Potential Speaker-Discriminating Power of Speaking Style: Application of Discourse Information Analysis to Forensic Speaker Recognition

Xin Guan

Abstract: Not only are there differences in the voices that are from different speakers of the same language, but also in the voices that are produced by the same speaker under different conditions or on different occasions. Such nature of human voice makes forensic speaker recognition possible, but difficult. Because in order to link a questioned voice in contact with criminal activity to a known suspect, the forensic speaker recognition expert has to correctly attribute the inevitable differences between two voice samples to either between-speaker differences or within-speaker differences. The parameters that are currently used in forensic speaker recognition are phonetic features, and their speaker-discriminating power has been tested in laboratories with lab-recorded audio materials. However, the forensically realistic conditions are far more complex than ideal laboratory conditions. Moreover, forensically realistic conditions have dramatic effects on forensic phonetic parameters. That is, there is a gap between FSR research and practice as far as the efficacy of forensic phonetic parameters is concerned. To bridge the gap, the study aims to explore non-phonetic features that have the potential to discriminate speakers and at the same time are resistant to within-speaker variability in voice and effects of forensically realistic conditions

Keywords: phonetic parameters; non-phonetic features; speaker-discriminating power; Discourse Information Analysis; individual speaking style

1 Introduction

When audio recordings of an unknown speaker are involved in a

legal case, usually an expert's opinion will be consulted on whether the audio recordings are produced by a known suspect. The expert's opinion may assist in investigation or be admitted as evidence, and the process for the expert to make a decision is forensic speaker recognition (FSR). FSR is the application of theories and methods in forensic phonetics that is an important branch of forensic linguistics (Du 2004: 61).

With the development of computer science, a variety of handy equipment and software to record and edit voice are becoming more and more popular. As a result, more and more audio recordings are getting involved in legal areas, and demands for FSR are increasing.

The fact that different speakers of the same language or dialect have different voices makes FSR possible. But it is also a fact that "the voice of the same speaker will always vary" (Rose 2002: 9) as a consequence of change in the speaker's age, health and emotional state, and communicative intend etc. (Alexander & McElveen 2007). This phenomenon is termed as *within-speaker variability* in voice.

Due to within-speaker variability in voice, what an FSR expert needs to do is to decide that the inevitable differences between voice samples to be compared are more likely to be between-speaker differences or within-speaker differences (Rose 2002: 9). However, in forensically realistic conditions, it is hardly possible to know such information as the questioned speaker's age, his health and mental state while he was speaking. Within-speaker variability in voice has become the main factor to restrict the development of FSR currently (Zhang 2009: 19).

Subject to the nature of human voice, there is a gap between FSR research and practice. The parameters that have been being explored in research and then used in practice are mainly phonetic features. Those phonetic FSR parameters are usually tested with audio materials intentionally recorded in ideal laboratory conditions so that the factors resulting in within-speaker variability are known and under control. But in practical casework total control and knowledge of these factors are impossible (Rose 2002: 18-20). Consequently, it becomes extremely difficult to attribute inevitable differences between voice samples in practical casework.

With a view to the increasing demands for FSR, it is necessary

and urgent to bridge the gap between FSR research and practice. Rose (2002: 92) suggests simulating real-word conditions in experiments by using as many as possible similar-sounding subjects' non-contemporaneous natural conversations when testing the efficacy of forensic phonetic parameters.

Following Rose's suggestion and considering the nature of human voice, this study makes an attempt on the adoption of natural conversations in the experiment in order to explore potential non-phonetic FSR parameters that are resistant to the within-speaker variability in voice and effects of forensically realistic conditions.

2 Relevant literature

2.1 Criteria for FSR parameters and requirements of forensic comparison sciences

On the basis of the criteria for an ideal acoustic FSR parameter set out by Nalon (1983: 11), Rose (2002: 51) sums up the following six criteria for an ideal FSR parameter that are applicable to any type of FSR parameters:

1) show high between-speaker variability and low within-speaker variability;

2) be resistant to attempted disguise or mimicry;

3) have a high frequency of occurrence in relevant materials;

4) be robust in transmission;

5) be relatively easy to extract and measure;

6) each parameter should be maximally independent of other parameters.

