

母語非母語話者間の会話支援のための

テキスト音声会話の分析

—日本人と中国人による日本語会話を題材として—

2022年2月

埴 裕美

母語非母語話者間の会話支援のための

テキスト音声会話の分析

—日本人と中国人による日本語会話を題材として—

筑波大学

図書館情報メディア研究科

2022年2月

埴 裕美

## 概要

### 母語非母語話者間の会話支援のためのテキスト音声会話の分析

#### —日本人と中国人による日本語会話を題材として—

2019年に改正出入国管理法が施行されてから、日本に中長期間在留する外国人が増加している。在留外国人に最も多い国籍は中国人であるため(出入国在留管理庁 2021)、日本人と中国人が日本語会話をする機会が多い。本研究は日本人と中国人の会話を対象とした分析の研究を行った。

外国人との会話では、内容が伝わりにくく共通理解を得にくいなどの問題がある。中国人と日本人の共通理解が困難である原因として、会話の際に中国人は日本語をリアルタイムで理解する必要があることが挙げられる。幼いころから周囲の環境によって習得した言語の使用者は母語話者(NS)と呼ばれ、母語を習得したのちに学習によって習得した母語以外の言語使用者は非母語話者(NNS)と呼ばれる。本研究では、NSの代表として日本人(日本語を母語とする)を採用し、NNSの代表として中国人(日本語を非母語とする)を採用して、日本語会話における共通理解に至る過程を中心に検討した。

本研究では、会話の一部分をテキスト(文字による情報伝達)でも発信する会話を試みた。会話の音声だけでなく、会話の主題や対象など一部分でもテキストで発信されれば、NNSが会話を理解する補助となると考えられる。NSが無理なくできる手助けの範囲でNNSの負担軽減につながる支援として、NSが会話の一部分をテキスト化して、それをNNSに提示するという手法を用いて、NNSとの非対称なコミュニケーションの問題を解決することを目標とした。NNSとの会話支援に用いられるコミュニケーション・メディアにはビデオとテキスト(Echenique et al. 2014)、対面とテキスト(Chen et al. 2018b)、音声とテキスト(Gao et al. 2014)などの従来研究があるが、本研究はNSが会話音声の一部をテキスト化することにより、音声とテキストを組み合わせた会話支

援の検討を行った。

本研究は代表的な従来研究(Gao et al. 2014) (Pan et al. 2017) と同じく自由度の高い課題を用いて、NS が会話音声の一部をテキスト化し NNS に提示するという、音声とテキスト併用の効果を検討した。会話支援研究では会話を実践してコミュニケーション・メディアの支援効果を測定することから、課題とメディアの最適な組み合わせを調べた研究や、最適な組み合わせの仮説を検証する研究も行われた(McGrath 1993) (Mennecke et al. 2000)。初期の NNS との会話支援の研究では、会話上の役割を固定し NNS が NS からの指示に従って正確に実行する課題や、発話内容の制限がある課題が多く用いられた。その後、音声とテキストを用いた代表的な従来研究(Gao et al. 2014) (Pan et al. 2017) で自由度の高い課題で NS と NNS のコミュニケーション支援が行われたが、自由度の高い課題における共通理解が示されていないため、本研究は自由度の高い課題を用いて音声とテキストによる支援を検討した。

本研究では NS と NNS の会話実験を行い、NS が会話音声の一部をテキスト化し NNS に提示することによる共通理解達成への効果と、音声のテキスト化がどのようになされ、それがどのように会話に影響するかという、テキスト音声会話の 2 つの側面について明らかにする目的を達成するため、それらに共通する 1 つの実験を行った。実験は NS と NNS による 2 人 1 組の音声会議で、16 組の参加者に対して実施した。NNS は日本語能力試験 N1 の中国人留学生で、日本人は日本生まれで日本語 NS の大学生であった。音声会議は非公式のディベートとし、すべての組が原発問題と死刑問題の 2 つの議題について「音声とテキスト条件」と「音声のみ条件」の 2 つの条件に参加する被験者内計画で行った。音声会話終了後、参加者は会話についての質問紙調査と自由再生課題に回答した。自由再生課題では、参加者は会話後に何も参照せずに会話中の単語や文章を記入した。

本研究の 1 つ目の目的として、音声とテキストを併用することによる共通理

解達成への効果を明らかにするため、NS と NNS が自由度の高い課題を行う日本語音声会議における共通理解を調査した。会話の協調モデルでは、会話は情報の提供を行う提示と、聞き手が情報を理解したことを示す受理によって構成される。会話のやりとりを継続することで相手と情報の共同解釈を構築(grounding)する。共同解釈によって一致した情報や知識などを共有し、相互に相手の発話を理解できていると確信できる基準を満たすことで、相手と共通の理解を達成(common ground)する。共通理解の達成は会話の最終段階であるという視点から、本研究では NS と NNS の共通理解の分析を行った。音声とその一部のテキストが共通理解を促進するかどうかについて、(1)音声とその一部のテキストの利用は共同解釈の構築を増加させ、(2)結果として共通理解を向上するため、「わかった」という感覚を得られ (3)NNS の了解性と快適性が向上する、という 3 つの仮説を検証した。共同解釈の構築の分析では、実験会話の音声の書き起こしから、共同解釈を構築する会話(Clark 1996)の件数を数えた。共通理解の達成の分析では、NS と NNS の自由再生課題回答の内容の一致件数を数えることで共通理解の達成を検証した。了解性と快適性については、参加者の質問紙への回答に対して分析を行った。

本研究の 2 つ目の目的として、音声のテキスト化がどのようになされ、それがどのように会話に影響するかを明らかにするため、NS による会話音声のテキスト化が、(1)いつ、(2)どのくらいの量で、(3)どのような内容で行われ、会話については、(4)テキスト化によってどのような話し方に変化するか、という 4 つの分析を行った。会話音声のテキスト化のタイミングについては、NS によるテキストを発話に対応付けた上で、発話とテキストの時刻差を算出した。また、テキストについては、形式的な特徴から 6 種類の名詞句 (Clark and Wilks-Gibbs 1986) に分類することで検討した。話し方の変化についても同様に、6 種類の名詞句に分類して検討した。

本研究が明らかにしたことの 1 つ目は、NS と NNS の音声会話における補足的なテキストの有効性である。NS によって作成された音声の一部テキス

トを NNS に提示することにより、共通理解が向上し、NNS の了解性と快適性が向上することを示した。これらは、従来研究(Echenique et al. 2014)に対し日本語の会話分析を行い共同解釈を構築する会話例を示した点、自由度の高い課題を用いた音声会話における補足テキストの有効性を示した点で、本研究分野に貢献した。

本研究が明らかにしたことの 2 つ目は、詳細な調査による、NS による音声テキスト化のタイミングと内容と、話し方の変化を示したことである。テキストは NS 発話から 2 秒後に表示され、NS と NNS の発話を平均すると 3 秒後に表示され、5 字程度の説明を補足する名詞句が表示されたことを示した。つまり、音声から NS により作成されたテキストは、会話中でも読める量が十分早く提供され、有効に利用できるために必要な水準を満たすことを示した。また、参加者の話し方が、早く短い発話から補足が多い発話に変化することを示した。これらは、従来研究とコミュニケーションの状況と分析対象が異なるため、本研究分野に貢献した。

本研究で行ったテキスト音声会話のテキストを自動化する会話支援ツールのデザインとして、音声のすぐ後に補足説明などを 5-6 字でテキスト表示させる支援が考えられる。本結果は、NS と NNS とのコミュニケーションを支援するツールの評価に、また、支援ツールの結果予測などに利用できると考えられる。多人数遠隔環境の音声会話で同様の補足テキストの効果を検証するのは、今後の課題である。本研究を足掛かりとして、今後は他の題材、参加人数、実験環境などのコミュニケーション環境を変え、これらの要因が NS と NNS を含む音声会議にどのような影響を与えるのかを明らかにしていく。

## **Abstract**

An analytical study of text-enhanced speech conversation for supporting conversation between native and non-native speakers: A case of Japanese conversation between Japanese and Chinese

Since the Immigration Control and Refugee Recognition Act was revised in 2019, opportunities to talk with foreign residents have increased in Japan. According to Government Statistics of Japan, especially Chinese people had the largest population among other foreign residents in Japan. This study focuses on analyses of the effectiveness of conversation to support communication between Japanese and Chinese.

In fact, communication problems with foreign residents have increased in schools and work places. Mutual understanding between native speakers (NS) of Japanese and non-native speakers (NNS) of Japanese has become a critical issue in Japanese society. During communicative acts between NS and NNS, mutual understanding is not always assured, and NNS may not retain details of the conversation. Previous studies including Gao et al. (2014) and Pan et al. (2017) state that NNS need extra time to understand second language utterances while conversing with NS, which adds an extra burden and a significant disadvantage to their cognitive load.

This study aims to offer a solution to these communication issues between native and non-native speakers of Japanese by exploring and analyzing the use of text transcripts from utterances in audio conference settings. Text transcripts from utterance are similar to subtitles on news programs, and may help NNS to understand conversations with NS better. Text transcripts show some portions of speech by NS during audio conference to balance the unevenly distributed cognitive load.

This study conducted experiments to examine the effects of voice and text from utterances typed by NS during a free-flowing debate task, during which interruptions are common and where all participants state equally valid opinions. Previous research focusing on audio and text including Gao et al. (2014) and Pan et al. (2017) investigated the effect of voice and text in idea creation and discussion. While the studies found that idea creation and discussion these pose a low-demand in terms of achieving task success, mutual understanding in free-flowing discussion and the investigation of voice text transcript usage was not adequate. Other research (McGrath 1993) (Mennecke et al. 2000) explores suitable combinations of conversation tasks and communication media, known as task-media fit hypothesis. These research focused on conducting conversation experiments with conversation tasks and multiple communication media. Communication media is the means of delivering and receiving information, which is video, voice, and text in communication research. However, these studies only examined communication between native speakers, and it is still unclear what occurs during communication tasks involving NNS and the effects of voice and essential text typed by NS remain rare. Thus, this study investigates the use of voice and text transcript from utterances made by NS in communication with NNS hopes to fill this gap.

In this study, conversation experiments focusing on the two research objectives were conducted. First, the effect of voice and portions of the conversation typed by NS in text was explored with the use of a free flowing conversation task in an audio conferencing setting, and the extent of the mutual understanding between NS and NNS was probed. Secondly, the data obtained from this experiment was used to analyze the quality and delay of text transcript from utterances made by NS. Experimental design was a single factor two-level and between subjects. NS and NNS were randomly distributed to organize dyads. Sixteen pairs of NS and NNS participated in

both of audio-only condition and audio-and-text conditions. The conversation task was debate of pros and cons of death penalty and nuclear power generation. Order of debate tasks was balanced between subjects, and every single pair participated in both of conditions within a day. NS were all Japanese who were born and grew up in Japan, and NNS were international students from China whose Japanese Language Proficiency Test's N1 level score were 118.9 on average. Post-experimental surveys were also conducted. Participants assessed their perception of conversation on likert scales questionnaire surveys, and answered free recall tests that ask comprehension and retention of the conversation content from memory of both of NS and NNS.

The first objective is to investigate the effect of voice and text technology. The present study focuses on the mutual understanding that NS and NNS establish in conversational interaction. Clark and Brennan (1991) defined conversation that speakers deliver knowledge and information during a presentation phase, and listeners receive the information at an acceptance phase. For communication to succeed, speakers and listeners have to coordinate the content and process to share information effectively. Speakers and listeners aim to reach common ground, that is mutually shared knowledge, mutual beliefs and assumptions, and endeavor to update their common ground continuously. Common ground is thus the final phase of mutual understanding. However, achieving a continuing state of common ground is challenging for NS and NNS. This thesis aims to analyze the role of voice and text technology in terms of supporting NS and NNS to attain and maintain common ground during video conferencing. Three hypothesis tests to examine the use of voice and text to increase the construction of joint interpretation to improve mutual understanding and add to the quality of the conversation were conducted.

The second objective of this study is to clarify speed, content and quantity of text transcript from utterances made by NS, and showing evidence of actual transformation of speech made by text transcript from utterances. The effects of voice and text depend on what part of the utterance the NS types during audio conference. An integrative analysis of utterance and typing to investigate timing of text transcript typed by NS was also conducted. Voice and text from important parts of conversations are strongly correlated with each other, and are bidirectional and flexible. The data collected from the conversation experiment was combined using text annotation software to investigate timing and content of NS participants' utterances and typing. In order to analyse the timing of these texts, they were matched with related utterances, and the time difference between the typing and utterances were calculated. The analysis of utterance and typing compares text and conversation content. Noun phrases (NP) have become the core of information exchange in a collaborative model of conversation. Clark and Wilkes-Gibbs (1986) define six types of noun phrases that present speaker's collaborative effort referring to utterances such as elementary, episodic, installment, provisional, dummy, and proxy, which have been used in this analysis. This study conducts an empirical study of participants' expressions by analysing data from voice and text.

This thesis investigated the effect of voice and text, and analysed the text timing and content through a conversation experiment. Results indicate that:

1. The use of text transcripts from utterances improved mutual understanding between NS and NNS.
2. NS and NNS constructed joint interpretation of conversations in an audio conferencing setting.

3. The text transcript of utterances showed up on the screen about 3 seconds after the utterance was made, which met the criteria of effective transcript.

4. Text transcripts from utterances provided supplemental information within 6 characters.

5. NS and NNS unintentionally changed their forms of speech with the use of text transcripts with adding paraphrases, examples and episodes.

6. Due to the NS adding explanations, the NNS felt more comfortable and understood on conversations.

Both objectives of the study were achieved. First, these findings highlight that voice and text have a positive effect of supplemental text transcripts on communication between NS and NNS. Data analysis of the conversation experiment indicates that NS and NNS constructed joint interpretation of conversations and improved mutual understanding, thus the NNS felt more understood and comfortable. In contrast to previous work (Echenique et al. 2014) the current study focused on the construction of joint interpretation, and improved mutual understanding in a free-flowing and flexible conversation task that posed a low-demand of task success. Especially from the viewpoint of task-media fit hypothesis, this type of conversation style increases the potential richness of communication. The findings may apply to other research contexts and thus contribute to the research field of voice and text technology to assist communication.

The second objective of the thesis is to present factual evidence of timing, quantity and content of text typed by NS as well as forms of utterances during audio conference. The text showed up on the screen about 3 seconds after the utterance to provide explanation within 6 characters, which met the criteria of 5 seconds as previous work reported (Lasecki et al.

2014). Participants' speech was clarified with the parenthetical phrases using voice and text, thus NNS felt more understood and comfortable with the use of audio and text. Detailed analysis of timing and content of text transcript from utterances was investigated for the first time to contribute new insights and findings in this research field. These findings may contribute to the design of automated text transcript from utterance applications/software, and assist in their evaluation to support communication between native and non-native speakers of a language.

Lastly, this research may act as a future technological support to enhance understanding between NS and NNS in free-flowing conversations in our rapidly globalizing world.

## 目次

概要 .....	i
Abstract .....	v
図目次.....	xv
表目次.....	xvi
第1章 序論.....	1
1.1 研究背景.....	1
1.1.1 非母語話者とのコミュニケーション .....	1
1.1.2 CMCにおける会話 .....	3
1.2 本研究の目的.....	5
1.2.1 音声とテキストの効果 .....	5
1.2.2 音声をテキスト化するタイミングと内容 .....	6
1.3 本論文の構成.....	7
第2章 関連研究 .....	8
2.1 会話による共通理解の形成.....	8
2.2 テキスト化のタイミングと内容.....	11
2.2.1 テキスト化のタイミング.....	12
2.2.2 テキスト化の内容 .....	13
2.3 課題別のコミュニケーション .....	14
2.3.1 自由度の低い課題のコミュニケーション .....	16
2.3.2 自由度の高い課題のコミュニケーション .....	20
2.4 コミュニケーション・メディアの組合せ .....	24
2.4.1 ビデオとテキスト .....	26

2.4.2	音声とテキスト .....	27
2.4.3	対面とテキスト .....	28
2.5	本研究の位置づけ .....	30
2.5.1	従来研究の整理 .....	30
2.5.2	本研究の位置付け .....	32
第3章	NSとNNSの音声会議におけるテキストの効果.....	35
3.1	実験.....	35
3.1.1	実験参加者 .....	35
3.1.2	実験手続き .....	36
3.1.3	音声会議の議題 .....	38
3.1.4	実験機器.....	38
3.1.5	参加者の日本語能力 .....	41
3.2	日本語音声会議における音声テキスト化の実験結果 .....	42
3.2.1	仮説と検証方法 .....	42
3.2.2	音声とテキストによる共同解釈の構築の分析結果 .....	43
3.2.3	音声とテキストによる共通理解の調査結果.....	48
3.2.4	音声とテキストに対する質問紙調査結果 .....	52
3.3	日本語音声会議の実験結果の検討 .....	56
3.3.1	音声とテキストによる共通理解の形成 .....	56
3.3.2	音声とテキストによる記憶への影響.....	58
3.3.3	音声とテキストに対する参加者の評価 .....	59
3.3.4	従来研究を支持する結果と異なる結果 .....	60
3.4	本実験の限界.....	61

3.5	3章のまとめ	61
第4章	音声会議中に音声を書き起こすタイミングと内容	63
4.1	音声会議のテキスト化の分析	64
4.1.1	音声をテキスト化するタイミング分析	65
4.1.2	テキスト化された内容の分析	70
4.2	音声会議における音声テキスト化の分析の検討	79
4.2.1	テキスト化による遅延	79
4.2.2	テキストの量と内容	80
4.2.3	音声会話のテキスト化についての総合的な検討	83
4.2.4	音声テキスト化の分析の限界	85
4.3	4章のまとめ	85
第5章	総合考察	87
5.1	本研究の成果	87
5.1.1	本研究の整理	87
5.1.2	本研究の目的の達成	89
5.2	本研究の成果の位置付け	90
5.2.1	共通理解の達成	90
5.2.2	タイミングと内容	93
5.2.3	課題別のコミュニケーション	94
5.2.4	コミュニケーション・メディアの組み合わせ	95
5.3	本研究の限界	97
5.4	今後の研究展開	98
5.4.1	課題解決会話の共通理解の研究	98

5.4.2  多人数遠隔コミュニケーションの共通理解の研究 .....	98
第6章  結論.....	100
謝辞 .....	103
引用文献.....	104
全研究業績 .....	119
付録 A.....	121
参加者基本情報記入用紙 .....	121
付録 B.....	122
自由再生課題 .....	122
付録 C.....	123
原子力発電問題用紙 .....	123
付録 D .....	125
死刑制度問題用紙 .....	125
付録 E.....	127
音声とテキストのタイミングと内容例 1.....	127
付録 F .....	128
音声とテキストのタイミングと内容例 2.....	128

## 図目次

図 1	共通理解の概念モデル(西田 1996).....	10
図 2	実験手続き .....	37
図 3	実験装置図 .....	39
図 4	音声会議において NS が参加者の発話から重要な部分や理解しにくい部分をテキスト化する「音声とテキスト条件」の映像.....	40
図 5	会話音声をテキスト化せず普段通りの音声会議をする「音声のみ条件」の映像.....	41
図 6	共同解釈を構築(grounding)する会話の件数(件組).....	48
図 7	会話後の自由再生回答(記憶)件数(件組).....	49
図 8	参加者間の一致語句(common ground)の平均件数(件組) .....	52
図 9	質問紙調査結果 .....	55
図 10	ELAN のアノテーションの出力結果.....	67
図 11	タイピングと対応する発話のタイミングの差.....	68
図 12	話者ごとの発話とテキスト化のタイミングの差.....	70
図 13	テキストと対応する発話を会話的特徴から 6 種類の名詞句に分類する方法 .....	74
図 14	6 種類の話し方の出現回数(件組).....	78

## 表目次

表 1	NNS を含む会話支援でよく使われる課題の一覧 .....	15
表 2	NNS 支援に用いるコミュニケーション・メディアの一覧 .....	25
表 3	従来研究の NNS との会話支援ツールに使用されるメディアと課題の組み合わせ .....	32
表 4	本研究の位置づけ .....	34
表 5	仮説と検証方法 .....	43
表 6	共同解釈構築(grounding)の証拠となる会話 .....	46
表 7	組ごとの共同解釈を構築する会話件数 .....	47
表 8	語句の分類根拠と例(田中 et al. 1998) .....	50
表 9	参加者間で一致している語句の例(埴 et al. 2020) .....	51
表 10	質問紙調査結果 .....	55
表 11	4 章の分析項目とデータ取得 .....	65
表 12	テキストと対応する発話の分類根拠と例 .....	67
表 13	6 種類の名詞句の定義と例 .....	72
表 14	6 種類の名詞句分類結果 .....	75
表 15	6 種類の話し方の分類 .....	76

## 第1章 序論

### 1.1 研究背景

#### 1.1.1 非母語話者とのコミュニケーション

国際化と情報化の進展に伴い、日本社会は多国籍の住民によるグローバルな社会へと変化している。2019年に法務省が新たな外国人材受入のための在留資格を創設し、改正出入国管理法(法務省 2019)が施行された。その結果、日本における中長期在留者とその家族が増加し、日常的に外国人と日本人が日本語で会話する機会が増えた。

外国語を含む言語習得に関して、山元(2019)によれば、母語を習得した後に学習によって使用できるようになった母語以外の使用者を非母語話者(NNS)と呼ぶ。これまでの第二言語教育では、NNSがNS(母語話者)のように話せることを理想としてきた。そのためNNSが増加する現在でも自己卑下に陥るNNSは多く、NNSを取り巻く環境は依然として困難である。Novinger(2013)は米国の国際ビジネスの視野から、メキシコや日本でのコミュニケーションが困難である例を多数紹介し、藤田(2009)は観光業発展にはコミュニケーション力や外国語力育成が急務であると述べている。日本の在留外国人に最も多い国籍は中国人である(出入国在留管理庁 2021)。本研究では、日本語を母語とする日本人をNS、日本語を非母語とする中国人をNNSとする。中国語は漢字を使用するが発音が異なり、他の言語と同様に日本語音声会話で意思疎通は困難である。

NS同士の会話支援研究が多数行われている一方で、NNSとのコミュニケーションに役立つ従来研究はまだ数少ない。NS同士の会話を調査した研究では、

NS のみの会話では音声でもテキストでも同様に意思疎通が可能であり、会話支援ツールを使用することで会話を頻繁にかつ均等にすることもできる(倉本 et al. 1998)。一方、NNS とのコミュニケーションでは共通理解が難しい。会話している瞬間は意味が通じたように感じられても、後になってみると話が相手に伝わっていなかったり、相手は誤解していたりということはよく起きる。NNS を含むコミュニケーションは NS 同士のコミュニケーションとは異なり、NNS を含む会話はまだよく知られていないため、本研究は NNS とのコミュニケーションの分析を行う。

NNS とのコミュニケーションでは会話内容をテキスト化する支援がよく行われる。テキストによる情報伝達をする会話の中でも、紙に書きながら会話する方法は筆談と呼ばれる。医療用カウンセリングでは会話の音声を紙面にテキストでメモしながら行われる(西阪 2008)。筆談のような手書きメモの欠点として、読みづらいこと、使用相手、使用環境が限定されることなどが挙げられる。手書きメモの代わりとしてテキストを共有できれば、NNS を含む会話において NNS が会話を理解する補助になる。例えば、NNS がニュース映像を見る際にトピックをテキストでも確認できれば、NNS がニュース内容を把握しやすくなる。同様に、NNS とリアルタイムで双方向の音声会話をする際に一部をテキスト化することがコミュニケーションの支援になると考えられる。

NNS を含む会話音声に対する書き起こしに自動テキスト化技術を用いた研究は数多く行われている。それに対し本研究では、会話音声の一部をテキスト化し、NNS との会話を通じた共通理解への有効性を調査する。また、テキストの共通理解への影響を調べるため会話内容を書き起こしたテキストの速度や内容についても検討を行う。本研究は以下の 2 つの分析を行う目的のため 1 つの実験を行い、実証的な研究を行った。

- 1, 音声のテキスト化はどのように NS と NNS の会話を助けるか
- 2, 音声はいつ、どのようにテキスト化されるか

### 1.1.2 CMC における会話

CMC における会話では、支配的な話者がいなくなり、発話が均衡化する利点があることが、様々な研究から知られている(Hampel et al. 2004)(倉本 et al. 1998)。NS 同士のテキストチャットではテキスト化することで会話への影響があるとする研究がある(Walther 2007)。同様に、NNS との会話では、CMC によって話者交代が頻繁になり、NS が発言を控えて NNS に発話させるようにしたことが知られている(Li and Rosson 2012)。従来研究から、CMC の利点が NS と NNS のコミュニケーションでも同様にあることがわかる。

Zoom Cloud Meetings(ズーム)や Microsoft Corp. Team (チーム) などオンライン会議が一般的になった現代社会では、NNS を含むオンラインコミュニケーションはよく行われている。オンラインの会議では、ビデオ映像をオフにして音声のみの会話や、テキストチャットやインスタントメッセージなどを使ってテキストのみの会話が行われる。CMC ではテキストによる情報伝達を活字で送信でき遠隔地の相手ともコミュニケーションが可能である。スマートフォン、タブレット端末、ノートパソコンの普及に伴い、CMC は増加している。

NNS との会話支援は大きく 2 つの方法に分けられる。1 つ目の方法は、直接的に NNS の聴解・理解を支援する方法で、2 つ目の方法は NS が NNS のために行う簡単な作業が NNS 支援となるような方法である。非母語で会話する NNS は、ワーキングメモリに負荷がかかっている状態にある(Pan 2017)。その結果として、相手との共通理解が果たせなかったり会話内容が記憶に残らなかったりする。また、NS は母語、NNS は非母語で話すため、NS と NNS は言語的に不均衡な状態にある。そのため従来研究では NS が簡単な作業をすることで NS と NNS のバランスを取る支援を行った (Inoue et al. 2015)。

1 つ目の方法の例として、会話音声に短い無音区間を挿入する支援ツールによって、NNS の聴解力や理解力を高めようとする研究がある(Yamashita et al.

2013)。この研究と近い例として、音声会議に無音区間を挿入したのちに聞こえなかった部分を高速再生して理解を調べた研究もある(Hautasaari and Yamashita 2014)。無音区間の挿入は、短いポーズ(音声の停止時間)によって NNS が理解や会話参加をしやすくなるという想定に基づいている。

2 つ目の方法の例として、Inoue ら(2015)は NS が作業をすることで、間接的に NS の発話を NNS に合わせるように試みた。NS がテキスト編集作業を行う会話支援では、NS の発話を妨げない程度の負荷であるため、NS の負担をふまえてもなお会話支援できるメリットの方が大きいことが想定された。また、NS が自動音声認識後のテキストに強調表示を付与し、NNS とバランスをとった従来研究がある(Pan et al. 2017)。その中で NS は強制によってではなく NNS のために進んで強調表示を付与する作業を行った。さらに、自動音声認識による会話の重要部分の自動表示は精度が低いため(Shukla, 2020)、NS が会話中に重要部分をテキスト化する支援も行われた(Echenique et.al. 2014)。

NS がテキストの強調・編集などの簡単な作業を行う方法は本来のコミュニケーションを失うことなく会話支援になると考えられる。NS と NNS のリアルタイムのコミュニケーションでテキストを自動化すると、精度が低く誤りが多いため NNS の理解をかえって妨げ会話を困難にしていた。NS がテキストを編集する作業は会話を妨げるほどではない。また、NS が誤りを含むテキストを読んで編集することは大きな負荷ではないが、NNS が会話中にテキストを読むことは、NNS をさらに過負荷の状態に置くと考えられる。

そのため、従来研究において NS がテキスト編集を行う会話支援が行われた。また、テキストは正確に全文を表示するのではなく、重要部分を表示する方法において一定の有効性が示された(Gao et al. 2013)(Gao et al. 2014)。このことから、NS と NNS とのコミュニケーションにおいては、CMC によって NS と NNS のバランスを取り、NS が簡単な作業を行うことでテキスト化する支援に有効性があると考えられる。

## 1.2 本研究の目的

本研究では、NS が音声を書き起こすテキストの有効性と、音声とテキストが共通理解にどのように影響するかを明らかにすることを目的とする。そのため、2 つの分析が実施できる 1 つの実験を行った。以下で、2 つの分析の概要と目的を述べる。

### 1.2.1 音声とテキストの効果

NS と NNS の音声会議における音声とテキストの効果を検証するため、会話内容、会話後の共通理解、了解性と快適性についてそれぞれ 3 つの仮説を構築し、仮説の検証を試みる。具体的には、NS と NNS の音声会議の実験を行い、実験から得たデータから、会話分析、会話後のテスト結果、会話後の質問紙調査結果の分析を行う。

発話内容の制限と正解がない自由度の高い課題を用いた音声会議において音声とテキストの効果の検証を行う。これは、音声とテキストを用いた代表的な従来研究(Gao et al. 2014)と同じ状況である。NS と NNS の会話支援の従来研究において共通理解の測定は、正解がある課題で指示通りに図形並べに正解した数を数える方法や(Yamashita et al. 2009)(Echenique et al. 2014)、指示通りに地図上に経路を再現した点数を数える方法 (Diamant et al. 2008)(Gao et al. 2015)(Gao et al. 2015)がこれまで行われてきた。このような状況を踏まえ、本研究は自由度の高い課題における NS と NNS の共通理解の測定を行い、音声とテキストの有効性について検討する。

従来研究では課題(会話の題材)とマルチメディアの組み合わせは複数あり、いくつかの最適な組み合わせが検討されてきた(McGrath 1993) (Mennecke et

al. 2000)。検討された対面、音声、ビデオ、テキストなどのメディアと課題の組み合わせは NNS とのコミュニケーションを対象にしておらず、本研究と検討内容は異なる。NS 同士とは異なり、NS と NNS の共通理解は困難であり現代社会では喫緊の課題である。

### 1.2.2 音声をテキスト化するタイミングと内容

次に、音声とテキストがどのように共通理解に影響するか検討するため、NS と NNS の会話におけるテキスト化のタイミングと内容に着目した。音声とテキスト会話の実験から、NS が(1)音声をテキスト化するタイミングと、(2)テキスト化した量と内容を取り上げ、音声とテキスト会話において、テキスト音声会話が NS と NNS に与える影響を検討する。

テキストを用いた従来研究では、機械翻訳や自動音声認識などの自動テキスト化技術を用いたテキストの速度や精度は十分に調査されており、多くの知見がある(宮部 et al. 2007)(Yao et al. 2011)。NNS との会話音声に対する書き起こしに自動テキスト化技術を使用した結果、遅延や誤りがあり、かえって NNS の理解を妨げることが分かった。従来研究をもとに、自動テキスト化技術によって書き起こしたテキストに編集や強調が追加された。その結果、NNS の了解性や快適性が向上した(Gao et al. 2014)。

本研究は、会話音声をテキスト化するタイミングと内容を調査し、音声会話と音声をテキスト化するという異なるモダリティを用いる会話の詳細を調査する。そのために NS による会話音声のテキスト化のタイミングと内容の分析を行い、会話内容からテキストの詳細を検討する。また、データ分析の結果から、テキスト音声会話の会話支援ツールのデザインについても解説する。

### 1.3 本論文の構成

本論文は、6章から成る。1章では本研究の背景、目的を示す。2章では本研究の目的に至る背景となる関連研究について概説する。まず、NSとNNSの会話において困難とされる共通理解の形成について述べる。次に、音声とテキストを用いた従来研究のテキスト化のタイミングと内容をまとめる。また、NSとNNSの会話の従来研究において用いられた課題について述べ、複数のメディアの組み合わせを用いた既存のNNSとのコミュニケーション支援方法について述べる。3章ではNSとNNSによる会話実験の手続きについて述べたのち、音声とテキストの有効性の分析と結果について述べる。4章では音声をテキスト化するタイミングと内容の詳細について述べる。5章では2章、3章、4章を踏まえた総合考察を述べ、今後の研究展開を述べる。最後に6章の結論で本研究をまとめる。

