Variability and Consistency in the Idiosyncratic Selection of Fillers in Japanese Monologues: Gender Differences

Shunichi Ishihara
The Australian National University
Canberra, Australia
shunichi.ishihara@anu.edu.au

Abstract

This study is a linguistic study on idiosyncrasy using speaker classification technique as an analytical tool. The goals of this study are to find out 1) to what extent Japanese filler words (e.g. *um*, *you know* in English) carry individual idiosyncratic information; 2) if there are any differences in the degree/nature of idiosyncrasy between the sexes; and 3) what contributes to the identified gender differences, if there are any. Based purely on the individual selection of fillers, we report in this study that 1) speaker discrimination performance was better in the male (ca. 85% accuracy) than the female (ca. 75% accuracy) speakers by approximately 10%, and 2) the poorer performance of the female speakers was due to the larger within–speaker differences in the female speakers than the male speakers. That is, the selection of fillers by female speakers is more variable, speech by speech, than that by male speakers, even under similar conditions (e.g. same type of audience and the same degree of formality). We also discuss that the findings of the current study agree with the previously–reported differences between the sexes in language use.

1 Introduction

We intuitively know that different people talk/write differently, even when they try to convey the same message. We also know that people tend to use their individually selected preferred words despite the fact that in principle they can use any word at any time from the vocabulary built up over the course of their lives. This is due to the idiosyncratic choice of words, expressions and so forth. Every speaker of a given language has their own distinctive and individual version of the language—which is often referred to as *idiolect* (Halliday et al., 1964; Coulthard and Johnson, 2007). This idiolect manifests itself in various aspects of communication, such as the choice of words, expressions, or even grammar, morphology, semantics and discourse structure. The idiosyncratic nature of word selection between speakers/writers has been studied in different fields. For example, it has been used to understand speaking styles of political leaders (Slatcher et al., 2004), to identify the authors of literary works (Thisted and Efron, 1987), to detect plagiarism (Woolls, 2003) and to enhance the performance of automatic speaker recognition (Doddington, 2001). In the domain of text (in contrast to speech), it has been demonstrated that word category usage is very stable across time and writing topics (Pennebaker and King, 1999).

Besides the idiosyncrasies of individual speakers, men and women speak/write differently (Koppel et al., 2002). This has been well reported in various linguistic and non–linguistic aspects of speech (Lakoff, 1975; Coats, 1993). Particularly in terms of linguistic styles, it has been argued that women tend to be more stylistically flexible and varied than men in language use (Holmes, 1997; Holmes, 1998; Chambers, 1992). Thus, the current study investigates:

- 'to what extent we are idiosyncratic' in selecting certain words rather than others, keeping in mind that there may be some differences in the degree/nature of idiosyncrasy between the sexes; and
- if there are any differences between the sexes, 'what contributes to the identified gender differences' in these instances.

This study focuses on the use/selection of fillers in Japanese as several existing studies identify

Shunichi Ishihara. 2010. Variability and Consistency in the Idiosyncratic Selection of Fillers in Japanese Monologues: Gender Differences. In *Proceedings of Australasian Language Technology Association Workshop*, pages 9−17

subjectively (Furui et al., 2002; Sato, 2002; Yamane, 2002) and empirically (Ishihara, 2009) that preference of fillers exists across speakers. Fillers are unique to spoken language. They are a sound or a word (e.g. *um*, *you know*, *like*) which is uttered by a speaker to signal that he/she is thinking or hesitating. It is reported that 6% of the total number of words spoken in Japanese are fillers (NIJL, retrieved 2008). It is also reported that speakers' attributes, such as age and gender, affect the choices of fillers in Japanese and English (NIJL, retrieved 2008; Watanabe et al., 2006). Studies on speech corpora show that males tend to use fillers more frequently than females in English and Japanese (Shriberg, 1994; NIJL, retrieved 2008).

In order to answer the above research questions, we will conduct a series of speaker discrimination tests—separately between male and female speakers—based solely on fillers. The hypothesis is that the more consistent the individual speaker's selection of fillers is, and the more significantly fillers selected by one speaker differ from those selected by another, the more accurately speaker discrimination can be performed.

We demonstrate first of all that fillers bear the idiosyncratic information of speakers to the extent that the accuracy of the speaker discrimination based solely on fillers can be as high as approximately 85% for male speakers and approximately 75% for female speakers. As can be seen in this difference in accuracy between the sexes, we also report that the speaker discrimination performance is better in the male than the female speakers by approximately 10%. Four reasons can be speculated: 1) the idiosyncrasy was not well modelled for the females due to the fact that less female speakers were used in the current study; 2) the between–speaker difference is larger in the female speakers; 3) the within–speaker difference is larger in the female speakers or 4) any combination of the above three. Further investigation of the data revealed that the poorer performance of the female speaker discrimination compared to the male speaker discrimination is due to the tendency of the female speakers to have larger within–speaker differences than the male speakers. That is, the selection of fillers is more variable or less consistent across the non–contemporaneous speeches of the same speakers for female than male speakers even under very similar conditions.