Rose (2002) points out that there is no ideal parameter that meets all six criteria and the most important criterion is a high-ratio of between-speaker to within-speaker variation.

Researchers agree (Nalon 1983: 101; Pruzansky & Mathews 1964; Rose 2002: 17; Wolf 1972) that the common way of selecting potentially useful FSR parameters is to inspect the ratio of between-speaker to within-speaker variation, that is, the *F*-ratio. *F*-ratio is usually a by-product of the Analysis of Variance. Thus, classical statistical discrimination analysis can be used to determine the discriminating power of FSR parameters (Rose 2002: 17).

Now it is in the midst of a *paradigm shift* in the evaluation and

presentation of evidence in the forensic comparison sciences (Morrison 2009). The shift requires that forensic evidence should be evaluated and presented in a logically correct manner. As a result, if FSR expects hope to achieve the degrees of reliability needed to serve the goals of justice, the likelihood-ratio framework that has been used as standard for DNA profiles since 1990s has to be adopted. Parameters used within the likelihood-ratio framework should be quantifiable (2009).

In short, effective FSR parameters in an FSR system should not only meet the six criteria summarized by Rose above, but also be quantifiable in order to meet the requirements of the ongoing paradigm shift in the forensic comparison sciences.

2.2 Types of currently-employed FSR parameters

Forensic phonetic parameters are the currently-employed FSR parameters in FSR practice which are complemented with such linguistic features as regional and social accents. Regional and social accents are usually used to profile a speaker according to his/her group identity, while it is phonetic features that are used to be FSR parameters to recognize a speaker.

Rose (2002: 32) categorizes forensic phonetic parameters into four main types: linguistic auditory-phonetic, non-linguistic auditory-phonetic, linguistic acoustic-phonetic and non-linguistic acoustic-phonetic. Generally, auditory-phonetic parameters are qualitative, and acoustic-phonetic features are quantifiable.

Linguistic auditory-phonetic parameters reflect the speaker-specific features with respect to the speaker's sound system and the way the sound system is realized. For example, how a consonant or vowel is realized. Non-linguistic auditory-phonetic parameters usually reflect such speaker-specific features as phonation type, and pitch range wide, which do not have to relate directly to individual speech sound as the linguistic auditory-phonetic features do. For example, whether a speaker's phonation type is whispery.

Linguistic acoustic-phonetic parameters reflect the acoustic features relating to speech sounds. For example, the acoustic features of a certain vowel. Non-linguistic acoustic-phonetic parameters usually reflect the features of the speaker's vocal apparatus, that is,

the features reflecting the shape and size of vocal tract.

In addition, Rose (2002: 39-41) classifies acoustic parameters into traditional and automatic. The linguistic and non-linguistic acoustic parameters introduced above are traditional acoustic parameters. Different from traditional parameters, automatic parameters do not relate to the linguistic auditory or articulatory properties of speech sounds, which are the mathematical abstraction of certain acoustic features of sound signal and used in automatic methods.

2.3 The nature of human voice and the problems of forensic phonetic parameters

Human voice is different from such biometric characteristics as DNA or fingerprints that are considered to be unique and in direct contact with the individual (Alexander & McElveen 2003; Morrison 2009; Nalon 1997).

But as a matter of fact, different speakers of the same language do differ in some aspects of their speech (Rose 2002: 325), which makes it possible to recognize a speaker from his voice. On the other hand, "the same speaker can differ in some aspects of their speech on different occasions, or under different conditions" (2002: 333). The existence of within-speaker variability means that there are always differences between two speech samples no matter whether or not they are produced by the same speaker (Coulthard & Johnson 2007: 148; Morrison 2009; Rose 2002: 10). It makes FSR difficult and controversial in practice (2002). Because due to the nature of human voice no one hundred percent match can be achieved between any two voice samples even if they are of the same origin.