## 第2章 関連研究

本章では序論で述べた本研究の背景や目的について、関連文献を引用しながら従来研究の動向を詳しく説明する。2.1 節では NNS とのコミュニケーションにおいて最も重要である共通理解の形成についてまとめる。2.2 節では音声とテキストを用いた従来研究によるテキスト化の詳細を述べる。会話支援研究では会話を実践することから、2.3 節では従来研究で行われた課題ごとに異なるコミュニケーションの特徴について具体的に述べる。また、2.4 節では従来研究におけるコミュニケーション・メディアの組み合わせについて述べる。それらを踏まえ、2.5 節では従来研究を整理し、本研究の位置づけを明確にし本研究の意義を確認する。

### 2.1 会話による共通理解の形成

近年はさまざまな言語・非言語コミュニケーションが行われている。代表的な非言語コミュニケーションとしては、ジェスチャ、絵、図、デザインなどがある。外国人観光客にわかりやすいような案内記号が日本中のコンビニや自動販売機等で使用されており、NNS を含むコミュニケーションに対する意識が高まっていることが伺える。言語コミュニケーションは、他者とのコミュニケーションの中心であり、心理言語学、言語学、社会言語学、情報学、認知科学など様々な側面からの研究が行われている。

言語学の研究対象は言語であり、言語表現の一部について深く分析する。言語学の語用論の分野では、オースティン(Austin 1962)やサール(J.R.Searle, 1969)らによって直接的にあるいは間接的に発話によって行為を実行したり、社会的効果をもたらす解釈が可能になった。また、チョムスキー(Chomsky

1986)の文法や文構造の研究は言語を数理的に処理することを可能にした。これらの言語学の研究は近年の音声認識や音声翻訳などの研究の発展にも生かされている(速水 2013)。

言語学の研究を踏まえ、クラーク(Clark 1996)は、会話を発話の受け手の反応、会話相手による受理、相手と共通理解を構築するという視点から相互作用に焦点を置いた。従来の会話分析は社会学の観点からサックス、シェグロフ、ジェファーソンら(Sacks, Schegloff, and Jefferson 1974)によって研究されてきたが、ゴフマン(Goffman 1981)とクラーク(Clark 1996)は認知的な観点から、会話への参加を構造的に明らかにしようとした(吉田 2014)。クラークによる会話の聞き手と受け手の相互作用の研究は、情報学の会話支援研究における参加者どうしの協調行為の分析に広く用いられている。

Clark(1996)は、言語によるコミュニケーションを協働行為であるとし、会話に参加することが協働行為であるとした。Clarkら(Clark and Wilkes-Gibbs 1986)によって示された協調モデルでは、会話における共通理解の構築は、情報の提供を行う提示と、受け手が情報を理解したことを示す受理によって行われると定義された。協調モデルでは、話し手が意図する対象を聞き手が想定するという相互認識の達成はコミュニケーション成立のための不可欠な要素であるとし、実際の会話を用いた理論の検証が行われた。

共通理解は、図 1 のような複数のレイヤで表現される。一番下の協働行為(joint activity)のレイヤでは、会話という協働行為を通じ他者と話題の空間を共有している状態である。協働行為では会話に参加しているだけで、共通理解を達成しているとは言えない。次に、会話参加者たちは下から二番目の共同解釈のレイヤで、共通理解の基盤を蓄積する過程として共同解釈を構築、つまりグラウンディング(grounding)する。共同解釈の構築は基盤化とも呼ばれ、共通理解を得るための重要な過程である(西田 1996)。共同解釈の構築(grounding)の段階では、聞き手が話し手の意図を目的に沿って正しいタイミングで理解し

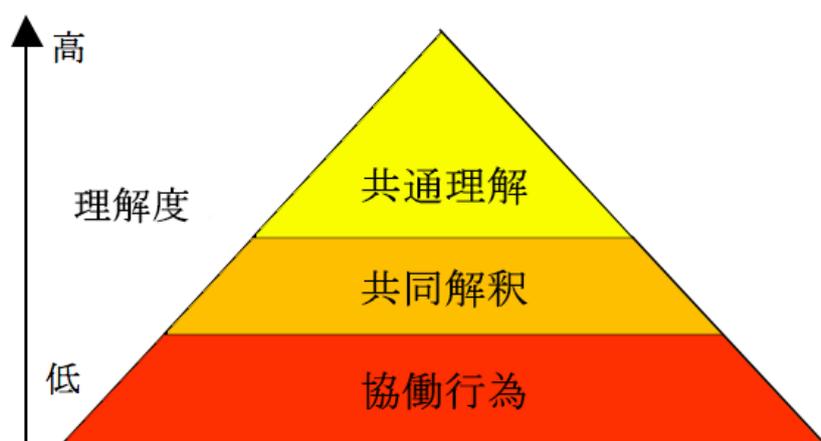


図 1 共通理解の概念モデル(西田 1996)

たことを表明し、会話相手と相互に会話を理解できていると確信を持てる基準を満たすことを共同解釈の構築(grounding)と呼ぶ。

共同解釈の構築を積み重ね、共通理解の最終段階、すなわちコモングラウンド(common ground)のレイヤに到達する。共通理解を共有することをコモングラウンド(common ground)と呼ぶ。最上位のレイヤである共通理解を得ることで会話の理解度が最高潮になる。共通理解の達成は、話し手と聞き手が相互に相手の発話を理解できていると確信し、一致した情報や知識を共有することで得られる(伝 2009)。

NNS とのコミュニケーション支援が研究され始めた時からこれまでにわたって、共通理解は長い間調査されている重要な指標である。本研究では、音声とテキストを用いた NS と NNS の日本語音声会議において、共通理解を得ることを中心にテキストの有効性を検討する。Gao ら(2014)は、自動音声認識技術によって NS と NNS の会話音声を書き起こしたワードエラー率による了解性の調査と、質問紙による快適性の調査によって、了解性と快適性の正の相関関係を示した。また、Pan ら(2017)は自動音声認識によるテキストの重要部分

に NS が強調表示を付与した研究において、質問紙による了解性と快適性の向上と、会話後の記憶から共通理解を示した(Pan et al. 2017)。

本研究で行う会話音声の一部のテキスト化によって、NNS の了解性が向上することが類推される。図 1 に示したように、共通理解は段階を踏んで形成されるため、本研究では音声とテキストの有効性として共通理解を以下の 2 つの仮説に分けて検討する。共通理解に至る過程である共同解釈の構築を仮説 1 において検討し、その後到達する共通理解を仮説 2 において調査する。

仮説 1: 音声とテキストの利用は共同解釈の構築を増加させる。

仮説 2: 音声とテキストの利用は共通理解を向上させる。

また、NS が会話音声の一部をテキスト化する支援によって NS と NNS が共通理解を達成するため、NNS は NS と十分に会話している感覚を得られ、NNS の主観評価が向上すると考えられる。従来研究(Pan et al. 2017)は、客観的指標である会話後の記憶と区別して、主観的指標である質問紙による了解性と快適性の調査を行った。本研究において客観的な理解を仮説 2 で検証し、主観的にわかるという感覚を仮説 3 で調査する必要があると考えられる。また、従来研究(Gao et al. 2014)で示された了解性はワードエラー率によるが、快適性と正の相関関係があることから、主観評価においても了解性と快適性は共起すると考えられる。従って音声とテキストに対する参加者の了解性と快適性の主観評価を以下の仮説 3 において検討する。

仮説 3: 音声とテキストの利用は会話における NNS の了解性と快適性を向上させる。

## 2.2 テキスト化のタイミングと内容

本節では、音声をテキスト化するタイミングに関する従来研究と、テキスト化の内容に関する従来研究を紹介する。音声をテキスト化するタイミングと内容の従来研究はほとんどないため、近年よく用いられている発話内容をテキスト化するいくつかの方法のタイミングに関する研究をまとめる。また、テキスト化の内容についての従来研究はほとんど見当たらないため、テキストチャットの内容を調べた従来研究をまとめる。

### 2.2.1 テキスト化のタイミング

NNS とのコミュニケーションにおける共通理解の問題を解決するため、従来研究では音声をテキスト化して表示することで NNS の理解を補う支援が行われている。音声のテキスト化は、機械翻訳や自動音声認識などの自動テキスト化技術を活用したもののほかに、キーボードを使ってタイピング入力するもの、またはそれらを組み合わせたハイブリッド方式などがある。

Lesecki ら(2014)が行った比較研究では、テキスト化の速度と精度が以下のように示されている。ニュアンス社(米国)ドラゴンスピーチによる自動音声認識テキスト化のみを用いる方式では、複数人の会話音声を 7.9 秒の遅延と 75%の精度でテキスト化したとされる。次に、自動音声認識後に熟達したタイピストが修正する方式では、95%の精度で 7.9 秒の遅延が生じるとされる。次に、熟達したタイピストがテキスト化する方式では、最も精度の高いテキスト化を実現し 4.2 秒の遅延が生じる。複数人で連携しテキスト化する方式では、一般の大学生 20 人が連携すると 3.9 秒遅れで 84.8%の精度を示した(Lasecki et al. 2014)。また、自動音声認識によるテキストの誤りを修正するよりも、はじめからクラウドワーカがテキスト化するほうが遅延は少ないとする研究もある(Gaur 2016)。また、ドラゴンスピーチ(米国)とイージースピーク(ドイツ)を使った Zuchowski ら(2022)による調査では、複数人の NS の会話音声を平均 6.8

秒の遅延と平均 85%の精度でテキスト化し、最高値では 2.3 秒で 100%の精度でテキスト化した。この調査により、自動音声認識はタイピングするよりも早く正確であるという知見が示された(Zuchowski et al. 2022)。

従来研究(Lasecki et al. 2014) (Zuchowski et al. 2022) は、自動音声認識ソフトウェアのトレーニング後に NS のみの音声を、静かな環境下で認識した場合の結果である。本研究で行う音声のテキスト化は自動音声認識を用いずに会話音声のテキスト化を行う。タイピストがテキスト化する方法と、会話参加者がテキスト化する方法は異なるため、遅延の程度も異なると考えられる。従来研究からは会話参加者がテキスト化するタイミングは明らかになっていない。本研究では NS による音声のテキスト化のタイミングを調べることで、NS によるテキスト化が共通理解を支援する理由を明確にし、テキスト音声会話が自動化されるとき会話支援ツールのデザインに言及する。

### 2.2.2 テキスト化の内容

NNS を含むコミュニケーションの従来研究ではテキストによる支援が数多く行われているが、会話音声のテキスト化の量を調べた比較はあまり行われていない。従来研究におけるテキスト化の量とは、音声の全文、一部分、重要単語のみなどがある。NNS が第二言語を処理して読解できる量はリーディングスパンとも呼ばれ、テキストの量が多いほど第二言語処理の負荷は大きくなることが知られている。本研究では会話音声のテキスト化の量を調査することで、テキストチャットや自動音声認識を利用したテキストの量と内容の違いを明確にする。

機械翻訳や自動音声認識などによって会話の全文を自動テキスト化すると、精度の低さから会話が更に困難になることが知られている。機械翻訳を意味伝達可能なレベルに修正するために 6 回程度の修正作業が必要であるとする研究

(宮部 et al. 2007)や、機械翻訳を介したテキストチャットでは意思疎通が困難であるとする研究(Yamashita et al. 2009)がある。また、テキストによる NNS とのコミュニケーションの困難さを示した例として E メールでのコミュニケーションの調査研究(He et al. 2017)や、ソーシャル・ネットワーク・サービス (SNS) のコミュニケーションの研究(Hautasaari et al. 2019)などがある。

そのため、従来研究では自動音声認識後にテキストを編集する研究や(Gao et al. 2014)、機械翻訳と自動音声認識後のテキストの重要部分を強調表示する支援が行われた(Gao et al. 2013)(Pan et al. 2017)。しかし、テキスト編集や重要部分を強調表示する支援において、修正した結果の精度、テキスト編集の量や、何をどのように編集したかなどは調査されておらず、テキストの内容や発話の変化はよく知られていない。

自動音声認識によるテキスト化の従来研究(Gao et al. 2014)において、NS の音声は明瞭になり NNS の了解性が高まることは、音声認識されやすいように NS が話し方を変えたとも推察される。しかし、音声のテキスト化によって話し方がどのように変化するかは明らかになっていない。音声をテキスト化する会話の話し方の変化を検討することは、テキスト化がどのように理解に影響するかを明確にする。したがって本研究ではテキスト化の内容を調査する。

### 2.3 課題別のコミュニケーション

NS と NNS のコミュニケーションの従来研究において、課題は正解のあるものからないものへと変遷し、会話上の役割は固定されたものから徐々に自由度の高いものへの変遷がみられる。表 1 に NNS を含む会話支援でよく使われる課題の一覧をまとめる。NNS を含むコミュニケーション支援において、正解と役割の固定された会話が多く行われた背景には、自動テキスト化技術の認

表 1 NNS を含む会話支援でよく使われる課題の一覧

自由度	課題	コミュニケーション方法	分析方法	出典
低	図形同定課題	対象指示	正解を数える	(Yamashita et al. 2009) (Echenique et al. 2014)
	生き残り課題	交渉	差を数える	(Yamashita et al. 2013) (Hautasaari and Yamashita 2014) (Inoue et al. 2015)
	地図課題	対象指示等	正解を数える	(Diamant et al. 2008) (Gao et al. 2015) (Gao et al. 2015)
	売買取引	交渉	金額を数える	(Hautasaari 2010)
高	政策決定	交渉	質問紙調査	(He et al. 2017) (Duan et al. 2019)
	ブレインストーミング	発想法	質問紙調査	(Gao et al. 2013)
	ストーリーテリング	物語産出ほか	質問紙調査	(Gao et al. 2014)
	ディベート	討論	会話の件数を数える	(Hanawa et al. 2016)
	アイスブレイク	共通点探し	共通点を数える	(Pan et al. 2017)

識精度を高める目的があると考えられる。自動テキスト化技術は、自由度の高い会話に対応できない問題が以前から指摘されている(近藤 et al. 2017)。自由度の高い会話とは、雑多なテーマで自由に話題が移り変わるような会話を指す。自由度の低い課題から高い課題への変遷に伴い、了解性の分析方法は正解した得点を調べることから会話を通じた相互理解に至り、さらに NNS の発話を調べる分析方法に変化している。

### **2.3.1 自由度の低い課題のコミュニケーション**

本節では、NNS とのコミュニケーション支援でよく使われる課題を紹介し、課題ごとのコミュニケーションの特徴を具体的に解説する。自由度が低い課題は参加者の役割を固定し、正解があるため一律に採点しやすく、定量化しやすいためにこれまで長く用いられてきた。一方で、自由度が高く NS と NNS が対等に情報交換する課題はこれまであまり扱われていない。以下では、従来研究において多く用いられた自由度の低い課題を整理し、それぞれの課題における効果の測定方法を具体的に述べる。さらに、従来研究において測定されたそのほかの項目を概説して従来研究における現状を確認する。

代表的な課題のひとつは、中国のパズルを起源とする図形同定課題である。図形同定課題には指示者と照合者がおり、指示者の説明の通りに照合者が図形を並べる。お互いの図形を見せ合うことは禁止である。図形を正しい順番に並べる課題と、複数の図形を組み合わせて正しい図形を作る課題がある。

図形同定課題のコミュニケーションでは指示者が会話をリードする形で図形の説明を長く発話し、照合者は指示者からの指示語を繰り返したり、「はい」や「わかりました」などの受理を示す発話を行う。指示者と照合者のコミュニケーションは徐々に図形同定のための指示語の応酬に収束していく。これを対

象指示コミュニケーションと呼ぶ(稲葉 et al. 2009)。NNS と図形同定課題を行う際は NS が指示者になることが多い。

NS が指示者になると多様な発話はほとんどなくなるが、コミュニケーションが定型化するため定量的な分析が可能になる。不完全な機械翻訳によるテキストチャットで図形同定課題を行った研究では、発話の順番をあらかじめ決めておき図形の正解を数えて機械翻訳の精度を評価したり(Yamashita et al. 2009)、共通の指示語を用いた発話の数と合意の数など発話数で分析を行った(稲葉 et al. 2009)。ビデオと音声に端的なテキストを付与して図形同定課題を行った研究では、NS が指示者で NNS が照合者となり、NS と NNS 間の共通指示語やジェスチャなどの分析も行われた(Echenique et al. 2014)。図形同定課題では統制すべき変数はなく、対象指示コミュニケーションの形式が明らかである。

よく行われている課題として、サバイバルシミュレーションゲームつまり生き残り課題もある。ヒューマンファクターの研究分野だけでなくアメリカのボーイスカウトや軍隊の訓練校で実際のトレーニングにも使用される。生き残り問題には月面、砂漠、北極、ワニ川などの困難な状況のシナリオがあり、そこで生き残るための品物のリストに順位づけを行う。はじめに一人で品物のうち重要と思われるものから順に順位をつけ、のちにグループで話し合いを経て品物の順位に変更を加え、制限時間内に最終的な順位を決定する。話し合い前後の順位の相関を調査した研究(Yamashita et al. 2013)や、エントロピーを用いて測定した研究もある(Chen et al. 2018b)。

生き残り課題はコンセンサスゲームとも呼ばれ、コミュニケーションでは合意に至るために交渉を行う。チーム内の参加者は対等な立場で順位付けを行い、指示者と照合者のような役割は課されない。2 人で行う場合は発話は均等であるが、複数人で行う場合は積極的にリードする者と追従する者がいると考えられる。そのため参加者の性格、課題への慣れ、文化的習慣が変数となり結果に

影響を与える可能性がある。例えば、NS と NNS が参加すれば交渉のリーダーシップを取ろうとする国民性や本人の性格があるものがいれば、そうでない者に従属的役割が決定する場合や、交渉が苦手な国民同士であれば時間内に順位決定しない場合もある。NS と NNS のコミュニケーションの支援に生き残り問題を用いる場合には性格、国民性、文化、習慣が変数となりやすい。

生き残り課題の場所である砂漠、北極圏、ワニの住む川などは日本にはなく、品物の名称、大きさ、使い方が日本人にとって重要性があるかわからないものが含まれている。例えば、銃、小麦加工品のクリスコ社、サングラスなどである。これらは日本において知名度がなく常用されていない。体積の単位クォート、長さの単位フィートなども異なるため、生き残り問題の日本語版では品物の使い方や大きさに解説がつくことが一般的である。NS と NNS の話題として一方に優位な知識がある話題は結果に影響を与え、参加者を発話しにくくさせ、正確な支援効果の測定を困難にする。NS と NNS のコミュニケーションの支援に生き残り問題を用いる場合には変数の影響がないように参加者を統制する必要があり、誰でも参加しやすい課題とは言えない。

また、イギリスで原型が作られた地図課題もよく使われる課題である。地図課題には案内者と追従者がおり、案内者のもつ地図上の経路を追従者の持つ地図上に再現する課題である。案内者と追従者の持つ地図は、出発地点、到達地点、目標物、道などが異なっており、お互いの地図を見せることは禁止である。会話で地図の違いを確認しながら、経路探索を解決する。

地図課題のコミュニケーションでは指示役が会話をリードする形で現在地点と目的地点までの経路の説明を長く発話し、照合者は指示者からの指示を聞き返して確認したり、「わかりました」などの受理を示す発話を行う。NS と NNS が行う場合は NS が案内者で NNS が追従者となることが一般的である。音声会議に自動音声認識したテキスト、辞書機能、視線追尾機能を備えて地図課題を行った従来研究では NS が案内者となり、地図課題の正解数と快適性が

調査された(Gao et al. 2015)。地図課題には英語版の会話コーパスが作成されており、日本語版も作成された。地図課題のコミュニケーション分析では、会話コーパスに従い共通理解が作られる会話形式の件数を数えることで行われる。

しかし、地図認識には心的回転能力が必須である。心的回転は頭の中で地図を回転させて方角を認識する空間認知能力と似通った能力であり、一般的には女性よりも男性のほうが得意である傾向がある。日本語版地図課題の会話コーパスの作成では、男女の統制が行われている(吉田悦子 2001)。NNS とのコミュニケーション支援は、性別を調査する目的ではないため性別調査は一般的には行われない。日本語版地図課題を NNS とのコミュニケーション支援に用いる場合、実験計画で性別が変数となるため、事前に予備調査を行い性差による影響の調査が必要になると考えられる。また、地図には複数の種類があり、それぞれ難易度が異なる。経路が複雑なものほど難易度が上がる。実験で複数の地図を用いる場合は、地図の難易度を揃えることも必要である。さらに、コミュニケーションに文化の違い、課題への慣れの影響が出ると考えられ、コントロールすべき変数が多い。誰にでも参加しやすい課題ではないと考えられる。

比較的新しい課題として、シェイプファクトリーシミュレーションについても解説する。シェイプファクトリーシミュレーションは売買取引ゲームであり、実世界型コミュニケーションとして扱われる(Bos et al. 2006)。10人が参加し、各参加者は1種類の図形を6個ずつもち、同じ種類の図形は各2人ずつ持つ。一回のゲームで、各参加者はそれぞれ異なった集めるべき図形順を3種類与えられる。制限時間内に図形を参加者同士で売買し、8種類を揃えた時点の合計スコアで競う。集めるべき図形を集められなければ、合計スコアにペナルティが課される。初めは制限時間20分で、2回目以降は15分になり、5回行う。図形は三角、丸、ハート型、プラス(十字型)、矢印型、箱型がある。

シェイプファクトリーシミュレーションのコミュニケーションは交渉である。参加者間で誰に売り誰から買うという選択をしながら図形を時間内に揃える競

争は、協調的ではなく競技的であるという指摘もある(Bos et al. 2004)。シェイプファクトリーシミュレーションは記号、価格、個数という数字が主体の交渉で、ゲームの勝敗が決まる。NS と NNS のグループでシェイプファクトリーシミュレーションを行った研究では、ゲームのスコアが測定された(Hautasaari 2010)。売買取引に必要な情報を得ることがコミュニケーションの目的であるため、NS と NNS の間で話しにくさ、役割や話し手の偏りなどもほとんどない。記号、価格、個数などの最小限の情報がわかれば成り立つコミュニケーションであり、機械翻訳の誤訳を含んでも取引に問題はないと推察される。コミュニケーションが定型化しやすく、NS と NNS の多様な発話はほとんど見られないと考えられ、NS と NNS との間で自由度の高い会話を行うコミュニケーション課題であるとは言えない。

### **2.3.2 自由度の高い課題のコミュニケーション**

自由度の高い課題には、ブレインストーミング、ストーリーテリング、ディベート、共通点探しゲームなどがある。以下ではそれぞれを詳細に述べる。

まず、ブレインストーミングについて述べる。ブレインストーミング課題は、集団討議によって個人の持つアイデアの総和以上アイデアが創出されることを目的に、集団での創造的活動の際に行われる(三浦 2001)。ブレインストーミングの参加者は、品質にかかわらずできるだけたくさんアイデアを口頭で表出する。参加者は課題以外のことはしゃべらず、他人の意見を評価や批判しないことが求められる。ブレインストーミングをグループで行う場合、アイデア交換の段階とアイデア拡張の段階がある。

意見を多く出すために、ブレインストーミングのコミュニケーションは拡散的なものになる。テキストチャットでアイデア創出を支援した研究では、コミュニケーション例が次のように示されている(倉本 et al. 1998)。例えば、ブレ

ーンストーミングのテーマが「温泉を楽しむための景色」である場合の会話の例として、「雪景色がいい、そうですね、川が流れてるのもいい、景色が楽しめるところがいいですね、空気のきれいな場所がいい、そうですね、夜景がきれいなところもいい、そうですね、高いところにある温泉、星空とかネオンがきれい、そうですね、」というような会話が続く。ブレインストーミングは協調作業のひとつであり、意見へのコメントが必要とされないため発話とあいづちが繰り返される。

ブレインストーミングのコミュニケーション分析では、意見数を数えることが多い。マルチメディア通信を用いた発想支援のグループウェアの研究では、意見数が数えられた(宗森 et al. 1995)。意見数が多ければコミュニケーション支援グループウェアとして有効であると判断される。しかし、ブレインストーミングでは意見を言わない者、手抜きする者、ただ乗りする者も出やすいという指摘や、批判されることを恐れるあまり「壊れやすい」話題という指摘もある(Paulus and Yang 2000)。NS と NNS のコミュニケーションでは NNS は発話しにくい傾向があるため、NS と NNS がブレインストーミングを行う場合には意見数を数えるのではなく、異なる分析が行われる。NS と NNS で機械翻訳によるテキストチャットに強調表示を付与してブレインストーミングを行った研究では、快適性が調査された(Gao et al. 2013)。NS と NNS の会話は NS 同士の会話とは異なっており、ブレインストーミングは NNS が参加しやすい課題とは言えない。

次に、ストーリーテリングについて概説する。ストーリーテリング課題は研究者によって実践方法が様々であり、総称してこう呼ばれる。具体的には、ビデオ等を視聴してそれについて語る課題(中井 2004)、物語を覚えて口頭で語る課題(浅木 2014)、絵を見て話の内容を説明する課題(栃木 1990)、実際に起こった一連の出来事を話し手が順を追って口頭で再現する課題(相場 et al. 2010)、絵を手掛かりに物語を作る課題(西川 1995)などがストーリーテリング課題と呼ばれるものであり、定義や目的はそれぞれ異なっている。

NS と NNS がストーリーテリングをする際の具体的特徴を調べた研究では、NNS のストーリーテリングで以下のような具体例が示されている。「サルたちは、木うえに、遊びます。男の人、ろじんです。木のところで、すわって、います。たくさん、帽子、あります。老人は、寝ました。サルは、帽子を、取った。サルは、帽子を、かぶりました。老人は起きました。見ました。老人は、サルを、帽子を、か、かぶ、かぶって、います。老人は、おこるになりました(以下略)」。

例のように、NS と比べて NNS のストーリーには構造上の違いはないが詳細さが欠ける。ストーリーテリングには物語の構造的知識、論理的一貫性、詳細な描写力が必要であるといわれる。そのため、ストーリーテリング課題のコミュニケーション分析は、正解はないにもかかわらず、NS と NNS の差が大きくなりやすい。物語産出の得点化は、物語中の単語、文間の連結、因果的もしくは時間的連結、物語の構造などから行われる。NS と NNS がストーリーテリングで話すコトの数と文章の数を調べた研究では NNS の話すコトの数は NS の半数程度で、コトの羅列が多く、明示的な接続詞がなくコトの関係が見えず、文章になっていないコトの数が多いとされる(栃木 1990)。コトとは、「外界の様子、物や人の状態や変化、働きを表す述語とその述語を中心として描かれる事象や心象に登場する人、物、概念などを表す保護から成る」(寺村 1982)。また、物語の理解と物語の産出を調査した研究では、物語産出に事象の連鎖と体制化のレベルによって得点を付与し成績化している(山本 et al. 1989)。

ストーリーテリング課題で NS と NNS の発話の量や内容を比べた場合に NS のほうが多くなることは必然である。そのため NS と NNS でストーリーテリングの話す順番をあらかじめ決めておき、コミュニケーション支援の快適性を調査した研究もある(Gao et al. 2014)。NNS によるストーリーテリングは、協同学習やスピーキング指導との関連があるとされ、NNS のストーリーテリングに関する研究は少なく、今後も研究の余地が残されている(伊勢 2014)。ス

トリーテリングは、相手の理解を確かめずに進む一方的な情報伝達であり、一般的な会話のような相互の情報交換とは言えず、コミュニケーション分析に適しているとは言えない。

次に、ディベート課題についてまとめる。ディベートは、参加者が意見を述べ妨害も頻繁に入る自由な流れの会話である(Ennis and O'Sullivan 2010)。ディベートは、公的な主題について肯定側と否定側の異なる立場に分かれて議論することを指す(池本 2015)。ディベートは論理的思考力やコミュニケーション力を高める教育的効果を得ることを目的として、中学校・高等学校の新学習指導要領においても総合学習、国語、外国語の時間での討論が推奨されている。参加者は自分の意見や考え、価値観などに関係なく、肯定側と否定側のどちらかに割り振られる。発言の順番や方法、制限時間が決められ、主張、質疑、反論で構成される。審判がいる場合は、肯定側と否定側のどちらの主張が論理的で説得力があったかを判定し勝敗を決める。ディベートの形式は実施目的、時間、参加者の技術によって変更できる。

ディベートのコミュニケーションは幅広い観点からの多角的な考察が必要となり相互理解に有効な課題であると考えられる。ディベートのコミュニケーションでは発話の回数、長さ、衝突回数など発話しやすさが調査される(Sellen 1992)。ディベートでは発話の順番が決められており発話の時間制限のためどちらかに偏ることはなく、NNS も参加しやすい課題である。Pellettieri(2000)によれば、実社会の問題について対立する立場で議論することは、NS と NNS がそれぞれの持つ語彙、理念、概念を超えて言語で表現し、均衡して会議に参加する機会となる。ディベートを行うまでには調査資料などに長い準備期間が必要であることや、ディベートを行う練習が必要であることなどが教育上の課題となっている。そのため、専門知識がない人でも議論に参加できるよう代表的意見を提示してディベートを支援するシステムも研究されている(大林 et al. 2002)。従ってディベートは協調学習やコンピュータによる学習支援システムとの親和性もあるといえる。

最後に、共通点探しゲームについて述べる。共通点探しゲームは、お互いの共通項を見つけるゲームである。しばしば初対面でアイスブレイクに使われるゲームで、決まった時間内に初対面の2人1組もしくは3人以上で自己紹介しながら共通点を探す(渡辺 2020)。自己紹介をするよりも時間が節約になることや、人数が多くても実施できるという利点がある。参加者は時間内にできるだけたくさんの共通点をグループ内で見つけることが目的である。共通点を探す際の話題は出身地、住居、家族の話題が探りやすいとされる。一般的には、朝食を食べた、靴下をはいているなどでも共通点としてよい。禁止事項がある場合はその関連事項で共通項が見つかりやすいとされる。

共通点探しは他者との共通点を知ることによる親近感の向上が狙いとされる。仕事上の対人関係とトラブル処理能力を高めるためのソーシャルスキル・トレーニングでは、共通点探しを含む複数のコミュニケーション課題が行われた。共通点探しで例のような会話を行う場合、初対面で人間関係を形成するソーシャルスキルが必要である(藤原 et al. 2017)。共通点探しのコミュニケーションは、参加者が初対面で自己紹介する際にたくさん話すことが求められる。共通点探しには固定された役割や正解がなく、かつ一方通行の情報伝達でもない。時間内に目的をもって行う協調課題として適切と考えられる。対面で NS と NNS が共通点探しを行い、会話の音声認識と機械翻訳によるテキストで支援した研究では、了解性、記憶、快適性などが調査された(Pan et al. 2017)。テキストを用いて了解性を高めることができれば、NS と NNS のコミュニケーションにおいて相互理解を支援する課題として適している。

## 2.4 コミュニケーション・メディアの組合せ

本節では、一般的なコミュニケーション・メディアに関する従来研究についてではなく、本研究で扱う NNS とのコミュニケーションを支援する際にテキ

ストとともに使用されたメディアについて概説する。会話支援で使用されるメディアは、情報量の多いものから順に対面、ビデオ、音声、テキストなどであり、メディアの組み合わせの研究は多数行われている。

本節ではテキストとともに NNS 支援で使用されたコミュニケーション・メディアについて比較・整理し、従来研究の変遷を概観する。続いて次節では、本章で取り上げた従来研究を整理し、従来研究における問題点を明らかにしたのちに本研究の意義・目的と位置づけについて述べる。

