Factors Affecting Speakers' Choice of Fillers in Japanese Presentations

Michiko Watanabe[1], Yasuharu Den[2], Keikichi Hirose[3], Shusaku Miwa[3] and Nobuaki Minematsu[1]

[1]Graduate School of Frontier Sciences, University of Tokyo, Japan
[2]Faculty of Letters, Chiba University, Japan
[3]Graduate School of Information Science and Technology, University of Tokyo, Japan
{watanabe, hirose, shusaku, mine}@gavo.t.u-tokyo.ac.jp, den@cogsci.l.chiba-u.ac.jp

10.21437/Interspeech.2006-257

Abstract

Disfluencies are intrinsic in spontaneous speech. Although it is known that there is a wide range of frequencies and types of disfluencies among speech, little is known about factors affecting speakers' choice of disfluency types. We first conducted a correspondence analysis using ratios of seven types of fillers, and other disfluencies, in 174 presentations and quantified the data. We conducted a cluster analysis using 4 dimension scores from the correspondence analysis, and extracted five filler-type groups. We then examined frequent types of presentations (formal or casual) and speaker attributes (gender and age) in each group. The results indicate that speakers' choice of filler types is affected by speech levels, speakers' gender and age, and that relevant factors differ depending on the type of fillers.

Index Terms: disfluency type, fillers, sociolinguistic factors, speaker variation, Japanese

1. Introduction

Disfluencies such as fillers, repetitions, false starts, word segments and prolongations are intrinsic in spontaneous speech. It is reported that 9 % of words (disfluencies incl.) are disfluent in presentations in Japanese. The ratio of fillers amounts to 6 % of the total number of words. Fillers are the most frequent type of disfluencies in Japanese [1].

Fillers have been claimed to be relevant to on-line speech production and management of communication. When speakers need some extra time for speech planning, they may utter fillers. By uttering fillers, speakers can not only buy time for planning but also let listeners know their mental processes and maintain communication [2], [3]. It has been reported that there is a wide variety in use of disfluencies [5], [6]. Therefore, mean values of data cannot always be assumed as representative in disfluency studies. We aim at finding characteristic features of speech, speakers, and types of disfluencies by grouping similar kinds of speech together according to frequent types of disfluencies.

Although it has been pointed out that variation is large in use of disfluencies, not much is known about factors which affect speakers' choice of disfluency types. Speech levels, individual speakers' speaking styles and characters of disfluency types can be considered as factors affecting speakers' choice of disfluencies in a given context.

Regarding speech levels, Yokobayashi [4] claimed that Japanese filler, *e*, is typical at the beginning of formal speech.

Concerning speakers' speaking styles, some regularity has been reported. Shriberg [6] pointed out that some speakers can be called "deleters" because deletions are common in their speech rather than repetitions whereas others can be called "repeaters" as repetitions, rather than deletions, are frequent. Sociolinguistic factors such as gender and age may affect speaking styles and choice of disfluency types. It was found both in English and Japanese speech corpora that male speakers produce fillers at higher ratios than female speakers [1], [6]. Total disfluency ratios were also higher for male speakers than for female speakers in Japanese [1].

Regarding characters of fillers, Clark and Fox Tree [7] claimed that English fillers *uh* and *um* have different functions: while *uh* tends to help listener comprehension of speech by signaling short delays of the upcoming speech and heightening listeners' attention, *um* does not have such functions because *um* signals longer delays.

Sadanobu and Takubo [2] argued that Japanese filler, *eto* is used when speakers are conceptualizing a message whereas *ano* is used when speakers are seeking for suitable linguistic forms. Yokobayashi [4] described usage of four frequent types of fillers for learners and teachers of Japanese as a foreign language as follows: use of *ano* gives polite impression to the interlocutor because *ano* indicates that speaker is trying to choose expressions suitable for the situational contexts; *eto* is used when one tries to concentrate on thinking. *Eto* can be uttered when one is on one's own, but not *ano*. Unlike *ano*, it is not polite to use *eto* when one asks for information or requests something; *E* is typical at the beginning of formal speech; *ma* is used when speaker gives comments on something. However, Ito *et al.* [8] argued that no listeners in their experiments reported any unnaturalness when original fillers in speech material were substituted with different types of fillers. It is not clear to what extent functions of different types of fillers actually differ.