2 Methodology

Two kinds of comparisons are involved in speaker discrimination tests. One is called *Same Speaker Comparison* (SS comparison) where two speech samples produced by the same speaker need to be correctly identified as the same speaker. The other is, *mutatis mutandis*, *Different Speaker Comparison* (DS comparison).

The series of speaker discrimination tests that we conducted can be categorised into two experiments: Experiments 1 and 2. Detailed procedures of Experiments 1 and 2 are explained in §4 and §5, respectively.

2.1 Database and Speakers

For speech data, we used the Corpus of Spontaneous Japanese (CSJ) (Maekawa et al., 2000), which contains recordings of various speaking styles such as sentence reading, monologue, and conversation. For this study we used only the monologues, categorised as Academic Presentation Speech (APS) or Simulated Public Speech (SPS). APS was mainly recorded live at academic presentations, most of which were 12-25 minutes long. For SPS, 10-12 minute mock speeches on everyday topics were recorded. We selected our speakers from this corpus based on three criteria: availability of multiple and non–contemporaneous recordings, spontaneity (e.g. not reading) of the speech, and speaking in standard modern Japanese. Spontaneity and standardness of the language were assessed on the basis of the rating the CSJ provides. Thus, only those speech samples which are high in spontaneity and uttered entirely in Standard Japanese were selected for this study. This gives us 416 speech samples (= 208 speakers: 132 male and 76 female speakers x 2 sessions).

2.2 Fillers

In CSJ, a filler tag is assigned to one of the pre–selected words given in Table 1 which have the function of 'filling up gaps in utterances'. Some of the words given in Table 1 can also be used as lexical words. If it is uncertain as to whether a given word is used as a lexical word or a filler, additional information is embedded in the tag indicating this uncertainty. In such cases, the word was removed from speaker discrimination tests.

In the selected speech samples, we observed 44

- a(-), i(-), u(-), e(-), o(-), n(-), to(-)†, ma(-)†
- u(-)n, a(-)(n)no(-)†, so(-)(n)no(-)†
- u(-)n(-)(t)to(-)†, a(-)(t)to(-)†, e(-)(t)to(-)†, n(-)(t)to(-)†
- one of the above + {∼desune(-), ∼ssune(-)}
- one of the above with † + {ne(-), sa(-)}

Table 1: Pre–selected fillers in CSJ. "-" stands for the prolongation of the preceding segment.

different filler words for the male speakers and 42 for the female speakers, which are listed in Table 2. As the previous studies which report that more fillers appear in formal speech than informal speech predict (Watanabe, 2009; Nishimura et al., 2010), a large number of fillers could be identified in the selected speech samples as the setting was relatively formal.

Although there are some ranking differences between the sexes in terms of frequency, it can be seen from Table 2 that very similar fillers are used across the sexes.

In order for the choice of filler words to be useful as a speaker classifier, it has to satisfy two criteria. First it has to be consistent within a speaker. The second criterion is the relative frequency of use of fillers compared to other speakers. In order to capture these characteristics, we have to model each speech in terms of the use of fillers, and it needs to be compared against another. We describe our method below.

2.3 Vector space model

Using the frequency counts of the identified fillers, each speech is modelled as a real–valued vector in this study. If n different fillers are used to represent a given speech S, the dimensionality of the vector is n. That is, S is represented as a vector of n dimensions ($\overrightarrow{S} = (F_1, F_2 \ldots F_n)$, where F_i represents the i^{th} component of $\overrightarrow{S}$ and F_i is the frequency of the i^{th} filler). For example, if 5 fillers are used to represent a speech (X), and the frequency counts of these fillers are 3, 10, 4, 18 and 1 respectively, the speech X is represented as $\overrightarrow{X} = (3, 10, 4, 18, 1)$.