表 2 に、NNS 支援に用いるコミュニケーション・メディアの一覧を示す。表 2 の自動テキスト化技術とは、機械翻訳や音声認識などを利用して、音声に対する書き起こしを追加する方法である。表 2 の音声テキスト化とは、会話参加者が会話中に会話内容をテキスト化する方法を指す。表 2 のテキスト編集とは、自動テキスト化技術によるテキスト化の後に、会話参加者がテキストに強調表示や修正などのテキスト編集作業を加える方法を意味する。

表 2 NNS 支援に用いるコミュニケーション・メディアの一覧

支援 メディア	自動テキスト化技術	音声テキスト化	テキスト編集
ビデオとテキスト	(Pan et al. 2009) (Pan et al. 2010) (Yao et al. 2011)	(Echenique et al. 2014)	
音声とテキスト	(Hautasaari and Yamashita 2014) (Cao et al. 2018)	(Inoue et al. 2015) (Hanawa et al. 2016)	(Pan et al. 2017) (Gao et al. 2014)
対面とテキスト	(岡本 et al. 2011)		(Chen et al. 2018a) (Chen et al. 2018b)

### 2.4.1 ビデオとテキスト

ビデオに関する研究は古くからあり、NS 同士のコミュニケーションにおいてビデオの効果を調べた研究が一般的である。ビデオ、音声のみ、音声とテキストを比較して、ビデオのコミュニケーションが音声とテキストと同程度のコミュニケーションが可能であるとした研究(Chapanis 1976)や、メール、電話、

ボイスメッセージなどのシステムを用いて、ビデオ会議と複数の機能の組み合わせを検討した研究がある(Isaacs and Tang 1993)。これらは NS 同士のコミュニケーションにおいてメディアの比較を行った研究である。

ビデオ会議やウェブ会議が広く行われている現代では、様々なビデオ会議システムが開発され、ビデオ会議は様々な検討が行われている。Pye ら(1976)によれば、ビデオ映像の表情、動作などの非言語情報には一定の意味がなく、重要な情報は音声会話される。そのためビジネスなどの会議においては非言語情報が重要とは考えられておらず、音声会議のほうが良い場合もある。一方で、表情を確認することは感情的な面で好ましいと考えられ、ビデオ会議やウェブ会議で表情を見せて会話することは感情に働きかけるコミュニケーションに適している。つまり、ビデオ会議は、その特徴から用途に応じた使い分けが行われている(Pye and Williams 1977)。

一般的にビデオ会議は対面会議より情報量が少なく、映像があるため音声会議より情報量が多いと考えられている。そのため音声会議には映像やテキストを追加して情報を補う支援が多く行われている。近年にはビデオとテキストを用いて NNS のコミュニケーション支援が行われるようになった。2009 年に Pan ら(Pan et al. 2009) はビデオ映像に自動音声認識技術を用いて音声を書き起こしたテキストを付与した実験を行い、NNS の了解性を調査した。その際、事前に正確なテキストを用意してリアルタイムのテキストと比較している。また、テキストの精度に差をつけた研究では、95%以下の精度では NNS 支援に

はならないとしている(Pan et al. 2010)。同様に、テキストの速度に差をつけた研究では、2 秒程度の遅れならば NNS 支援になるとしている(Yao et al. 2011)。

#### 2.4.2 音声とテキスト

音声会議は、映像を用いず音声のみによって遠隔地間で行われる会議のことで、ウェブ会議で映像をオフにしたときの状況も音声会議に含まれる。サンビジコム(2021)によれば、音声会議は海外では電話網の広がりや品質の安定性のために幅広く利用されており、企業間や国家間などで用いられる。日本においては、コロナ禍をきっかけにリモートワークにおけるウェブ会議が一般に急速に広まった。ウェブ会議で家の中のプライバシーを守るため、ビデオ映像をオフにし音声のみで会議を行う人が増加した。ウェブ会議の普及によって、音声会議に着目させる状況が起きている(サンビジコム 2021)。

従来研究で行われてきたマルチメディアの比較は、主にビデオ、テキストチャットに関連するものが多く、音声会議についての研究はほとんど行われていない。音声会議に関する研究としては、文書共有、チャット機能、プレゼンテーション機能などの会議ツールを備えたものや、インターネット電話である Voice over Internet Protocol(VoIP)を利用したものなどにおいて快適性を調査した研究がある(Ding et al. 2007)。これらは NS 同士の通話の場合であり、NNS を対象とした研究は近年まで行われなかった。

2014 年に NS と NNS の音声会議に自動音声認識技術による書き起こしを付与し、編集を加えたテキストを用いて NNS の理解度を調べた研究(Gao et al. 2014)が行われた。NS と NNS の音声会議においては、発話が NS に偏りやすい問題や、NNS が会話全文のテキストを会話中に読むことは大変な時間と負担がかかるため本来のコミュニケーションの聞く・話すが確実ではなくなる問

題が指摘されている。従来研究では、自動テキスト化技術から会話の重要部分を自動的に表示する技術はまだ精度が低く(Shukla, 2020)、誤りの多いテキストはかえって NNS の理解を妨げる問題がある。そのため、2015 年には Inoue らによって NS が音声会議中に会話音声をテキスト化する試みが行われた。その後、宋(2018)によって NS によるテキスト編集と発話の関わりも調査されている。

また、自動音声認識技術によるテキストに強調表示をつけ理解度を調査した研究も行われている(Pan et al. 2017)。NS と NNS の音声とテキストによる会話支援のテキスト編集(Gao et al. 2014) (Pan et al. 2017)は、従来の自動テキスト化よりもテキストがインタラクティブつまり双方向性があると考えられる。自動音声認識によるテキストは一方通行に情報伝達するだけであったが、NS が会話中に編集するテキストは参加者ごとに異なるため、会話内容や相手の理解に応じて内容を変化させる。すなわち、テキストは会話中に音声と相互に連携し合い、結果として NS の音声は明瞭になるなどの積極的な効果があると考えられる。

そのほか、NNS の聴解を支援する研究や、聴解の程度を調査する研究は数多く行われている。音声会議の聞こえない部分に対して、1.6 倍速で再生し自動音声認識によるテキストをつけて NNS の理解度を測定した研究(Hautasaari and Yamashita 2014)や、音声に自動音声認識によるテキストを付して、NNS の理解度とテキストの遅延の快適性を調査した研究(Cao et al. 2018)などが行われている。

### **2.4.3 対面とテキスト**

対面会話は最も基本となるコミュニケーションである。対面での NNS とのコミュニケーションは異文化コミュニケーションと呼ばれ、社会・文化の観点

からしばしば観察されてきた。観光招致のために異文化コミュニケーションの重要性を述べた研究や(Uzama 2012)、対人距離、ジェスチャ、視線などの非言語情報に依存することの危険性や、異文化コミュニケーションにおける誤解を指摘した研究もある(Novinger 2001)。岡本ら(岡本 et al. 2011)によれば、異文化コミュニケーションでも自動音声認識によるテキストと画像を表示することにより、相互理解の支援ができるとされる。これらの研究では、NNS との対面のコミュニケーションには文化の違いという要素が含まれるため、コミュニケーションは言語のみによらないことが示されている。従って NNS との対面のコミュニケーション分析では文化差に着目した測定を行う必要がある。

対面会話支援を行った従来研究のほとんどは、NNS 支援を目的としたものではない。古くからある対面のコミュニケーションを基本として、従来研究ではそのほかのコミュニケーション手段の効率を調査した。対面、音声、手書き、テキストの 4 種類を比較して使用単語数と所要時間で比較した結果、音声と対面では課題解決にかかる時間がほぼ同じでタイピング経験は問題解決の時間に影響がなかったとする研究がある(Chapanis 1976)。

近年では、携帯電話やタブレットなどの普及に伴い NNS との対面のコミュニケーションに計算機による支援が行われることもある。自動テキスト化技術を用いてコミュニケーション支援をした研究では、誤りを含むテキストに編集と強調を加え会話量と使用言語を比較すると、テキストがある場合に NS と NNS の会話量が等しくなる傾向があるとされている(Chen et al. 2018b)。このほかは、NNS を含む対面コミュニケーションをテキストで支援する研究はあまり行われていない。対面のコミュニケーションで計算機を用いる場合は、会話相手の顔ではなく携帯電話やタブレットなどのモニタを見る。対面で計算機による支援を用いる会話は、従来のコミュニケーションの様相とは異なると考えられ、今後の研究が期待される。

## 2.5 本研究の位置づけ

本節では、従来研究を整理し、本研究分野における本研究の位置づけを述べる。

### 2.5.1 従来研究の整理

これまでに従来研究について整理する。コミュニケーション支援研究は課題によってコミュニケーションが異なり、課題や参加者の会話上の役割と切り離すことができない。そのため、NS 同士の会話において、課題とメディアの最適な組み合わせを調べた研究や(McGrath 1993)、最適な組み合わせの仮説を検証する研究もある(Mennecke et al. 2000)。同様に、NNS とのコミュニケーションでも、課題によってコミュニケーションが変わるため、課題の自由度とメディアの組み合わせを順に整理する。

2008 年には異文化コミュニケーションの研究として、母国語が異なる 2 者が共通言語を用いて地図課題 (NS と NNS の役割が固定されるもの)が行われ、コミュニケーションの快適性が調査された(Diamant et al. 2008)。異文化コミュニケーション研究の研究として文化差が主な分析項目となる研究は数少ない。これ以降は言語コミュニケーションに焦点があてられた。

2009 年以降の約 5 年間にわたり自由度の低い課題を用いて NNS とのコミュニケーション支援が行われた。その中の了解性の分析は、心理的な主観評価によるものもあるが、正解のある課題では得点化による理解の測定が行われた。これらの研究にはテキストによる NNS 支援のため自動テキスト化技術が用いられている。また、機械翻訳の精度や自動音声認識などの自動テキスト化技術の精度や速度を詳細に調査した研究がほぼ同時期に行われた。

自動テキスト化技術のみでは了解性が低いことが従来研究の課題であることから、2013 年以降の約 5 年間は自動音声認識後にテキスト編集作業を加え、了解性や快適性を高める支援が多く行われた。また、Inoue ら(2015)は NS による音声テキスト化の先駆的な試みを行い、簡単な作業が NS と NNS の会話支援となることを示した。

続いて 2017 年からはより自由度の高い NS と NNS の会話によるコミュニケーションが行われている。従来研究のテキストによる支援において、快適性の向上は NS の話し方の変化によるものであることに基づき、2019 年には Inoue らによって会話順や役割を決定せず直接に話速を変化させ快適性や了解性が調査されている(Inoue 2019)。テキストによる支援は自動テキスト化技術のみを用いたものから始まり、発話と連動するテキスト支援に続いて発話を変える支援へと変遷した。

表 3 について以下に詳述する。Mennecke ら(Mennecke et al. 2000)は NS 同士の会話支援ツールに使用されるマルチメディアの組み合わせと課題の組み合わせのテストを行い、それを表にまとめた。本研究では NS と NNS のコミュニケーション支援ツールの従来研究を表 3 にまとめた。従来研究では自由度の低い課題から自由度の高い課題への変遷が見られる。本研究では従来研究を課題の自由度によって分けた。表 3 に示したように、NNS を含むコミュニケーション支援は課題によって NS と NNS の会話上の役割、会話の自由度、コミュニケーションの特徴、支援効果を測定する方法が異なる。また、NNS とのコミュニケーションは使用するメディアで区分できる。コミュニケーション・メディアの組み合わせを、音声とテキスト、ビデオとテキスト、対面とテキストのコミュニケーションに分けてまとめる。

表 3 従来研究の NNS との会話支援ツールに使用されるメディアと課題の組み合わせ

メディア 自由度	音声とテキスト	ビデオとテキスト	対面とテキスト
低	(Yamashita et al. 2013) (Inoue et al. 2015) (Hautasaari and Yamashita 2014) (Gao et al. 2014) (Gao et al. 2015) (Cao et al. 2018)	(Diamant et al. 2008) (Echenique et al. 2014)	(Chen et al. 2018b)
高	(Gao et al. 2014) (Hanawa et al. 2016) (Pan et al. 2017)	(Gao et al. 2013) (Chen et al. 2018a) (Duan et al. 2019)	(岡本 et al. 2011) (Gao and Fussell 2017) (Guo and Inoue 2019) (Inoue and Kawai 2019) (Sasaki 2021)

### 2.5.2 本研究の位置付け

従来研究では NNS とのコミュニケーションを様々な方法を用いて支援し、了解性や快適性を調べた。テキスト化には自動テキスト化技術を用いる方法や、音声を書き起こす方法や、それらを同時に用いる方法があった。本研究は NS と NNS が課題解決のため多様な発言を行う自由度の高い議論において、NS と NNS が共通理解に至る過程を説明したうえで、NS が音声をテキスト化することによる共通理解への効果を検討する。

同時に、本研究は音声テキスト化する速度と内容を分析し、テキストが会話に与える影響を検討する。テキストチャットの従来研究において、NS 同士のテキストチャットにおけるテキスト化の量と内容などが示されたが(Walther 2007)、NNS を含む会話における音声テキスト化のタイミングと内容とは異なる。NNS を含む音声会議において音声テキスト化するタイミングと内容は音声と関連があり、本研究で検討する音声テキスト化のタイミングと内容の分析は、会話とテキストというモダリティが異なる 2 つが連携する会話の性質を明らかにする。分析から、テキスト音声会話の会話支援ツールの自動化のデザインを検討し、NNS とのコミュニケーションを支援するツールの評価に活用できる知見とする。

表 4 にテキストの効果を調べる本研究の位置づけを示す。表 4 の自動テキスト化技術とは、機械翻訳や音声認識などを利用していることを意味し、下部には自動テキスト化技術の会話支援への有効性を示した研究の出典一覧を記載した。次に、テキスト編集とは自動テキスト化技術によるテキストに会話参加者が修正や強調を加えたことを意味し、会話参加者が会話音声をテキスト化する支援の研究の一覧も記載した。

表 4 本研究の位置づけ

調査対象 支援技術	テキストによる会話支援の有効性	タイミングと内容
自動テキスト化技術	(Yamashita et al. 2009) (Hautasaari 2010) (Yao et al. 2011) (Pan et al. 2010) (Pan et al. 2009) (Hautasaari and Yamashita 2014) (Cao et al. 2018) (岡本 et al. 2011)	(Yao et al. 2011) (Pan et al. 2010) (Pan et al. 2009) (宮部 et al. 2007)
テキスト編集 (音声のテキスト化を 含む)	(Li and Rosson 2012) (Gao et al. 2013) (Gao et al. 2014) (Inoue et al. 2015) ( <u>Hanawa et al. 2016</u> ) ( <u>埜 et al. 2020</u> ) (Pan et al. 2017) (Chen et al. 2018b)	( <u>Hanawa et al. 2017</u> )

## 第3章 NS と NNS の音声会議におけるテキストの効果

2 章で述べたように、現代は NNS を含むオンラインコミュニケーションが頻繁に行われており、テレビ会議やウェブ会議時に家の中や身だしなみを画面に映さないためにビデオをオフにして音声のみで会話することを求める人が多い。音声のみの会話で共通の理解を得るためには、NNS を含む音声会議の会話支援ツールが役立つかもしれない。本章では、オンラインの音声会議を想定した実験を行い、会話の音声を NS がテキスト化する支援の効果を検証した。

### 3.1 実験

NNS を含むコミュニケーションは NS 同士のコミュニケーションとは異なり、相互理解を得にくかったり、会話が困難であったりすることが多い。このような問題を解決するため、本研究では、NNS が参加する音声会議場面を設定し、NS が発話を部分的にテキスト化する会話支援の実験を行い、その効果と会話への影響を調査した。

#### 3.1.1 実験参加者

本実験は筑波大学大学院図書館情報メディア研究科で倫理申請を行い、承認を受けて実施した(承認番号: 第 27-129 号)。日本人 16 名と留学生 16 名の 32 名が 2 人で 1 組を作り、16 組が実験に参加した。留学生は日本語能力試験(日本語能力試験とは 2020)の N1 の中国人留学生で、日本人は日本生まれで日本語 NS の大学生であった。参加者の平均年齢は 25.3 歳であった。各組の参加者

は実験前に面識がない初対面の大学生同士であった。実験は研究室で行った。音声会議は非公式のディベート形式とし、原発問題と死刑問題の 2 つの議題について以下の 2 つの条件で実施した。

1 つ目が、NS が参加者の発話内容のうち重要な部分や理解しにくい部分をテキストに書き起こしながら NNS と音声会議する「音声とテキスト条件」で、2 つ目が普段通りに音声会議をする「音声のみ条件」である。

実験は NS と NNS の 2 人 1 組の音声会議を 16 組実施した。すべての組が「音声とテキスト条件」と「音声のみ条件」の 2 つの条件に参加する被験者内計画で行った。2 条件は同日に実施した。音声会議の議題と実験条件は順序効果を考慮し、組間で 2 つの議題と 2 つの条件についてカウンターバランスをとった。

### **3.1.2 実験手続き**

図 2 に実験手順をチャートで示す。音声会議のシミュレーションとして、実験参加者は相互に背面を向けて着席した。実験参加者は実験参加同意書と参加者基本情報記入用紙に記入したのち、手荷物や携帯電話、腕時計などを置き実験に参加する準備をした。実験者が口頭で実験の教示を行った。まず実験手順については、異なる内容の 2 つの議題について 2 つの手法を使用して音声会議

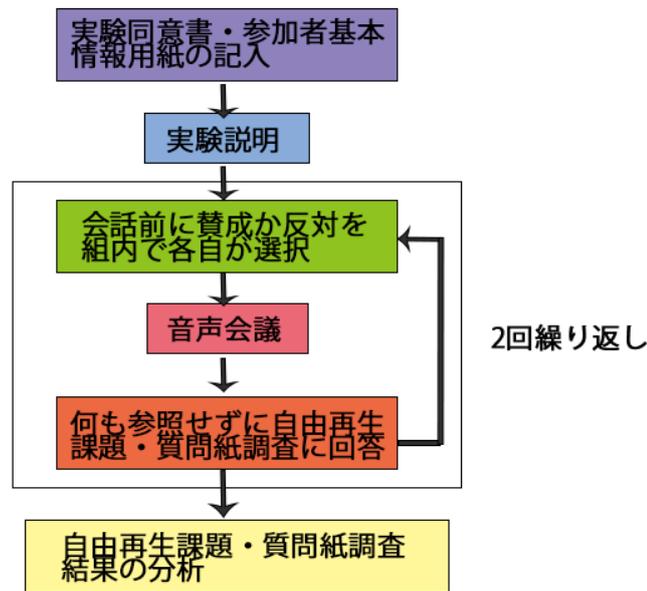


図 2 実験手続き

をすると説明した。2つの手法とは、1つは音声会議中にNSがテキスト化を行い、もう1つはテキスト化をしない普段通りの音声会議をすることであることを説明した。ただし、実験結果に影響を与えないために、2つの手法の違いについては説明せず、NS、日本人、NNS、非母語話者などの呼び方は使用しなかった。

次に、実験に慣れてもらうため30秒の練習会話を行った。PCの使用法は実験を始める前に実験者が口頭で説明した。テキスト化の教示は、次のように口頭で教示した。「テキストは重要な部分や理解しにくい部分をキーボードでタイプして下さい、全文をタイピングすることではありません、NSはタイピングしながら話し、NNSはタイピングしないで話して下さい」。実験説明は、2、3分で終了した。NNSには実験者が適宜中国語でも実験説明を行った。実験者は説明のあとで参加者が説明を理解したか確認した。

次に、2人1組の参加者内で原子力発電と死刑制度に関して賛成か反対のどちらか一方ずつになるよう選び、7分間の音声会議を行った。実験参加者は自

分の意見を賛成か反対から自由に選択した。NS と NNS が同意見を持つ場合は、ランダムにどちらか一方を割り当てられた。

各会議が終了した直後、参加者は自由再生課題(福田 2012)にキーボードでタイピング入力により回答した。本研究の自由再生課題の教示は、音声会議後の NS と NNS に「何も参照せず、自分の話が伝わったと思う単語や文章と、相手の話から理解できた単語や文章をできるだけ多く正確に記入してください」と説明した。自由再生課題の回答記入に時間制限はなかった。音声会議と自由再生課題回答を 2 回ずつ繰り返した後、参加者は研究室を退出した。

### **3.1.3 音声会議の議題**

音声会議の議題は、自由度の高い会話で NS がテキスト編集を行い音声とテキストの効果を検討した代表的な従来研究(Gao et al. 2014)と同じ状況とするため、審判と進行役のない非公式のディベートとした。ディベートの議題は、参加者の文化・習慣によって考え方に違いが出やすく、よく知られる死刑制度と原子力発電の是非とした。音声会議の資料として、原子力発電と死刑制度について定義、賛成・反対それぞれの一般的に考えられる代表的な意見が日本語で記載されていた。一般的にディベートでは開始前に資料やデータを準備する時間があるが、実験会話では資料収集の時間がないため、会話前に参加者に資料を与え 1 分間読む時間をとった。参加者は資料に編集を加え、自分の意見にしてディベートを行った。音声会議中は自分の資料のみを見ることができ、自分の資料は相手に共有されなかった。

### **3.1.4 実験機器**

研究室の机の上に NS と NNS が使用する参加者用ノート PC(Lenovo B590 15.6 インチ)を背面合わせになるよう 1 台ずつ設置し、LAN ケーブルを接続した。参加者用ノート PC から延長した 39 インチモニタとキーボードを 1 台ずつ設置した。実験の撮影は、1 台のビデオカメラで実験の全景を、2 台のビデオカメラで実験参加者の上半身を横から撮影した。また、NS と NNS が使用する参加者用ノート PC のモニタ画面をデスクトップキャプチャソフトウェアを用いて撮影した。図 3 に実験装置図を示す。

音声会議の資料を Microsoft Word 2013 で開き、参加者用 PC のモニタ右側に配置した。さらに参加者用 PC 上で白紙の文書を Microsoft Word 2013 で開き、NS がテキストをタイピングするためのテキストウィンドウとしてモニタ左側に配置した。NS と NNS のそれぞれの PC で Skype を起動し、音声通話を開始した。本研究は音声会議のシミュレーションであるため、Skype の音声はオフにし、参加者は背面を向けたままネットワークを介さず肉声で音声会議した。

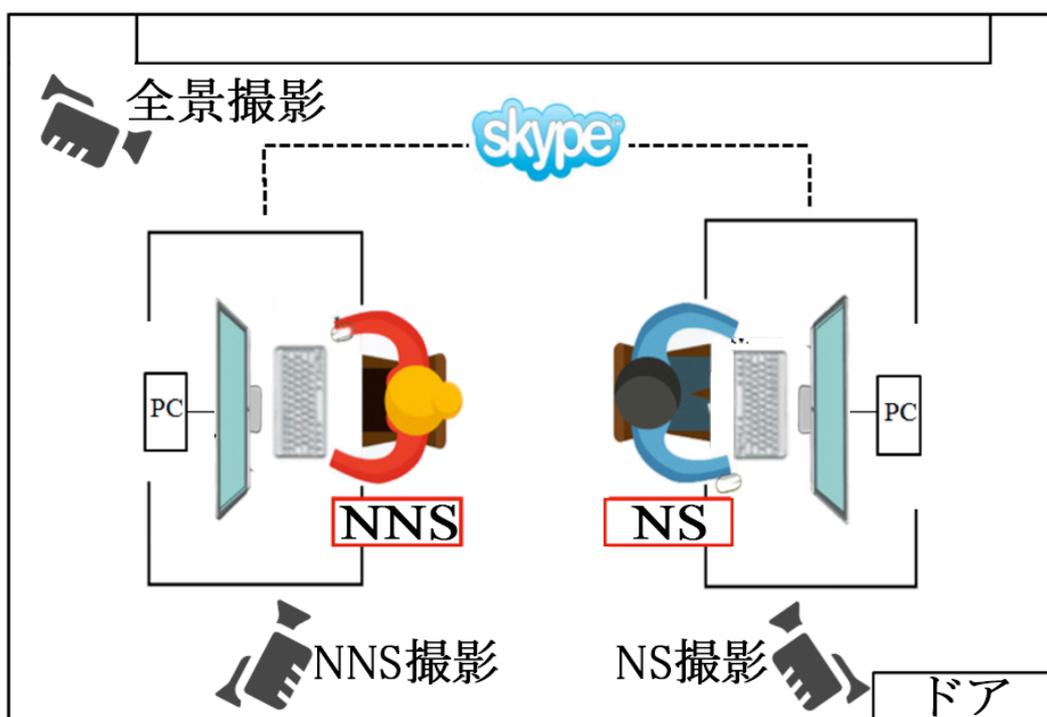


図 3 実験装置図

「音声とテキスト条件」では、Skype の画面共有機能を用いて NS のテキストウィンドウ(モニタ左側)が相手に共有されるようにした。Skype の画面共有機能は遅延がほぼ無く、リアルタイムに相手とモニタを共有できる(Sirintrapun 2012)。NS のテキストウィンドウ(モニタ左側)では NS のみがテキストの編集作業をすることができ、NNS がテキストを編集することはできない。音声会議用の参考資料(モニタ右側)を配置し、モニタ右側は相手に共有されないようにした。図 4 はビデオカメラとデスクトップキャプチャソフトウェアで撮影した映像から成る、音声会議において NS が参加者の発話から重要な部分や理解しにくい部分をテキスト化する「音声とテキスト条件」の映像である。

「音声のみ条件」では、NS のテキストウィンドウ(モニタ左側)が Skype の画面共有機能を用いて相手に共有される。音声のみ条件の NS のテキストウィンドウにはタイピングによる編集を行わないため、NS のテキストウィンドウは白紙のまま相手に共有される。モニタ右側は音声会議用の自分の参考資料を配置し、相手に共有されない。図 5 は会話音声テキスト化せずに普段通りに音声会議をする「音声のみ条件」の映像である。

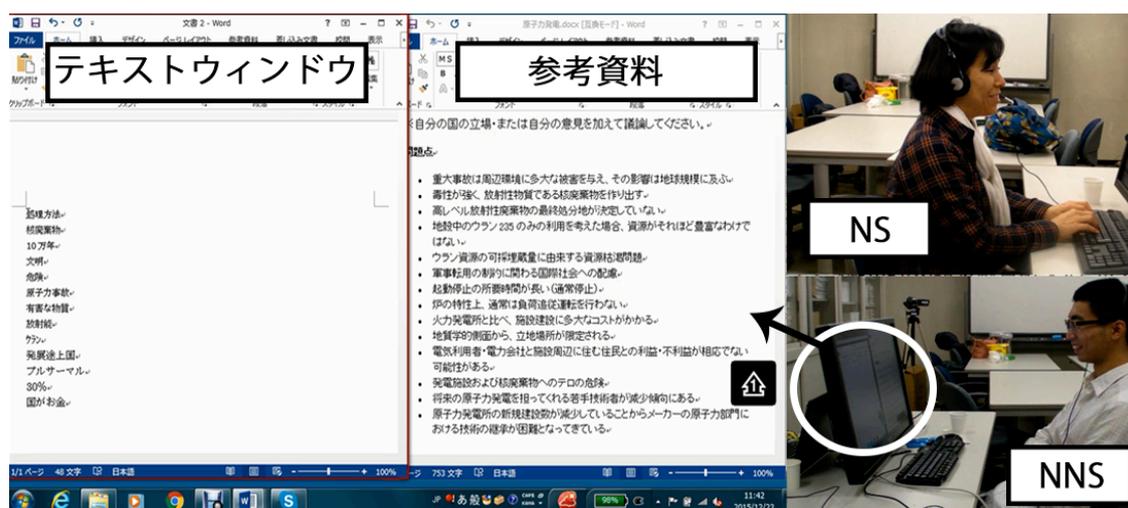


図 4 音声会議において NS が参加者の発話から重要な部分や理解しにくい部分をテキスト化する「音声とテキスト条件」の映像

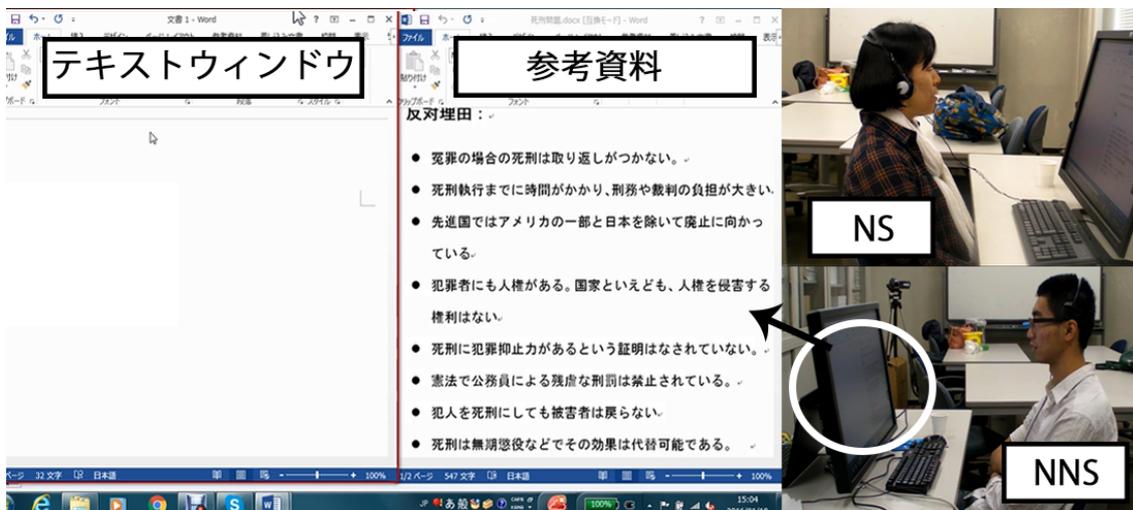


図 5 会話音声テキスト化せず普段通りの音声会議をする「音声のみ条件」の映像

### 3.1.5 参加者の日本語能力

本研究における NNS の日本語能力は、日本語能力試験の N1 レベルを基準とした。日本語能力試験は日本語の試験としては最も受験者数が多い試験である。N1 の読解力は、幅広い話題に関する論理的・抽象的な文章を読んで内容を理解することができ、聴解力は自然なスピードのまとまりのある会話を聞いて、流れや内容を理解することができる(日本語能力試験 2020)。N1 レベルは本研究で必要とされる日本語会話を聞いて理解する能力と日本語を読んで理解する能力を満たす。参加者の日本語能力試験の点数は N1 合格基準が 180 点満点中 100 点以上のところ、平均 119 点だった。参加者の日本語学習平均年数は 4 年であった。

N1 合格者の実践的な日本語コミュニケーション能力が十分ではないことが、ビジネス日本語能力テストと日本語能力試験の比較(BJT の特徴のご紹介 ビジネス日本語能力テスト「日本語能力試験 (JLPT) との比較」 2020)によって指

摘されている。日本語能力試験は日本語コミュニケーション能力を直接的に測るものではないため、この実験では自己評価による能力の推定を行った。会話実験前に参加者基本情報を記入する際、「日本語自己評価」という質問項目に 7 段階尺度の中から最もあてはまる程度を選んでもらったところ、とても低い=1、低い=2、やや低い=3、どちらともいえない=4、やや高い=5、高い=6、とても高い=7 のうち、参加者の自己評価はやや低かった( $M=3.50$ 、 $SD=1.00$ 、 $MIN=2.00$ 、 $MAX=6.00$ )。N1 合格者の日本語コミュニケーション能力は必ずしも高いとは言えないため、N1 の参加者に本研究のコミュニケーション支援をすることでも支援効果は測定できる。

## 3.2 日本語音声会議における音声テキスト化の実験結果

### 3.2.1 仮説と検証方法

本節では 3.2.1 節で述べた仮説を検証するため、仮説と検証方法について述べる。表 5 に仮説と検証方法を示した一覧を示す。仮説は会話分析による共同解釈の構築、自由再生課題による共通理解の調査、質問紙調査による調査の 3 つから成る。分析方法は 3.4.2 節以下で詳述する。