In the present study we investigate occurrence of fillers, and other disfluencies, in two types of presentations in Japanese, using "the Corpus of Spontaneous Japanese (CSJ)" [1]. First, we conduct a correspondence analysis to examine whether there is some similarity in frequencies of different types of fillers among presentations. We then conduct a cluster analysis using 4 dimension scores from the correspondence analysis and classify presentations into 5 groups. Finally, we examine frequent types of presentations and speaker attributes (gender and age) in each group to speculate about factors influencing speakers' choice of fillers. We report the analyses in detail in the following sections.

2. Quantification of presentations

2.1. Material

We examined 177 presentations in CSJ. The corpus comprises sound files, transcripts and morphological analyses of 660 hours of 3302 speeches with 7.7 million words. Fillers are treated as words in the morphological analysis of the corpus, and given "F" tags. The list of speech segments classified as fillers in the corpus is found in [9]. About 90% of compiled speeches are academic and casual presentations. Academic presentations were recorded in several academic conferences, majority of which were those of science and engineering fields. The majority of speakers of academic presentations were young male researchers. The presentations can be classified as a formal type of speech. Speakers of casual presentations (called "simulated public speaking" in the corpus) were paid volunteers. They were given general topics such as "the happiest (or saddest) experience in my life", "my town" beforehand and told to prepare for notes for 10 to 15 minutes' talks. The audience was small and the presentations were given in a relaxed atmosphere. They are less formal than academic presentations [10].

We analysed the two types of presentations in a subset of the corpus, called "the Core". Speech data in the Core contain more precise and detailed annotation than the others. The speakers in the Core are limited to speakers of dialects of in and around Tokyo [10].

2.2. Procedure

Three presentations were excluded from analysis because the recording staff of the presentation noted that the speaker read the manuscript. Thus, 174 presentations (68 academic and 106 casual) were retained for further analysis.

First, we classified fillers in 7 types as follows, disregarding difference in the length of vowels or consonants in the transcripts: *ano*, *e*, *eto*, *ma*, *sono*, vowel sounds other than *e* (i.e. *a, i, u, o*) together with a nasal *n*, and others. *E* was separated from other vowel sound fillers because the frequency of *e* is far higher than the others. *Ano* and *sono* are derived from determiners similar to "that" and "the" in English respectively. *Eto* exclusively functions as filler. *Ma* has functions as interjection expressing surprise as well as adverb.

We counted each type of filler in each presentation. As the cumulative ratio of the first 6 types amounted to 99% of the total fillers, we excluded the last category, "others", from analysis. We also merged *sono* with *ano* group because of the low frequency of *sono* (6 samples per presentation on average). We included two other types of disfluencies instead in the analysis: non-lexical prolongations of vowels and consonants (tagged as "H" and "Q" in CSJ, respectively), and word segments and substituted function words (tagged as "D" and "D2" in CSJ, respectively. We call the two types "word segments" for simplicity). We assumed that prolongation of speech segments serves for purposes similar to those of fillers because they also allow speakers time for planning. Word segments can be results of repetition, substitution or discard of parts of words. Table 1 shows the mean frequencies of five types of fillers, prolongations and word segments.

Then, we conducted a correspondence analysis using frequency data of five types of fillers, prolongations and word segments in 174 presentations.

2.3. Results

We selected the first four dimensions based on the cumulative proportions of inertia (see Table 2). These four dimensions explained 84% of the data distribution. Figure 1 shows a symmetric map of the first two dimensions for filler types. The first dimension separated *e* from other types (63% of *e* explained). The second dimension set apart *eto* and prolongations in opposite directions from other fillers (43 % of *eto* and 38% of prolongations explained). Similarly, the third dimension separated *ano/sono* and *eto* in opposite directions from other types (40 % of *ano/sono* and 38% of *eto* explained), and the forth dimension put vowel type fillers apart from other kinds (71% of vowel type fillers explained). The cumulative explanatory ratio for word segments was low (3%) and not reliable.