2.4 Term frequency inverse document frequency weighting

The usefulness of particular words is determined by their uniqueness as well as how frequently they are used. Different weights were given to different filler words depending on their uniqueness in the pooled data. The *tf · idf* (term frequency inverse document frequency) weight (Formula 1) is used to evaluate how unique a given filler word is in the population, and a weight is given to that filler to reflect its importance to the speaker discrimination (Manning and Schütze, 2001).

$$W_{i,j} = tf_{i,j} * log(\frac{N}{df_i}) \quad (1)$$

In Formula 1, term frequency ($tf_{i,j}$) is the number of occurrences of word *i* (W_i) in the document (or speech sample) *j* (d_j). Document frequency (df_i) is the number of documents (or speech samples) in the collection in which that word *i* (W_i) occurs. *N* is the total number of documents (or speech samples).

2.5 Cosine similarity measure

The difference between two speech samples, which are represented as vectors ($\overrightarrow{x}$, $\overrightarrow{y}$), is calculated based on the cosine similarity measure (Formula 2) (Manning and Schütze, 2001). This particular method was selected as the durations of the speech samples are all different, assuming that the direction of a vector should be constant if the speech sample is long enough.

$$\begin{aligned} diff(\overrightarrow{x}, \overrightarrow{y}) = cos(\overrightarrow{x}, \overrightarrow{y}) &= \frac{\overrightarrow{x} \cdot \overrightarrow{y}}{|\overrightarrow{x}||\overrightarrow{y}|} \\ &= \frac{\sum_{i=1}^{n} x_i * y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} * \sqrt{\sum_{i=1}^{n} y_i^2}} \end{aligned} \quad (2)$$

The range of the difference in two vectors (*diff*($\overrightarrow{x}$, $\overrightarrow{y}$)) is between 1.0 (=$cos(0°)$) for two vectors pointing in the same direction and 0.0 (=$cos(90°)$) for two orthogonal vectors.

Please note that in the experiments of this study (§4 and §5), the length of the vectors were standardised by only looking at the X most frequent fillers ($X = (5, 10, 15, 20, 25, 30, 35, 40)$) as cosine similarity measure requires vectors of equal length.

3 Method for Speaker Discrimination

In this study, the performance of speaker discrimination is assessed on the basis of the probability distribution functions (PDFs) of the difference for two contrastive hypotheses. One is the hypothesis that two speech samples were uttered by the same speaker (the same speaker (SS) hypothesis)

Male: (52588 fillers)						Female: (20575 fillers)					
F	%	F	%	F	%	F	%	F	%	F	%
e-	31.71	u-	1.08	nto	0.03	e-	24.24	e-to-	0.83	to-	0.15
ma	10.84	n-	1.00	to-	0.03	ano-	15.20	u	0.71	eto-	0.13
e	10.62	o	0.83	nto-	0.02	ano	12.86	a-no	0.67	n-to	0.07
ma-	7.85	etto	0.74	u-n	0.01	e	12.07	etto-	0.61	u-n	0.05
ano	6.41	i-	0.59	a-to	0.01	ma	8.27	un	0.45	nto	0.02
ano-	6.19	e-to-	0.57	n-to-	0.008	ma-	5.45	a-no-	0.42	n-to-	0.01
sono	3.31	i	0.35	nto-	0.008	sono	2.74	to	0.35	u-nto	0.005
e-to	3.17	eto	0.30	a-to-	0.006	n	2.45	e-tto-	0.32	nto-	0.005
a	2.52	etto-	0.24	ntto	0.002	a	2.37	o	0.29	ntto-	0.005
o-	2.09	to	0.22	n-tto-	0.002	e-to	2.17	eto	0.28	so-no-	0.005
n	2.07	e-tto-	0.21	u-nto	0.002	sono-	1.29	u-	0.28	unto	0.005
a-	2.01	a-no-	0.21	a-tto	0.002	e-tto	1.22	o-	0.28	so-no	0.005
e-tto	1.59	a-no	0.19	e-ttodesune	0.002	n-	1.13	i	0.25		
sono-	1.49	un	0.09	so-no-	0.002	a-	1.02	i-	0.16		
u	1.20	eto-	0.05			etto	0.88	at	0.15		

Table 2: Fillers and their frequencies (%) of occurrence given separately for the different sexes. F = fillers. "-" stands for the prolongation of the preceding vowel.

and the other is that two speech samples were uttered by different speakers (the different speaker (DS) hypothesis). These probabilities can be formulated as $P(E|H_{ss})$ and $P(E|H_{ds})$ respectively, where E is the difference, H_{ss} is the SS hypothesis and H_{ds} is the DS hypothesis. In this study, the PDF of the difference assuming the SS hypothesis is true is called the SS PDF (PDF_{ss}) and assuming the DS hypothesis is true the DS PDF (PDF_{ds}). Please note again that in this study, the difference of two speech samples refers to the cosine difference between the two vectors representing the two speech samples.