表 5 の 3 つの仮説を検証するため、実験中にビデオカメラで撮影した実験映像から音声を書き起こし、音声会議の全文を取得した。音声会議後の自由再生課題回答から NS と NNS の回答件数つまり記憶の件数と、回答内容つまり参加者間で一致した語句の件数を取得した。分析対象は参加者 16 組の音声とテキスト条件と音声のみ条件の 7 分の音声会議(全 224 分)である。

表 5 仮説と検証方法

仮説	検証方法
音声とテキストの利用は共同解釈の構築を増加させる。	会話音声の書き起こしにより会話の全文を取得し、共同解釈を構築する会話(Clark 1996)の件数を数える。
音声とテキストの利用は共通理解を向上させる。	自由再生課題に記入した記憶の件数と内容を取得し、NS と NNS の自由再生課題回答内容の一致件数を数える。
音声とテキストの利用は、了解性と快適性を向上させる。	質問紙調査に回答した内容を取得し、リッカート尺度による回答の代表値を算出する。

### 3.2.2 音声とテキストによる共同解釈の構築の分析結果

従来研究(Pan et al. 2017)では自動音声認識技術を利用して音声に対する書き起こしを行うことで会話の了解性が高まることが示されたが、NS と NNS の共通理解は示されていない。本研究では、共通理解へ向かう過程となる共同解釈の構築を説明し、仮説 1:「音声とテキストの利用は共同解釈の構築を増加させる」を検証する。

Clark ら(Clark and Brennan 1991)によれば、会話の参加者は相手の発話を理解したことを示し、共同解釈の構築(grounding)を適切に蓄積する過程を経て、会話の共通基盤(common ground)すなわち共通理解を得る。共同解釈を構築する会話では、聞き手の理解は以下の 3 種類の反応によって示される。(1) 相槌・主張による話し手への反応 (2) 前提や提示となる発話、次発話を発話 (3) 視線・ジェスチャによる非言語的な反応(伝 2009)。これらは話し手や聞き手と

して相手の発話への受理を示す行動として認定されている。これらの発話や反応があれば、共同解釈を構築する会話と認められる。

会話分析のための音声の書き起こしは録画した実験映像から会話分析の手法に基づいて行った。日本語話し言葉の節単位のうち「～です」などの文末表現と、「けれども～」など構造的に大きな切れ目をなす並列の節境界を用いて、発話節に区切って書き起こしを行った(坊農 et al. 2009)。

会話分析は、従来研究(Clark and Brennan 1991)で示された共同解釈構築の証拠となる会話の種類、定義と会話例に基づいて行った。表 6 に共同解釈を構築する証拠となる会話の例を示す。

表 6 の相槌とコメントは、情報交換中に聞き手の連続したたくさんの相槌の後に「なるほど」、「わかった」などの主張による話し手への反応をすることによって聞き手の受理を表明する。表 6 の会話例のように、相槌とコメントはたくさんの「はい」に加え、「なるほど」という発言の締めくくりで、共同解釈を構築する会話の件数を 1 件と数えた。情報交換中のたくさんの相槌は無限に続くが、なるほど、わかったなどの発言の締めくくりで終了する。相槌の回数による定量的な判断はなかった。

表 6 の一貫した内容での情報交換は、隣接ペアと繰り返しと相槌から成り、前提となる発話に対し次発話を発話することで聞き手の受理を表明する。隣接ペアは質問と応答のような組があれば、共同解釈を構築する会話が 1 件と数えた。繰り返しと相槌は、一貫した内容部分での情報交換中に発言の繰り返しと相手の相槌の組があれば、共同解釈を構築する会話が 1 件と数えた。

表 6 のジェスチャは視線や動作などの非言語的な反応で聞き手の反応を表明する。ジェスチャは、会話中にじっと相手を見る、対象を指さすなど会話に注意を払っていることを表す行動である。実験会話は相手の行動が見えない音声会議のため、視線やジェスチャは数えないこととした。実験者 2 名が 7 分間の会話に対して独立に会話分析を行った。2 名の判断の一致率を示すカッパ係数

は高かった( $\kappa=.85$ )。参加した組ごとの共同解釈を構築する会話の件数の平均値を表 7 に示す。音声のみ条件で十分に共同解釈を作れる組はなく、音声とテキスト条件ではほぼすべての組が共同解釈を構築していた。

Shapiro-Wilk 検定にて正規性を認めないことを確認し、Wilcoxon の符号付き順位検定を行ったところ、共同解釈を構築する会話の件数は音声とテキスト条件( $M=6.4$ ,  $SD=4.1$ )で音声のみ条件( $M=2.6$ ,  $SD=2.1$ )より有意に多かった( $Z=3.4$ ,  $p<.01$ )。図 6 に条件ごとの共同解釈を構築する会話の平均件数(件/組)を示す。エラーバーは標準誤差を示す。この結果から仮説 1: 「音声とテキストの利用は共同解釈の構築を増加させる」ことが支持された。

表 6 共同解釈構築(grounding)の証拠となる会話

種類	定義	実験会話例
相槌とコメント	連続したたくさん の相槌の あとの主張	<p>NNS ん：ひところ(.)人殺したものは：</p> <p>NS はい</p> <p>NNS ん：あの犯罪者：あの犯罪者は：他人の：命を奪って：あの：この：この人の命じゃなくて：</p> <p>NS はい</p> <p>NNS だけじゃなくて：彼の家族の：ぜん(.)全部の：すごく(.)あの：あ：不幸になります(.)なりました：[あの：</p> <p>NS はい]</p> <p>NNS あとは：ん(.)：ん：ん：死刑を賛成というのは：あの：すごく：ん：(1) 残酷の：すごくん：悪い人に：対する(.)ん：あとは：この：(.)ん：犯罪者の存在は：あと：社会の安全：に：あの：危険な：存在だと思えます(.)</p> <p>NS はい</p> <p>NNS こくろくみ：の：人々にとって：</p> <p>NS はい</p> <p>NNS すごくきけん(.)危険な存在だとも思います</p> <p>NS はい</p> <p>NNS 人々の：安全のために：あの：この：あの：危険が：ん：つ：ん：この：ん：犯罪(.)事件を：</p> <p>NS はい</p> <p>NNS ん：(.)防犯ために：あのすごく(.)あの：必要だと思えます</p> <p>NS はい (1) あ(.)なるほど：</p>
一貫した内容での情報交換	隣接ペア	<p>NNS たとえばどのような刑罰ですか？</p> <p>NS 日本では無期懲役があります</p>
	繰り返すと相槌	<p>NNS 実施は難しいだと思います</p> <p>NS 実施が難しい？</p> <p>NNS はい”</p> <p>NS ん：なるほど</p>
ジェスチャ	注意を払い続ける	×

会話分析に用いられる記号[19]

： 長音はコロンで表す。

( ) 発話中に短い間がある場合につける。

[ 複数の話手の音声重なっている場合、重なるの始まりを角括弧で表す。

] 重なるの終わりを角括弧で表す。

1秒以上の間は沈黙として、丸括弧中にその秒数を表す。

? 語尾の音が上がっていることは疑問符で表す。

表 7 組ごとの共同解釈を構築する会話件数

組	音声とテキスト	音声のみ
1	5.0	2.0
2	6.5	3.0
3	9.0	6.5
4	4.5	0.0
5	4.0	1.0
6	8.0	4.0
7	11.5	3.0
8	4.5	2.0
9	18.	5.5
10	3.5	1.0
11	8.0	0.0
12	3.0	2.0
13	9.0	7.0
14	1.0	2.0
15	3.0	2.0
16	3.5	1.0

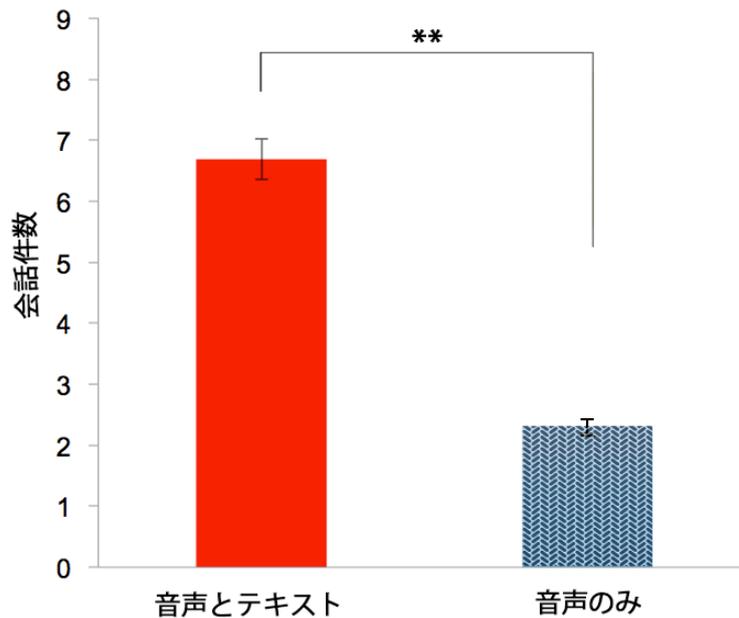


図 6 共同解釈を構築(grounding)する会話の件数(件/組)

N=16, \*\*:  $p < .01$

### 3.2.3 音声とテキストによる共通理解の調査結果

音声とテキストによる会話支援の従来研究では、参加者の共通理解を会話内容の記憶から調べた。共通理解の測定において、参加者間で会話後に一致した情報を持つことは重要な指標となるため、仮説 2:「音声とテキストの利用は、共通理解を向上させる」を検証する。

自由再生課題の分析ではまず、NS と NNS のそれぞれの記憶による回答件数を条件ごとに数えた。PC 上で回答された NS と NNS の自由再生課題回答内容は空白や改行を用いて明瞭に分割されていた。自由再生課題回答内容が長い場合は、単語や文章を意味の切れ目で区切り、NS と NNS それぞれの回答数を条件ごとに数えた。例として「資源が枯渇しそうだということと、風力のようないわゆるクリーンな発電方法は日本の自然環境には向いていない」は、

「資源が枯渇しそうだということ」と「風力のようないわゆるクリーンな発電方法は日本の自然環境には向いていない」に区切り、回答数は2件と数えた。

NS の記憶による自由再生課題回答件数に対し、Shapiro-Wilk 検定にて正規分布していることを確認し、対応のある t 検定を行ったところ音声とテキスト条件( $M=4.8$ ,  $SD=1.3$ )と音声のみ条件( $M=4.7$ ,  $SD=1.7$ )で有意差が見られなかった。同様に、NNS の記憶による自由再生課題回答件数を Shapiro-Wilk 検定にて正規性の確認後、対応のある t 検定を行い音声とテキスト条件( $M=4.5$ ,  $SD=1.7$ )と音声のみ条件( $M=4.3$ ,  $SD=1.5$ )で有意差がなかった。NS と NNS の自由再生課題回答件数の平均(件/組)を図7に示す。

次に、参加者間で自由再生回答内容を比較し、一致した語句の件数を数えた。一致しているかどうかは、従来研究(田中 et al. 1998)で示された意味解釈の基準と例に基づいて行った。表8に語句の分類根拠と例を示す。一致語句の分類根拠は、刑務所という言葉に対しては、囚人を拘留する所、悪い

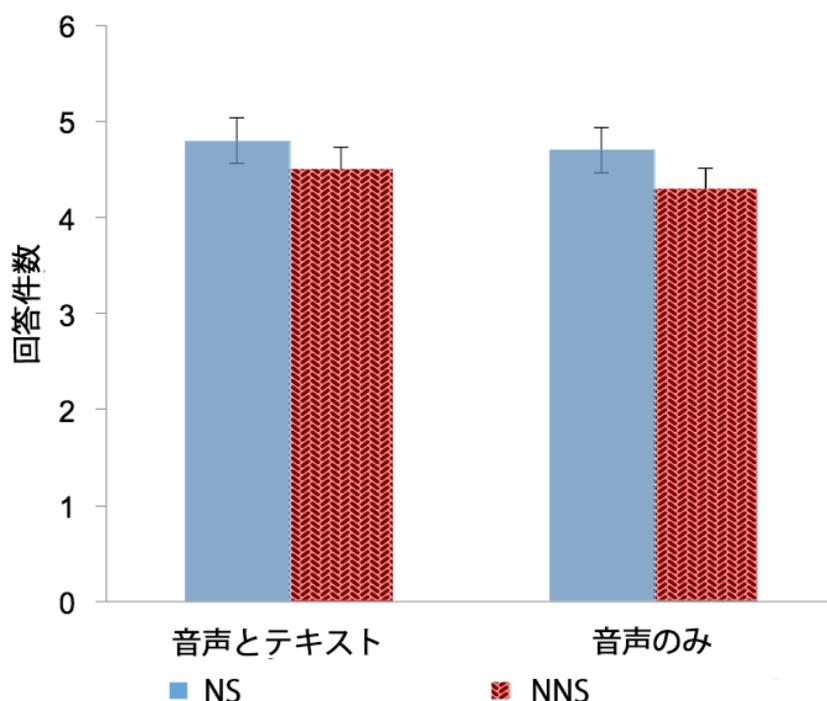


図7 会話後の自由再生回答(記憶)件数(件/組)

人を閉じ込める建物、犯罪者を拘束する機関は、類似の意味と他者が一般に意味づけできるので同じ語句と言える。南極大陸とベースボールは、他者から見て一般的に意味づけが異なるため同じ語句とは言えない。全く同じ、または意味が同じ語句を用いている場合に、一致語句を一件と数えた。参加者間で一致語句と分類できる語句の例を表 9 に示す。

参加者間で語句を比較し、他者から見て類似の意味と理解できる語句を一致語句に分類し条件ごとに件数を数えた。分析では実験ビデオを参照して自由再生課題回答内容を確認し、理解できない回答は分析対象外とした。

NS と NNS の自由再生課題回答内容の分析は、2 名が独立に実施した。採点者間の一致を示すカッパ係数を調べたところ信頼係数は高かった( $\kappa = .75$ )。NS と NNS の共通理解の件数は Shapiro-Wilk 検定にて正規分布に従わないことを確認し、Wilcoxon の符号付き順位検定を行ったところ、音声とテキスト条件( $M=3.3, SD=1.3$ )で音声のみ条件 ( $M=1.8, SD=0.8$ )より有意に多かった( $Z=3.2, p<.01$ )。Cohen の効果量を算出した結果、大きな効果が見られた( $d=1.4$ )。参加者間の一致語句(common ground)の平均件数(件/組)を図 8 に示す。この結果から仮説 2:「音声とテキストの利用は共通理解を向上させる」が支持された。

表 8 語句の分類根拠と例(田中 et al. 1998)

分類	根拠	語句例	語句例
同じ語句	類似の意味と他者が一般に意味づけできる	刑務所	囚人を拘留する所
			悪い人を閉じ込める建物
			犯罪者を拘束する機関
異なる語句	他者から見て一般的に意味づけが異なる	南極大陸	ベースボール

表 9 参加者間で一致している語句の例(嶋 et al. 2020)

分類	議題	NS	NNS
一致 語句	原発 問題	環境への影響を考えるとよくない	地球環境に悪い
		原子力発電は軍事転用の危険性がある	軍事転用
		日本は狭いため、事故が起きたときの被害が大きい	日本は狭い国だから、原子力発電の事故があったら、大変な問題になる可能性もある
		原発事故	原子力発電での事故
		原子力発電危険性	原子力発電も危ない
		太陽光や風力は、日光が当たらなかつたりしたら使えない	太陽光や風力の発電方法は安定的な電力を提供できない
		原子力発電は環境へのリスクが大きい	地球の環境と人間にとってよくないだともう
		地震の少ないところで発電	被害問題について、天災が少ない地域におけるなら問題が少なくなる
		放射能	原子力発電による放射能
		技術力があることを世界に示せる	技術力をアピール
	地球温暖化などの環境問題	今国際性で問題になる環境問題がますます深刻している	
	死刑 問題	無期懲役で代替できる	死刑の代わりに、無期罰役のほうがより適合だと思う
		犯罪者の意識が、罪を償う間に変わるかもしれない	犯罪者の意識は変わらないわけじゃない
		無期懲役というものがある	無期懲役
		犯罪者が世の中に出ると危ない	犯罪者の危険性
		再犯のおそれ	服役後にまた殺人を犯して人がいる
		人間が捜査を行うので必ず間違いが起こる	冤罪になる可能性がある
		税金が犯罪者に使われるのは快く思わない	納税人として、自分の税金が人を殺す犯人に使うのは納得できない。
		犯罪者にも人権があるということ	個人の人権とほかの人の人権の問題
		国であろうとも人の命を奪う権利はないということ	犯罪者にも人権があり、尊重すべきだと思います
死刑によって犯罪を抑制する効果が実証されていないのであれば死刑をやる必要はない		死刑を使うことと犯罪抑止の関係は未だ証明されていなかった	

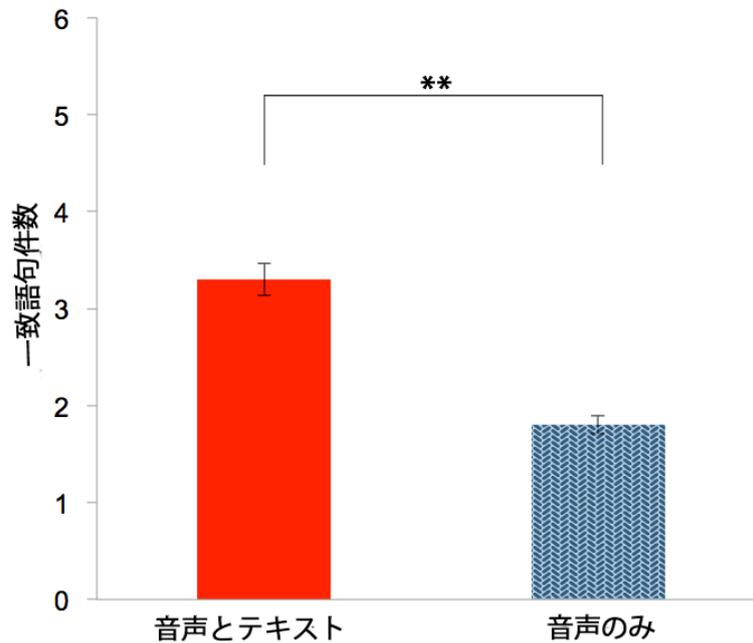


図 8 参加者間的一致語句 (common ground) の平均件数 (件/組)

N=16, \*\*:  $p < .01$

### 3.2.4 音声とテキストに対する質問紙調査結果

音声とテキストによる会話支援の従来研究では、支援ツールを使用する会話と使用しない会話を質問紙によって調査した。実験参加者の主観評価は会話支援ツールの効果を測定する研究において重要である。仮説 3: 「音声とテキストの利用は NNS の了解性と快適性を向上させる」を検証する。

会話についての質問項目は 16 項目とし、類似した質問項目は質問項目数の異なる 4 つの分類に分けた。16 項目の質問は、「相手の発言から豊富な情報が受け取れた」などの了解性に関する 4 つの質問、「相手と自然にコミュニケーションできた」などの会話に関する 4 つの質問、「自分はゆっくりできた」などの快適性に関する 5 つの質問、「テキストが見やすかった」などテキストに関する 3 つの質問から成る。それぞれの質問の提示順序は順序効果を相殺する

ためランダムに提示された。回答は、全く当てはまらない、当てはまらない、あまり当てはまらない、どちらともいえない、やや当てはまる、当てはまる、よく当てはまる、の 7 段階のリッカート尺度から最も当てはまるものを参加者が一つ選択した。以降では全く当てはまらないに 1 点、よく当てはまるに 7 点を割り当てて間隔尺度として扱った。質問内容は従来研究(Liu et al. 2014)(Inoue 2014)(西村 et al. 2006)(高橋 et al. 2001)(岡崎 et al. 2014)を参照した。

表 10 に質問の内容、調査結果の平均値と標準偏差を示す。分類別に見るため、関連のある質問項目間の内的整合性となる信頼係数を調べたところ、了解性 ( $\alpha=.73$ )、会話 ( $\alpha=.76$ )、快適性 ( $\alpha=.73$ )、テキスト( $\alpha=.78$ )と概ね高かった。快適性は逆転処理を行い、テキストは「テキストが見やすかった」などの条件間で対応のある質問項目のみ整合性を求めた。質問紙調査結果については、「音声とテキスト」と「音声のみ」の 2 水準を「条件」要因とし、「NS」と「NNS」の 2 水準を「話者」要因として、分散分析を行った。条件については被験者内計画、話者については被験者間計画であるため、対応がある要因(条件)と対応がない要因(話者)の 2 要因で平均値の比較を行う混合モデルの二元配置分散分析を行った。図 9 に質問紙調査結果の平均値を示す。エラーバーは標準偏差を示す。

了解性について条件の主効果は有意であり [ $F(1, 252)=5.39, p<0.05$ ]、条件と話者に有意な交互作用が認められた [ $F(1, 252)=5.83, p<0.05$ ]。2 要因の分散分析の交互作用が有意である際の下位検定にはいくつかの方法がある。本研究は各要因の水準数が 2 であるため多重比較は行わず、一元配置分散分析か  $t$  検定のいずれかのうち、2 つの水準間の平均値の差を結論する方法である  $t$  検定を行った。交互作用が有意であるため、条件の主効果を話者要因の処理水準ごとに検定した。その結果、NNS では音声とテキスト条件( $M=5.11, SD=1.28$ )が音声のみ条件( $M=4.66, SD=1.39$ )よりも有意に高い了解性を示した( $t(15)=2.06, p<0.05$ )。また、話者要因の単純主効果について条件要因の処理水準ごとに検

定した結果、音声とテキスト条件において NNS( $M=5.11$ ,  $SD=1.28$ )は NS( $M=4.31$ ,  $SD=1.38$ )よりも有意に高い理解性を示した( $t(30)=2.55$ ,  $p<0.01$ )。

会話について、条件と話者の交互作用が有意であった [ $F(1, 252)=3.77$ ,  $p<0.05$ ]。交互作用が有意であるため、条件の主効果を話者要因の処理水準ごとに検定した結果、NS と NNS の音声とテキスト条件と音声のみ条件の間に有意差は見られなかった( $p>0.05$ )。話者要因の単純主効果について条件要因の処理水準ごとに検定した結果、音声のみ条件の会話の評価が NNS( $M=4.11$ ,  $SD=1.50$ )よりも NS( $M=4.67$ ,  $SD=1.21$ )において有意に高い傾向が見られた( $t(30)=-1.71$ ,  $p<0.1$ )。

快適性について条件と話者の交互作用が有意であった [ $F(1, 316)=23.74$ ,  $p<0.001$ ]。交互作用が有意であるため、条件の主効果を話者要因の処理水準ごとに検定した結果、NNS の快適性では音声とテキスト条件( $M=5.30$ ,  $SD=1.53$ )が音声のみ条件( $M=3.68$ ,  $SD=1.84$ )よりも有意に高かった( $t(15)=2.92$ ,  $p<0.01$ )。話者要因の単純主効果について条件要因の処理水準ごとに検定した結果、音声とテキスト条件における NNS の快適性( $M=5.30$ ,  $SD=1.53$ )では NS( $M=4.54$ ,  $SD=1.76$ )よりも有意に高かった( $t(30)=2.59$ ,  $p<0.01$ )。

テキストについては話者による主効果が有意であり [ $F(1,188)=13.66$ ,  $p<0.001$ ]、条件と話者の交互作用が認められなかった [ $F(1, 188)=1.09$ ,  $p>0.05$ ]。平均から NS( $M=5.40$ ,  $SD=1.31$ )は NNS( $M=5.32$ ,  $SD=1.78$ )よりも高かった。

質問紙調査の結果から、音声とテキスト条件における NNS の理解性は NS よりも有意に高く、音声のみ条件よりも有意に高いことが判明した。音声のみ条件における NS の会話の評価は、NNS よりも有意に高い傾向がある事が示された。音声とテキスト条件における NNS の快適性は NS よりも有意に高く、音声のみ条件よりも有意に高いことが判明した。また、NS は NNS よりも有意にテキストをよく見ている事が示された。仮説 3 は NNS の理解性と快適性についてであった。図 9 の理解性と快適性に示したように、NNS の主観評価

表 10 質問紙調査結果

分類	質問	音声とテキスト		音声のみ	
		NS	NNS	NS	NNS
了解性	相手の話を理解しやすかった	4.31 (1.38)	5.11 (1.28)	4.67 (1.29)	4.66 (1.39)
	ディベート中、相手の主張に納得できた				
	相手の発言から豊富な情報が受け取れた				
	会話に集中できた				
会話	自分の言いたい事が話しやすかった	4.19 (1.36)	4.28 (1.28)	4.67 (1.21)	4.11 (1.50)
	相手と自然にコミュニケーションできた				
	相手とまた話したいと感じた				
	相手が反対しても自分の意見を主張できた				
快適性	相手との会話は不快であった	4.54 (1.76)	5.30 (1.53)	4.71 (1.48)	3.68 (1.84)
	相手との会話が退屈だと感じた				
	相手と会話していて疲れた				
	自分はゆっくりできた				
	自分は気楽にできた				
テキスト	テキストが見やすかった	5.69 (1.12)	6.00 (0.91)	5.10 (1.41)	4.65 (2.14)
	PC モニタは見やすかった				
	PC モニタをよく見た				

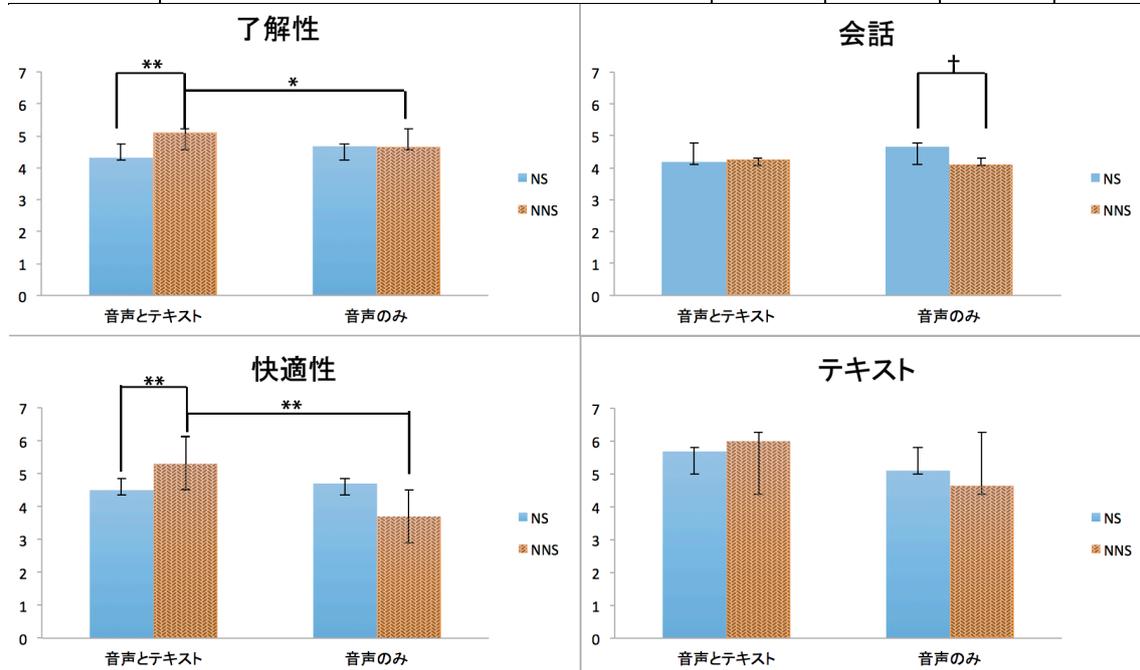


図 9 質問紙調査結果

† :  $p < .1$ , \* :  $p < .05$ , \*\* :  $p < .01$

の了解性と快適性は音声のみ条件より音声とテキスト条件において有意に高く、音声とテキスト条件の NS よりも有意に高い事が認められた。したがって、仮説 3:「音声とテキストの利用は NNS の了解性と快適性を向上させる」が支持された。

### 3.3 日本語音声会議の実験結果の検討

#### 3.3.1 音声とテキストによる共通理解の形成

これまでの研究では情報伝達が重視され、共通理解に至るまでの過程の会話が分析されたことはなかった。本研究では共通理解に至る過程である共同解釈を説明し、共同解釈を構築(grounding)する会話を定義して件数を数えた。図 6 は音声とテキストを利用する場合に NS と NNS が共同解釈を構築する証拠となる会話件数が増えたことを示す。

本結果は、従来研究(Gao et al. 2013)で自動音声認識技術を利用した音声に対する書き起こしに強調表示を付与することで NS と NNS の会話の了解性が高まることと同様の結果となった。NS による発話内容のテキスト化が、NS 自身に発話内容を深く考えさせるきっかけとなり、テキストと発話内容をわかりやすくしたことが、発話内容の変化に繋がったと考えられる。また、テキストチャットの研究では、参加者は状況や相手に応じて話し方を変更したり言語形式を変容したりする行動がみられた(Walther 2007)。本研究はこの行動を支持する結果であり、会話参加者がテキストの利用によって言語形式を変容したことを示す。

音声会議中には NS が話したトピックの 78%と NNS が話したトピックの 51%が NS によってテキスト化された。この結果は、NS は NS と NNS の発

話内容をテキストに表示したことを意味する。音声会議では、NS が音声とテキストで NNS の発話に関して確認を行い共有解釈を構築する発話が見られた。従って本結果は日本語会話に不慣れな NNS が NS と双方向に発話する場面で発話と理解を支援するコミュニケーション支援になることを示唆する。

図 7 は自由再生課題の回答件数が NS と NNS の間で有意差がないことを示した。会話後の記憶は、NS と NNS の共通理解の測定に重要な指標とされ、従来研究(Echenique et al. 2014)(Pan et al. 2017)でも調査された。本研究は、自由再生課題の回答件数は記憶の件数に過ぎず、その中で両者に一致する語句数が共通理解の指標となり得ると考える。従って、本研究は音声会議後の NS と NNS のそれぞれの記憶と、一致語句を調べ、一致語句を NS と NNS の共通理解として測定した。NS と NNS のそれぞれの記憶の件数に有意差がないことは、従来研究(Echenique et al. 2014)を支持する結果となった。会話後の一致語句が音声とテキスト条件において多い結果は、従来研究においては調査されていない。本結果は、NS と NNS の共同解釈を構築する会話の件数が音声とテキスト条件で増加することと整合する結果となった。

図 8 は会話後の一致語句が音声とテキスト条件において多いことを示した。本結果は、NS と NNS の共同解釈を構築する会話の件数が音声とテキスト条件で増加することと整合する結果となった。言葉の理解において、意味づけがずれを含むことは避けられず、その場の限られた状況においてほぼよい程度で理解される(Ferreira et al. 2009)。図 8 で音声とテキスト条件で共通理解が増加したことは、音声とテキストの利用によって会話の大きなずれが減少したためと考えられる。

音声のみ条件で共通理解が少ない理由は、会話音声を表示したテキストがない場合に NNS は正確に単語認知できず、聴解が出来なかったことが原因と考えられる。聴解と単語認知の考え方は、音声と文字の単語認知モデルから説明されている(平井 2000)。音声からの単語認知は音声から即座に意味解釈にいた

る。一方で、文字情報からの単語認知は、文字から入った情報が即座に意味解釈にいたるルートと、音声化をたどって意味解釈にいたるルートがある。音声と文字からの複数の単語認知経路を活用できれば単語認知しやすくなり、意味解釈に至り易くなる。つまり、音声とテキスト条件では NNS は NS の発話の音声からとテキストからの複数の単語認知経路を活用でき、意味解釈に至りやすくなると考えられる。