Rose (2002: 270) emphasizes that to be able to accurately attribute the differences between voice samples, the internal composition of a voice must be understood. However, the internal composition of a voice is complex and so far it is difficult to understand all the complexities (2002: 270-95).

Nalon (1997: 749) defines a speaker's voice as the "interaction of *constraints* imposed by the physical properties of the vocal tract, and *choices* which a speaker makes in achieving communicative goals through the resources provided by the various components of

his or her linguistic system". Figure 1 shows the components of a voice. In the course of the interaction, the two mechanisms (linguistic and vocal) process two inputs (communicative intent and intrinsic indexical factors) and output a speaker's voice.

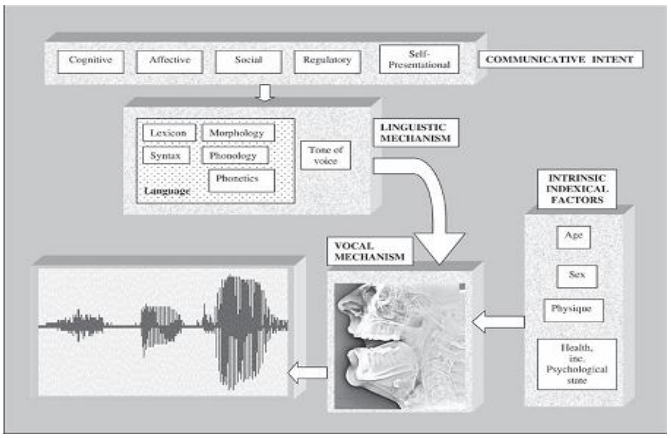

Figure 1. A voice model (from Rose 2002: 278)

Rose (2002: 295) points out that within-speaker variations are a function of a speaker's communicative intent and the dimensions and condition of his individual vocal tract. Communicative intent decides what is conveyed and reflects the effects of contexts on the speaker. A speaker's vocal tract imposes limits, instead of absolute values, to the ranges of phonetic features that his language makes use of.

That is, within-speaker variations result from the interaction between a speaker and the contexts in which he is speaking, and the complexities of the interaction have not been totally understood yet.

Therefore, lab-recorded audio materials have been being used as experimental materials in FSR research to test the efficacy of forensic phonetic parameters so that the sources of within-speaker variability can be under control and known to ensure the correct attribution of the differences between samples.

But, in practice not all sources of within-speaker variability are known or under control. For instance, there usually lacks the information about the questioned speaker's intrinsic indexical factors, like age, sex, health and psychological state. On the other hand, even in ideal laboratory conditions, as for the known speaker, the total

control of his communicative intent is impossible. Figure 2 demonstrates the factors that can cause within-speaker variability in voice in real-world conditions and their interactive relationship.

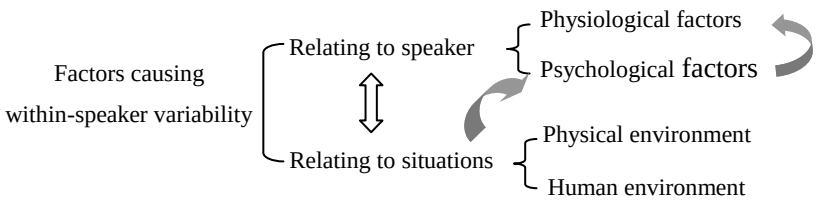

Figure 2. Factors Causing Within-Speaker Variability and Relationship among them (from Guan, 2014a)

Figure 2 shows that all factors relating to either speaker or situations may result in within-speaker variability in voice. It is also obvious that the total control of all factors relating to either speaker or situations in practice is hardly possible. Consequently, there is little guarantee that the lab-tested forensic phonetic parameters still show high between-speaker variability and low within-speaker variability, remain resistant to attempted disguise or mimicry, and keep robust in transmission under the effects of many unknown or uncontrolled factors when they are used to compare voice samples in practice. Moreover, Rose (2002: 20-30) summarizes that the forensically realistic conditions reduce the number of available parameters and distort the parameters.