Each PDF was modelled using the kernel density function (`KernSmooth` library of R statistical package). Examples of PDF_{ss} and PDF_{ds}, which are based on all of the male speakers with the dimensions of 40, are given in Figure 1. As can be seen from Figure 1, those PDF_{ss} and PDF_{ds} do not conform to a normal distribution. This is the motivation of the use of the kernel density function in this study.

As can be seen from Figure 1, PDF_{ss} and PDF_{ds} were not always monotonic, resulting in more than a single crossing point, particular when the dimension of a vector is less than 5. Thus, the performance of the system with the length of a vector being less than 5 is not given.

These two PDFs also show the accuracy of this

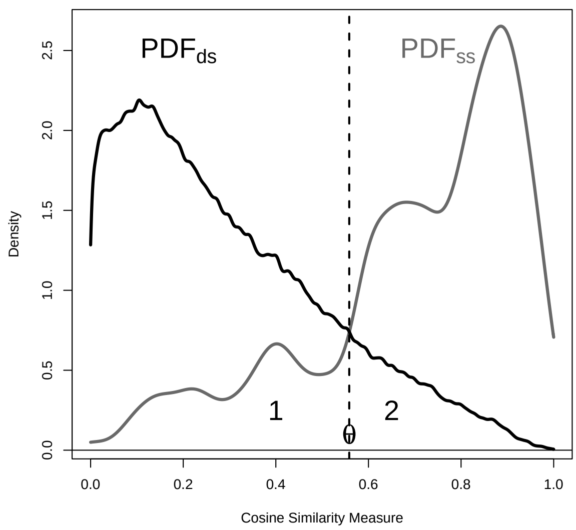

Figure 1: Examples of PDF_{ss} (the grey curve) and PDF_{ds} (the black curve). The vertical dashed line ($x = \theta$) is the crossing point of PDF_{ss} and PDF_{ds}.

particular speaker discrimination system. If the crossing point (θ) of the PDF_{ss} and the PDF_{ds} is set as the threshold, we can estimate the performance of this particular speaker discrimination system from these PDFs. Area 1 in Figure 1—the area surrounded by the grey line (PDF_{ss}), the vertical dotted line of $x = \theta$ and the line of $y = 0$—is the predicted error for the SS comparisons, and Area 2 of Figure 1—the area which is surrounded by the black line (PDF_{ds}), the vertical dotted line of $x = \theta$ and the line of $y = 0$—is the predicted error for the DS comparisons. Therefore, the accuracy of the SS ($ACCURACY_{ss}$) and DS comparisons ($ACCURACY_{ds}$) can be calculated by Formulae 3 and 4, respectively.

$$ACCURACY_{ss} = \left(1 - \int_{0}^{\theta} PDF_{ss}(x)dx\right) * 100 \quad (3)$$

$$ACCURACY_{ds} = \left(1 - \int_{\theta}^{1} PDF_{ds}(x)dx\right) * 100 \quad (4)$$

The accuracy of a speaker classification system (both in SS and DS comparisons) was estimated in this study.

4 Experiment 1: Discrimination Performance and its Difference between the Sexes

In Experiment 1, a series of speaker discrimination tests were conducted separately for the male and the female speakers (132 male and 76 female speakers). Out of the 264 speech samples of the 132 male speakers, 132 SS comparisons and 34584 DS comparisons are possible. Likewise, for the female speakers, 76 SS comparisons and 11400 DS comparisons are possible.

The performance of a speaker classification system is assessed separately for the male and the female speakers as explained in §3, with different numbers of the dimensions of a spacial vector. The spacial vectors of 5, 10, 15, 20, 25, 30, 35 and 40 dimensions are used in Experiment 1. That is, for example, the spacial vector of 5 dimensions means that the frequency counts of the 5 most frequently used fillers are used to represent a speech sample. In Figure 2, the accuracy of a speaker classification system is plotted separately for the male (solid lines) and the female speakers (dotted lines) as a function of the different numbers of dimensions (= fillers). The grey and black lines represent the SS and the DS comparisons, respectively, in Figure 2.

Despite the fact that the techniques used in the speaker discrimination tests are standard and fairly simple, the performance of speaker discrimination is fairly good, particularly for the male speaker discrimination of which accuracy is as good as approximately 85%. It can be observed from Figure 2 that when 20 or more fillers are included in the vectors, 1) the performance of the SS and the DS comparisons becomes stable; 2) the speaker discrimination of the male speakers outperforms that of the female speakers by approximately 10% and 3) the performance of the SS and the DS comparisons becomes comparable. A trade–off between the performance in the SS comparison and that in the DS comparisons is evident if less than 20 fillers are used. The third point above is important that the comparable performance between the SS and the DS comparisons means that the result is well calibrated.