NS がテキスト化の作業を負担した先駆的な従来研究(Inoue et al. 2015)では、作業負担と比べても会話支援として得られるものが上回るため作業する意義があると述べた。本結果は Inoue らの主張した NS の作業による会話支援のメリットの根拠を示した。NS が作業負担する場合、テキスト化のメリットとして、共通理解できること、発話内容がわかりやすくなること、NNS の快適性が増すことが挙げられる。テキスト化のデメリットとしては NS の作業負担がある。一方でテキスト化しない通常の会話の場合、1章で述べたようにテキスト化の負担はないことがメリットだが、共通理解が困難であることと、NNS の負担が大きいことがデメリットとなる。NNS と困難なコミュニケーションを続けることよりも、NS がテキスト化を負担することでコミュニケーションを改善することに意義があると考えられる。

### **3.3.2 音声とテキストによる記憶への影響**

記憶に関してはコミュニケーションだけでなく、認知科学からの研究が多数行われている。音声のみを用いる場合と、文字のみを用いる場合と、文字を付した音声を用いる場合を比較して児童の単語記憶が調査された(Menne 1972)。その結果、文字を付した音声を用いる場合に最も単語の記憶テストの点数が高かった。これらは、メディアの数が多いほど記憶への手がかりが多くなる効果

(cue summation)と呼ばれ、認知科学や心理学の観点から説明が与えられている。

音声からの単語認知は弱く、視覚を通す方がより正確に多くの情報が記憶に残りやすい。また、音声言語は一過性であり時間の経過につれて保存と再現性に弱い。文字情報はこれらに強い(鹿島 1997)。以上から、NS と NNS は音声のみの会話では共通理解を得にくいと考えられる。

音声とテキスト条件において参加者の発語数が大きくなれば一致語句(common ground)の平均件数(件/組)も大きくなる可能性がある。そこで、発話を調べたところ、音声とテキスト条件の全語数は平均 1908 語/組、音声のみ条件の全語数は平均 1909 語/組であり、差がなかった。従って、音声とテキスト条件における一致語句の増加は音声あるいはテキストによる変化であり、発語数が増えたためではないと考えられる。

### 3.3.3 音声とテキストに対する参加者の評価

質問紙調査では音声とテキストに対する NS と NNS の主観的評価を調べた。NNS は音声とテキスト条件において“相手の話を理解しやすかった”“相手の発言から豊富な情報が受け取れた”などの了解性と、“自分は気楽にできた”“テキストが見やすかった”などの快適性に関する質問で有意に高く評価した。NNS が音声のみで震災、放射能、無期懲役、冤罪などについて会話をするには集中力を要し、それが音声とテキストによって軽減されることを示唆した。

一方 NS は“自分の言いたいことが話しやすかった”など会話についての質問において音声とテキスト条件を有意に低く評価している。また、NS は音声のみ条件において“PC モニタをよく見た”などのテキストの評価が NNS よりも有意に高い傾向が見られた。NS の了解性に対し音声とテキスト条件と音声

のみ条件の間で有意な差がないことから、NS の負荷はさほど大きくないことが裏付けられた。

### 3.3.4 従来研究を支持する結果と異なる結果

本研究の結果から NNS を含む音声会話における NS による音声のテキスト化は NS と NNS が相互に会話への理解を向上させることを示した。本結果は従来研究(Pan et al. 2017)で自動音声認識技術を利用した音声に対する書き起こしに強調表示を付与することで NNS の会話の理解性が高まることを支持する結果となった。

また、会話後の質問紙調査において NNS の快適性が向上した。質問紙調査結果は、従来研究(Pan et al. 2017)で NS がテキストに強調表示を付与し、NNS のみの快適性が向上したことを支持する結果となった。自動音声認識技術を利用したテキストの誤りに対し NS がテキスト編集を行った従来研究(Gao et al. 2014)では、NS の発話音声に対する NNS の理解性が高まることで NNS の快適性が向上するという、理解性と快適性の相関関係が示されている。本研究で NNS の理解性と快適性の向上を示した結果は、従来研究(Gao et al. 2014)を支持する結果となった。

本研究において会話とテキストについても調査した。その結果、NS は音声のみ条件の会話を高く評価し、NNS は音声とテキスト条件を高く評価した。また、音声とテキスト条件において NNS の快適性は高く、NS の快適性は低くなった。従来研究(Gao et al. 2014)と比較すると、NS がテキストに強調表示を付与する場合には NS の快適性は低下せず、この点において異なる結果となった。本研究の音声のテキスト化は従来研究(Gao et al. 2014)の強調表示とは異なるため、異なる結果になったと考えられる。

### 3.4 本実験の限界

本実験の限界として、会話参加者が会話音声を書き起こすテキストは、参加者の主観と課題によって変わる可能性がある。現時点ではテキスト表示を自動化する技術にはまだ問題があるものの(Rebai et al. 2019)、今後は自動音声認識や機械翻訳技術が進歩する可能性がある。今後進歩するメディアを活用できるように環境を設計すれば、本研究の結果は他の研究でも活用できると考えられる。

背面合わせの会話実験環境は、表情や動作が見えないため音声会議のシミュレーション環境とした。しかし、現実の遠隔音声会議の環境では通信ノイズや遅延の影響があるため、今後は遠隔音声会議の環境で実験を行う必要がある。

また、本研究は実験参加者にのみ情報を伏せて行ったが、分析は盲検化が困難のため非盲検で行った。結果には 2 名の独立な分析の平均値を用いたが、非盲検によってバイアスが生じる可能性がある。

本研究の限界として、参加者の会話後の共通理解を条件間で比較した。しかし、参加者間で会話前に一致した情報や知識を持つことも考えられるため、会話前後の比較がより正確な共通理解の測定になると考えられる。

### 3.5 3章のまとめ

NS と NNS の音声会議でテキストを表示する会話支援の有用性について 3 つの仮説を立て検証した。本実験の結果から、仮説 1:「音声とテキストの利用は共同解釈の構築を増加させる」の検証では、音声とテキストを利用する場合に共同解釈を構築する証拠となる会話が多く見られることが実証された。また、仮説 2:「音声とテキストの利用は共通理解を向上させる」の検証では、音声と

テキストを利用する場合に参加者間で一致した記憶件数が増加し、音声とテキストの利用が共通理解の増加に有効であることが実証された。その結果として参加者への質問紙調査では、NNS の了解性と快適性が向上した。この結果から仮説 3:「音声とテキストの利用は、会話における NNS の了解性と快適性を向上させる」が実証された。本結果は NS と NNS の音声会議における共通理解の達成を示し、音声とテキストによる NNS との会話支援ツールとしての有用性を示唆する。

## 第 4 章 音声会議中に音声を書き起こすタイミングと内容

第 3 章では、NNS との音声会議において NS が発話音声の一部をテキストに書き起こす支援により、共通理解が向上するなどの知見を得た。第 4 章では、第 3 章の実験データを用いて、組ごとに音声をテキスト化したタイミングと内容について分析を行う。

会話音声のテキスト化の量と内容を調査することで、テキスト音声会話が共通理解を向上させる理由を明確にすることができる。同時に、テキストチャットや自動音声認識を利用したテキストとの違いを実証的に明らかにし、テキスト音声会話の自動化のデザインに言及できると考えられる。従って本研究では音声をテキスト化する支援で実際に行われた音声テキスト化のタイミングと内容を調査する。

2 章で述べたように、自動テキスト化技術のタイミングと内容には詳細な調査が行われた一方で、テキスト編集や音声をテキスト化するタイミングと内容に対しては調査がほとんど行われなかった。NS と NNS のテキストチャットに関する従来研究においては、機械翻訳による会話支援の効果が検討され (Yamashita et al. 2009)(Hautasaari 2010)、ほぼ同時期に自動音声認識の精度や速度を調べた研究が多く行われた (Pan et al. 2009)(Pan et al. 2010)(Yao et al. 2011)。その結果、自動テキスト化技術には誤りが多く含まれ、誤りを含むテキストはかえって NNS の理解を損なうことや、遅延は少ないほうが NNS に好まれるなどの知見が明らかになった。そこで、誤りを含むテキストに編集や強調を加え、NNS の理解を促す試みが行われた (Gao et al. 2013) (Gao et al. 2014) (Pan et al. 2017)。テキスト編集は、NNS の了解性や快適性が向上し会話の雰囲気良くなるなどの一定の効果が示された。NS が音声をテキスト化するタイミングのおおよその分析は示されたが (宋 2018)、詳細なタイミングと内容は明らかでない。

データ分析の結果から、NS 発話のテキスト化は発話から約 2 秒遅れ、全体として発話のテキスト化は約 3 秒遅れることが多かった。また一回のテキストは平均 5 字程度で、2～3 語から成り、会話中で繰り返し説明された名詞句がテキスト化されることが多かった。さらに、音声テキスト化の支援がある会話を分析した結果、話し方の形式の変化と補足説明の追加が認められた。音声のみの会話よりも説明やエピソードが追加され、詳しく説明する話し方が行われることを示した。すなわち、本研究で行った音声のテキスト化は音声の直後に補足するテキストが表示されたため、話し方を変化させ、NNS の理解力を補うことで NS と NNS の共通理解を向上させることを示した。この結果により、会話支援ツールのデザインは、音声の 2 秒後に補足内容をテキスト表示し、話し方の話し方を変化させることで共通理解を促すことを示した。

#### 4.1 音声会議のテキスト化の分析

3 章の音声とテキストによる会話支援の実験データから、音声会議中の NS による発話内容のテキスト化のタイミングと内容を実証的に検討する。分析項目はビデオ分析による音声とテキストのタイミングの差、テキスト化の量、名詞句のカテゴリ分類によるテキストの内容、音声のテキスト化による話し方の変化の 4 つである。表 11 に 4 章の分析項目とデータ取得をまとめる。

データは、16 組の音声とテキスト条件の実験ビデオ映像合計 112 分と、参加者用 PC のモニタをデスクトップキャプチャソフトウェアで録画したビデオ映像と、NS がテキスト化したスペースや改行の切れ目で区切られた 248 個の語句と、音声を書き起こした音声会議の全文から取得した。

表 11 4章の分析項目とデータ取得

分析項目	データ取得方法	分析方法
遅延	実験ビデオと、参加者用 PC のモニタをデスクトップキャプチャソフトウェアで録画したビデオから取得	テキストとそれが含まれる直近の発話の時間差を分析
定量	NS がテキスト化した、スペースや改行で区切られた語句	字数, 回/分, 字/分, 語/分を算出
内容	実験ビデオと、参加者用 PC のモニタをデスクトップキャプチャソフトウェアで録画したビデオから取得	テキストと対応する発話を会話的特徴から 6 種類の名詞句に分類
話し方	会話音声を書き起こし会話の全文を取得	会話の書き起こし上の名詞句を 6 種類に分類

#### 4.1.1 音声をテキスト化するタイミング分析

本節では、NS が発話内容をテキスト化するタイミングを調査する。音声とテキスト条件の実験ビデオ映像合計 112 分に、アノテーションソフトウェアである ELAN(Eudico Linguistic Annotator)を使って発話とタイピングのタイミングと内容をラベリングした。ELAN は同時に複数の映像に対して数種類のタイムラインにラベリングをし、ラベルのカテゴリごとに各種形式のファイルに出力できる。本研究において発話のタイミングは間休止単位に基づいて行った。間休止単位 (IPU) は一定時間以上の無音区間により発話を区切る単位である。NS がタイプしたタイミングは、テキストが画面に表示された時点を開始時間

とし開始から終了までをラベリングした。テキストとそれが含まれる直近の発話を対象として発話とテキストの時間差を分析した。テキストは NS がテキスト化する際に挿入したスペースや改行の切れ目で区切られた 248 個の語句を対象とした。

ELAN を使ったアノテーションの手順を以下に述べる。図 10 に ELAN のアノテーション結果を示す。まず、NS と NNS を考慮せずに発話の開始時刻とタイピングの開始時刻の差を調べた。次に、タイピングされたテキストと発話を比較し、テキストと対応する発話を対応づけた。表 12 にテキストと対応する発話の分類根拠と例を示す。対応する発話の判断基準は全く同じ語句が発話された場合に対応する発話とし、あるいはテキストが含まれる発話に対応する発話とした。テキストと対応する発話が 2 回以上発話された場合は発話開始時刻がタイピング開始時刻に最も近い発話を選んだ。具体的には、図 10 の 8 行目のテキスト“幼少期”は、図 10 の 9 行目“青年期”と対応させるのではなく、図 10 の 7 行目の発話と対応させ時間差を分析した。図 10 の 11 行目テキスト“心理的な障害を持っている”と対応する発話は図 10 の 13 行目ではなく、11 行目の発話とした。12 行目テキスト“ケア”は図 10 の 10 行目の発話との時間差をとった。

	A	B	C	D	
1	ティア	開始時間	終了時間	間隔	内容
2	NS	00:01.2	00:01.4	00:00.2	あ
3	NS	00:03.4	00:08.2	00:04.8	えとじゃあ先に私が()反対の理由()としては:
4	NS	00:10.0	00:18.1	00:08.2	えっと:日本の今の:若い人()たちが()死刑になるような: 重大な犯罪を犯す時
5	テキスト	00:18.1	00:22.1	00:03.9	日本の若者
6	NS	00:22.0	00:23.5	00:01.5	っていうのは:
7	NS	00:24.5	00:26.1	00:01.6	なんていう:幼少期
8	テキスト	00:26.1	00:31.0	00:04.9	幼少期
9	NS	00:29.0	00:37.4	00:08.4	に()寂しい思いをしていたりとか:あとは青年期に愛を受け 取れないって言う()データが出ていて
10	NS	00:39.4	00:49.2	00:09.8	心理面の:ケアが必要なことが多くて:( )死刑よりも:そういう ケア()のほうに目を向けたほうがいいか なって思うからです
11	テキスト	00:49.1	00:59.5	00:10.4	心理的な障害を持っている
12	テキスト	01:00.8	01:02.4	00:01.6	ケア
13	NNS	01:36.0	01:43.5	00:07.5	しかし:えと:心理:じょうたい障害は: あの:ん:病気なので
14	NNS	01:45.0	01:45.4	00:00.4	え:
15	NNS	01:49.0	02:00.8	00:11.9	その一回の:犯罪は:え:死刑が: 必要です() また:え:と 法律ですが:え:と
16	NNS	02:01.8	02:17.3	00:15.6	)法律を:守るべきとを:守るため: ( )え:と() 法律い() え: 法律を違反する人を:まも() まもることが必要ない と思います
17	NS	02:22.8	02:28.6	00:05.8	法律を守らない人() に対する処罰?刑罰?
18	テキスト	02:29.1	02:36.4	00:07.3	刑罰
19	NNS	02:29.1	02:29.4	00:00.3	刑罰です
20	NS	02:29.3	02:30.5	00:01.2	刑罰

図 10 ELAN のアノテーションの出力結果

表 12 テキストと対応する発話の分類根拠と例

分類	根拠	テキスト例	音声例
対応する 語句	全く同じ語句	罪が誤ったもの	罪が誤ったもの:
	テキストが 含まれる発話	幼少期	なんていう:幼少期
対応しない 語句	テキストと 異なる	地震	ま:原子力発電(.)を:しなければ:

発話の開始時刻とタイピングの開始時刻の差を算出する手順を以下に述べる。図 11 にタイピングと対応する発話のタイミングの差を示す。まず、図 10 に示したタイピング開始時刻と対応する発話開始時刻の差を計算した。次に、発話とテキストが 0-1 秒差であるテキスト、1-2 秒差であるテキスト、と 1 秒ごとに音声のテキスト化の回数を数えた。次に、音声のテキスト化の回数が全体のテキストの個数に占める割合を 1 秒ごとに算出した。その結果を、タイピングと対応する発話のタイミングの差としてヒストグラムに表した。横軸はタイピングの開始時刻から発話の開始時刻を引いた時間である。縦軸は秒ごとのテキスト化の回数が全体のテキストの個数に占める比率である。分析対象とした 248 個のうち外れ値を除いた 236 個が図 11 中に存在する。発話の開始時刻とタイピングの開始時刻の差は平均 2.5 秒となり、全体の 84%が  $M \pm \sigma$  の範囲に分布していた ( $M=3.0$ ,  $SD=5.6$ )。すなわち大部分のタイピングは発話開始の 3 秒前から 8 秒後の間に開始されていた。結果から、音声とテキストを使用する会話では発話から 3 秒遅れて発話内容がテキスト化されることが示された(宋 2018)。

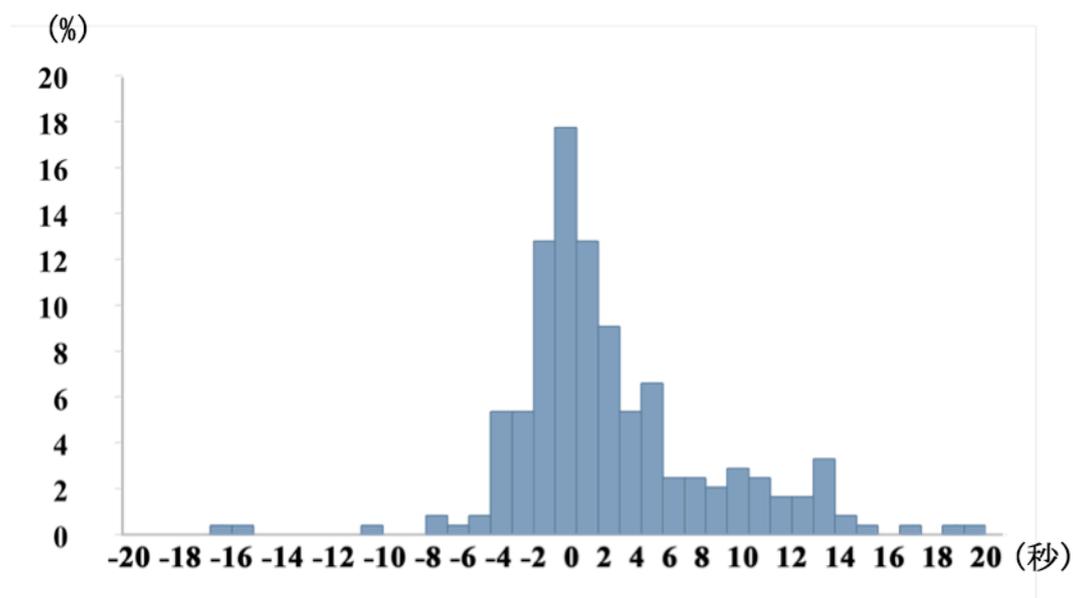


図 11 タイピングと対応する発話のタイミングの差

次に NS と NNS 別に発話の開始時刻とタイピングの開始時刻の差を調べた。まず、NNS の発話とテキスト化のタイミングを調べるため、タイピングと発話を対応付けた。テキストと対応する NNS 発話は 60 個であった。タイピング開始時刻と NNS 発話開始時刻の差を取り、1 秒ごとに音声のテキスト化の回数を数えた。音声のテキスト化の回数が全体のテキストの個数に占める割合を 1 秒ごとに算出した結果を、タイピングと対応する NNS 発話のタイミングの差としてヒストグラムに表したものが図 12 左である。横軸はタイピングの開始時刻から発話の開始時刻を引いた時間である。縦軸は秒ごとのテキスト化の回数が全体のテキストの個数に占める比率である。NNS 発話の開始時刻とタイピングの開始時刻の差は平均 5.8 秒となった。分析対象とした 60 個のうち外れ値を除いた 59 個が図 12 中に存在する。全体の 73%が  $M \pm \sigma$  の範囲に分布していた( $M=5.8$ ,  $SD=4.1$ )。つまり、NNS 発話は、発話から 1.8 秒後から 9.9 秒遅れてテキスト化された。図 12 左は NS が NNS の発話をテキスト化するタイミングは発話より絶対的に遅れることを示す。

次に NS の発話とテキスト化のタイミングを調べるため、テキストと NS の発話を対応づけた。テキストと対応する NS 発話は 188 個であった。タイピング開始時刻と NS 発話開始時刻の差を取り、1 秒ごとに音声のテキスト化の回数を数え、音声のテキスト化の回数が全体のテキストの個数に占める割合を 1 秒ごとに算出した結果を、タイピングと対応する NS 発話のタイミングの差としてヒストグラムに表したものが図 12 右である。NS 発話の開始時刻とタイピングの開始時刻の差は平均 2.3 秒となった。外れ値は除き、分析対象とした 188 個のうち、167 個が図 12 右に存在する。全体の 76%が  $M \pm \sigma$  の範囲に分布している( $M=2.3$   $SD=5.8$ )。図 12 は音声会話中の NS がタイピングによってテキスト化を行うと、0-1 秒を最頻値として発話よりテキスト化が早い場合と、発話よりテキスト化が遅い場合が分布することを示す。

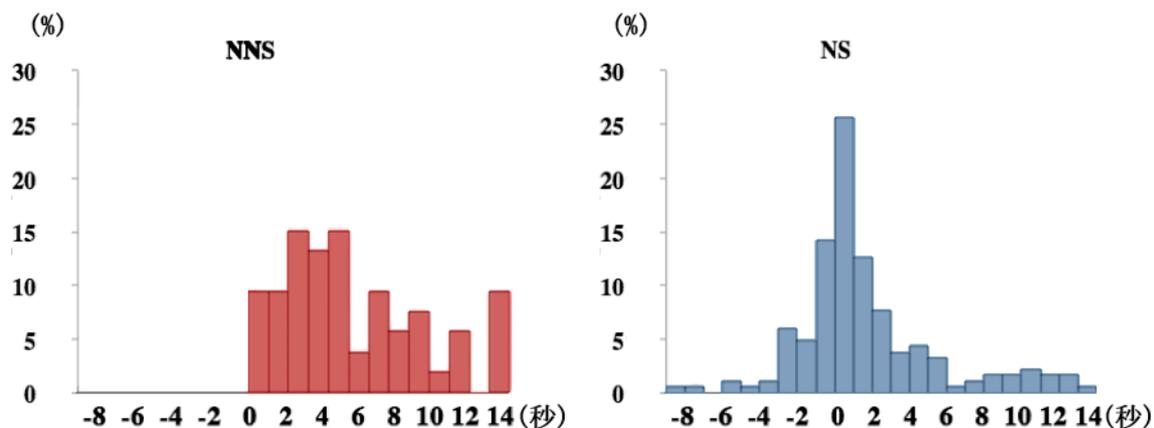


図 12 話者ごとの発話とテキスト化のタイミングの差

#### 4.1.2 テキスト化された内容の分析

本節では音声テキスト化された量、内容と話し方の変化を調べる。まず、テキスト化された量を調べたところ、音声のテキスト化は一回あたり平均 5.7 字/回、テキスト化の回数は平均 2.2 回/分で、テキストの量は平均 CPM=12.7/分、テキストの語句数は平均 WPM=6.7/分となった。音声のテキスト化は一回あたり平均 5.7 字/回のため、音声会話を妨げず、会話中に読める量であると考えられる。

テキスト化された語句には、3 章において自由再生課題回答として会話後の記憶に残ったテキストと、自由再生課題回答に記入されなかった、つまり記憶に残らなかったテキストがある。自由再生課題回答に記入されなかったテキストは 176/248 で全テキストのうち 70%で、平均字数は 6.1 字であった。形態素解析を用いて語の最小単位である品詞に分解して調べたところ、それぞれのテキストを構成する語数の平均は 3.1 語であった。また、自由再生課題回答に記入されたテキストは 72/248 で全テキストのうち 30%で、平均字数は 5.4 字で

あった。テキストを構成する単語数の平均は 2.9 語であった。自由再生課題で記入されたテキストの平均字数と自由再生課題で記入されなかったテキストの平均字数の間に、有意差は見られなかった( $p>.05$ )。

テキストと理解度にどのような関係性が存在するかを検証するために、3.2.3 節の参加者間の一致語句件数と音声テキスト化の文字数のピアソンの積率相関係数を算出したところ、有意ではないやや負の相関が見られた( $r=-0.31$ ,  $p>0.05$ )。また、一致語句件数と音声テキスト化の回数のピアソンの相関係数を算出したところ、有意な相関は見られなかった( $r=-0.07$ ,  $p>0.05$ )。従ってテキスト化の量が多ければ理解が向上するわけではないと考えられる。従来研究 (Pan et al. 2010) で音声認識技術によるテキスト化よりも、テキストの重要部分に強調表示を付与した方が理解と快適性が向上した。本結果は、従来研究を支持する結果を示し、音声のテキスト化は単にテキスト化する量が多いほど良いわけではなく、音声の重要部分がテキスト化され理解性を高めることが理解に効果があると考えられる。

次に、テキストと対応する発話を会話的特徴から 6 種類の名詞句に分類した。表 13 に 6 種類の名詞句の定義と例を示す。6 種類の名詞句とは、Clark ら (Clark and Wilkes-Gibbs 1986) が対象指示コミュニケーションにおいて名詞句の意図、働き、理解に効果的な表現を会話参加者同士の話し方にに基づき発話まとまりを 6 種類の名詞句として分類し定義したものである。6 種類の名詞句の会話的特徴を以下に述べる。

一括提示は、話者が最も簡潔にポイントを示す名詞句のまとまりで、一気に発話され、正確で標準的な早く短い発話である。補足は、2 つ以上のまとまりで発話され、不十分な発話をもう一度補足したり説明を加えたりして拡張する。分割提示は 2 つ以上のまとまりで発話される。間にポーズがあり相手の返事を挟み相手の理解を確認し、発話を拡張する。補足と分割提示は話者が前の内容を補足・拡張し相手の理解を補う名詞節であるため、非標準型の構成に分

表 13 6 種類の名詞句の定義と例

6 種類の名詞句	会話的特徴	音声とテキスト条件における会話例
一括提示	話者が最も簡潔にポイントを示す名詞句のまとまりで、一気に発話され、正確で標準的で早く短い	私は原子力発電に反対しています
補足	2 つ以上のまとまりで発話され、不十分な発話をもう一度補足したり説明を加えたりして拡張する	犯人を：え：死刑に(.)したとしても：その：被害者は戻らないという(.)ことがあって：その(.)犯人も：いくら死刑にしても：お(.)え：まあそれによって何かが変わるわけではないかなあと
分割提示	2 つ以上のまとまりで発話され、間にポーズがあり相手の返事を挟み相手の理解を確認しながら発話を拡張する	NS：え：と死刑判決 NNS:うん NS:が出た後に： NNS:うん NS:その：やっぱり間違いでしたと
非明示	不十分な情報の発話の後に後続発話がある	ですけどあの：そうですね安全性もあるんですけど：そこまで地震が起きた時にこの：地震ですよ
理解不能	それだけでは意味をなさない	その：じ(.)す(.)
先取り	発話の間に十分な間と相手の確信があり、相手にこれから言うべきことを先取りされる	NNS：その：(1)ん：原子力は(.)発電：発電の周りは：(2)ん；あの：ほうし(.)あの： NS:放射能？ NNS:放射性の：

類される。非明示は不十分な情報の発話の後に後続発話がある。理解不能はそれだけでは意味をなさない発話を指す。先取りは発話の間に十分な間と相手の確信があり、相手にこれから言うべきことを先取りされる。理解不能、先取りは不適切な情報や現実にあまり見られない名詞節まとまりに当たる(Clark and Wilkes-Gibbs 1986)。

図 13 にテキストと対応する発話を示し、対応する発話の会話的特徴からテキストを 6 種類の名詞句に分類する方法を、表 13 を参照しながら以下に述べる。図 13 の 18 行目のタイピングによるテキスト“刑罰”は図 13 の 17 行目の発話に対応づけられる。“法律を守らない人(.)に対する処罰？刑罰？”は一気に発話され正確で標準的で早く短い発話であるという特徴から、18 行目のテキスト“刑罰”は会話的特徴から一括提示に分類される。図 13 の 11 行目のテキスト“心理的な障害を持っている”と図 13 の 12 行目のテキスト“ケア”は図 13 の 10 行目に対応づけられる。9,10 行目は 2 つの発話のまとまりで発話されたため、テキスト“ケア”と“心理的な障害を持っている”は 2 つの発話のまとまりで発話されるという会話的特徴から補足に分類される。図 13 の 8 行目のテキスト“幼少期”は図 13 の 7 行目で“なんていう”が先に発話されたため非明示に分類される。

	A	B	C	D	
1	ティア	開始時間	終了時間	間隔	内容
2	NS	00:01.2	00:01.4	00:00.2	あ
3	NS	00:03.4	00:08.2	00:04.8	えとじゃあ先に私が()反対の理由()としては:
4	NS	00:10.0	00:18.1	00:08.2	えっと:日本の今の:若い人()たちが()死刑になるような: 重大な犯罪を犯す時
5	テキスト	00:18.1	00:22.1	00:03.9	日本の若者
6	NS	00:22.0	00:23.5	00:01.5	っていうのは:
7	NS	00:24.5	00:26.1	00:01.6	なんていう:幼少期
8	テキスト	00:26.1	00:31.0	00:04.9	幼少期
9	NS	00:29.0	00:37.4	00:08.4	に()寂しい思いをしていたりとか:あとは青年期に愛を受け 取れないって言う()データが出ていて
10	NS	00:39.4	00:49.2	00:09.8	心理面の:ケアが必要なことが多くて:( )死刑よりも:そういう ケア()のほうに目を向けたほうが いいかなって思うからです
11	テキスト	00:49.1	00:59.5	00:10.4	心理的な障害を持っている
12	テキスト	01:00.8	01:02.4	00:01.6	ケア
13	NNS	01:36.0	01:43.5	00:07.5	しかし:えと:心理:じょうたい障害は: あの:ん:病気なので
14	NNS	01:45.0	01:45.4	00:00.4	え:
15	NNS	01:49.0	02:00.8	00:11.9	その一回の:犯罪は:え:死刑が: 必要です() また:え:と 法律ですが:え:と
16	NNS	02:01.8	02:17.3	00:15.6	)法律を:守るべきとを:守るため: ( )え:と() 法律い() え: 法律を違反する人を:まも() まもることが 必要ない と思います
17	NS	02:22.8	02:28.6	00:05.8	法律を守らない人() に対する処罰?刑罰?
18	テキスト	02:29.1	02:36.4	00:07.3	刑罰
19	NNS	02:29.1	02:29.4	00:00.3	刑罰です
20	NS	02:29.3	02:30.5	00:01.2	刑罰

図 13 テキストと対応する発話を会話的特徴から 6 種類の名詞句に分類する方法

テキストと発話を対応づけ、対応する発話の特徴からテキストを 6 種類の名詞句に分類する。8 行目のテキスト“幼少期”は 7 行目の“なんていう:幼少期”と対応づけられ、7 行目は不要な語が先に発話されたため 8 行目のテキスト“幼少期”は非明示に分類される。17 行目は一気に発話され正確で標準的で早く短い発話のため、18 行目の“刑罰”は一括提示に分類する。

音声とテキスト条件のテキストを 2 名が独立に分類したところ、2 名の一致度は高かった( $\kappa=0.75$ )。図 13 の方法でテキストと対応づけられた発話を表 13 のカテゴリに分類した結果を表 14 に示す。表 11 はテキストの発話的特徴が一括提示よりも補足が多く、その他の不適切な発話は少ないことを示す。すなわち、テキストは発話内容を補足・拡張する名詞句であることが多いといえる。

表 14 6 種類の名詞句分類結果

カテゴリ	発話形式	計	平均	割合
標準形	一括提示	68	4.3	28%
非標準形	補足	103	8.4	55%
	分割提示	32		
その他	非明示	26	1.6	15%
	理解不能	4	0.3	
	先取り	8	0.5	
	その他	7	0.4	2%
合計	全て	248	15.4	100%

音声とテキストによって音声情報からだけでは分からない部分をテキストの視覚的補助によって NNS の理解を補ったことを示す。

従来研究で音声のテキスト化によって話し方が変化することが示されているが、どのように変化するかは明らかでない。そのため、音声とテキスト条件と音声のみ条件の会話を書き起こし、会話の書き起こし上で名詞句を数え 6 種類に分類することで話し方を比較した。表 15 に 6 種類の話し方の分類を示す。音声とテキスト条件の名詞句は表 15 に基づいて音声のみ条件と同じようにテキスト化の有無に関係なくすべての名詞句を数えた。