In addition, as far as auditory-phonetic parameters are concerned, it is difficult to make them quantifiable. The evaluation of these parameters depends on the individual expert's knowledge of linguistics and phonetics, experience, his familiarity with the language/dialect as well as his listening ability (Hollien 1990: 205). As far as acoustic-phonetic parameters are concerned, they are quantifiable, but they are more sensitive to real-world conditions (Broeders 2001; Bijhold *et al.*, 2007; Jession 2010; Rose 2002: 36-41).

In summary, owing to the nature of human voice, neither qualitative auditory-phonetic parameters nor quantifiable

acoustic-phonetic parameters are immune to the effects of forensically realistic conditions. The effects of forensically realistic conditions lead to within-speaker variability in voice. As a result, the within-speaker variability reduces the validity and reliability of forensic phonetic parameters in practice in that their efficacy has been tested in the controlled lab conditions instead of in real-world conditions.

2.4 Available solutions and the inadequacy

So far, two solutions to the problems of forensic phonetic parameters have been suggested. One way is to try best to simulate real-world conditions when the efficacy of forensic phonetic parameters is tested. The other is to explore non-phonetic parameters and then test them with natural audio materials that occur in real-world conditions.

2.4.1 Simulating real-world conditions when testing phonetic parameters

The literature reviewed above illustrates that forensic phonetic parameters are affected by the factors relating to the speaker and the contexts in which he is speaking, and these factors are far more complex and far more difficult to control in forensically realistic conditions than in ideal lab conditions. Rose (2002 92-93) suggests that experiments to test the efficacy of forensic phonetic parameters should attempt to simulate real-world conditions as closely as possible through using non-contemporaneous natural conversation and as many as possible similar-sounding subjects.

Rose (2002) suggests Eliiott's map task to elicit natural conversation. The map task (Elliott 2001) required the caller to guide his friend through a predetermined route that had been marked on the map. Because the caller and his friend used two similar but not identical maps, they had to negotiate the differences between their maps. In the course of their negotiation, the tokens to be examined were elicited. The caller, as the subject to be examined, was recorded in the laboratory.

Morrison, Rose, and Zhang (2012) suggest information exchange task over the telephone to elicit natural conversations. The task required one speaker to confirm with another speaker the

information of numbers and letters that is illegible in their faxes. Both of them were recorded with specific equipment at least in quiet rooms.

However, *simulating* is a kind of control. It is obvious that either map task or information exchange task had imposed control on the speakers' communicative intent as well as the contexts in which they were speaking. Figures 1 and 2 display that any degree of control will give the butterfly effect. In other words, the problems of forensic phonetic parameters cannot be solved by simulating real-world conditions.

2.4.2 Exploring non-phonetic parameters

Guan (2014a) suggests cross-validation method that validates comparisons of speech sound and individual speaking style. She thinks that speech as a whole should be taken as the object of investigation in FSR instead of speech sound only in that speech can be compared in terms of both phonetic parameters and non-phonetic parameters representing a speaker's individual speaking style, and their outcomes can validate each other.

2.4.2.1 The object of investigation in FSR – speech

Rose indicates that the linguistic mechanism refers primarily to the aspects of the structure of the speaker's language including phonetics, phonology, morphology, and syntax (see Figure 1). He overlooks the aspect of semantics when describing the components of a voice. However, it is a fact that voice carries information what a speaker intends to convey. Moreover, the primary function of language is for communication. That is, in the process of producing voice, what is finally output is not just only voice, but is speech that reflects the speaker's communicative intent. The speech is composed of the voice and information, see Figure 3, and the voice is kind of the container of information.