Table 1: *The mean frequencies of five types of fillers, prolongations and word segments in 174 presentations.*

e	70
ano + *sono*	36
eto	13
ma	31
vowel sounds (*a, i, u, o*) and *n*	18
prolongation	52
word segment	31

Table 2: *Each and cumulative proportions of inertia of the first 4 dimensions.*

	Proportion of Inertia	
Dimension	Accounted for	Cumulative
1	0.43	0.43
2	0.15	0.58
3	0.14	0.72
4	0.12	0.84

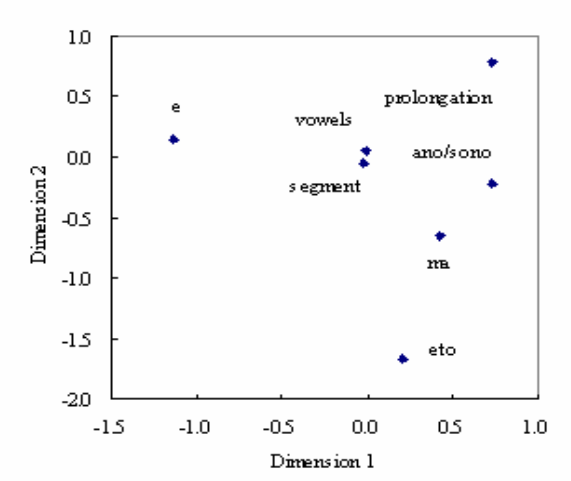

Figure 1: *A symmetric map of the first two dimensions for filler types. The first dimension separates e from other types and the second dimension set apart eto and prolongations in opposite directions from other fillers.*

3. Cluster analysis of presentations

Using the four dimension scores of each presentation obtained through the correspondence analysis, we conducted a cluster analysis by Ward method and extracted five groups. Figure 2 shows the mean frequency of each type of disfluencies per 100 words in the 5 groups. Figure 2 demonstrates patterns of combinations of frequently used disfluency types. According to frequent disfluency types in each group, we named the five groups as follows:

1: *e* type	(53 presentations)
2: *eto* type	(23 presentations)
3: *ma-ano* type	(59 presentations)
4: vowel type	(23 presentations)
5: prolong type	(16 presentations)

In *e* type, *e* occurs at a far higher ratio than any other filler. Similarly, in prolong type prolongations appear far more frequently than any type of fillers. In contrast, there is not such dominant filler in the remaining groups. Although the ratios of fillers which the groups were named after were considerably higher than in the other groups, four or five types of fillers are used at fairly similar ratios in *eto*, *ma-ano* and vowel type groups.

4. Features of filler type groups

4.1. Presentation types

We examined attributes of presentations and speakers in the five filler type groups. Figure 3 shows ratios of academic and casual presentations in each group. The leftmost bar indicates the ratios in all groups. Figure 3 demonstrates that *e*-type is far more frequent in academic presentations than in casual presentations whereas *ma-ano* type and prolong type are more common in casual presentations. *Eto* type and vowel type appear roughly equally in the two kinds of presentations.

Figure 4 describes the ratios of filler type groups in academic and casual presentations. It is obvious from Figure 4 that *e*-type is the most common in academic presentations whereas *ma-ano* type is the most frequent in casual presentations. Figure 4 also shows that prolong type are rare in academic presentations. These results indicate that speech levels strongly affect speakers' choice of fillers. As Yokobayashi [4] claimed, *e* seems a filler for formal settings. *E* appears every 19 words in more than half of the academic presentations. In contrast, the small ratio of prolong type in academic presentations indicates a low speech level of non-lexical prolongations.

