった。一方、s06, s07, s08 については P1 と変わらず P2 でも LB の方が高い結果となった。

表 4 : P1 と P2 の平均 F0 最大値 (Hz) と SD

branch	P1_mean	P1_sd	P2_mean	P2_sd
LB	268.935866	66.1996155	229.066025	58.4823642
RB	263.887496	65.5283445	234.037874	63.4389042

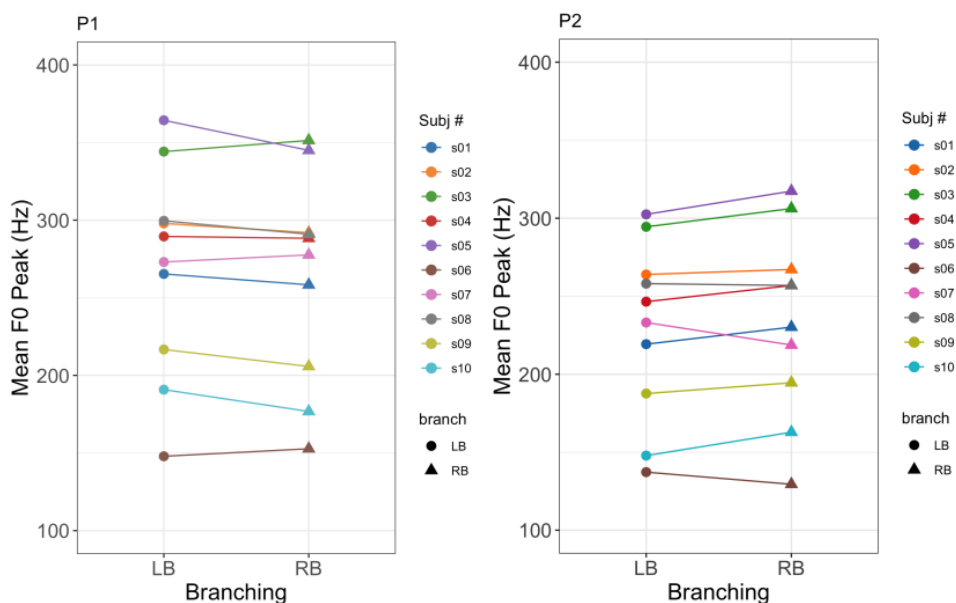


図 4 : 各参加者の P1 (左)、P2 (右) の平均最大 F0 値

さらに、P2 の平均を文の構造ごとに比較すると、RC の方が NP に比べて全体的に F0 が高く、RB 条件が LB 条件よりも高い傾向にある (図 5)。また、P2 は文ごとにやや異なった傾向を示していた (図 6)。7 種類中、6 種類の文では RB の方が高い傾向にあったが、*cap* のみ RB 条件の方で P2 が低い。