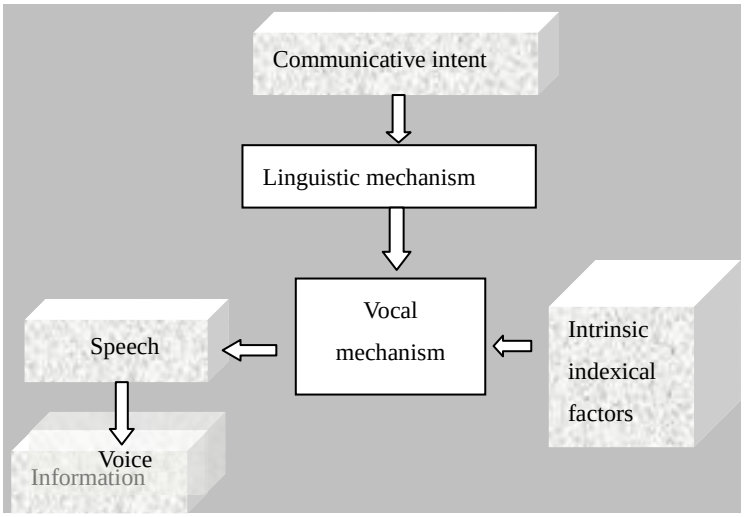

Figure 3. The Components of Human Voice

As far as speech is concerned, speaker-specific characteristics reflecting a speaker's identity are embedded at all levels of speech (Alexander & McElveen 2007). Sapir (1927) treats speech as human behavior and defines five levels of speech. The five levels are the voice as such, the speech dynamics, the pronunciation, the vocabulary, and the individual style of connected utterance. So far, the application of speaker-specific features at the first four levels of speech to FSR has been documented thanks to their direct contact with voice. The features at the last highest level have not been dealt with due to their non-direct correlation with voice.

The highest level, the *individual style of connected utterance*, is defined as "an individual method of arranging words into groups and of working these up into larger units" (1927). In the light of the definition, this level appears to have nothing to do with voice, the container, but have something to do with information. Speaker-specific features at this level tend to be immune to within-speaker variability in voice. Natural conversations are proper experimental materials used to investigate information contained in voice.

Considering the distribution of speaker-specific features in speech, it is reasonable to define speech as the object of investigation in FSR. With speech to be the object of investigation, it is possible to

explore the non-phonetic parameters concerning the individual style of connected utterance through adopting natural conversations in experiments.

2.4.2.2 The nature of individual style of connected utterance and the approach to the analysis of speech

Individual style of connected utterance is in fact a speaker's individual speaking style in light of Sapir's definition. Sapir (1927) announced that it was theoretically possible to analyze individual speaking style but it would be a very complicated problem to disentangle social determinants from the individual ones.

Guan (2014a) argues that discourse analysis methods are appropriate to analyze speech to explore individual speaking style features based on the comments on discourse analysis methods by Johnstone (see Johnstone 1996: 24), Du (see Du 2008) and Qian (see Qian 2006).

Individual speaking style represents a speaker's individuality that he displays through his talk (Johnstone 1996: 7) and it appears "more or less consistent over time and situation" (1996: 5). Moreover, it characterizes a speaker just as gaits, facial expressions, and ways of dressing do (1996: 129), and none of social and psychological factors as well as changes in rhetorical situation "causes people to talk one way or another" (1996: 55). That is, speakers can seldom impose conscious control on individual speaking style, and it tends to remain consistent under different conditions and on different occasions. It means that it is relatively stable and is not affected by the forensically realistic conditions. Therefore, individual speaking style parameters are resistant to attempted disguise or mimicry and robust in transmission. To be specific, they meet the two criteria for ideal parameters in nature according to which the performance of phonetic parameters are reduced and doubted in practice.

Similarly, there is a layer in discourse that speakers can seldom impose conscious control whose structure is relatively stable compared with the flexible language forms (Du 2011). The layer is defined as *discourse information* by one new discourse analysis method, *Discourse Information Analysis* (DIA). Logically, it is hoped that the application of DIA to the analysis of speech can explore some

individual speaking style features that have the potential to be FSR parameters.

2.5 Discourse Information Analysis

DIA has grown out of the Tree Model of Discourse Information (cf. Du 2007). The Tree Model defines *discourse information* as proposition. Proposition is also the minimal and complete cognitive meaning unit. Each proposition is an *Information Unit*, which is the minimal and complete communicative meaning unit with a relatively independent structure.