The fact that speaker discrimination performance peaks with half of the dimensions available is not surprising. The feature vectors were based on the frequencies of occurrence of a given filler word, and we first picked ones with higher frequency to be included in the feature. So vectors in the later orders have very low frequencies, such as 0. This means that the latter part of longer vectors tends to include very similar low numbers across speakers, not contributing as a strong unique feature of speakers.

In Experiment 2, we will look into what contributes to the difference in performance between the male and the female speakers.

5 Experiment 2: Why is Female Discrimination Worse?

In Experiment 1, it was demonstrated that the male speaker discrimination outperforms the female speaker discrimination by approximately 10%. Four reasons for this seem to be possible. The first possible reason (R1) is a simple technical and statistical reason. As the number of female speech samples (152 = 76 speakers x 2 sessions) is less than that of male speech samples (264 = 132 speakers x 2 sessions), the idiosyncratic use of fillers was not modelled as well for the female as for the male speakers, resulting in a poor performance for the female speaker discrimination. The second possible reason (R2) is that the between–speaker differences are smaller and less significant in the female than the male speakers. That is, the female speakers behave more uniformly than the male speakers, making the speaker discrimination of the female speakers more difficult. The third possible reason (R3) is that the within–speaker difference is larger in the female than the male speakers. That is, the female speakers are less consistent with their idiosyncrasy in selecting fillers than the male speakers, giving rise to the poorer perfor-

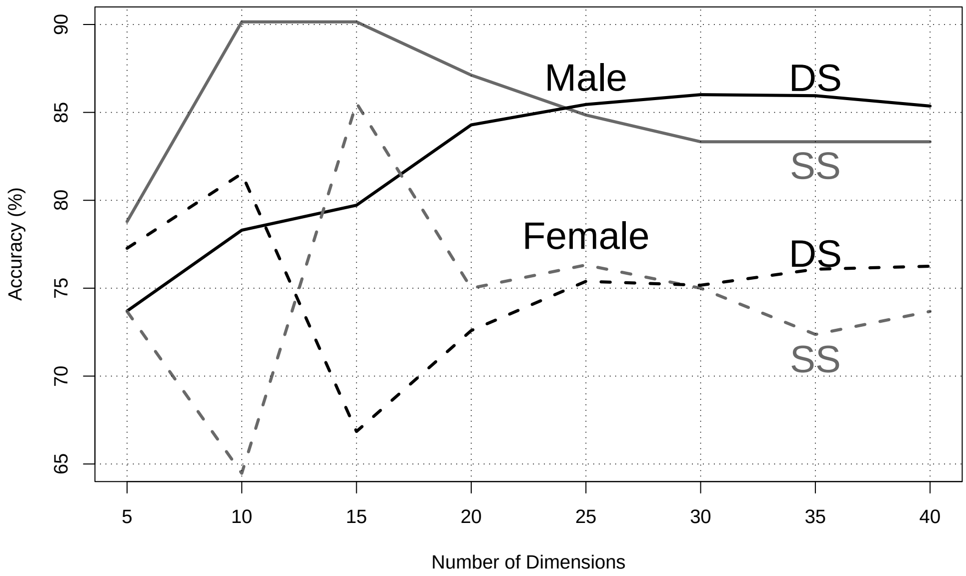

Figure 2: Speaker discrimination performance. The solid lines denote male speakers and the dashed lines denote female speakers. The black lines indicate DS comparisons and the grey lines indicate SS comparisons.

mance of the female speaker discrimination. The fourth possible reason (R4) includes any combination of the above three.

As a first step for identifying which is the true picture contributing to the difference in speaker discrimination performance between the male and female speakers, we need to conduct speaker discrimination tests under the same conditions for both the male and female speakers, equalising the number of speakers. Therefore, 3 male speaker groups, each of which consisted of 76 speakers, were randomly created from the 132 male speakers. Three different speaker discrimination tests were conducted separately using these 3 male speaker groups. If the speaker discrimination performance of these 3 groups of 76 male speakers is similar to that of the 132 male speakers obtained in Experiment 1, we can eliminate R1. In Experiment 2, the spatial vectors of 40 dimensions were used for the speaker discrimination tests. The results of these speaker discrimination tests are summarised in Table 3, together with those of the previous tests.

As can be seen from Table 3, the performance of the male speaker discrimination remains as accurate with only 76 male speakers as with 132 male speakers, the male speaker discrimination outperforming the female speaker discrimination. This result indicates that R1 can be eliminated as a possible reason.