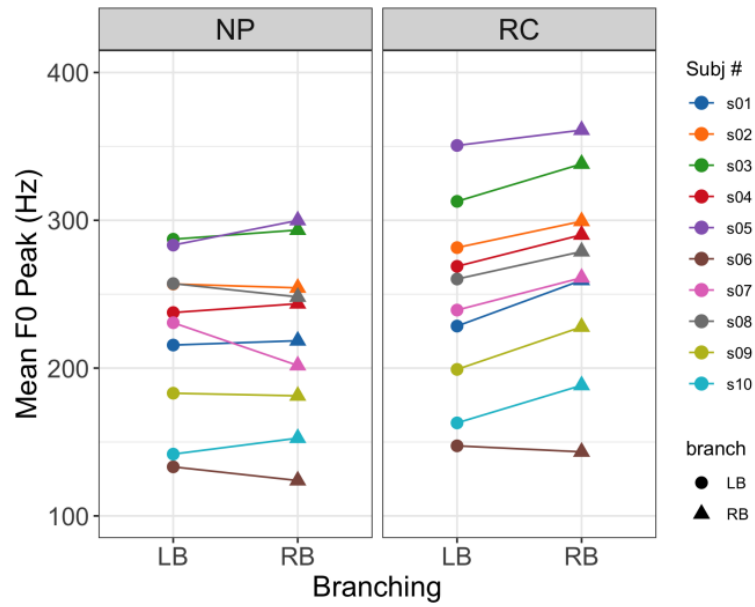


図 5：各参加者の P2 の平均値

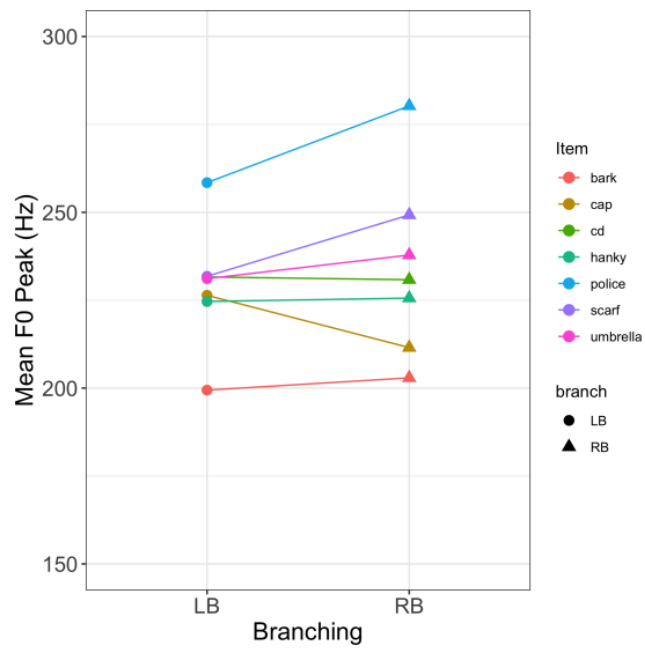


図 6：各文の P2 平均値

F1、F2 の平均と SD については、F1 においてのみ、LB 条件で 6Hz 程度高い結果であった。表 5 にそれぞれの平均と SD を示す。

表 5 : F1、F2 の平均最大 F0 値と SD

branch	F1_mean	F1_sd	F2_mean	F2_sd
LB	255.816796	68.6974929	215.179945	57.6841675
RB	249.210112	67.9194163	214.03654	60.1214944

P1、P2、F1、F2 の平均 F0 について項目間と参加者間でウェルチの t 検定を行った。検定に使用した F0 値は被験者ごとに z スコアに変換されている。表 6 に項目間 t 検定の文条件ごとの結果を示す。F0 に関連した 4 つの変数のうち、全てにおいて有意差のある文はなかった。最も多い場合でも、*cd* と *police* の条件で 2 変数のみであり、その組み合わせは文によって異なっている (*cd* では P1, F1、*police* では P2, F2)。*bark*、*umbrella* については有意傾向も全く見られなかった。一方、参加者ごとの t 検定で有意差が確認されたのは s07 の F2 のみであり ($t = 2.89$, $df = 12.0$, $p < 0.05$)、それ以外は有意な差は現れなかった。

表 6 : P1、P2 の項目間 t 検定の結果 (0 ‘****’ 0.001 ‘***’ 0.01 ‘**’ 0.05 ‘.’ 0.1 ‘.’ 1)

stcName	P1	P2	F1	F2
<i>bark</i>				
<i>cap</i>		.		*
<i>cd</i>	*		**	
<i>hanky</i>				.
<i>police</i>		*		*
<i>scarf</i>		*	*	.
<i>umbrella</i>				

t 検定に加え、それぞれの独立変数の効果量を調べるためにロジスティック回帰分析を行った。 t 検定同様、全ての F0 値は z スコアに変換されている。ロジスティック回帰分析ではそれぞれ独立変数の異なる 3 つのモデルを作り、ステップワイズロジスティック分析によって AIC を比較することで、観測データに最もよく当てはまるモデルを検討した。目的変数は枝分かれ条件であり、RB を 1 とした。独立変数には、4 箇所ピークの値と P1 と P2 の差 (zP_diff)、F1 と P1 の差 ($z1_diff$)、F2 と P2 の差 ($z2_diff$) に加えて、文の構造や AP1・AP2 のアクセントに関する変数を異なった組み合わせで入れた。下の表 7 は基本の 3 モデルに用いられた F0 に関連する変数と、それぞれの AIC を示している。なお、全てのモデルに文構造やアクセントに関する変数は含まれているが、紙面の都

合上、表 7 では省略している。交互作用項を含むモデル 3 に比べると、モデル 1、モデル 2 はやや AIC が小さい結果となった。

表 7：基本のモデルと AIC

Model #	F0 Independent Variables	AIC
1	P1_z + P2_z + F1_z + F2_z	200.3348
2	zP_diff + z1_diff + z2_diff	200.3968
3	P1_z + P2_z + F1_z + F2_z + P1_z:P2_z + P1_z:F1_z + P2_z:F2_z	201.0203

これら 3 つのモデルにそれぞれステップワイズ・ロジスティック回帰分析を行い、3 つのモデルの中から AIC の最も小さいモデルを検討した。その結果、最良モデルは P2、F1、F2 に加えて、P2 と F2 の交互作用項が含まれているモデルであった (AIC = 192.0737)。