音声とテキスト条件の会話の書き起こし上で表 15 の 6 種類の名詞句を数えた結果、261 個であった。音声のみ条件の会話の書き起こし上で、6 種類の名詞句は 250 個であった。音声とテキスト条件と音声のみ条件の名詞句の数に大きな差がなかったため、テキスト化によって発話が増えるわけではないと考えられる。

表 15 6 種類の話し方の分類

分類	音声とテキスト条件	音声のみ条件
一括提示	日本以外のほとんどの国で死刑制度は廃止されています。	ほとんど中東から石油を輸入しなきゃいけない
補足	犯人を：え：死刑に(.)したとしても：その：被害者は戻らないという(.)ことがあって：その(.)犯人も：いくら死刑にしても：お(.)え：まあそれによって何かが変わるわけではないかなあと	技術：が発展してなくて：あの(.)安全性がもししっかりとちよつとあの：確保できていない(.)と思うんですけれども：その安全性をしっかりと確保できるまでは(2)ん：極力しないほうが良い
分割提示	NS：え：と死刑判決 NNS:うん NS:が出た後に： NNS:うん NS:その：やっぱり間違いでしたと	NNS:日本は地震が起こりやすい国 NS:うん NNS:とされています(.)地震の時原子力はひかえることは難しく NS:うん
非明示	ですけどあの：そうですね安全性もあるんですけど：そこまで地震が起きた時にこの：地震ですよね	え：人は：え：なんですか(3)え人類は：あ：エネルギー(.)エネルギーの：あ：え：なに要求は：そんなに：多くない
理解不能	その：じ(.)す(.)	あ：すき(.)あ：自然 (1) ん：あ：忘れた (10)
先取り	NS：その：(1)ん：原子力は(.)発電：発電の周りは：(2)ん；あの：ほうし(.)あの： NS:放射能？ NNS:放射性の：	NNS：ん：その人は(.)とりかえして：かえしう：あ：とりかえしが： NS：つかないですか？ NNS：つかない

表 15 の分類に基づき 2 名が独立に分類したところ、2 名の一致度は高かった( $\kappa=.65$ )。音声とテキスト条件において一括提示は 3.5 件/組、補足は 8.8 件/組、分割提示は 2.1 件/組、非明示は 1.2 件/組、理解不能は 0.3 件/組、先取りは 0.1 件/組であった。表 15 でその他である非明示、理解不能、先取りはほとんど見られなかった。図 14 左に音声とテキスト条件における 6 種類の名詞句の出現回数を示す。

音声とテキスト条件の話し方を確認するため、音声とテキスト条件における 6 種類の名詞句の出現回数が正規性のないデータであることを確認し、6 種類の名詞句の出現回数に対してフリードマン検定を行ったところ有意差が見られた( $\chi^2[5]=57.5, p<.001$ )。多重比較(ボンフェローニ修正後のウィルコクソン検定)の結果、一括提示と補足の間で有意差が見られ( $p<.01$ )、補足と分割提示の間で有意な差があった( $p<.001$ )。同様に補足と非明示などのその他の間にも有意差が見られた( $p<.001$ )。したがって音声とテキスト条件では補足が多く行われる話し方であることを示した。

同様に、表 15 の分類に基づき 2 名が独立に分類した平均値をとったところ、音声のみ条件において一括提示は 11.件/組、補足は 1.7 件/組、分割提示は 1.4 件/組、非明示は 0.5 件/組、理解不能は 0.2 件/組、先取りは 0.2 件/組であった。表 15 でその他である非明示、理解不能、先取りはほとんど見られなかった。図 14 右に音声のみ条件における 6 種類の名詞句の出現回数を示す。

音声のみ条件の話し方を確認するため、音声のみ条件における 6 種類の名詞句の出現回数が正規性のないデータであることを確認し、フリードマン検定を行ったところ有意差が見られた( $\chi^2[5]=55.8, p<.001$ )。多重比較(ボンフェローニ修正後のウィルコクソン検定)の結果、一括提示と補足の間で有意差が見られ( $p<.01$ )、一括提示と分割提示の間で有意な差があった( $p<.001$ )。同様に、一括提示と非明示、理解不能、先取りなどの間にも有意差が見られた( $p<.001$ )。し

たがって音声のみ条件の話し方は一括提示が多く、簡潔で一気に話す、早く早い話し方が多いと言える。

3章で示したように、NSとNNSが共同解釈を構築する会話の件数は音声とテキスト条件( $M=6.4$ ,  $SD=4.1$ )で音声のみ条件( $M=2.6$ ,  $SD=2.1$ )より有意に多かった。音声のみ条件で共同解釈を構築できない理由は、音声のみ条件の話し方が不明瞭な場合や、早すぎる場合があったことが、この研究結果から裏付けられた。音声とテキスト条件では、分からない部分を補足する会話音声とテキスト化し視覚情報で補助することによって共通理解が補われたと考えられる。結果から音声テキスト化によって話し方は一括提示を減らし補足が多い話し方に変化したといえる(宋 2018)。

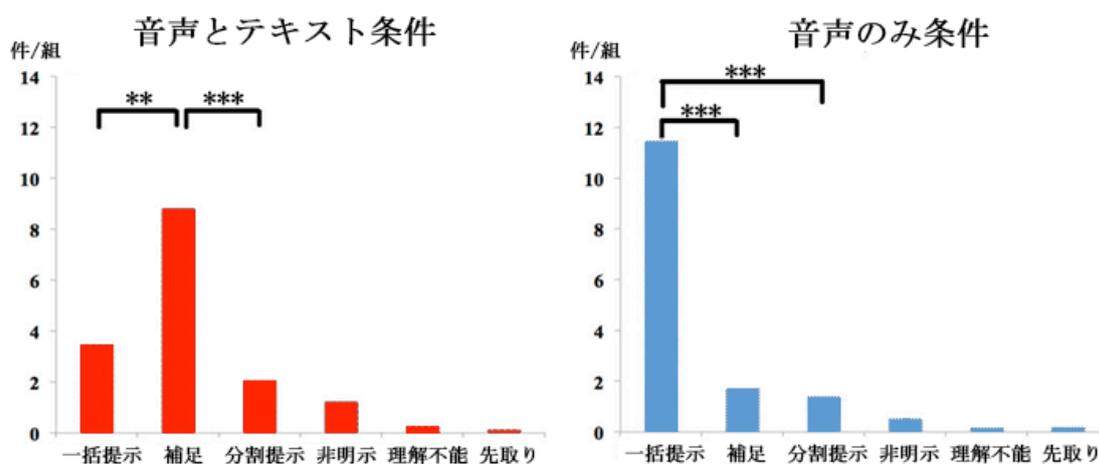


図 14 6 種類の話し方の出現回数(件/組)

(\*\*: $p<.01$ , \*\*\*: $p<.001$ )

## 4.2 音声会議における音声テキスト化の分析の検討

### 4.2.1 テキスト化による遅延

図 11 は従来研究(宋 2018)で示されたように、音声とテキストの時間差を示した。NS が音声をテキスト化する時間差の平均は発話より約 3 秒遅れることが多かった(Hanawa et al. 2017)。この結果から、従来研究ではテキスト提示のタイミングと発話のずれについて検討を行った(宋 2018)。本研究では、テキスト音声会話がどのように共通理解に影響するか検討するため、音声テキスト化の速度を NS と NNS に分け、従来研究と比較して詳しい検討を行った。

発話のテキスト化は時間的にあまり大きな遅れはなくタイピングが発話より早い場合もあった。実験参加者はほとんどがブラインドタッチではなくキーボードを見ながらタイピングしていた。発話をテキスト化する会話は個人のタイピングレベルが影響しないというテキストチャットの従来研究(Walther 2007)と同様の結果となった。従来研究でタイピストがタイピングするテキストの遅延は 5 秒以内であれば参加支援になるとされている(Lasecki 2014)。音声の自動テキスト化は遅延が少ないほど良いとされており、本研究においてもテキスト化の遅延に積極的な効果はないと考えられる。本研究の音声テキスト化は 5 秒以内の水準を満たすことを示し、NS と NNS の音声会議を支援できたことから、従来研究を支持する結果となった。

図 12 は NS が NS の発話をテキスト化する場合の遅延と、NNS の発話をテキスト化する場合の遅延に分けてタイミングを示した。NS が NNS の発話をテキスト化する場合は必ず発話より遅れることと、大部分は NNS の発話から約 2 秒から 10 秒遅れることを示した。

一方、図 12 の NS が NS の発話をテキスト化する場合のタイミングは、NS の発話開始の 3.5 秒前から開始される場合があった。NS による発話内容のテ

キスト化は 0-1 秒を最頻値として、発話より早い場合と、8 秒程度遅いタイミングで行われる場合が分布していた。本結果から会話参加者によるテキスト化は他の参加者の発話をテキスト化することも可能だが、本人の発話の方が早くテキスト化できることが示された。

NS による音声テキスト化により会話中に「ポーズ(音声の停止時間)」が生じる場合がある。本研究において自然に発生する微小なポーズは、それぞれ長さが異なり、音声テキスト化の時間内に含まれる場合もあればそうでない場合もある。ポーズに関する従来研究からは、ポーズの長さによって話者交代、理解や記憶への異なる影響が報告されている。微小なポーズを挿入して NNS の会話を支援した研究(Hautasaari and Yamashita 2014)によれば、0.2 秒の短いポーズは NNS の支援となるが、それ以上の長さは話者交代を困難にし NS と NNS の不満を増大させる。また、会話中に数秒のポーズを挿入しテキストの効果と比較した研究(Inoue et. al. 2015)では、話者交代や発話量においてポーズの効果はテキストに及ばないと述べている。また、音声の再生後に音声と同じ長さのポーズの影響を調査した平井(2000)の研究によれば、内容が次に進む前にとる短いポーズと、テキストを見て音韻化する「リハーサル」には、NNS の負担軽減、単語の記憶を保持する効果があるとされる。本研究で自然に発生した長さの異なるポーズが参加者に与える影響を判断することは困難であるが、テキスト音声会話の共通理解が向上する原因の一つに、NS が会話中に音声をテキスト化する間にポーズが発生することや、NNS が会話中にテキストを音韻化することが考えられる。

#### 4.2.2 テキストの量と内容

会話中のテキスト化は一回あたり 5 字程度で、回数は平均 2.2 回/分であった。発話と比較して、テキスト内容に誤字は見られなかった。また、NS がタイピ

ングしたテキストは、会話中に読める量であると考えられる。音声聞きながらテキストを読むことは黙読とは異なり、読み手が話し手の予測できない速度とリズムで読むことを強制され、早すぎて読みきれない場合がある(平井 2000)。この点において、本研究のテキストは発話と連動しており、NNS が読める量のため音声とテキストの会話支援になると考えられる。

また、自由再生課題で記入された会話後の記憶に残らなかったテキストは176/248 で全テキストの 70%で、会話後の記憶に残ったテキストは全テキストの 30%であった。テキストの平均字数と自由再生課題で記入されなかったテキストの平均字数の間に、有意差は見られなかったため、テキストが会話後の記憶に残るかはテキストの長さによらないと考えられる。

表 14 では、NS はどのような会話内容をテキスト化したかを調べるため、テキストに関連する発話を 6 種類のカテゴリに分類した。従来研究(Hanawa et al. 2017)では、テキスト内容の概要を示し、テキスト化された内容は補足が最も多い結果を示した。本研究では、日本語会話の文献を参照し 6 種類の名詞句について詳細な検討を行い、NS の発話が不十分な部分や音声だけでは意味が間違われる可能性のある部分がテキスト化されたことを示した。

図 14 では話し方の分類結果を示し、音声とテキスト条件では補足が多いことを示した(宋 2018)。Hanawa ら(2017)はテキスト音声会話中の話し方の変化について調査するため初めて会話分析を行った。その後、テキスト音声会話中の発話を分析する観点から、NS と NNS に分けた分析が行われ、NS と NNS の話し方から一括提示が減少することが示された(宋 2018)。本研究ではこれらの従来研究を踏まえ、テキスト音声会話の共通理解に有効である理由を調べる目的で会話分析方法と検定方法を見直し、話し方の変化を再検討した結果を示した。

本結果は、第 3 章で NNS の了解性と快適性が向上したことと整合する。従来研究(Gao et al. 2014)では、自動音声認識後のテキストに NS が編集を加え

ることで NS の話し方が明瞭になり、NNS の了解性と快適性に正の相関が示された。本研究の補足は、話者が不十分だと判断した発話に自身による説明を加え、節を分けて発話を修復する。補足を加えることで、話し手と聞き手の両方が誤解や不適切な判断に即座に対応することができる。そのため補足は一括提示よりも即応性と双方向性が高いと考えられる。NNS との会話は NS 同士の会話とは異なり NNS が対話中に聴解できない場合や、音声情報のみでは意味理解に到達できない場合がある。NS 同士は簡便な話し方を好むが、NNS との会話では理解を促す補足の多い話し方は有益と考えられる。また、第 3 章で共通理解が向上したことから、補足の多い話し方は NS にとっても好都合と考えられる。

音声のみ条件で一括提示が多い理由は、一般的に最もよく使われ、話者の最も好む方法であるためと考えられる。一括提示において話し手は、名詞句を 1 つの節を 1 回しか使用せずに短く早く言いたいことを発話し、補足説明を伴わず次の話に移行する。一括提示を行う話し手は、聞き手が発話内容を当然のように正確に理解して受理すると思い込んでおり、聞き手は 1 回聞くだけで即座に正確に理解することが求められる。

音声とテキスト条件と比較すると、音声とテキスト条件では音声はテキスト化され、また補足が多い話し方が行われる。音声とテキスト条件では参加者は発話を修復し、聞き手の理解に応じて話し方を変えやすいと考えられる。会話中に理解できていないと気づいたら話し方を変えるか、発話修復などの補足・拡張を用いることで共通理解が達成される。

音声のみ条件では話者の最も好む方法である一括提示が多用されるため、参加者は聞く・話す行為を短時間で処理しなければならず、NNS の負担は大きくなり、正確な理解ができなくなると考えられる。NNS は理解に時間がかかったり、2,3 回繰り返さないと確実に理解できなかったりすることはよく起こる。また、NS と NNS の通常の音声会話では NS ばかりが話しがちになるこ

とが知られている。音声のみ条件で一括提示が多いことは NS と NNS の共通理解が十分に果たされない理由の一つと考えられる。

テキストチャットを含むウェブ会議システムにおいて第二言語習得が促進されることが従来研究(Wang 2004)(Li and Rosson 2012)に示されていることから、本研究で音声会話中の音声テキスト化が NNS の第二言語習得に役立つと考えられる。補足が多い会話ではエピソードや文脈の広がりを利用した理解や記憶の可能性が示唆される。心的辞書とは体験を通じて得る語の定義を指す(Fitzpatrick and Barfield 2009)。NNS が NS から複数の言葉の使い方や意味などの心的辞書を獲得し第二言語習得に有益である可能性が示唆される。

#### **4.2.3 音声会話のテキスト化についての総合的な検討**

テキスト化のタイミングは NS 発話から平均して約 2 秒後に行われ、内容は発話を補足・拡張する内容を 5 字程度であった。発話内容のテキスト化は 1 分間に約 2 回程度行われた。また、参加者の話し方は、テキスト化によって早く短い方法から、不十分な説明を補足によって補う方法へと変化した。

音声とテキストによる会話支援を時系列でみると、NS が会話中に発話した内容を遅れてテキスト化し、NNS は音声とテキストによって音声会話の内容を確認し、NS と NNS が補足説明によって理解上の問題に対処しながら会話が続くというインタラクションであるといえる。そのため、音声会話の話し方が音声とテキストの会話支援によって変化することは、共通理解の達成に大きな貢献をしているといえる。

一方、通常の NS と NNS の音声会話ではテキストがなく、NS が意識的に調整を加えて NNS 向けに話す状況がほとんどである。従来研究(筒井 2008)によれば、NS が適切かつ簡略な言語調整を加えて NNS 向けに話す言語方略は

「フォーリナー・トーク」と呼ばれる。筒井(2008)の調査によれば、日本語教師群が NNS に向けて行う「ティーチャー・トーク」では具体例が豊富で非常にわかりやすいのに対し、一般の日本人 NS が NNS に行う「フォーリナー・トーク」はゆっくりでも語彙・文構造の改変が少ないためさほど理解に繋がらないことが示されている。本研究は、音声テキスト化によって NS が意識することなく語彙・文構造の面から話し方を NNS の理解に応じて変化させる会話支援ツールであることを分析によって示し、テキスト音声会話の会話支援ツールとしての有効性を示した。

本研究のテキスト音声会話では、NS のテキスト化によって共通理解が向上するだけでなく、話し方が変化するため NNS の理解力が最大に発揮されることを示した。本結果は NNS の了解性と快適性が向上することとも整合性があり、テキスト化のタイミングと話し方の変化には関連があると考えられる。音声テキスト化のタイミングは早い方が好ましいことが従来研究で示されている(Cao, X., et al. 2018)。したがって、NS の音声はテキスト化が早く NS の話し方に影響するため、共通理解に貢献すると考えられる。

本研究の音声とテキストが有効である条件は、NNS が日本語をある程度理解できる場合であると考えられる。本研究において N1 程度の NNS を対象とした実験を行い音声とテキストの有効性を示した。NNS が N2 程度の場合を対象として、音声とテキストの有効性を検証する実験を行った研究(宋 2018)においては、音声のみの会話では NS の発話が多い傾向があるが、音声とテキスト会話において NS と NNS の発話時間に差が見られなかった。結果として、N2 程度の場合においても音声とテキストによる NNS の会話支援が有効であることが示された(宋 2018)。NNS の日本語能力が極めて高い場合には、NS 同士の会話同様に、音声のみによる NS と NNS の共通理解の形成が可能であると考えられる。したがって、音声とテキストは日本語をある程度理解できる NNS の参加を前提として有効であると言える。

本研究は音声会議において発話内容の補足をテキスト化することが、共通理解を支援することを示した。従来の協調課題や問題解決会話の研究が多数行われており、会話に相互作用が活用されている。会話に参加しコミュニケーションを学ぶことは、インプットやアウトプットのみでなく会話上で使用しながら学ぶことが必要である(Swain 2000)。会話内容の理解力と記憶力には個人差があるため、本研究のテキスト化は多様な言語使用の機会を体験する会話支援として理解や記憶を促進できると考えられる。本研究によって音声のテキスト化が NS と NNS の会話支援となることを示した。

#### **4.2.4 音声テキスト化の分析の限界**

一般的にテキストチャットの研究はテキストのみの定量的な分析が行われる。本研究のテキスト内容の分析は発話と相互依存的な関係があり、発話とテキストを分けて分析することは困難である。テキスト内容の分析と議論を含めて、分析があいまいさを常に内包していることは否定できない。また、本研究のサンプル数の少なさと、テキスト化は参加者ごとに異なることにより、テキスト化を一般化することは容易ではない。

本研究では音声とテキストによる会話支援のテキスト化のタイミングと内容に実証的な分析を行った。従来研究では発話とテキストのタイミングと内容の分析が行われていない。発話の後にテキストがあり、また発話があるという事象を時系列でモデル化することが難しいため、今後はこれらの時系列の変化を示す分析モデルが求められる。

### **4.3 4章のまとめ**

これまで音声会議におけるテキストの利用によってコミュニケーションに積極的な効果や影響があることが従来研究によって示されている。本研究では NS がテキスト化するタイミングと内容を発話との関わりから調べた。結果から、NS と NNS の音声会議においてテキスト化は発話から平均 2-3 秒遅れて行われた。テキスト内容は 5 字程度で補足する内容の名詞句が多かった。

本研究ではテキスト音声会話によって、音声のすぐ後のタイミングで音声の補足がテキスト化されることを示した。また、NS のテキスト化によって NS と NNS の話し方が変化し、テキストで補足された内容が NNS の理解を補い快適性が高まることを示した。従って、共通理解が向上した理由は、NS からの話を 2-3 秒の間に音声とテキストで繰り返して NNS が確認することができ、確実な理解と快適性に繋がったためと考えられる。

## 第5章 総合考察

本研究は NNS を含む音声会議の実験を行い、NS が発話内容をテキスト化することが会話に与える影響について検討している。本章では、各章の総括をしたのち、本研究の知見をまとめ、最後に音声会議におけるテキスト利用の今後の研究展開を検討する。

### 5.1 本研究の成果

本章ではまず本研究を整理する。次に、本研究で検討した NS が音声を書き起こすテキストの有効性と、音声とテキストが共通理解にどのように影響するかを明らかにするための 3 つの仮説を検証した成果について述べる。次に、本論文で検証した研究成果について考察する。

#### 5.1.1 本研究の整理

1 章では、現代社会における NNS とのコミュニケーションの現状と、本研究を始める動機を述べた。具体的には、日本において在留外国人が増加し、特に日本人と中国人の日本語によるコミュニケーション機会が増加していることを述べた。また、NNS との会話支援で自動テキスト化技術を使用するにあたって問題とされていた自由度の高い会話に対応できない点について、会話上の役割と正解、NS 同士のコミュニケーションとの違いについてまとめた。

そのうえで、本研究の目的である音声とテキストによる会話支援の詳細を調査する意義について述べた。NS と NNS の音声会議において発話内容をテキスト化する支援の有効性と、音声テキスト化のタイミングと内容を調査する 2

つの分析を行うために、2つの分析が共通する1つの実験を行うことを述べた。また、NSとNNSが共通理解を達成するためには共同解釈の構築が欠かせない点について詳しく述べた。

2章では、2000年はじめごろから開始された情報学分野のNNSとのコミュニケーション支援について、5年ごとに区切り研究の動向をさらに詳しくまとめた。その中で、実験で行われる会話が自由度の低い課題から徐々に自由度の高い課題へと変遷していくことについて、課題の詳細に触れながらまとめた。

さらに、従来研究の会話とコミュニケーション・メディアの最適な組み合わせの仮説を立てた従来研究と、それを検証した従来研究を取り上げ、すでに行われたメディアの組み合わせを具体的な研究事例をあげながらまとめた。最後に、これらの従来研究を踏まえ本研究の位置づけについて述べ、従来研究における問題点を明確にした。本研究が示すべき研究への貢献をまとめた。

3章では、NNSとの音声会議において、NSが会話音声の一部分をテキスト化するテキスト音声会話の実験を行い、テキスト音声会話の有効性を調査した。実験の結果から、音声とテキストの利用はNSとNNSの共通理解を支援し、NNSの了解性と快適性が向上することを示した。

4章では、3章で述べた実験における音声のテキスト化のタイミングと内容を、NSとNNSの発話内容との関わりから調べた。調査の結果から、NNSとの音声会議において、NSがタイピングによって発話内容をテキスト化するタイミングは、NS発話から平均2秒遅れることが示された。テキスト化された内容は、平均5字程度で補足する内容であった。また、発話をテキスト化することによって補足説明を行う話し方に変化することが示された。この結果は、3章において共通理解と会話後の一致した記憶が増加することを裏付ける結果となった。

5章では、2章、3章、4章を踏まえ本研究を整理する。本研究の成果として本研究の目的の達成と、本研究の成果の位置づけを述べる。本研究の成果の位

置づけは、共通理解、タイミングと内容、課題別のコミュニケーション、コミュニケーション・メディアの組み合わせのそれぞれについて従来研究と比較して本研究の成果を述べる。5章の最後に今後の研究展開を述べる。

6章では本研究の結論を述べる。

### 5.1.2 本研究の目的の達成

本研究の目的は、NS が音声を書き起こすテキストの有効性を明らかにすることと、音声とテキストが共通理解にどのように影響するかを明らかにすることであった。本研究において NS と NNS の音声会議実験を行い、実験から得たデータを分析した結果から、音声とテキストの有効性に関する 3 つの仮説が支持された。従って、本研究の一つ目の目的である NS が音声を書き起こすテキストの有効性を示す目的を達成したと言える。また、音声とテキストが共通理解にどのように影響するかを検証するため実験データを分析した結果、発話の直後に発話を補足する内容のテキストが表示され、参加者の話し方が補足の多い話し方に変化するため理解が深まることが示された。本結果から、本研究の二つ目の目的である音声とテキストがどのように会話に影響するかが示された。

本研究では NS と NNS の音声会議の実験を行い、音声とテキストの有効性のうち客観的な理解と理解する過程について仮説 1 と 2 で検証した。実験の結果から、仮説 1:「音声とテキストの利用は共同解釈の構築を増加させる」の検証では、音声とテキストを利用する場合に共同解釈を構築する証拠となる会話が多く見られることが実証された。また、仮説 2:「音声とテキストの利用は共通理解を向上させる」の検証では、音声とテキストを利用する場合に参加者間で一致した記憶の件数が増加し、音声とテキストの利用が共通理解の増加に有効であることが実証された。

音声とテキストがどのように共通理解に影響するかを調べるため、NS と NNS の音声会議の実験から得たデータを分析した。分析の結果から、テキスト化は NS 発話から平均 2 秒遅れ、テキストの内容は 5 字程度で補足する内容の名詞句が多いことが示された。また、NS と NNS の発話内容を NS がテキスト化するタイミングと内容を、発話との関わりから調べた結果、音声会議のテキスト化によって補足の多い発話が促進されることが示された。したがって、テキスト音声会話で共通理解が向上する理由は、発話の 2-3 秒後に発話内容の補足説明がテキスト化され、参加者が音声とテキストで内容を確認することで理解と快適性が高まるためであることが示された。

さらに、音声とテキストによって主観的にわかるという感覚が生じると考えられるため、仮説 3 で参加者の主観評価を調査した。結果として仮説 3:「音声とテキストの利用は会話における NNS の了解性と快適性を向上させる」の検証では、音声のみよりも音声とテキストの利用時に質問紙調査で NNS から“相手の話を理解しやすかった”“自分は気楽にできた”など了解性と快適性に対し有意に高い評価が得られた。本結果は音声とテキストが NNS の了解性と快適性を向上することを示した。

## 5.2 本研究の成果の位置付け

本節では本研究の成果を 2 章の従来研究と比較し、本研究の成果の位置付けと、研究分野に対する貢献について述べる。

### 5.2.1 共通理解の達成

本研究では NS と NNS が音声会議において発話内容の一部を NS がテキスト化する実験を実施し、音声とテキストの有効性を調査した。従来研究(Echenique et al. 2014)ではウェブ会議と音声とテキストを比較し、課題の正解率によって共通理解を調査した。それに対し本研究では、共通理解を達成する過程となる共同解釈を構築する会話を会話分析によって示し、会話後の記憶を調査し、一致した語句数の増加によって共通理解の達成を示した。具体的には、共同解釈を構築する会話の基本形式を日本語会話に応用し、共通理解を達成する過程となる会話の形式を示した。また、何も参照せずに会話後の記憶を調査する自由再生課題によって、音声とテキストの共通理解への有効性を示した。

従来研究(Yamashita et al. 2009)(Echenique et al. 2014)(Gao et al. 2014)では、会話の参加者人数と会話上の役割のため、共同解釈を構築する会話の分析は行われていない。共同解釈を構築する会話の原型は英会話で、参加者2名の自由会話から定義されている(Clark and Brennan 1991)。そのため3人以上の会話や役割の決まった会話分析には適用されなかったと考えられる。伝(2009)によれば共同解釈構築(グラウンディング)すなわち基盤化は、情報学分野では広く認められた概念であり、様々な研究に応用されている。本研究は英会話で定義された会話例を日本語に応用して会話分析を行っており、同じ例は他に見当たらない。本研究で示した共同解釈を構築する会話例は、他の研究でも同様に活用できると考えられ、本研究分野に対する貢献となる。

また、本研究では NS と NNS の会話後の記憶の調査によって共通理解の達成を示した。従来研究(Echenique et al. 2014)では、図形同定課題の対象指示コミュニケーション(稲葉 2009)において類似に収束した語句数を数えた。また、従来研究(Pan et al. 2017)では共通点探しの課題で用いられた共通の語句を数えた。つまり、従来研究(Echenique et al. 2014)(Pan et al. 2017)は NS と NNS が共通の語句を用いる課題において共通の語句を数えた。それに対し本研究は自由度の高い音声会議で記憶に残った語句のうち一致した語句を

数えた。自由度の高い会話でも共通理解を調査できることを示したことは、本研究分野に対する貢献となる。

従来研究から導かれた仮説 1、仮説 2、仮説 3 は、共通理解に関連する仮説であった。山鳥(2021)によれば、理解とは個人の持つ記憶(メンタル・イメージ)を重ね合わせることであり、たとえ話による置き換えやイメージ増大によってイメージが一つにまとまり正確に理解でき、「わかった」という感情が働き快感が生じる(山鳥 2021)。つまり、理解には客観的なテストで調査する「理解」と感情としての「理解」があると考えられる。本研究の仮説 1: 共同構築の解釈では一貫した内容での情報交換の中で情報を置き換えたり増やしたりしてははっきりさせ、仮説 2: 共通理解では正確にわかったことを自分の言葉で表現し、仮説 3: 了解性と快適性に「わかった」心の働きが表現されたと考えられる。Gao ら(2014)の従来研究では了解性の主観的調査は行われず、Pan ら(2017)の従来研究では理解の過程は調査されていない。本研究は音声とテキストの有効性として共通理解を 3 つの視点から考察した結果を述べ共通理解の達成を示した点で、本研究分野への貢献となる。

本研究は音声とテキストを音声のみと比較し、自由度の高い会話における NS と NNS の理解度が音声とテキストの利用時に高まることを実験結果から示した。NS 同士の自由度の高い会話を検討した従来研究では、映像音声ありとテキストのみを比較し、映像音声ありのコミュニケーションの所要時間が 206 分、テキストのみが 204 分で内容には差がなかったことを示した(倉本 et al. 1998)。従って、NS 同士の会話ではコミュニケーション・メディアが異なっても共通理解が形成できることがわかる。本研究は NS と NNS の 7 分間の会話を行い、音声とテキスト (3.3 件)と音声のみ(1.8 件)の共通理解に有意な差があることを示した。従って、NS 同士とは異なり、NNS を含む会話ではコミュニケーション・メディアが異なると共通理解の形成が困難になることを示した。

### 5.2.2 タイミングと内容

本研究では NS が発話内容をテキスト化するタイミングと内容を発話との関わりから調査した。分析結果から、NS と NNS の音声会議において NS がテキスト化するタイミングは、NS 発話から 2 秒遅れて行われ、内容は 2-3 語の補足説明を 5 字程度でテキスト化することが示された。従来研究(Lasecki et al. 2014)では、会話参加を助けるテキストの水準として遅延 5 秒以内、精度 95% 以上とした。本研究結果から、発話内容の補足のテキスト化は 5 秒以内の水準を満たし、理解と快適性が高まり、共通理解が向上することが示された。1 章で述べた NS が発話内容をテキスト化する CMC で、NNS の理解力を引き出すことを示した。

本結果は、数ある支援方法の一つとしてテキスト音声会話を用いた会話支援ツールのデザインへの言及と、本研究分野の会話支援ツールの評価に活用することができる。本研究のテキスト音声会話を自動化する会話支援ツールでは、会話音声を認識し、会話の補足説明となる言い換え、具体例、意味定義などを 5 字程度で 5 秒以内にテキスト表示させるデザインが考えられる。会話支援ツールの自動化によって共通理解が向上するかどうかは、別途検討が必要である。

また、会話支援ツールの評価では、快適性と理解の向上が期待できるかについての評価を根拠とともに示すことができると考えられる。本研究分野における今後の会話支援ツールに本研究結果を活用することで、結果予測と、根拠を提示できると考えられる。