4.2. Speaker attributes

Figure 5 demonstrates ratios of male and female speakers in each filler type group. The leftmost bar indicates the ratios in all groups. Figure 5 reveals that vowel type is more frequent among male speakers whereas prolong type is far more common among female speakers. Ratios of male and female speakers are approximately the same in the other filler type groups. These results suggest that speakers' gender also affects choice of disfluency types.

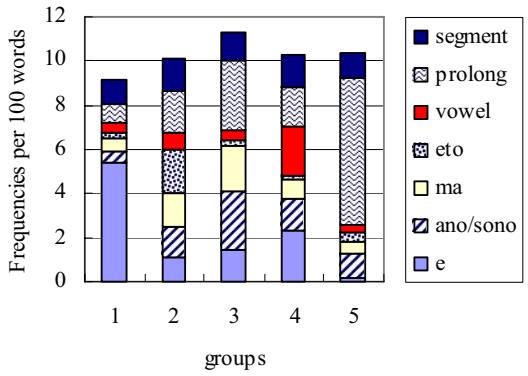

Figure 2: *The mean frequencies of five types of fillers, prolongations and word segments per 100 words in the five groups extracted through a cluster analysis.*

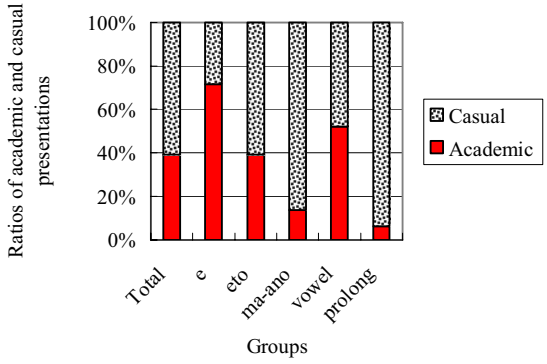

Figure3: *Ratios of academic and casual presentations in total and in each filler type group.*

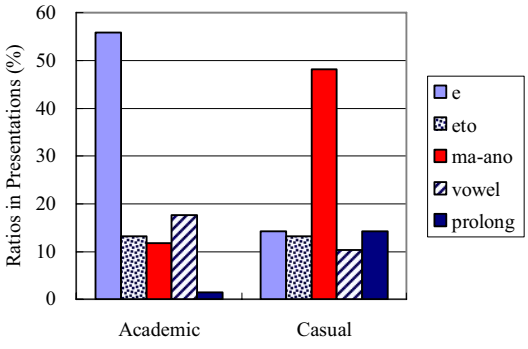

Figure4: *Ratios of filler type groups in academic and casual presentations.*

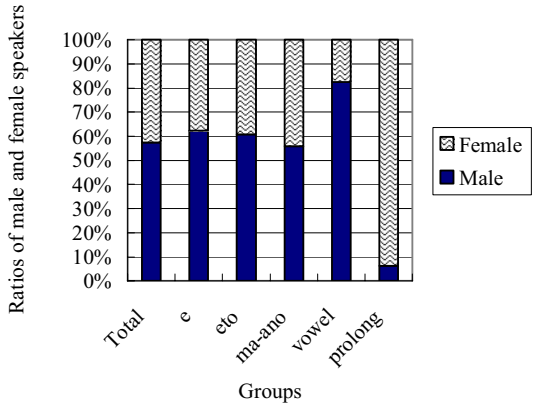

Figure 5: *Ratios of male and female speakers in total and in each filler type group.*

.

Table 3 describes the mean age of speakers in five groups. The rightmost column indicates the mean age of the speakers in all groups. Table 3 shows that speakers in vowel type group tend to be older than in the other groups. In contrast, speakers in *e* type and *eto* type tend to be younger than in the other groups. Figure 6 presents ratios of filler type groups in three age groups, age of 20 to 32 (63 speakers), 33 to 42 (56 speakers) and over 42 (55 speakers). Figure 6 confirms that the ratio of vowel type is higher in the oldest group than in any other group. The ratio of *ma-ano* type is also highest in the oldest group. By contrast, the ratios of *eto* type and *e* type are much lower in the oldest group than in the other groups. There is not large difference in ratios of prolong types among three age groups. These results suggest that speakers' age also affects choice of fillers.