この最良モデルを用いて、4 つの変数のうちどれが最も重要な変数であったかを係数を用いて検討する。表 8 にこのモデルの係数表を示す。 β 値の絶対値を比較すると、P2 が最も大きいことがわかる ($\beta_{P2} = 0.998$)。ロジスティック回帰の β 値の指数は、その変数が 1 単位変わった際にどれだけ目的変数の事象の確率が変化したのかを示す。この場合、P2 が 1SD 分上昇すると、RB と分類される確率が約 2.7 倍になる。また、交互作用項では P2 と F2 がそれぞれ 1 SD ずつ上昇すると、RB と分類される確率が約 1.9 倍となる。一方、F1 や F2 は係数が負の値であるため、F0 が上昇するほど RB となる確率が下がる。このモデルを用いた予測オッズを図 7 に示す。青い線で示された P2 の F0 変化の効果は他の 2 つの変数に比べて非常に大きいことがわかる。例えば、P2 の F0 が 2SD 分上昇すると、RB と分類される確率が約 7 倍になる。

表 8：最良モデルの係数表

term	estimate	std.error	z-value	Pr(> z)
(Intercept)	-0.0431905	0.30523335	-0.1414999	0.88747507
P2_z	0.99783902	0.378013	2.639695	0.00829807
F1_z	-0.491403	0.23810667	-2.0637935	0.0390373
F2_z	-0.2572184	0.31086294	-0.8274336	0.40799134
P2_z:F2_z	0.62143122	0.32144446	1.933246	0.0532059

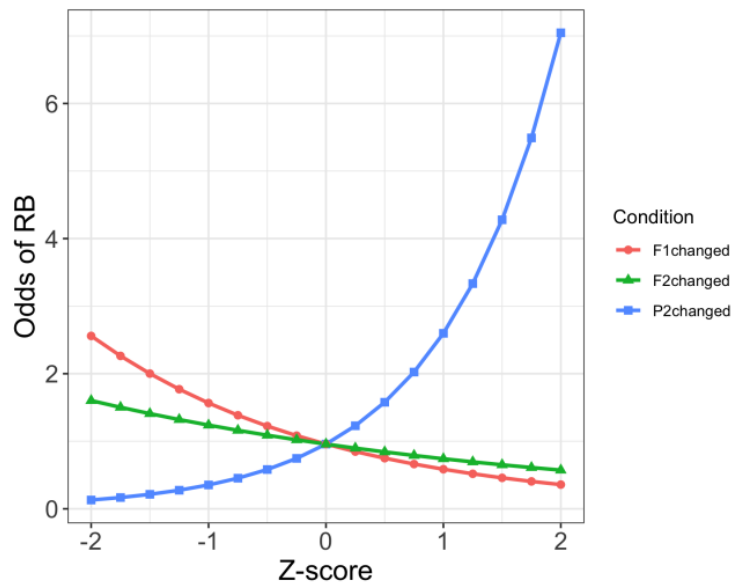


図 7：最良モデルを用いた予測オッズ。●は F1 以外の F0 が 0 の場合の F1 の変化によるオッズの変化を示す。▲はそれ以外の F0 が 0 の時の F2 の変化によるオッズの変化を示す。■はそれ以外の F0 が 0 の時の P2 の変化によるオッズの変化を示す。

3.2. ポーズ長

F0 に加えて、本研究ではポーズ長についても分析を行った。しかし、ターゲット文ごとに語頭の音素が異なっており、一部のポーズ長には後続する語の閉鎖子音の閉鎖区間長も含まれている。そのため、本研究では文ごとの t 検定のみを行った。まず、文ごとのポーズ長を表 9 と図 8 に示す。

文によって、かなりポーズ長の分布にばらつきがあることがわかる。例えば、*umbrella* は T1, T2 どちらもポーズと呼べる十分な長さはなく、一方で *police* のようなポーズ長のばらつきが大きい文もある。全体的に P1 は P2 に比べて平均ポーズが長い条件が多く、文ごとの T1 を確認すると RB 条件でポーズがやや長い傾向にある。また、T2 については *bark*, *police*, *scarf* など、LB 条件でやや長い傾向が見られる文もある。

次に、T1・T2 それぞれの項目ごとの t 検定では、T1 の *scarf* ($t = -3.16$, $df = 7.68$, $p < 0.05$)、T2 の *police* ($t = 2.39$, $df = 8$, $p < 0.05$) において有意な差が確認された。具体的には、*scarf* については T1 において RB の方が有意に長く、*police* においては LB の方が有意に T2 のポーズが長いことを示している。しかし、それ以外の組み合わせでは有意な結果は示されなかった。