In order to examine the validity of R2$\sim$R4, the differences of paired speech samples that were calculated for the SS and the DS comparisons were scrutinised for the male and the female speakers. Table 4 contains the average differences of pairs of speech samples for the SS and the DS comparisons, which were calculated separately for the male and the female speakers. Let us remind the reader that the value of the cosine similarity measure becomes smaller if the difference of two speech samples is larger. It can be seen in Table 4 that for the DS comparisons, the male (0.29) and the female speakers (0.31) show very similar values, while for the SS comparisons, the average difference of compared speech samples is larger for

Sex	Male				
n	76–1	76–2	76–3	**76–Ave.**	**132**
SS	88.2	89.5	77.6	**85.1**	**83.3**
DS	84.7	82.1	89.0	**85.3**	**85.3**

Sex	Female
n	**76**
SS	**73.7**
DS	**76.2**

Table 3: Comparison of speaker discrimination accuracies (%) under the same conditions for the male and the female speakers. n = number of speakers. The discrimination accuracies when 132 male speakers are pooled together are given as references. The numerals in bold are the values of most concern for the sex comparisons.

the female (0.62) than the male speakers (0.73).

Sex	Male			
n	76–1	76–2	76–3	**76–Ave.**
SS	0.75	0.70	0.75	**0.73**
Skew	-0.97	-0.65	-1.14	**-0.92**
DS	0.29	0.27	0.30	**0.29**
Skew	0.83	0.87	0.83	**0.84**

Sex	Female
n	**76**
SS	**0.62**
Skew	**-0.53**
DS	**0.31**
Skew	**0.81**

Table 4: The average differences of pairs of speech samples for SS and DS comparisons in cosine similarity measure and the degree of skewness for each PDF. n = number of speakers. The numerals in bold are the values of most concern for the sex comparisons.

The above gender difference, i.e. that the female speakers have greater differences than the male speakers for the SS comparisons, can also be seen from the different patterns observed between the PDF_{ss} of the female speakers and those of the male speakers. Figure 3 contains the PDF_{ss} and the PDF_{ds} plotted for the female speaker discrimination test (solid line) and those plotted for the 3 male speaker discrimination tests (dotted lines) conducted in Experiment 2.

Figure 3–1 shows that the PDF_{ds} of the female speakers (solid line) is very similar to those of the male speakers (dotted line)—with the PDF_{ds} of the male speakers being slightly more positively skewed (the average male skew: 0.84; the female skew: 0.81). On the other hand, as for the PDF_{ss} (Figure 3–2), the male speakers (dotted lines) show more negative skewness than the female speakers (solid line) (the average male skew: –0.92; the female skew: –0.53). Statistically speaking as well, three sets of two–sided two–sample Kolmogorov–Smirnov tests (Male 1 vs. Female; Male 2 vs. Female and Male 3 vs. Female) confirm that the distributional pattern is different between the male and female speakers in their PDF_{ss} ($p \leqq 0.04432$).

Thus, it can be concluded that the larger within–speaker difference in the female speakers than the male speakers—which is R3—contributed to the poorer performance in the female speaker discrimination than the male.

6 Discussion

We have demonstrated in Experiment 1 that Japanese fillers carry individual information to the extend that we can discriminate speakers with approximately 85% and 75% accuracy for the male and the female speakers, respectively. 75%∼85% accuracy is not too bad, but not so great as a speaker discrimination task. However, we would like to remind the reader that the techniques we employed are very simple.

In Experiment 2, it has been demonstrated that the male and the female speakers exhibit a very similar pattern for DS comparisons, whilst for SS comparisons they are different in that the female speakers generally have greater differences than the male speakers for compared speech samples. This indicates that the within–speaker difference is larger for the female speakers than the male speakers. In other words, the selection of fillers is more flexible and variable in the female speakers, even under fairly controlled and similar situations, whereas male speakers tend to be more consistent with their selection of fillers.

However, it is not clear at this stage if this is a general tendency observed in many languages or unique to Japanese. Furthermore, we do not know why female speakers are more variable in selecting fillers across non–contemporaneous occasions in comparison to male speakers.

Judging from what has been researched on gender differences in languages, the flexibility/variability of women's speech appears to be one of the universals (Holmes, 1998). Holmes (1997, p. 198) remarks that "women tend to use a wider range of linguistic variants than men, and that their usage varies according to identical contextual factors". A very similar statement can be found in Chambers (1992, p. 199) that "... they [women] command a wider range of linguistic variants ... they have the linguistic flexibility to alter their speech as social circumstances warrant." The flexibility/variability of women's speech in linguistic styles has been empirically supported by various studies (Nichols, 1983; Ide, 1982; Escure, 1991). In speech perception as well, it has been reported that females are more sensitive to variations in speech styles (Wiley and Eskilson, 1985).