Table 3: *The mean age of speakers in total and in each filler type group*

	e	*eto*	*ma-ano*	vowel	prolong	Total
Age	34	34	42	47	38	39

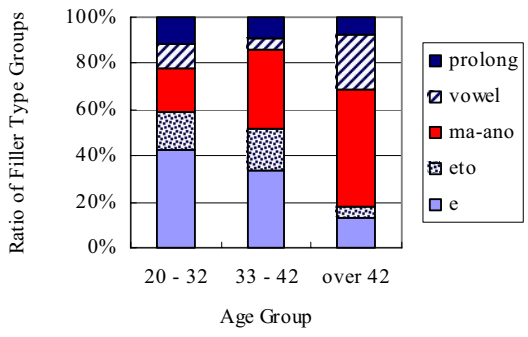

Figure 6: *The ratio of each filler type group in three age groups, age of 20 to 32 (n = 63), 33 to 42 (n = 56) and over 42 (n = 55).*

5. Conclusions

Our study indicates that both speech levels and speaker attributes affect choice of frequent types of fillers. The results suggest that relevant factors differ depending on filler and disfluency types. The speech level seems relevant to the use of *e*, *ma*, *ano* and prolongations. *E* is frequently used in formal presentations, whereas *ma*, *ano* and prolongations are common in casual ones. Speakers' gender is likely to affect the use of vowel type fillers and prolongations. Vowel type fillers are preferred by male speakers than by female speakers. In contrast, prolongations are far more frequent in female speakers' speech than in male speakers'. Speakers' age seems relevant to most types of fillers, but not to prolongations. It appears that vowel type fillers are more for speakers over 40, whereas *e* and *eto* are more for speakers under 40. As a next step, we plan to examine interactions among different factors.

6. Acknowledgements

This research was partly supported by the 21st Century COE (Center of Excellence) Program in Electronics for Future Generations, the University of Tokyo.

7. References

[1] The National Institute for Japanese Language, Homepage of the Corpus of Spontaneous Japanese, http://www2.kokken.go.jp/~csj/public/6_1.html, retrieved in April, 2006.

[2] Sadanobu, T., and Takubo, Y., "The monitoring devices of mental operations in discourse – a case of 'eeto' and 'ano (o)' – ," *Gengo Kenkyu* 108, 74-93, 1995.

[3] Shriberg, E. E., "Spontaneous speech: How people really talk, and why engineers should care", *Proc. Eurospeech*, Lisbon, 1781-1784, 2005.

[4] Yokobayashi, S., 1994. "What are the functions of *eto, ano, ma*, and *e*?" *The Monthly Nihongo*, 5, 68-69.

[5] Itagaki, T., Shinoda, K., and Sagayama, S., "Language modelling using speaker dependency of fillers for spontaneous speech recognition", *Proc. the Second Spontaneous Speech Science and Technology Workshop*, Tokyo, 79-84, 2002.

[6] Shriberg, E. E., *Preliminaries to a Theory of Speech Disfluencies*, PhD thesis submitted to the University of California at Berkeley, 1994.

[7] Clark, H. H., and Fox Tree, J. E., "Using *uh* and *um* in spontaneous speaking", *Cognition* 84: 73-111, 2002.

[8] Ito, T., Minematsu, N., and Nakagawa, S., "Analysis of filled pauses and their use in a dialogue system", *J. Acoust. Soc. Japan*., 55(5), 333-342, 1995.

[9] Koiso, H., Mabuchi, Y., Nishikawa, K., Saitou, M. and Maekawa, K., "Specifications for transcription, Version 1.0", in *CSJ*, 2004.

[10] Maekawa, K., "Outline of the corpus of spontaneous Japanese, Version 1.0", in *CSJ*, 2004.