表 9 : T1、T2 の平均ポーズ長 (秒) と SD

ID	T1_mean	T1_sd	T2_mean	T2_sd
<i>bark</i>	0.06629166	0.08077952	0.02635679	0.04145098
<i>cap</i>	0.07960206	0.07618827	0.05519835	0.02529502
<i>cd</i>	0.01514593	0.04589282	0.07365349	0.02924432
<i>hanky</i>	0.08670842	0.07177871	0.01987249	0.04507629
<i>police</i>	0.15129014	0.14559312	0.07454243	0.15432112
<i>scarf</i>	0.15564527	0.13725364	0.01842233	0.05825652
<i>umbrella</i>	0.01600032	0.04928183	0.0658333	0.02034513

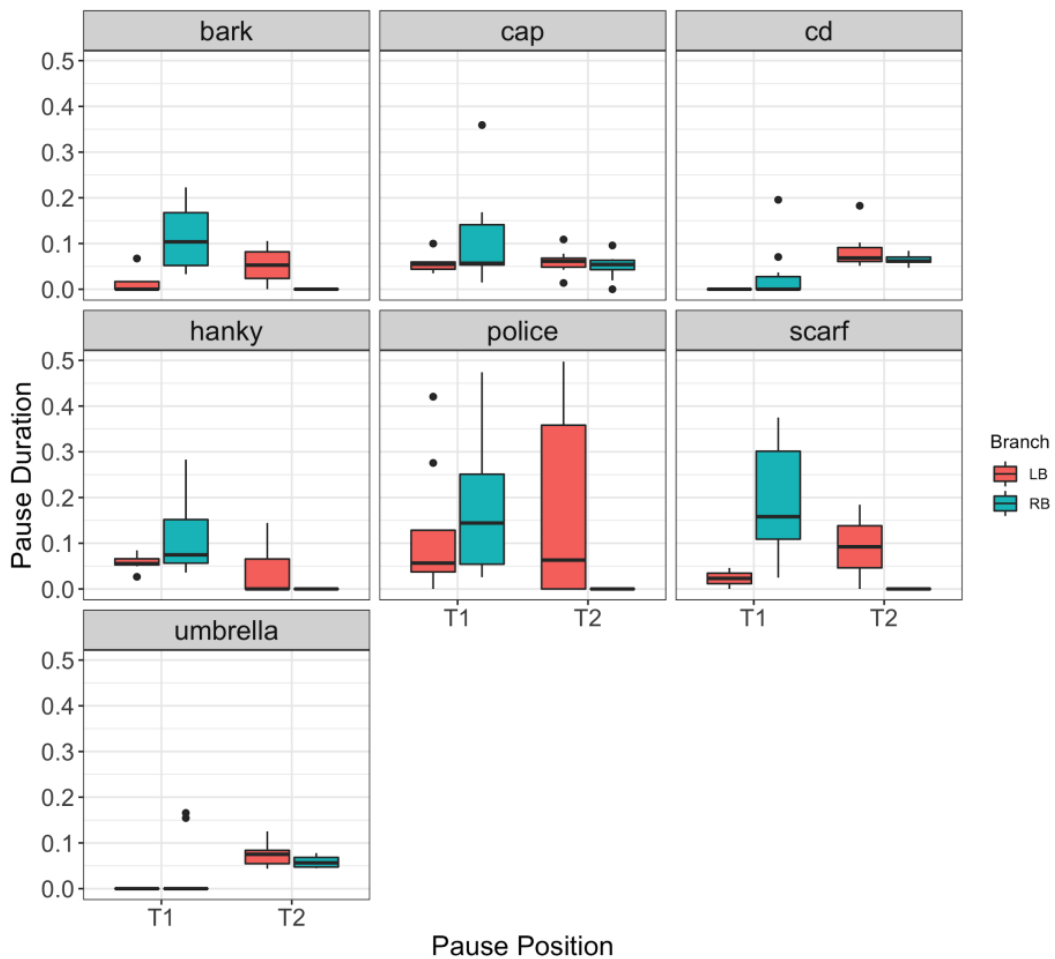


図 8 : 各文の T1 と T2。それぞれ右側の赤い箱が LB 条件のポーズ長、左側の青い箱が RB 条件のポーズ長を示す。y 軸は秒数を示す。

4. 考察

4.1. 全体のまとめ

本研究では F0 についてアクセント核の F0、文節末 1 モーラの F0、時間長についてはそれぞれ文節の間のポーズ長を計測し、枝分かれの違いがどのように反映されているのかを検討した。項目ごとの *t* 検定の結果を以下の表 10 にまとめる。