さらに、本研究で分析した NS がテキスト化するタイミングと内容は、音声認識後に NS が編集するタイミングと内容を類推する手助けとなる。従来研究は、音声を全てテキスト化するタイミングと内容を調査し、音声認識技術によるテキスト化(Pan et al. 2010)、タイピストによるテキスト化(Lasecki et al. 2014)、クラウドワーカによるテキスト化、音声認識後のタイピストによる修

正、音声認識後のクラウドワーカによる修正(Gaur et al. 2016)などを比較調査した。多人数音声会議の自動音声認識速度は 6.8 秒(Zuchowski et al. 2022)であるため、NS と NNS の音声を認識した後に NS が編集をする従来研究(Gao et al. 2014)の速度は 6.8 秒以上になると考えられる。また、NS によるテキスト化に誤りはほぼないが、音声認識後に NS が編集したテキストは誤りを含むと考えられる。音声認識後の NS によるテキスト編集は、会話参加を助ける程度の速度と精度の水準以下であると推察されることから、発話を助ける支援にはならないと考えられる。以上のように、NS がテキスト化する詳細なタイミングと内容を調査したことは、本研究分野に対する貢献となる。

### 5.2.3 課題別のコミュニケーション

本研究では、NS がテキスト編集を行う従来研究(Gao et al. 2014)で扱われたストーリーテリングと同じ状況とするため、自由度が高く参加者の文化・習慣などの影響がなく、よく知られている課題である非公式のディベートを取り上げた。NS と NNS のコミュニケーション研究において、課題によってコミュニケーションは異なるため、課題の選択は重要である。従来研究では自由度の低い課題から高いものへと変遷し、了解性の分析方法は正解した得点を調べることから会話を通じた相互理解や NNS の発話を調べる分析方法に遷移した。従来研究で扱われた自由度の高い課題はまだ少数のため、本研究で自由度の高い課題としてディベート課題を提供したことは本研究分野への貢献となる。

従来研究でよく扱われた自由の低い課題には、図形同定課題(Yamashita et al. 2009)(Echenique et al. 2014)、生き残り課題(Yamashita et al. 2013)(Hautasaari and Yamashita 2014)、地図課題(Diamant et al. 2008)(Gao et al. 2015)などがある。音声とテキストの効果をビデオ会話と比較した Echenique ら(2014)の研究では、NS が指示役で NNS が追従役であったが、

本研究では音声会議と比較し、自由度の高い課題で音声とテキストの有効性を示した。

自動テキスト化技術が自由度の高い会話に対応できない問題があるため(近藤 et al. 2017)、自由度の高い課題では誤りを含むテキストを NS が編集する支援が活用された。従来研究で扱われた数少ない自由度の高い課題には、政策決定課題(He et al. 2017)(Duan et al. 2019)、ブレインストーミング (Gao et al. 2013)、ストーリーテリング (Gao et al. 2014)、共通点探し (Pan et al. 2017) がある。NS がテキスト編集を行う従来研究(Gao et al. 2014)のストーリーテリング課題では、NS と NNS でストーリーテリングの話す順番があらかじめ決められている。また、相手の理解を確かめず会話が進むため、一般的な会話のような相互の情報交換とは言えないなどの欠点もある。

本研究のディベートは、準備時間が必要であることや、訓練が必要であることが問題である。そのため、誰でも参加できるような資料を提示すればコンピュータシステムとも相性の良い課題である(大林 et al. 2002)。本研究で NS と NNS のコミュニケーションで自由度の高い課題にディベートを加えたことは、本研究分野に対する貢献となる。

#### **5.2.4 コミュニケーション・メディアの組み合わせ**

本研究は NS と NNS が言語的に不均衡な状態から、CMC によってバランスを取る支援を行い、本来のコミュニケーションを失うことなく会話支援できることを実証した。具体的には、Inoue ら(2015)の従来研究で述べられたように、NS がテキスト化する作業負担と比べても会話支援として得られるものが上回るため作業する意義があるという主張の根拠を示したことが、本研究分野への貢献となる。

音声テキスト化のメリットとして、NS と NNS が共通理解できること、NNS の了解性と快適性が増すこと、発話内容がわかりやすくなることが挙げられる。テキスト化のデメリットとしては NS の作業負担がある。一方でテキスト化しない音声のみの会話の場合、テキスト化の負担はないことがメリットだが、デメリットとして共通理解が困難になること、NNS の負担が大きく、話し方がわかりにくいことがデメリットとなる。上記から、NS が簡単な作業を行うデメリットよりもメリットが大きいことから、音声テキスト化の価値を示した。

また、本研究は音声とテキストの組み合わせが音声のみの会話よりも有益であることを示した。従来研究の多くは、機械翻訳と音声認識を利用して、NS と NNS の会話音声の全文を書き起こしてテキストを追加した(Gao et al. 2014) (Pan et al. 2017) (Chen et al. 2018)。特に、音声の重要部分に NS が強調表示を追加した従来研究(Pan et al. 2017)は、会話の全文と重要部分によって了解性と快適性を向上させた。本研究は、会話音声の一部のみをテキスト化することで共通理解を向上させた。音声とテキストの組み合わせにより、テキストで部分的に表示する有効性を示したことが本研究分野への貢献となる。

さらに、ビデオとテキストの従来研究は多いが、音声会議に関する研究は少ない。コミュニケーション・メディアは対面、ビデオ、音声の順に情報量が少ないため、NS 同士のビデオ会議や音声会議ではチャット、プレゼンテーション、文書、画面の共有機能などが併用される(Ding et al. 2007)。本研究は数少ない音声会議を扱い、音声とテキストの有効性を示した点が、本研究分野に対する貢献となる。近年広く行われているウェブ会議で映像をオフにして、音声のみで行う会議は音声会議の一つと考えられる。テレワークの普及によって NNS を含む音声会議の機会が増加し、音声とテキストの有効性が示される場面は今後増加すると予想される。本研究で示したテキスト音声会話の有効性は、NNS とのオンラインコミュニケーションで役立つと考えられる。

### 5.3 本研究の限界

本研究は NS が発話の一部をテキスト化するタイミングと内容を調査した。本研究の結果から音声認識後の誤りを NS がテキスト編集する速度は 3 秒以上の遅延になり精度は低下することを予測した。従来研究(Gao et al. 2014)では音声認識後の NS による修正のタイミングと精度は検証されていないため、音声認識後の誤りを NS が修正する速度と精度は今後検証が必要である。

また、本研究では音声とテキストの利用時に短いポーズの有無と、NNS がリハーサルを行ったかなどの検証は行われていない。音声とテキストの利用時に共通記憶が多い原因の一つは自然に発生するポーズの活用が考えられるが、ポーズの長さによって影響が異なると考えられるため、検証されていない。また、NS が会話中に音声をテキスト化したことで、NNS がテキストを視覚で見るだけでなく、会話中にテキストを音韻化することも共通理解に有効と考えられるが、音韻化は検証されていない。本研究におけるポーズとリハーサルの効果については今後さらに検証が必要と考えられる。

本研究で示した音声とテキストの有効性は、ディベート課題で N1 レベルの中国人留学生と日本人の会話実験結果から示された。本研究の音声とテキストが有効である条件は、日本語能力が N1 以下の NNS の参加であると考えられる。NS 同士の会話では、音声のみによる共通理解が可能であるため、音声のテキスト化は行われない。同様に、NNS の日本語能力が極めて高い場合も音声のみによる共通理解が可能で音声のテキスト化は行われないと考えられる。したがって、音声のテキスト化は N1 以下の NNS が参加する会話の条件で有効と考えられる。また、他の意見交換課題においては、音声とテキストの有効性は検証されていないため今後検証が必要である。

## 5.4 今後の研究展開

本研究を実施して明らかになった課題をもとに、今後の研究展開について述べる。

### 5.4.1 課題解決会話の共通理解の研究

本研究では、多様な発話を行う音声会議における共通理解の過程に着目した分析を行った。その結果、NS がテキストで補足説明を加えながら、NNS と共通理解を構築していたことが示された。また、音声とテキストを使用した会話では共通理解が会話後の記憶に残ることも明らかになった。

本研究のような多様な発話を行う課題は他にもいくつかあり、課題が会話に与える影響は大きいと考えられる。2 章で紹介した Mennecke ら(2000)の研究では、NS 同士の会話支援ツールに使用されるマルチメディアの組み合わせと課題の組み合わせがテストされている(Mennecke et al. 2000)。よく利用される課題としては、地図課題、意思決定会議、ポンプ組み立て、経路探索などの課題解決会議などがある。NNS を含む会話支援では同様の結果になるかは不明のため、今後はほかの課題において音声とテキストの効果を検討する実験を行い、有効性を明らかにする必要がある。

### 5.4.2 多人数遠隔コミュニケーションの共通理解の研究

本研究では、音声会議のシミュレーションとして会話参加者が 2 名の実験室内で背面を合わせて実験を行った。実際の遠隔環境下の状況では、通信ノイズや遅延の影響があり、通信環境が実験結果に影響すると考えられる。また、実

際の遠隔環境の会議は複数人で行われることがあり、多人数遠隔環境下での会議支援を検討した研究が行われている(Gao et al. 2014)。音声とテキストによる会話支援の有効性をさらに検証するため、実際の遠隔会話状況を設定し、多人数の遠隔環境下で実験を行う意義があると考えられる。

これまでの NNS を含む会議支援の研究では、役割を固定せずに多様な発話を行う状況下の研究はそれほど多くない。従って、今後は話題や参加人数や実験環境などのコミュニケーション状況を変え、これらの要因が NS を含む音声会議にどのような影響を与えるのかを明らかにしていく。

## 第 6 章 結論

本研究では NS として日本人(日本語母語)、NNS として中国人(日本語非母語)を採用し、会話の分析を対象とする研究を行った。NS と NNS の会話で共通の理解が得られないため、会話の一部をテキスト(文字による情報伝達)で発信する会話支援を行い、テキストの有効性を検討した。本研究の一つ目の目的として、日本語音声会議におけるテキストの効果を明らかにすることとした。本研究の二つ目の目的として、音声とテキストがどのように理解に影響するかを検証することとした。本研究では上記 2 つの目的の分析が共通する NS と NNS の日本語会話実験を行い、以下の 3 つの仮説を構築し、実験データから仮説の検証を行った。本研究の会話実験の課題(会話の題材)は、自由度の高い課題を採用した。

- (1)音声とテキストの利用は共同解釈の構築を増加させる。
- (2)音声とテキストの利用は共通理解を向上する。
- (3)NNS の了解性と快適性が向上する。

仮説 1 と 2 における会話の共同解釈と共通理解は、会話の協調モデルによる検討を行った。仮説 3 は質問紙調査の分析を行った。また、目的 2 のため NS による会話音声のテキスト化はいつ、どのくらい、どのような内容で、話し方に変化を及ぼしたかという分析を行った。テキスト化のタイミングは、音声会議の実験映像にラベリングしたテキストと関連する発話を対応付け、発話との比較から検討した。テキストの量と内容は、実験ビデオの音声の書き起こしと、画面キャプチャの映像から取得した。話し方の変化は、発話内容を 6 種類の名詞句に分類することで検討した。

実験から得たデータを分析した結果から、音声とテキストの有効性に関する 3 つの仮説が支持された。仮説 1:「音声とテキストの利用は共同解釈の構築を

増加させる」の検証では、会話音声を書き起こし会話分析を行った結果、音声とテキストを利用する場合に共同解釈を構築する証拠となる会話が有意に多く認められ、仮説 1 が支持された。仮説 2:「音声とテキストの利用は共通理解を向上させる」の検証では、音声とテキストを利用する場合に NS と NNS の一致した記憶の件数が有意に増加したため、仮説 2 が支持された。仮説 3:「音声とテキストの利用は会話における NNS の了解性と快適性を向上させる」の検証では、NS と NNS の質問紙調査を分析したところ、音声のみよりも音声とテキストの利用時に NNS の了解性と快適性が有意に高まったことから、仮説 3 が支持された。従って、本研究の目的 1 であるテキストの有効性が示され、目的 1 を達成したと言える。

また、目的 2 である音声とテキストが共通理解にどのように影響するかを検証した結果、テキストは NS の発話から 2 秒後に、NS と NNS の発話を平均すると 3 秒後に 5 字程度で発話内容を補足説明する名詞句が表示されることが示された。本結果から、NS による音声のテキスト化は字幕に必要な水準を満たし、会話中に読めることが示された。話し方の検討では、テキスト化によって早く短い発話から説明を補う発話の形への変化が認められた。検討の結果から、NS が発話の一部をテキスト化する会話では、発話の直後に発話を補足する内容のテキストが表示され、参加者が音声とテキストで会話内容を確認することで理解と快適性が高まることが示された。以上から、本研究の目的 2 である音声とテキストがどのように会話に影響するかを示す目的が果たされた。

本研究の貢献として、日本語会話分析による共同解釈を構築する会話例を示したことが挙げられる。また、自由度の高い課題としてまだ実践例が少数であるディベート課題の音声会議を行ったことも本研究分野への貢献となる。さらに、本研究は音声のみと音声とテキスト会話を比較して共通理解の差を示したことから、NNS との会話ではコミュニケーション・メディアによって共通理解の程度に差があることを示した。また、会話の一部分のみをテキスト化する

タイミングと内容の詳細な調査はこれまで実施例がほとんどなく、従来研究とコミュニケーションの状況と対象が異なるため、本研究分野への貢献となる。

本研究結果から、テキスト音声会話を自動化する会話支援ツールのデザインとして、会話中に言い換えや具体例などの補足説明を5-6字で表示する支援が示唆される。また、NS と NNS のコミュニケーション支援ツールの評価、結果予測や了解性と快適性の評価などの研究分野でも使用できると考えられる。

今後の課題は、会話題材、参加人数、実験環境などのコミュニケーション環境を変えて音声とテキストの有効性を検証することである。NNS を含む多人数遠隔環境の音声会話で、複数の会話題材を検証し、これらの要因がどのような影響を与えるのかを明らかにすることが必要である。

## 謝辞

博士論文をまとめるにあたり、指導教員である筑波大学図書館情報メディア系の井上智雄教授には多くのご助言を頂きました。2013年から研究生として研究室に迎えていただき、終始多大なご指導・ご助言に支えられ2022年までの長期間の博士後期課程を修了することができました。井上智雄教授には科学的な考え方、研究の進め方、研究発表の仕方等、多岐にわたるご指導を頂きました。同じ研究室の大学院生と共同で研究を進める機会や、ティーチング・アシスタントに就く機会を与えて頂いたことにより、研究や授業における学生指導に関する知識と経験を得ることができました。

副指導教員である筑波大学図書館情報メディア系の佐藤哲司教授、鈴木佳苗教授には時折研究や論文に貴重なご助言を頂くとともに温かい激励の言葉を頂きました。

審査にあたり、井上教授をはじめとして、佐藤哲司教授、鈴木佳苗教授とともに論文審査委員を務めていただいた高橋伸准教授、真栄城哲也教授には、貴重なご助言とご意見を頂きました。

研究論文の共著者である宋暁宇氏、唐夢苑氏、于睿政氏をはじめ、共に研究を進めた筑波大学図書館情報メディア研究科井上研究室の大学院生の皆様には、多大なご支援を頂きました。大学院生の皆様と共に、筑波大学図書館情報メディア研究科では充実した大学院生活を送ることができました。

筑波大学職員の方々、研究会や国際会議等の学外発表においてご意見を下さった方々、論文投稿にあたり貴重な査読コメントを下さった方々の支援もあり、本博士論文を出すに至りました。すべての方々に厚く御礼を申し上げます。

著者の家族には多大な精神的、経済的支援を頂きました。博士後期課程在学中にこの世に生を受けた息子、そして、博士論文執筆中に支えてくれた夫に感謝を送ります。

埴裕美

## 引用文献

- Austin, J. L. 1962. *How to Do Things with Words*. Harvard University Press.
- Bos, N. Sadat Shami, Judith S. Olson, Arik Cheshin, and Ning Nan. 2004. “In-Group/out-Group Effects in Distributed Teams: An Experimental Simulation.” In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, ACM, 429–436.
- Bos, N. Sadat Shami, Judith S. Olson, Arik Cheshin, and Ning Nan. 2006. “Collocation Blindness in Partially Distributed Groups: Is There a Downside to Being Collocated?” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1313–21.
- Cao, X., Yamashita, N., and Ishida, T. 2018. “Effects of Automated Transcripts on Non-Native Speakers' Listening Comprehension.” *IEICE Transactions on Information and Systems*, E101.D(3): 730–39.
- Chapanis, A. 1976. “Human Factors in Teleconferencing Systems.” *Johns Hopkins Univ. Final Report 1*.
- Chen, M. L., Yamashita, N., and Wang, H. C. 2018a. “Beyond Lingua Franca: System-Facilitated Language Switching Diversifies Participation in Multiparty Multilingual Communication.” In *Proceedings of the ACM on Human-Computer Interaction (CSCW)*, ACM, 1-22
- Chen, M. L., Yamashita, N., and Wang, H. C. 2018b. “Come Together: Facilitating Collocated Multilingual Group Discussion with a Language Support Tool.” *Conference on Human Factors in Computing Systems(CHI)*, ACM, 1-6
- Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. Greenwood Publishing Group.

- Clark, H. H. and Wilkes-Gibbs, D. 1986. "Referring as a Collaborative Process." *Cognition* 22(1): 1-39
- Clark, H. H. 1996. "Using Language: Ch12 Layering." *Using Language*.  
[http://web.stanford.edu/~clark/1990s/Using language/Old versions/Clark.UsingLanguage.Ch12.96.pdf](http://web.stanford.edu/~clark/1990s/Using%20language/Old%20versions/Clark.UsingLanguage.Ch12.96.pdf).
- Clark, H. H. 1996. *Using Language*. Cambridge university press.
- Clark, H. H. and Brennan, S. E. 1991. "Grounding in Communication." *Perspectives on socially shared cognition* 13: 127–149.
- Diamant, E. I., Fussell, S. R., and Lo, F. 2008. "Where Did We Turn Wrong?: Unpacking the Effect of Culture and Technology on Attributions of Team Performance." In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, ACM, 383–392.
- Ding, X., Erickson, T., Kellogg, W. A., Levy, S., Christense, J. E., Sussman, J. B., Wolf, T. V., and Bennett, W. E. 2007. "An Empirical Study of the Use of Visually Enhanced Voip Audio Conferencing: The Case of IEAC." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1019–1028.
- Duan, W., Yamashita, N., and Fussell, S. R., 2019. "Increasing Native Speakers' Awareness of the Need to Slow down in Multilingual Conversations Using a Real-Time Speech Speedometer." In *Proceedings of the ACM on Human-Computer Interaction* 3(CSCW), ACM, 1-25
- Echenique, A., Yamashita, N., Kuzuoka, H., and Hautasaari, A. 2014. "Effects of Video and Text Support on Grounding in Multilingual Multiparty Audio Conferencing." In *Proceedings of the 5th ACM International Conference on Collaboration across Boundaries: Culture, Distance & Technology*, ACM, 73–81.

- Ennis, C., McDonnell, R., and O'Sullivan, C. 2010. "Seeing Is Believing: Body Motion Dominates in Multisensory Conversations." *ACM SIGGRAPH 2010 Papers, SIGGRAPH 2010* 1(212): ACM, 1–9.
- Ferreira, F., Engelhardt, P. E., and Jones, M. W. 2009. "Good Enough Language Processing: A Satisficing Approach." *Proceedings of the 31st Annual conference of the Cognitive Science Society* (1): 413–418.
- Fitzpatrick, T. and Barfield, A. 2009. "*Lexical Processing in Second Language Learners: Papers and Perspectives in Honour of Paul Meara.*" Multilingual Matters.
- Gao, G., Yamashita, N., Hautasaari, A., Echenique, A., and Fussell, S. R. 2014. "Effects of Public vs. Private Automated Transcripts on Multiparty Communication between Native and Non-Native English Speakers." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 843–852.
- Gao, G., Xu, B., Hau, D. C., Yao, Z., Cosley, D., and Fussell, S. R. 2015. "Two Is Better Than One: Improving Multilingual Collaboration by Giving Two Machine Translation Outputs." In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM, 852–863.
- Gao, G., Hwang, S., Culbertson, G., Fussell, S. R., and Jung, M. F. 2017. "Beyond Information Content: The Effects of Culture on Affective Grounding in Instant Messaging Conversations." In *Proceedings of the ACM on Human-Computer Interaction (CSCW)*, ACM, 1–18.
- Gao, G. and Fussell, S. R., 2017. "A Kaleidoscope of Languages: When and How Non-Native English Speakers Shift between English and Their Native Language during Multilingual Teamwork", In *Proceedings of the*

2017 CHI Conference on Human Factors in Computing Systems(CHI), ACM, 760–772.

- Gao, G., Wang, H., Cosley, D., and Fussell, S. R., 2013. “Same Translation but Different Experience: The Effects of Highlighting on Machine-Translated Conversations.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 449–458.
- Gao, G., Yamashita, N., Hautasaari, A., and Fussell, S. R. 2015. “Improving Multilingual Collaboration by Displaying How Non-Native Speakers Use Automated Transcripts and Bilingual Dictionaries.” In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 3463–3472.
- Gaur, Y., Lasecki, W. S., Metze, F., and Bigham, J. P., 2016. "The effects of automatic speech recognition quality on human transcription latency." In *Proceedings of the 13th Web for All Conference*, 1-8.
- Goffman, C. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Guo, Z. and Inoue, T. 2019. “Using a Conversational Agent to Facilitate Non-Native Speaker’s Active Participation in Conversation.” *Conference on Human Factors in Computing Systems*, ACM, 5–10.
- Hampel, R. and Hauck, M. 2004. “Towards an Effective Use of Audio Conferencing in Distance Language Courses.” *Language Learning and Technology* 8(1), 66–82.
- Hanawa, H., Song, X. and Inoue, T. 2016. “Key-Typing on Teleconference: Collaborative Effort on Cross-Cultural Discussion”, In *International Conference on Collaboration Technologies (CollabTech2016)*, Springer, 74-88.

- Hanawa, H., Song, X. and Inoue, T. 2017. "Keyword Generation by Native Speaker Is Quick and Useful in Conversation between Native and Non-Native Speaker." *Proceedings of the 2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 145–150.
- Hanawa, H., Song, X., Tang, M. and Inoue, T. 2017. "Speakers' Empowerment with Keywords: The Speaking Competency Development for Multilingual to Contribute in a Real-Time Interaction." In *International Conference on Collaboration and Technology (CRIWG)*, Springer, 148-163
- Hautasaari, A. 2010. "Machine Translation Effects on Group Interaction: An Intercultural Collaboration Experiment." In *Proceedings of the 3rd ACM International Conference on Intercultural Collaboration*, ACM, 69–78.
- Hautasaari, A. and Yamashita, N. 2014. "Do Automated Transcripts Help Non-Native Speakers Catch up on Missed Conversation in Audio Conferences?" In *Proceedings of the 5th ACM International Conference on Collaboration across Boundaries: Culture, Distance & Technology*, ACM, 65–72.
- Hautasaari, A., Yamashita, N., and Gao, G. 2014. "'Maybe It Was a Joke'- Emotion Detection in Text-Only Communication by Non-Native English Speakers." *Conference on Human Factors in Computing Systems*, ACM, 3715–3723.
- Hautasaari, A., Yamashita, N., and Gao, G. 2019. "How Non-Native English Speakers Perceive the Emotional Valence of Messages in Text-Based Computer-Mediated Communication." *Discourse Processes* 56(1), 24–40. <https://doi.org/10.1080/0163853X.2017.1323184>.

- He, H. A., Yamashita, N., Wacharamanotham, C., Horn, A. B., Schmid, J., and Huang, E. M. 2017. "Two Sides to Every Story: Mitigating Intercultural Conflict through Automated Feedback and Shared Self-Reflections in Global Virtual Teams." In *Proceedings of the ACM on Human-Computer Interaction (CSCW)*, ACM, 1–21.
- Inoue, T. 2014. "Naturalistic Control of Conversation by Meal: Induction of Attentive Listening Attitude through Uneven Meal Distribution in Co-dining", In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, 601-606.
- Inoue, T., Hanawa, H., and Song, X. 2015. "With a Little Help from My Native Friends□: A Method to Boost Non-Native 's Language Use", In *Proceedings of the Ninth International Workshop on Informatics (IWIN)*, 223-226.
- Inoue, T. and Kawai, K. 2019. "Influence of the Representation of a Conversational Agent for Supporting a Non- Native Speaker to Talk with a Native Speaker." In *Proceedings of the 13th International Workshop on Informatics (IWIN)*, 213-217.
- Isaacs, E. A. and Tang, J. C. 1993. "What Video Can and Can't Do for Collaboration: A Case Study." In *Proceedings of the first ACM international conference on Multimedia*, ACM, 199–206.
- Lai, C. and Zhao, Y. 2006. "Noticing and Text-Based Chat." *Language Learning and Technology*, 10(3), 102–120. <http://dx.doi.org/10125/44077>
- Lasecki, W. S., Kushalnagar, R., and Bigham, J. P. 2014. "Helping Students Keep up with Real-Time Captions by Pausing and Highlighting." In *Proceedings of the 11th Web for All Conference*, ACM, 1-8.
- Li, N. and Rosson, M. B. 2012. "Instant Annotation: Early Design Experiences in Supporting Cross-Cultural Group Chat." In *Proceedings*

- of the 30th ACM International Conference on Design of Communication*, ACM, 147–155.
- Lim, H., Cosley, D., and Fussell, S. R. 2018. “Beyond Translation: Design and Evaluation of an Emotional and Contextual Knowledge Interface for Foreign Language Social Media Posts.” *Conference on Human Factors in Computing Systems - Proceedings* 2018-April: 1–12.
- Liu, R. and Inoue, T. 2014. "Application of An Anthropomorphic Dining Agent to Idea Generation", In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, 607-612.
- McGrath, J. E. and Hollingshead, A. B. 1993. “Putting the ‘Group’ Back in Group Support Systems: Some Theoretical Issues About Dynamic Processes in Groups With Technological Enhancements.” *Group support systems: New perspectives*, New York: Macmillan. 78–96.
- Menne, J. M. and Menne, J. W. 1972. “The Relative Efficiency of Bimodal Presentation as an Aid to Learning.” *AV communication review* (20(2)), 170–180.
- Mennecke, B. E., Valacich, J. S., and Wheeler, B. C. 2000. “The Effects of Media and Task on User Performance: A Test of the Task-Media Fit Hypothesis.” *Group Decision and Negotiation* 9(6): 507–529.
- Novinger, T. 2001. “Intercultural Communication: A Practical Guide.” University of Texas Press.
- Pan, M. H., Yamashita, N., and Wang, H. C. 2017. “Task Rebalancing: Improving Multilingual Communication with Native Speakers-Generated Highlights on Automated Transcripts.” In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*, ACM, 310–321.

- Pan, Y., Jiang, D., Yao, L., Picheny, M., and Qin, Y. 2010. "Effects of Automated Transcription Quality on Non-Native Speakers' Comprehension in Real-Time Computer-Mediated Communication." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1725–1734.
- Pan, Y., Jiang, D., Picheny, M., and Qin, Y. 2009. "Effects of Real-Time Transcription on Non-Native Speaker's Comprehension in Computer-Mediated Communications." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2353–2356.
- Paulus, P. B. and Yang, H. C. 2000. "Idea Generation in Groups: A Basis for Creativity in Organizations." *Organizational behavior and human decision processes* 82(1): 76–87.
- Pellettieri, J. 2000. "Negotiation in Cyberspace: The Role of Chatting in the Development of Grammatical Competence." *Network-based language teaching: Concepts and practice*, Cambridge University Press, 59–86.
- Pye, R. and Williams, E. 1977. "Teleconferencing: Is Video Valuable or Is Audio Adequate?" *Telecommunications Policy* 1(3), 230–241.
- Rebai, I., Ayed, Y. B., and Mahdi, W. 2019. "Spoken Keyword Search System Using Improved ASR Engine and Novel Template-Based Keyword Scoring." *Multimedia Tools and Applications* 78(2), 1495–1510.
- Sacks, H., Schegloff, E. A., and Jefferson, G. 1974. "A Simplest Systematics for the Organization of Turn-Taking for Conversation." *language*: 696–735.
- Sasaki, K. and Inoue, T. 2021. "Effects of the notification of native speaker's speech rate on asymmetric second language conversation." *Journal of Information Processing Society of Japan (IPSJ)* 62(1), 115–124.

- Searle, J.R.1969. *Speech Acts - an Essay in the Philosophy of Language*.  
Cambridge University Press.
- Sellen, A. J. 1992. “Speech Patterns in Video-Mediated Conversations.”  
*Conference on Human Factors in Computing Systems*, 49–59.
- Shukla, R. 2020. “Keywords Extraction and Sentiment Analysis using  
Automatic Speech Recognition.” ArXiv, abs/2004.04099.
- Sirintrapun, J. C. A. 2012. “Dynamic nonrobotic telemicroscopy via skype: A  
cost effective solution to teleconsultation”, *Journal of pathology  
informatics*, 3(1), 28.
- Swain, M. 2000. “The Output Hypothesis and Beyond.” *Sociocultural theory  
and second language learning*, 97–114.
- Uzama, A. 2012. “Yokoso! Japan: Classifying Foreign Tourists to Japan for  
Market Segmentation.” *Journal of Hospitality Marketing &  
Management* 21(2), 132–154.
- Walther, J. B. 2007. “Selective Self-Presentation in Computer-Mediated  
Communication: Hyperpersonal Dimensions of Technology, Language,  
and Cognition.” *Computers in Human Behavior* 23(5), 2538–2557.
- Wang, Y. 2004. “Supporting Synchronous Distance Language Learning with  
Desktop Videoconferencing.” *Language Learning & Technology* 8(3),  
90–121.
- Yamashita, N., Echenique, A., Ishida, T., and Hautasaari, A. 2013. “Lost in  
Transmittance: How Transmission Lag Enhances and Deteriorates  
Multilingual Collaboration.” In *Proceedings of the 2013 Conference on  
Computer Supported Cooperative Work (CSCW)*, ACM, 923–934.
- Yamashita, N. Inaba, R. Kuzuoka, H., and Ishida, T. 2009. “Difficulties in  
Establishing Common Ground in Multiparty Groups Using Machine

Translation.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 679–88.

Yao, L., Pan, Y., and Jiang, D., 2011. “Effects of Automated Transcription Delay on Non-Native Speakers’ Comprehension in Real-Time Computer-Mediated Communication.” In *IFIP Conference on Human-Computer Interaction*, Springer, 207–214.

Zuchowski, M., and Göller, A., 2022, "Speech recognition for medical documentation: an analysis of time, cost efficiency and acceptance in a clinical setting." *British Journal of Healthcare Management*, 28:1, 30-36

相場いぶき. 中井陽子. 2010. “会話授業におけるストーリーテリングの分析:—聞き手を引き込むために話し手が用いる言語的・非言語的・音声的要素—.” *日本語教育方法研究会誌* 17(1): 70–71.  
<https://ci.nii.ac.jp/naid/110009497335/>.

浅木尚実. 2014. “ストーリーテリング(お話)と国語教育「話す力」「聞く力」養成: 教員志望学生のコミュニケーション力向上に関する考察.” *淑徳短期大学研究紀要 = Bulletin of Junior College of Shukutoku* (53): 53–67.  
<https://ci.nii.ac.jp/naid/110009799422/>.

池本薫規. 2015. “相談援助演習における ディベート実践の方法と展開.” *福祉教育開発センター紀要* 第12号: 139–150.

伊勢恵. 2014. “ストーリーテリングを用いたスピーキング指導と協同学習の親和性の検証.” *国際文化研究* (20): 1–15.  
<https://ci.nii.ac.jp/naid/120005430839/>

稲葉利江子. 山下直美. 石田亨. 葛岡英明. 2009. “機械翻訳を用いた 3 言語間コミュニケーションの相互理解の分析.” *電子情報通信学会論文誌 D* 92(6): 747–757.