According to the tree model, the surface layer of discourse is language, the underlying layer is cognition, and information lies in between. Discourse information is more stable compared with the surface layer of language, and is more accessible compared with the underlying layer of cognition. Analysis of discourse information structure makes investigation of discourse producer's cognitive structure more direct and more reliable than analysis of flexible language forms (Du 2007; Du 2011).

The information units in discourse are hierarchically structured like an inverted tree, see Figure 4.

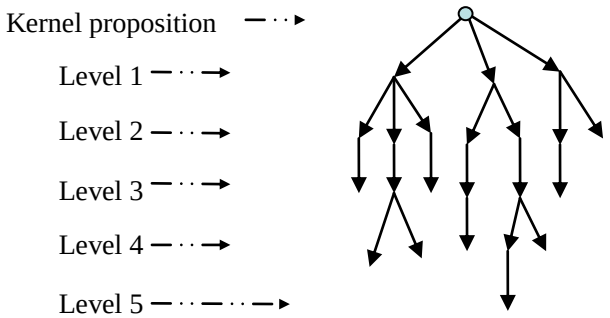

Figure 4. The Tree Structure of Discourse Information (from Du, 2013)

Each discourse has only one kernel proposition, which is developed by information units at different levels. Each information unit develops its superordinate information unit and their relationship is termed as *information knot* and represented by one of the 15 interrogative key words, see Table 1.

Table 1. Types of Information Knots

Abbreviation	***Interrogative Word***	***Abbreviation***	***Interrogative Word***
WT	What Thing	HW	How
WB	What Basis	WY	Why
WF	What Fact	WE	What Effect
WI	What Inference	WC	What Condition
WP	What Disposal	WA	What Attitude
WO	Who	WG	What Change
WN	When	WJ	What Judgment
WR	Where		

In addition, at the micro level, each information unit consists of information elements. There are three main types of information elements, *Process*, *Entity*, and *Condition*. Every type has its sub-types, see Table 2.

Table 2**.** Types of Information Elements

Type	***Abbreviation***	***Type***	***Abbreviation***	***Type***	***Abbreviation***
Process	**P**	**Entity**	**e**	**Condition**	**c**
State	S	Agent	a	Instrument	i
Quality	Q	Dative	d	Location	l
Relation	R	Patient	p	Source	s
Affect	A	Fractitive	f	Goal	g
Cause	C	Attribute	b	Commititive	c
Turn	T			Time	t
Behave	B			Affected	a
Negation	N			With	w
				Basis	b
				Manner	m
				Elaboration	e
				Situation	o

Further, in order to develop forensic application research, CLIPS (the Corpus for Legal Information Processing System) has been

constructed and put into use, in which such types of annotated data as texts, conversations, videos, images, and photographs are stored. In addition, a forensic linguistic laboratory has been established in which hardware and software systems to analyze speech signal are equipped with.

In short, DIA, along with CLIPS, offers strong support in the respects of theory, methodology, analysis tools, and qualified data to the analysis of natural conversations, which aims to discover non-phonetic parameters that represent the speaker's individual speaking style and are immune to within-speaker variability in voice.

To sum up, the literature reviewed above illustrates the problems of phonetic parameters and advantages of exploring non-phonetic parameters, and then suggests a perspective of exploring non-phonetic parameters. As the first step of exploring non-phonetic individual speaking style parameters, the following experiment is designed to demonstrate that individual speaking style features are possibly explored by the application of DIA to the analysis of speech and to test that they have the potential to discriminate speakers.

3 Experiment design and research procedures

3.1 Experiment design

As far as individual speaking style is concerned, it has been considered to be theoretically speaker-specific and analyzable. DIA is predicted to be an appropriate discourse analysis approach that can be used to analyze natural conversations to extract parameters reflecting individual speaking style. On this basis, the experiment intends to verify the assumption that individual speaking style is potentially speaker-specific.

To achieve the goal, two experiments were designed and conducted one after another in light of the nature of individual speaking style, see Figure 4.