表 10 : F0・ポーズ長の *t* 検定結果まとめ

ID	Type	P1	P2	F1	F2	T1	T2
<i>police</i>	RC		*		*		*
<i>cap</i>	NP		.		*		
<i>scarf</i>	RC		*	*	.	*	
<i>hanky</i>	NP				.	.	.
<i>umbrella</i>	NP						
<i>bark</i>	NP						
<i>cd</i>	NP	*		**			

枝分かれ構造の読み分けは、文によって位置が異なっていることがこの比較からわかる。ただし、F0 についてもポーズ長についても *umbrella*, *bark* の 2 文は枝分かれ条件間で全く違いがない。同じ「青い」を 1 文節目にもつ *cd* は P1 で有意差が確認されているため、有核・無核など語アクセントによる違いではないと考えられる。

また、文タイプごとに比較をすると、文タイプが関係節 (RC) である場合は F0 と時間長どちらも有意差が見られた項目があったのに対し、文タイプが名詞句 (NP) である場合は F0 情報のみに有意差が見られた。このことから、F0 はより汎用的に用いることのできる韻律的特徴であると考えられる。Azuma & Tsukuma (1990) では知覚実験の結果から、F0 が韻律情報としてより重要であることを示していたが、同様の傾向が発話にも認められる可能性を示している。

4.2. F0

P1 と P2 では P2 の平均 F0 が RB 条件において高くなる傾向であった。話者別では 10 名中 7 名がその傾向を示していた。次に、P2 の平均が関係節をもつ比較的長い文において高い傾向が示された。これは、文が長いことによって、話者がそもそも文頭の F0 を高く設定し発話していた影響であると考えられる。さらに、2 つ目の AP の後も AP が 3 つ以上続く発話であるため、RB 読みでは P2 の位置で一度 F0 をリセットし、高い位置から読み始めていたことも考えられる。

それぞれの平均 F0 最大値を比較するにあたり、項目ごとと参加者ごとに t 検定を行った。どの F0 計測箇所にも有意差がない文は 7 文中 2 文であった。

また、ロジスティック回帰分析により、P1 以外の F0 情報が含まれたモデルが採用され、P2 が最も重要な F0 情報であることがわかった。このことから、仮説 4 は支持されたと言える。P2 については仮説 1 通りの結果であったが、F1 については仮説 2 とは反対の結果であった。

4.3. ポーズ長

t 検定において有意な差が確認されたのは、T1 の *scarf*, T2 の *police* であった。*scarf* の T1 において RB の方が有意に長く、*police* の T2 では LB 条件において有意にポーズが長いことを示している。非常に限られるが、これらの結果は仮説 3 を支持している。T1 の *scarf* のみが仮説 3 を文字通り支持する結果であるが、韻律を用いて統語境界を強調するという観点では T2 において LB が長くなることもこの前提に沿った結果であると言える。加えて、この T2 の結果は、広瀬 (2006) を部分的ではあるが追認する結果である。

次に、有意差の 2 つの文はどちらも関係節を含むターゲット文であることにも注目したい。発話が長いほど統語的境界の位置にポーズを入れて、枝分かれ構造を読み上げていたことが考えられる。発話が長くなると韻律がより強調される傾向は Venditti (1994) でも指摘されていた。しかし、今回使用した関係節を含む曖昧文は 2 つのみであり、上述した傾向が一般化されるかどうかについてはさらなる実験が必要である。

5. 結論・今後の課題

本研究では、枝分かれ構造がどのように F0 や時間といった韻律的情報に反映されているかを検討した。F0 については、2 つ目の要素のアクセント核の F0 が重要であることが示された。時間長については、関係節を含む比較的長い文の一部のポーズにおいて有意な差が認められた。しかし、弁別されるポーズの位置は文によって異なっていた。

まず課題としてあげられるのは、サンプル数の少なさである。今回は 10 名の話者の 1 試行の発話を録音したデータを用いたが、発話間でのばらつきなどを考慮すると、項目数を減らした複数回発話の分析が必要となるであろう。2 つ目の課題はポーズ長も含めた包括的な検討が挙げられる。今回の実験でもポーズを計測したが、用いた刺激文では語頭子音をコントロールしなかったため、項目ごとの比較のみに留まった。そのため、今後の実験では刺激文に含まれる計測箇所の語頭子音をコントロールすることが求められる。そのような実験を通じて F0 情報と時間情報を比較することで、枝分かれの構造の韻律的実現に関するより包括的な議論を行うことが可能となる。

謝辞

本稿を仕上げるにあたり、丁寧に指導して下さった田嶋圭一先生に深く御礼申し上げます。また、分析方法について有益なアドバイスをくださった北原真冬先生に感謝いたします。さらに、今回の産出実験に参加していただいた 10 名の参加者の方にも改めて感謝申し上げます。