- 大林史明, 山本専, 伊藤京子, 下田宏, 吉川榮和. 2002. “総合学習支援システムのエネルギー学習への適用実験と考察.” *日本原子力学会和文論文誌 = Transactions of the Atomic Energy Society of Japan* 1(3): 253–262.  
<https://ci.nii.ac.jp/naid/10009757197/>.
- 岡崎光太郎, 神田智子. 2014. "人間とエージェント間におけるフィルターの有用性の検証", 第76回全国大会講演論文集, 2014(1), 237-238.
- 岡本健吾, 吉野孝. 2011. “会話中の名詞の関連情報を用いた対面型異文化間コミュニケーション支援システムの構築と評価.” *情報処理学会論文誌* 52(3): 1213–1223. [http://www.wakayama-u.ac.jp/~yoshino/lab/publication/pdf\\_download.php?file=IP SJ-JNL5203026.pdf](http://www.wakayama-u.ac.jp/~yoshino/lab/publication/pdf_download.php?file=IP SJ-JNL5203026.pdf).
- 鹿島英一. 1997. "文字の情報理論序説". 風間書房.
- 株式会社サンビジコム. 2022. “電話会議ドットコム.”  
<https://denwakaigi.com/?pid=84868511>.
- 倉本到, 宗森純, 由井蘭隆也, 首藤勝. 1998. “発想支援グループウェアの実施に及ぼすテキストベースコミュニケーションの影響.” *情報処理学会論文誌* 39(10): 2778–87. <https://ci.nii.ac.jp/naid/110002722315/>.
- 国立国語研究所. 2021. “形態素解析ツール.” *web茶まめ: 各種UniDicを使用した形態素解析ツール*. <https://chamame.ninjal.ac.jp/>.
- 近藤悠介, 石井雄隆. 2017. “英語学習者の発話自動採点システムの開発と英語教育プログラムへの導入可能性の検討.” *Language education & technology* (54): 23–40.
- 佐伯胖, 杉本美穂子, 楠房子, 鍛冶英紀. 2002. “GE-4 協調的科学的討論支援システムの開発.” *日本科学教育学会年会論文集* 26(0): 159–160.  
<https://ci.nii.ac.jp/naid/110002952621/>.

出入国在留管理庁. 2022. “在留外国人統計（旧登録外国人統計）統計表.”

[https://www.moj.go.jp/isa/policies/statistics/toukei\\_ichiran\\_touroku.html](https://www.moj.go.jp/isa/policies/statistics/toukei_ichiran_touroku.html)

総務省. 2022. “情報難民ゼロプロジェクト.”

[https://www.soumu.go.jp/menu\\_seisaku/kokumin/jyohonanminzero/index.html](https://www.soumu.go.jp/menu_seisaku/kokumin/jyohonanminzero/index.html)

宋暁宇. 2018. “非母語話者と母語話者の会話における母語話者テキスト入力の研究.”, 筑波大学修士(情報学)学位論文.

高橋徹. 濱崎雅弘. 武田英. 2001. "Avatar-likeエージェントを用いたWebコミュニティ支援システム", 人工知能学会全国大会論文集 JSAI01(0), 4.

田中茂範. 深谷昌弘. 1998. "< 意味づけ論 > の展開: 情況編成・コトバ・会話". 紀伊國屋書店.

筒井千絵. 2008. "フォリナー・トークの実際--非母語話者との接触度による言語調整ストラテジーの相違". 一橋大学留学生センター紀要 (11) 79-95.

寺村秀夫. 1982. “日本語のシンタクスと意味” 第1巻. くろしお出版.

伝康晴. 2009. “聞き手行動から見たコミュニケーション 聞き手行動の認知科学に必要なもの.” 認知科学 16(4): 475-480.

<http://ci.nii.ac.jp/naid/130004490971/>.

栃木由香. 1990. “日本語学習者のストーリーテリングに関する一分析□: 話の展開と接続形式を中心にして.” 筑波大学留学生教育センター日本語教育論集 (5): 159-174. <https://ci.nii.ac.jp/naid/110000528722/>.

中井陽子. 2004. “会話分析の視点を生かした会話授業: ストーリーテリングのテクニック指導の実践報告.” 日本語教育方法研究会誌 11(2): 4-5.

<https://ci.nii.ac.jp/naid/110009496976/>.

- 西川由紀子. 1995. “幼児の物語産出における「語り」の様式.” 発達心理学研究 6(2): 124–133.
- 西阪仰. 高木智世. 川島理恵. 2008. "女性医療の会話分析". ソキウス研究叢書 : 6 文化書房博文社
- 西田豊明. 1996. “会話情報学 2. 現象としての会話の諸相 その2” : 1–6.  
<https://www.ii.ist.i.kyoto-u.ac.jp/wordpress/wp-content/uploads/2012/10/ci-pub-2013-2b.pdf>
- 西村良太. 北岡教英. 中川聖一. 2006. "対話における韻律変化・タイミングのモデル化と音声対話システムへの適用", 言語・音声理解と対話処理研究会, vol.48, 37-42.
- 日本語能力試験とは. 2022. <https://www.jlpt.jp/about/>
- 日本語能力試験. “N1～N5:認定の目安” 2022.  
<https://www.jlpt.jp/about/levelsummary.html>.
- 花村博司. 2014. “日本語の雑談会話における話題転換研究の方法—話題転換はどこで行われ、どう分類されるか—.” 言語文化学研究 9(大阪府立大学人間社会学部言語文化学科): 71–99.
- 埴裕美. 井上智雄. 2020. “日本語非母語話者を含む音声会議におけるテキストの効果.” 情報コミュニケーション学会論文誌 16号, vol.16. No.s1&2, 12-23.
- 速水悟. 2013. “音声言語処理と自然言語処理, コロナ社, 日本音響学会誌 69(12): 655.
- ビジネス日本語能力テスト, “BJTの特徴のご紹介「日本語能力試験 (JLPT) との比較」.” 2022. <https://www.kanken.or.jp/bjt/about/feature.html>
- 平井明代. 2000. "文字言語情報と音声言語情報の処理モデルに基づく英語学習システムの開発研究". 科学研究費補助金(基礎研究C).

- 福田由紀. 2012. 言語心理学入門: 言語力を育てる. 培風館.
- 藤田玲子. 2009. “観光立国ジャパン--異文化コミュニケーション力に関する一考察.” コミュニケーション科学 (30): 3-14.
- 藤原和政. 河村茂雄. 川俣理恵. 2017. “保育士養成校に在籍する専門学校生に対するCATSの効果検証.” カウンセリング研究 50(1): 13-22.
- 坊農真弓. 高梨克也. 2009. “多人数インタラクションの分析手法 (知の科学).”
- 三浦麻子. 2001. “ブレインストーミングにおけるコミュニケーション・モードと目標設定の効果.” 対人社会心理学研究 (1): 45-57.  
<https://ci.nii.ac.jp/naid/120005666313/>.
- 宮部真衣. 吉野孝. 重信智宏. 2007. “折返し翻訳を用いた翻訳リペアの効果.” 電子情報通信学会論文誌 D 90(12): 3141-3150.
- 宗森純. 五郎丸秀樹. 長澤庸二. 1995. “発想支援グループウェアの実施に及ぼす分散環境の影響.” 情報処理学会論文誌 36(6): 1350-1358.  
<https://ci.nii.ac.jp/naid/110002721884/>.
- 山下直美. エチエニーケ アンディ. 葛岡英明. 石田亨. ハウタサーリ アリ. 2013. “国際電話会議の負担を軽減する手法の提案と評価.” 情報処理学会論文誌 54(6): 1794-1806.
- 山鳥重. 2021. "わかるとはどういうことかー認識の脳科学."筑摩書房.
- 山本博樹. 天沼聡. 杉原一昭. 1989. “物語産出の体制化に関する発達的研究: 体制化のレベルと物語の内容構造との関係.” 筑波大学心理学研究 (11): 79-84.  
<https://ci.nii.ac.jp/naid/110000258648/>.
- 山元淑乃. 2019. “非ネイティブスピーカー志向の第二言語習得: 日本人英語学習者の語りの分析を通して.” 言語文化教育研究 17: 360-382.

吉田悦子. 2001. “日英語対照研究のための名称なし日本語地図課題対話コーパスの作成.” 人文論叢(三重大学) 第18号: 173–195.

吉田睦. 2014. “日本語母語話者・学習者の会話構築に関わる質問表現の研究.”  
[http://tsukuba.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtIz0EUrE0xTkoGJI83EOMXQ3DDN2NzIzBhYURunpAGbtyZp4AlULz-jiGCjcCeLUCYG2P13aKcNgKobfchkqZk5MwOzsQFoEZebgR6w8VpSXJpdmpSIVDm4CTLwuCBNagsxMGUlijAEPpu-9NmcNS9WzXu2vhdIvli58UVD6-Pm3U\\_XLnu-fxKI3bj](http://tsukuba.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtIz0EUrE0xTkoGJI83EOMXQ3DDN2NzIzBhYURunpAGbtyZp4AlULz-jiGCjcCeLUCYG2P13aKcNgKobfchkqZk5MwOzsQFoEZebgR6w8VpSXJpdmpSIVDm4CTLwuCBNagsxMGUlijAEPpu-9NmcNS9WzXu2vhdIvli58UVD6-Pm3U_XLnu-fxKI3bj).

渡辺大地. 2020. “共通点を探そう.” 助産雑誌 74(7): 542–545.  
[http://tsukuba.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwtV0\\_T9wwFLcOKiQYUCuoSlskljd0uCN2Ev8ZOthHqu6AynZKOAcOdAcS3F5OgpGhY6WqUr8BI5\\_nRkt-iz7b4S5QBhiQosh6iX9O4pf3np349wiJWStq3rMJqEjMJl2aWy6UFZwXMuWWdqPCogPbtT43DNvZZF-M3G40ft4ujZkKn7XnUYZ971bSPqH3J6A](http://tsukuba.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwtV0_T9wwFLcOKiQYUCuoSlskljd0uCN2Ev8ZOthHqu6AynZKOAcOdAcS3F5OgpGhY6WqUr8BI5_nRkt-iz7b4S5QBhiQosh6iX9O4pf3np349wiJWStq3rMJqEjMJl2aWy6UFZwXMuWWdqPCogPbtT43DNvZZF-M3G40ft4ujZkKn7XnUYZ971bSPqH3J6A).

## 全研究業績

### 【学術雑誌論文】

1. 埴裕美、井上智雄 日本語非母語話者を含む音声会議におけるテキストの効果、情報コミュニケーション学会論文誌、Vol.16 No.s1&s2, 12-23, 2020

### 【査読付き国際会議】

1. Hiromi Hanawa, Xiaoyu Song, Mengyuan Tang, Tomoo Inoue, "Speakers' empowerment with keywords: the speaking competency development for multilingual to contribute in a real-time interaction", Collaboration and Technology (Proceedings of the 23rd International Conference, CRIWG 2017), LNCS 10391, pp.148-163, Springer, 2017.
2. Hiromi Hanawa, Xiaoyu Song, Tomoo Inoue, "Keyword generation by native speaker is quick and useful in conversation between native and non-native speaker", Proceedings of the 2017 IEEE 21th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2017), pp.145-150, 2017.
3. Hiromi Hanawa, Xiaoyu Song, Tomoo Inoue; Key-Typing on Teleconference: Collaborative Effort on Cross-Cultural Discussion, Collaboration Technologies and Social Computing Volume 647 of the series Communications in Computer and Information Science, Proceedings of the 8th International Conference, (CollabTech 2016), CCIS 647, pp.74-88, Kanazawa, Japan, September 14-16, 2016
4. Tomoo Inoue, Hiromi Hanawa, Xiaoyu Song; With a Little Help from My Native Friends: A Method to Boost Non-native's Language Use in Collaborative Work, Proceedings of the Ninth International Workshop on Informatics (IWIN 2015), pp.223-226, Amsterdam, Netherlands, September 6-9, 2015.
5. Hiromi Hanawa, Tomoo Inoue; Study of Face-to-face Dyadic Conversation Behavior on Uneven Meal Distribution Setting for

Designing an Attentive Listening Agent, *Communications in Computer and Information Science* 460, pp.77-85, Springer, 2014.9.

【査読無し論文】

1. 宋曉宇, 塙裕美, 井上智雄, 非母語話者と母語話者のディベート会話における母語話者テキスト入力の影響, 情報処理学会研究報告, Vol.2018-GN-104, No.10, pp.1-6, 2018.3.12.
2. 于睿政, 唐夢苑, 宋曉宇, 塙裕美, 井上智雄, 非母語話者を含む会話における母語話者テキスト入力の効果, 電子情報通信学会研究報告, Vol.116, No.524, pp.23-28, 2017.3.8.
3. 塙裕美, 宋曉宇, 井上智雄, 母語話者の文字入力による非母語話者との会話支援 -母語話者による会話中のテキスト入力が音声会話に与える影響-, 電子情報通信学会研究報告, Vol.116, No.31, pp.139-144

【学会発表】

1. 母語話者の文字入力による非母語話者との会話支援 -母語話者による会話中のテキスト入力が音声会話に与える影響-, 電子情報通信学会研究報告, Vol.116, No.31, pp.139-144 (ヒューマンコミュニケーション基礎 HCS2016-18 (2016-05)), 沖縄, 2016年5月18-19日, 2016年5月11日.
2. Hiromi Hanawa, Xiaoyu Song, Tomoo Inoue; Keyword generation by native speaker is quick and useful in conversation between native and non-native speaker, Proceedings of the 2017 IEEE 21th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2017), pp.145-150, Wellington, New Zealand, April 26-28, 2017.
3. 非母語話者と母語話者のディベート会話における母語話者テキスト入力の影響, 情報処理学会第104回グループウェアとネットワークサービス研究発表会, 茨城県, 2018年3月19日

## 付録 A

### 参加者基本情報記入用紙

No. \_\_\_\_\_

参加者基本情報 氏名 \_\_\_\_\_ 年 月 日

本実験でご記入いただいた個人情報は、本実験に関わる目的にのみ使用し、他の目的には使用いたしません。

1. 年齢 \_\_\_\_\_
2. 性別 \_\_\_\_\_
3. 国籍 \_\_\_\_\_
4. 母国語 \_\_\_\_\_
5. 日本語を学んだ年数（日本国籍以外の方のみ）  
\_\_\_\_\_ (年)
6. 日本で暮らした年数（日本国籍以外の方のみ）  
\_\_\_\_\_ (年)
7. 日本語レベル N1/N2 点数（日本国籍以外の方のみ）  
N1 \_\_\_\_\_  
N2 \_\_\_\_\_
8. 日本語自己評価（日本国籍以外の方のみ）

とても低い      低い      やや低い      どちらでもない      やや高い      高い      とても高い  
1 ----- 2 ----- 3 ----- 4 ----- 5 ----- 6 ----- 7

9. 以下について問題の知識がありますか？（原子力発電・死刑・動物実験）  
\_\_\_\_\_ はい ・ いいえ
10. 会話相手のことを知っていますか？  
\_\_\_\_\_ はい ・ いいえ



## 付録 C

### 原子力発電問題用紙

#### 問題用紙

#### 原子力発電(げんしりょくはつでん)

原子核反応時に出るエネルギーを利用した発電、あるいはその方法。

※問題用紙を追記や変更をしても構いません。

※自分の国の立場・または自分の意見を加えて議論してください。

#### 利点

- 安定した電力供給が可能
- 発電時に 地球温暖化の原因とされる二酸化炭素を排出しない
- 使用する燃料が極端に少なくて済む
- 燃料の中東依存度を減らすことができる
- 核燃料サイクル等によるウラン 238 の有効利用や海水からのウラン採取が実現すれば、燃料は非常に豊富
- 経済性が高い
- 再処理により準国産エネルギーを実現できる
- 技術力があることがアピールできる
- 酸素を必要としない

## 問題用紙

### 原子力発電(げんしりょくはつでん)

原子核反応時に出るエネルギーを利用した発電、あるいはその方法。

※問題用紙にメモを取っても構いません。

※自分の国の立場・または自分の意見を加えて議論してください。

#### 問題点

- 重大事故は周辺環境に多大な被害を与え、その影響は地球規模に及ぶ
- 毒性が強く、放射性物質である核廃棄物を作り出す
- 高レベル放射性廃棄物の最終処分地が決定していない
- 地殻中のウラン 235 のみの利用を考えた場合、資源がそれほど豊富なわけではない
- ウラン資源の可採埋蔵量に由来する資源枯渇問題
- 軍事転用の制約に関わる国際社会への配慮
- 起動停止の所要時間が長い(通常停止)
- 炉の特性上、通常は負荷追従運転を行わない
- 火力発電所と比べ、施設建設に多大なコストがかかる
- 地質学的側面から、立地場所が限定される
- 電気利用者・電力会社と施設周辺に住む住民との利益・不利益が相応でない可能性がある
- 発電施設および核廃棄物へのテロの危険
- 将来の原子力発電を担ってくれる若手技術者が減少傾向にある
- 原子力発電所の新規建設数が減少していることからメーカーの原子力部門における技術の継承が困難となってきている

## 付録 D

### 死刑制度問題用紙

## 問題用紙

### 死刑について考える

※問題用紙に追記・編集しても構いません。

※自分の国の立場・または自分の意見を加えて議論してください。

#### 反対理由：

- 冤罪の場合の死刑は取り返しがつかない。
- 死刑執行までに時間がかかり、刑務や裁判の負担が大きい
- 先進国ではアメリカの一部と日本を除いて廃止に向かっている
- 犯罪者にも人権がある。国家といえども、人権を侵害する権利はない
- 死刑に犯罪抑止力があるという証明はなされていない。
- 憲法で公務員による残虐な刑罰は禁止されている。
- 犯人を死刑にしても被害者は戻らない
- 死刑は無期懲役などでその効果は代替可能である。

## 問題用紙

### 死刑について考える

※問題用紙に追記・編集しても構いません。

※自分の国の立場・または自分の意見を加えて議論してください。

#### 賛成理由：

- ①人を殺すような犯罪者に人権を認める必要はない
- ②冤罪は慎重な裁判によって防げる。一回の裁判で死刑判決が言い渡されることはない。
- ③日本には独自の国民性があり、古来より報復、復讐を是としてきた。死刑は日本の国情に合うもので、他国の意見に惑わされる必要はない。
- ④自分の身内を殺されたら被害者家族は死刑を求める
- ⑤社会の安全を守るためには危険要素である殺人者を排除することが社会から求められる
- ⑥人を殺した者が自分の命で償うのは当然。

# 付録 E

## 音声とテキストのタイミングと内容例 1

#	A	B	C	D	E	F	G	H	I	J	K
1	開始時間	終了時間	間隔	時	テキスト	NS					NNS
2	00:00.9	00:02.2	00:01.3		はい( )えっと						
3	00:01.1	00:03.6	00:02.5		原子力発電						
4	00:02.4	00:04.4	00:02.0		原子力発電は:						
5	00:04.9	00:05.9	00:01.0		確かに:						
6	00:06.2	00:07.3	00:01.1		えっと						
7	00:07.7	00:09.3	00:01.6		新しい技術で:						
8	00:09.3	00:10.4	00:01.1		新しい技術						
9	00:10.1	00:12.0	00:01.9		えっと: 安全性						
10	00:12.6	00:16.0	00:03.4		安全性						
11	00:14.0	00:17.9	00:03.9		)安全性などを考慮した:あどで:						
12	00:18.1	00:18.9	00:00.8		えっと:						
13	00:19.1	00:21.1	00:02.0		運用されていると思いますが:						
14	00:21.6	00:22.4	00:00.9		えと実際に:						
15	00:22.4	00:23.5	00:01.1		日本では:						
16	00:24.1	00:24.6	00:00.6		えっと:						
17	00:24.1	00:24.4	00:00.3		日本						
18	00:24.6	00:26.0	00:01.3		原発事故が:						
19	00:25.1	00:29.1	00:03.0		原発事故						
20	00:26.4	00:27.7	00:01.3		起きていて:						
21	00:29.4	00:30.9	00:01.5		えその:あどの:						
22	00:31.3	00:31.8	00:00.5		えっと:						
23	00:31.8	00:32.9	00:01.0		復興						
24	00:32.1	00:32.3	00:00.2		復興						
25	00:33.2	00:34.5	00:01.3		)復興が:						
26	00:36.4	00:39.9	00:03.5		いまだに問題になって( )いるので:						
27	00:40.3	00:41.5	00:01.2		えっと:						
28	00:41.9	00:43.3	00:01.4		本当に安全なのかどうか						
29	00:41.6	00:44.9	00:03.3		ほんとに安( )ほん( )に安全なのかどうか						
30	00:46.0	00:49.6	00:03.6		が:問題があると( )思います( )あと						
31	00:49.6	00:50.1	00:00.4		えっと						
32	00:50.1	00:51.5	00:01.4		もともと:						
33	00:50.8	00:55.5	00:04.7		もともと任んでいた						
34	00:51.6	00:52.5	00:00.8		もともと						
35	00:52.5	00:53.3	00:00.8		任んでいた						
36	00:54.7	00:55.6	00:00.9		土地の人たち						
37	00:55.5	00:58.8	00:03.3		ところ						
38	00:57.1	00:59.4	00:02.3		のところに:その新しい						
39	00:59.9	01:01.0	00:01.1		原子炉が						
40	01:01.3	01:02.2	00:00.9		できる						
41	01:02.5	01:03.3	00:00.9		などの:						
42	01:03.9	01:06.0	00:02.1		問題もあると思います						
43	01:06.5	01:08.6	00:02.2		あどは:これが:						
44	01:09.2	01:11.0	00:01.7		その日本だけじゃなくて						
45	01:11.1	01:13.0	00:01.9		事故があったときに:						
46	01:13.2	01:14.0	00:00.8		その影響が:						
47	01:15.9	01:17.4	00:01.5		事故の影響が:						
48	01:16.4	01:18.3	00:01.9		事故の影響						
49	01:17.6	01:20.4	00:02.9		国際的な問題になるっていうのも:						
50	01:18.3	01:21.2	00:02.8		国際的な問題						
51	01:20.7	01:22.8	00:02.2		ある( )という風にもあります						
52	01:33.2	01:34.9	00:01.7		えっと( )以上						
53	01:35.7	01:36.1	00:00.4		です						
54	01:58.0	02:03.5	00:05.5			あ:あ:げん:ん:げん:げんしはつ原子力発電					
55	02:03.6	02:04.0	00:00.4			あ:に:					
56	02:04.0	02:05.4	00:01.4			あ:反対					
57	02:04.8	02:07.5	00:02.7			あ:反対					
58	02:05.5	02:07.4	00:01.8			あ:どうですか					
59	02:07.5	02:09.4	00:01.9			あ:私は:					
60	02:09.5	02:12.2	00:02.7			なんかげん( )原子力発電を:					
61	02:12.3	02:15.9	00:03.7			あ:利点がある( )から:賛成します					
62	02:14.5	02:18.9	00:04.4		利点があるから賛成						
63	02:16.6	02:17.5	00:01.0			あ:					
64	02:18.0	02:20.0	00:02.1			あ:今:あん:					
65	02:21.4	02:22.9	00:01.5			環境問題は:					
66	02:22.7	02:25.1	00:02.5		環境問題						
67	02:23.1	02:24.1	00:01.0			国際					
68	02:24.3	02:25.2	00:01.0			国際					
69	02:25.3	02:25.8	00:00.5			的:					
70	02:25.3	02:28.7	00:03.4		国際的な						
71	02:25.9	02:26.4	00:00.5			に:					
72	02:26.5	02:28.7	00:02.2			深刻している					
73	02:30.2	02:32.9	00:02.7			げんし( )力で発電					
74	02:30.8	02:33.9	00:03.1		問題						
75	02:33.3	02:35.2	00:01.9			あ:す( )す( )					
76	02:35.4	02:36.1	00:00.7			すれば					
77	02:36.4	02:37.1	00:00.7			あ:					
78	02:37.4	02:39.7	00:02.4			ま( ):あ:					
79	02:39.8	02:41.0	00:01.3			地球:も:					
80	02:41.3	02:42.4	00:01.1			温暖化( )					
81	02:42.6	02:45.7	00:03.1		地球温暖化						
82	02:43.2	02:44.1	00:00.9			あ:					
83	02:44.4	02:45.6	00:01.2			温暖化を:					
84	02:46.1	02:48.9	00:02.8			あ( )あ( )あえ:					
85	02:49.3	02:51.3	00:02.0			地球( )温暖( )					
86	02:51.6	02:53.4	00:01.9			化の:原因とされる:原因とされる:					
87	02:51.6	03:05.5	00:13.9		二酸化炭素排出						
88	02:53.4	02:55.1	00:01.6		二酸化炭素を						

# 付録 F

## 音声とテキストのタイミングと内容例 2

	A	B	C	D	E	F	G	H	I	J	K	L
1	開始時間 - 終了時間 - 間隔 - 時:テキスト				NS	NNS						
2	00:00.8	00:02.9	00:02.1			よろしくおねがいします						
3	00:02.9	00:03.5	00:00.6			はいよろしくおねがいします						
4	00:03.4	00:04.7	00:01.3			えっと:						
5						死刑の是非について						
6	00:04.8	00:07.0	00:02.2			死刑の是非について						
7	00:07.1	00:07.4	00:00.3			うん						
8	00:07.3	00:07.3	00:00.0			えっと(1)その:そちらがさん餐うん						
9	00:07.3	00:09.0	00:01.7			でえっと:						
10	00:09.5	00:11.6	00:02.1			その:そちらがさん賛成でしたっけ?						
11	00:11.9	00:13.2	00:01.3			はい私は賛成です						
12	00:13.2	00:18.3	00:05.1			あ:は:え:とじゃあ反対で:え:とじゃあまたそれぞれ理由を言っていますか?						
13	00:18.3	00:21.0	00:02.7			ん:理由?私が賛成の理由?						
14	00:21.0	00:22.6	00:01.6			あ:はいじゃあそれでお願います						
15	00:22.6	00:26.1	00:03.5			あ:じゃ:ん:						
16						賛成の理由						
17	00:26.8	00:28.7	00:01.9			家族にとても自分の:						
18	00:28.8	00:31.6	00:02.8			ん:親族とか:あの父とか母とか:						
19	00:32.0	00:32.4	00:00.4			はい						
20	00:32.2	00:33.6	00:01.5			大切な人を						
21	00:33.6	00:35.1	00:01.5			ん:殺人犯に						
22	00:35.1	00:35.9	00:00.8			殺されると						
23	00:36.0	00:36.3	00:00.3			はい						
24	00:36.2	00:37.6	00:01.3			あ:とても怒り						
25	00:37.9	00:39.0	00:01.2			ん:怒っている						
26	00:39.1	00:39.4	00:00.3			うんうん						
27	00:39.4	00:42.9	00:03.5			かもしれない怒ったら:ん:						
28						賛成の理由						
29	00:43.0	00:51.1	00:08.0			司法(?)とか:の法律に:(,)あ:殺人(?)え:殺人を:						
30	00:51.5	00:52.6	00:01.1			殺人犯						
31	00:52.7	00:52.9	00:00.2			うん						
32	00:53.0	00:53.9	00:00.9			の死刑を求めるのは						
33	00:53.9	00:55.5	00:01.7			当然だと思っています						
34	00:55.7	00:56.0	00:00.3			うんうん						
35	00:56.1	01:04.0	00:08.0			犯人に死刑を求める						
36	00:56.1	00:59.0	00:02.9			もしこのような:要求を満足						
37	00:59.6	01:01.2	00:01.7			ん:できなければ						
38	01:01.4	01:02.9	00:01.5			自分:あ:						
39	01:03.7	01:07.1	00:03.4			が殺人犯(?)のほうはこれから						
40	01:07.3	01:08.3	00:01.0			自分:ん:						
41	01:07.5	01:07.6	00:00.2			うん						
42	01:10.4	01:13.8	00:03.4			し(?)にしないでいいと思って:						
43	01:13.7	01:13.9	00:00.2			うん						
44	01:13.9	01:15.0	00:01.1			もっと:						
45	01:15.0	01:31.3	00:16.3			再犯の可能性						
46	01:16.1	01:17.1	00:00.9			もう一度:						
47	01:17.1	01:18.2	00:01.1			殺人する						
48	01:18.6	01:20.2	00:01.5			というくらいがある						
49	01:20.4	01:21.6	00:01.2			は:は:						
50	01:20.9	01:21.2	00:00.2			は:は:						
51	01:22.4	01:23.6	00:01.2			もう一度						
52	01:23.6	01:23.9	00:00.3			うん						
53	01:24.1	01:26.4	00:02.3			犯罪を(するかもしれない)ということですね						
54	01:25.1	01:25.2	00:00.1			はい						
55	01:26.0	01:26.3	00:00.4			はい						
56	01:28.8	01:29.6	00:00.9			は:は:は:( )						
57	01:31.2	01:34.1	00:02.9			賛成の理由として(は)とりあえず(?)そのくらいいいってことですか?						
58	01:34.4	01:35.8	00:01.4			あ:[まず						
59	01:35.9	01:37.3	00:01.4			じゃあ]え:(と)とりあえずじゃあ:						
60						反対の理由						
61	01:37.3	01:39.8	00:02.5			反対もいくつかありますね:						
62	01:39.1	01:39.4	00:00.3			はい						
63	01:41.5	01:43.2	00:01.7			反対としては:						
64	01:43.3	01:43.7	00:00.4			ん:						
65	01:46.3	01:47.9	00:01.6			えっとですね:その:						
66	01:48.3	01:53.1	00:04.8			裁判で:まその死刑になるわけですけども:						
67	01:53.1	01:53.1	00:00.0			その:間違っていた場合?						
68	01:53.1	01:53.8	00:00.6			その:						
69	01:53.2	01:53.5	00:00.2			ん:						
70	01:53.7	01:58.7	00:05.0			冤罪						
71	01:53.8	01:55.5	00:01.7			間違っていた場合						
72	01:56.3	01:56.6	00:00.3			ん:						
73	01:56.6	01:57.7	00:01.1			え:と						
74	01:57.7	01:58.1	00:01.5			冤罪って言葉で						
75	01:59.1	01:59.7	00:00.6			わかりますか?						
76	01:59.7	02:01.5	00:01.8			あ:冤罪(?)わかりました						
77	02:01.3	02:01.8	00:00.5			はい						
78	02:01.8	02:03.5	00:01.8			間違って						
79	02:03.6	02:03.6	00:00.0			死刑にしてしまった場合に:とりかえしがつかない						
80	02:03.6	02:05.8	00:02.2			死刑にしてしまった場合に:						
81	02:06.8	02:06.6	00:00.8			取り返しがつかない						
82	02:07.3	02:07.6	00:00.3			ん:						
83	02:07.7	02:12.8	00:05.1			てことがあるので:え:と死刑になるべくやめたほうがいいんではないかと						
84	02:09.9	02:10.1	00:00.2			ん:						
85						賛成の理由						
86	02:14.1	02:14.5	00:00.4			ん:そうです						
87	02:14.2	02:15.7	00:01.5			で:]その:						
88	02:16.5	02:18.1	00:01.6			取り返しがつかないと						
89	02:18.3	02:19.7	00:01.3			あ:とりかえし(?)						
90	02:19.9	02:20.4	00:00.5			つかない						
91	02:19.9	02:22.2	00:02.2			取り返してだい(?)わかりますか?						
92	02:22.4	02:23.4	00:01.1			あ:はい(?)わかりました						
93	02:24.4	02:26.3	00:01.9			はい(?)あとはその:						
94	02:26.4	02:28.5	00:02.1			ま実際にその:						
95	02:29.7	02:31.9	00:02.1			ま人を殺す						
96	02:32.4	02:33.9	00:01.5			とまその:						
97	02:34.0	02:36.7	00:02.8			ほかの人の権利を侵害してると:						