Thus, the result of the current study, which demonstrated the differences between males and

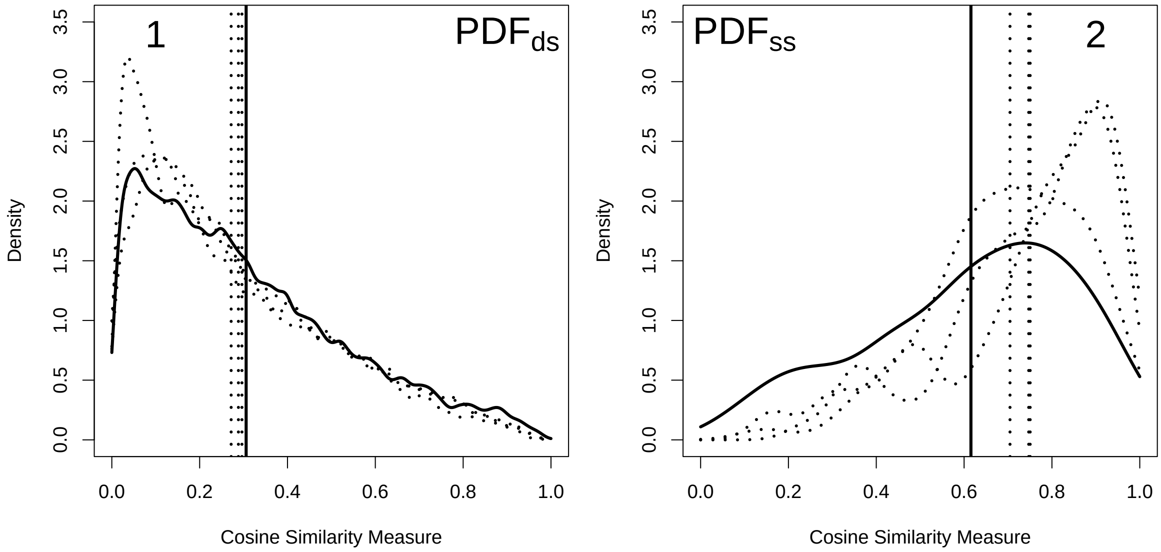

Figure 3: Differences between male and female in PDFs. Panels 1 and 2 are for PDF_{ds} and PDF_{ss}, respectively. The vertical lines are average cosine similarity values. The solid lines are used for the female speakers and the dotted lines are for the three groups of the male speakers.

females in their linguistic styles, contributes further evidence supporting the above assertion, namely that women are more variable and flexible in their linguistic usages. Yet, from the findings of this paper, it is difficult to explain linguistically why females behave in this way.

7 Conclusions

In this study, we have first demonstrated that Japanese fillers carry the idiosyncratic information of speakers. We have shown that the speaker discrimination performance is more effective in the male than the female speakers by approximately 10%. We also have demonstrated that the poorer performance of the female speaker discrimination compared to the male speaker discrimination is due to the tendency for female speakers to have larger within–speaker differences than male speakers. That is, the selection of fillers is more variable, or less consistent, for female than male speakers even under very similar conditions. We have also discussed that the result of the current study conforms to the previously reported differences between males and females in their speech; that women's linguistic use of their language is more variable and flexible in comparison to males.

8 Future Research

Japanese data was used for this study. Thus it is interesting to see if we can recognise the same sex–difference in other languages.

This study is part of a large study on forensic voice comparison (FVC). In FVC, the strength of evidence (or likelihood ratio) is equally important to the discriminability of the system. Therefore, it is interesting to see what sort of strength of evidence can be obtained from the idiosyncratic selection of fillers. Furthermore, FVC usually uses acoustic features, such as Mel–frequency cepstrum coefficients, formant–patterns, fundamental frequency (f0) and so on. The feature used in the current study is a non–acoustic feature (or a text–based feature) which is completely independent from the acoustic features. It thus has potential to make a significant contribution to improving the accuracy of speaker classification systems (Shriberg and Stolcke, 2008) by combining the non–acoustic feature of the current study and usual acoustic features. As a next step, therefore, we intend to extend this study by combining this feature with other, more conventional speaker classification features, such as formants or f0 related features.

Acknowledgements

This study was financially supported by the International Association for Forensic Phonetics and Acoustics and the College of Asia and the Pacific, the Australian National University. The author thanks anonymous reviewers for their valuable comments.