参考文献

- 郡史郎 (2020) . 『日本語のイントネーションーしくみと音読・朗読への応用』, 大修館書店.
- 広瀬友紀 (2006). 「話者の意図と聞き手の理解: 語彙アクセントの隠れた作用」, 『認知科学』, 13 (3) , 428-442.
- Azuma, J., & Tsukuma, Y. (1990) . Prosodic features marking the major syntactic boundary of Japanese: A study on syntactically ambiguous sentences of the Kinki dialect. In *First International Conference on Spoken Language Processing*.
- Hirose, Y. (2020) . Sequential Interpretation of Pitch Prominence as Contrastive and Syntactic Information: Contrast Comes First, but Syntax Takes Over. *Language and speech*, 63 (3) , 455-478.
- Igarashi, Y. (2018) . Chapter9: intonation, In Hasegawa, Y. (2018) . *The Cambridge handbook of Japanese linguistics*, 181-201.
- Ito, K., Arai, M., & Hirose, Y. (2015) . The interpretation of phrase-medial prosodic prominence in Japanese: Is it sensitive to context? *Language, Cognition, and Neuroscience*, 30 (1-2) , 167-196.
- Ladd, D. R. (2008) . *Intonational phonology*. Cambridge University Press.
- Pierrehumbert, J.B., & Beckman, M.E. (1988) . *Japanese Tone Structure*. Cambridge, Massachusetts: MIT Press.
- Selkirk, E., & Tateishi, K. (1991) . Syntax and downstep in Japanese. *Interdisciplinary approaches to language*, 519-543.
- Venditti, J. J. (1994) . The influence of syntax on prosodic structure in Japanese.

注

-
- ¹ Selkirk & Tateishi (1991) も同様の主張をしている。彼らは RB 条件の 2 つ目の文節に Major Phrase の境界があるため、ピッチのリセットが発生するのだとしている。
- ² 郡 (2020) は文節末のイントネーションを類型分けし、F0 曲線のパターンなどからそれぞれの機能を紹介している。郡の分類に則ると、パイロット実験で確認され

たのは強調型上昇調であると考えられる。この音調は「相手にわかってほしいという気持ちを込める」時に使う音調だと紹介されている。

³ *cd* は広瀬 (2006) , *umbrella* は Hirose (2020) , *scarf* は Venditti (1994) , *cap* と *hanky* は Ito (2014) で用いられた文である。

⁴ 「犬」は単語のみでは尾高型の語アクセントであるが、後ろに助詞「の」がつくと無核語のアクセントパターンになるため、無核語と同様の表記がされている。

⁵ 参加者によって録音の環境が異なるため、一部の音声には環境による雑音などが大きく含まれていた。録音環境の統一が現状では困難であったため、音声は最低限アナレーションができる程度にスペクトログラムの様相がはっきりとしており、F0 曲線を取得することができる状態で録音した。

⁶ 画像でのヒントが示されたのは、7つのターゲット文のうち *cap*, *hanky*, *cd*, *umbrella* の4つである。画像の有無による韻律への影響は非常に興味深い課題ではあるが、今回の研究では問題として扱わない。

⁷ ここで用いられている F1, F2 などはフォルマント周波数ではなく、基本周波数の計測箇所の名称である。

先進国から退きつつあるこの国で音声学を研究することについて： 「選択と集中」の過ち

北原 真冬

上智大学 外国語学部
E-mail: mafuyu@sophia.ac.jp

1. はじめに

ちょうど1年前、「COVID-19と教育・研究」というタイトルで、このワーキングペーパーに拙文を寄せました。それ以降、世界におけるこの感染症の蔓延は各地で悲惨な事態を招いています。日本においても死者・重症者が日々次々と報告されるのはもちろんのこと、非感染者であっても、恐怖に煽られながら強いストレスのもとで生きることを強いられています。政治・経済が右往左往する中、大学における研究・教育も強い抑圧を受け、ほとんどの学会はオンライン開催で命脈を保っているのが現状と言えましょう。

本稿はそのような COVID-19 の影響に直接言及するよりも、この日本という国における研究の役割、弱み、研究者として身を立てて行く上での困難点という点について私見を述べたいと思います。砂に覆われたレリーフが姿を表すように、社会が予想もしなかった強さで揺さぶられる中で、今まで見えにくかったところがはっきりと見えてきたのではないかと思うが故に、この機会を捉えて、大学院生や若手研究者の将来について展望したいと言うのが目的です。

非常に大きなテーマをいきなり掲げましたが、日本の研究力が年々低下しているということは、大学関係者にとっては飽きるほど聞かされている言説であり、文部科学省の資料[1][2]においても詳しいデータが挙げられています。ただ、いくら多数の指標で測ったとしても国家の総合的な研究力というのは研究者個人にとってはなかなか実感しにくいものです。また後に述べるように、「文系・理系」という悪名高い分断により、二つの系はそれぞれ研究力と呼ばれるものを全くの他人事として見る癖が付いているという印象があります。例えば研究費の額や論文のインパクトファクターという指標は系によって、さらには分野によって全く規模も受け止め方も違いすぎるので、それらを平均したり総合してランク付けをしたりというのは、分野の壁をまたいだ公正性を持っているとは言えません。研究者個人としては、「研究力と言っても他分野のことはそれぞれの事情があるので…」と言葉を濁すしかないのが現状です。しかしながら、人口も GDP も縮小し、賃金水準も下がっていく中で、政策の舵取りがどれほどうまくいったとしても、研究と言われるものの国単位としてのアウトプットはこれからも下がる一方でしょう。個人や分野の情勢に多少の浮き沈みがあっても、総和が減少するという趨勢に変化は望めません。「選択と集中」というよく取り沙汰される方法論は、研究の中の一部を突出させる意図を持ちますが、意図は(偶然に左右されるので)必ずしも実現するとは限らず、一方で総和は着実に減ることが約束されている、全く研究という営みに向かない方法論です。しかも、国家の規模が縮小する

につれて、限られたパイの中で「選択と集中」をする圧力は高まります。この負のフィードバックはこの国の一般的・構造的な問題であり、「先進国から退きつつある」というタイトルの意味するところです。

少し話が大きくなりすぎました。次節では言語学・音声学の分野に限って、学会の歴史と役割を振り返りながら「選択と集中」についてさらに掘り下げます。

表 1: 言語学・音声学に関連する古参の学会

名称	設立年	会員数
日本音声学会	1926	713
日本英文学会	1929	2800
日本音響学会	1936	3868
日本言語学会	1938	1799
日本語学会	1944	1467

2. 学会

言語学・音声学に関連する学会の中でも特に歴史が長いものを表 1 に示します。

これらの学会は、それぞれ戦前・戦中期に設立され、言語学・音声学研究の核となって日本における研究を発展させてきました。しかし、今やどの学会も新入会員の減少と収支の縮小に悩み、例えば音声学会では会費の値上げ、学会誌の発行回数の漸減や電子化に踏み切っています。また学会という組織の多くは、専従の職員を持たないか、持っていたとしてもごく少数の人員で事務作業を行なっていて、その業務の大半は教育・研究機関や企業において本業を持つ人々の献身によって成り立っています。

もちろん、学問・研究は歴史のある学会だけで行われるものではなく、新たな学問潮流やパラダイムの提唱に伴い、新たな研究会や学会が興り、パワフルなリーダーや運営陣によって盛んに発表や交流が行われることはあります。しかしながら、特定の学派や理論体系に立脚した学会や研究会は、構造上その成立が前節で述べた「選択と集中」に依存していて、表 1 に挙げた歴史ある学会に比べるといささか間口が狭いのも確かです。

また、歴史ある学会は、その学会において発表すること自体にいくばくかの権威が付与される、というのが一般的には期待されています。発表応募が多数に上り、有限の会場と時間の都合から、採択に激しい競争があれば、その権威にも根拠があることとなります。しかし、それだけではまたも間口が狭い問題に悩まされることになるわけで、学会は同時に若手を励まし、発表の機会を与えることが望まれます。そうしない限り、表に出る研究はベテランのものばかりになって、若手の研究者が育ちません。

そこで、様々な学会が、いくつかの手法を組み合わせる若手を奨励する試みを行なっています。若手を対象にした優秀論文賞、優秀発表賞などを設けるだけでなく、応募の手続きを可能な限り簡略化して、対象者の発表には必ずフィードバ

ックを与えるシステムもあります(音響学会)。また、会場と時間の有限性に対しては、口頭発表の他にポスター発表の機会を設けることで対処しています。

しかし、これらの権威や施策が本当に機能しているのかどうかは、各学会の中だけでなく、横断的に検証されるべき課題であると考えます。例えば音声学会は発表に応募する件数がそれほど多く揃うとは言いがたく、権威を保證するような十分な競争によって成り立っているとは言えない時もあります。表彰についても、なかなかハードルが高く、博士課程に在籍する学生が音声学会や言語学会で優秀賞を取ることは極めて困難です。実はここでも「選択と集中」の害が現れていると考えています。発表の採択にしても表彰にしても、それは文字通り「選択」を行いそこに名誉や業績を「集中」させるものです。しかし入り口のところで十分な応募がなければ、選択のしようもありません。会員数が多い時代であれば機能していた方策も、会員数が縮小しクリティカルマスを失ってしまった時には、もはや形骸化してしまう恐れが多分にあります。