References

M. A. K. Halliday, A. McIntosh and P. Strevens. 1964. *The Linguistic Sciences and Language Teaching*. London: Longman.

M. Coulthard and A. Johnson. 2007. *An Introduction to Forensic Linguistics: Language in Evidence*. London, New York: Routledge.

R. Slatcher, C. Chunga, J. Pennebakera and L. Stone. 2004. Winning words: Individual differences in linguistic style among U.S. presidential and vice presidential candidates. *Journal of Research in Personality*, 41(1):63–75.

R. Thisted and B. Efron. 1987. Did Shakespeare write a newly–discovered poem? *Biometrika*, 74(3):445–455.

D. Woolls. 2003. Better tools for the trade and how to use them. *Forensic Linguistics. The International Journal of Speech, Language and the Law*, 10(1):102–112.

G. Doddington. 2001. Speaker recognition based on idiolectal differences between speakers, *Proceedings of the Eurospeech*, Aalborg, Denmark, September 2001.

J. W. Pennebaker and L. A. King. 1999. Linguistics styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77(6):1296–1312.

M. Koppel, S. Argamon and A. R. Shimoni. 2002. Automatically categorizing written texts by author gender. *Literary and Linguistic Computing*, 17(4):401–412.

R. Lakoff. 1975. *Language and Women's Place*. New York: Harper and Row.

J. Coats. 1993. *Women, Men and Language*. New York: Longman.

J. Holmes. 1997. Women, language and identity. *Journal of Sociolinguistics*, 1/2:195–223.

J. Holmes. 1998. Women's talk: The question of sociolinguistic universals. In, J. Coates (ed.) *Language and Gender: A Reader*. Oxford, 461–483.

J. K. Chambers. 1992. Linguistic correlates of gender and sex. *English World–Wide*, 13:173–218.

S. Furui, K. Maekawa and H. Isahara. 2002. Intermediate results and perspectives of the project 'Spontaneous Speech: Corpus and Processing Technology', *Proceedings of the 2nd Spontaneous Speech Science and Technology Workshop*, 1–6.

Y. Sato. 2002. *'UN' and 'SO' in Japanese Casual Conversation between Native Speakers: The Use of Fillers*. MA thesis, Nagoya University.

C. Yamane. 2002. *Fillers in Japanese Conversation*. Tokyo: Kuroshio publisher.

S. Ishihara. 2009. How diverse are we? An empirical analysis of individual differences in the use of fillers, *The 11th International Pragmatics Conference*, Unpublished paper.

The National Institute of Japanese Language. Homepage of the Corpus of Spontaneous Japanese. http://www.kokken.go.jp/katsudo/seika/corpus/public/6_1.html, retrieved in August, 2008.

M. Watanabe, Y. Den, K. Hirose, S. Miwa and N. Minematsu. 2006. Factors affecting speakers' choice of fillers in Japanese presentations, *Proceedings of Interspeech 2006*, 1256–1259.

E. E. Shriberg. 1994. *Preliminaries to a Theory of Speech Disfluencies*, PhD thesis, University of California, Berkeley.

K. Maekawa, H. Koiso, S. Furui, and H. Isahara. 2000. Spontaneous speech corpus of Japanese, *The Second International Conference of Language Resources and Evaluation (LREC2000)*, Athens, 2000, 947–952.

M. Watanabe. 2009. *Features and Roles of Filled Pauses in Speech Communication, A Corpus–Based Study of Spontaneous Speech*. Tokyo: Hitsuji Shobo Publishing.

R. Nishimura, N. Kitaoka and S. Nakagawa. 2010. Analysis of factors to make prosodic change in spoken dialog. *Journal of Phonetic Society of Japan*, 13(3):66–84.

C. D. Manning and H. Schütze. 2001. *Foundations of Statistical Natural Language Processing*.

P. C. Nichols. 1983. Linguistic options and choices for black women in the rural South. In, B. Thorne, C. Kramarae and N. Henley (eds.) *Language, Gender and Society*. Cambridge: Cambridge University Press, 335–358.

S. Ide. 1982. Japanese sociolinguistics: Politeness and women's language. *Lingua*, 57:357–385.

G. Escure. 1991. Gender roles and linguistic variation in the Belizean Creole community. In, J. Cheshire (ed.) *English around the World: Sociolinguistic Perspectives*. Cambridge: Cambridge University Press, 595–609.

M. G. Wiley and A. Eskilson. 1985. Speech styles, gender stereotypes, and corporate success: What if women talk more like men? *Sex Roles*, 12(9/10):993–1007.

E. Shriberg and A. Stolcke. 2008. The case for automatic higher–level features in forensic speaker recognition, *Proceedings of Interspeech 2008*, 1509–1512.