3. 文系と理系

前節は主に規模の小さな「文系」の学会についての話であり、「理系」の学会は大規模でそのような悩みとは無縁であろうという意見もあるかもしれません。しかし2015年の音響学会の記事[3]によれば、工学・理学系の学会もほとんどは会員数の減少に悩まされているだけでなく、人口動態との相関から重回帰分析によって予測すると音響学会も会員数の激減が見込まれるとのこと。学会の衰退を防ぐ対抗策として(1)「定年退職後に入りたくなる学会」を目指す、(2)資格・規格の認定機能の整備、(3)海外会員の増加、の三つが提言されていますが、どれもなかなか困難だけでなく、劇的な効果を生むとは言いがたいのではないかと思います。そもそも、文系と理系という分断も高校生の「選択と集中」の結果ではないかと疑っています。少し古くなりますが、拙論[4]において以下のように述べました:『「暗記さえすれば確実に点が取れる」という理由で生物か化学しか履修しない生徒が多いと聞く。論理的に考えさせるような問題を含む数学や物理は、確実に得点できるかわからないから出来る限り回避される』(p.64)。人生のキャリアパスをデザインする上で、きわめて論理的に考えた結果の選択が、受験における「論理的」な科目を避けることにつながるというのは、非常によくできた皮肉だと言えます。問題は、高校生が「選択と集中」という方法によって社会にある既存の分断に身を委ねてしまうところであり、極めて間違っています。しかし、文系と理系の壁をとっばらうとか、文系の学生にもデータサイエンスを必修にするとか、そのような対症療法では、高校生の論理的な「選択と集中」に対する歯止めとはならないでしょう。誰にでも自分の得意・不得意を見極めて、不得意な分野では勝負をしない、という選択をする自由はあります。問題は社会も技術も刻々と変化し高度化していく中で、その見極めが将来に渡って有効である保証が全くないところです。

4. おわりに

音声学の中身にはあまり関係のない領域の話ばかりではありましたが、ここまで、社会、組織、そして個人というそれぞれのレベルにおける「選択と集中」の過ちを述べてきました。結局のところ、「選択」はあくまで過去の知見に基づく将来の予測によってなされるのであって、現実には手ひどく裏切られることは大い

にありえます。COVID-19はそのことを、悲しいほど鮮やかに示してしまいました。ほんの2年前には、世界がこんなにも翻弄され、移動にも食事にも人の集まりにも著しい制限が加わるなど全く予測できませんでした。しかも、予測が大きく裏切られたことに対して「選択」をやり直し、「集中」すべきところに力を注ぐように計画を変更するのはとても困難で、COVID以前と全く同じ世界に戻る、あるいはもう戻りつつあると信じて動いてしまっている社会・組織・個人の行動を日々見せつけられるばかりです。

だからこそ、私たちの最後のよりどころは、小賢しい将来予測に基づいた「選択」ではなく、純粋な知的好奇心を原動力として科学を追求するところにあると信じていたいと思います。それぞれのレベルで短期的なわかりやすい成果を出すための「選択」をすることは止められないかもしれません。資金を得るために、気の進まない「選択」を強いられることも多いでしょう。知的好奇心はそのような「選択」とは正反対のベクトルです。たとえ現実と折り合いをつけるための「選択」を繰り返していても、意識の中でそれを失わない限り、研究という営みは続けられる、続けていかねばならないと感じています。

参考文献

- [1] 文部科学省科学技術政策研究所 (2013). 「日本の大学における研究力の現状と課題(ver.2)」, <https://www.nistep.go.jp/wp/wp-content/uploads/NISTEP-booklet01.pdf>
- [2] 文部科学省 (2019). 「日本の研究力低下の主な経緯・構造的要因案: 参考データ集」, https://www.mext.go.jp/content/1407654_009.pdf
- [3] 伊藤彰則 (2015). 「会員数から見る日本音響学会の現在と未来」, 『日本音響学会誌』, 71, pp.5-6.
- [4] 北原真冬 (2008). 「書評: 藤村靖著『音声科学言論』岩波書店 2007」, 『音声研究』, 12, pp.63-69.



Sophia University
Phonetics Lab
上智大学音声学研究室

Sophia University Working Papers in
Phonetics

Vol.27 2021

2021年4月30日発行

編集発行 上智大学音声学研究室
代表者 北原 真冬

〒102 - 8554 東京都千代田区紀尾井町7 - 1
電話 03 - 3238 - 3492