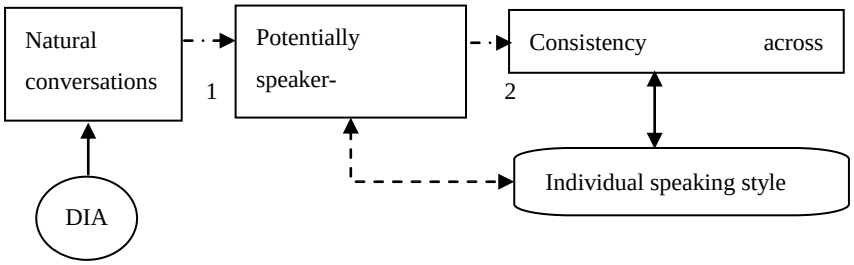

Figure 5. Experiment Flowchart

The first experiment intends to test the potential speaker-discriminating power of those features that are discovered through the application of DIA to the analysis of natural conversations. If the features were tested to be potentially speaker-specific, the second experiment would be activated and conducted that aims at examining the consistency of those discovered potentially speaker-specific features under different conditions and on different occasions.

According to the nature of individual speaking style, individual speaking style features should be speaker-specific, and at the same time stay consistent within a speaker regardless of the contexts. Thus, consistency is the proof that the discovered potentially speaker-specific features represent individual speaking style. In other words, only when consistency is available for the discovered potentially speaker-specific features, will the assumption be finally positively verified that individual speaking style is potentially speaker-specific. In both experiments, natural conversations that occurred and were recorded under real-world conditions were adopted as experimental materials.

3.2 Data

All natural conversations used in the experiments were sampled from CLIPS, where both audio file and annotated character-to-character transcribed text for every conversation are available.

The conversation data in CLIPS are face-to-face or telephone conversations occurring in real-world conditions. The conversations were recorded with mobile phone's build-in recording software

automatically, or with digital voice recorders. Before being input and stored in CLIPS, the speakers have affirmed that they were unaware of the recording process, which ensures the naturalness of the input conversations.

All speakers are the postgraduates from the School of English for International Business in Guangdong University of Foreign Studies and are at the age of 20-25 and speak good standard Chinese.

29 conversations from 13 speakers were randomly sampled as the experimental materials, whose basic information is displayed in Table 3.

The 29 conversations sampled from 13 speakers were numbered serially from 1 to 29, so were the 13 speakers from S1 to S13, as listed in the first two columns of Table 3. The information of each speaker's sex (*F* for female, *M* for male), age, and duration of each conversation are also listed in Table 3.

Table 3. Basic Information of Sampled Conversations

No.	***Speaker***	***Sex***	***Age***	***Duration (m:s)***	***Character number***	***Time***	***Medium***
1	S1	F	22	00:58	214	12/01/2013	F
2	S2	F	23	01:15	81	12/03/2013	F
3	S3	F	25	01:12	164	12/13/2013	F
4	S4	F	23	00:23	49	12/07/2013	F
5	S5	F	23	01:07	187	12/13/2013	F
6	S6	F	23	01:08	133	12/04/2013	T
7	S7	F	23	00:48	96	12/05/2013	F
8	S8	F	23	00:30	87	12/15/2013	F
9	S9	F	23	01:10	154	12/07/2013	F
10	S10	F	21	01:15	188	12/12/2013	F
11	S11	F	23	01:17	113	12/08/2013	F
12	S11			01:10	126	12/12/2013	F
13	S11			01:16	173	12/19/2013	F
14	S11			01:03	106	12/16/2013	F
15	S11			00:50	90	01/02/2014	T
16	S12	F	21	01:11	137	11/15/2013	F
17	S12			01:20	87	12/08/2013	T
18	S12			01:12	188	12/11/2013	F

19	S12			01:09	160	12/15/2013	F
20	S12			01:09	203	12/25/2013	F
21	S12			00:45	177	12/05/2013	T
22	S13	M	25	00:44	76	12/03/2013	F
23	S13			01:00	158	12/13/2013	F
24	S13			00:39	82	12/11/2013	F
25	S13			02:26	114	12/24/2013	F
26	S13			01:00	152	12/14/2013	T
27	S13			01:12	149	12/21/2013	T
28	S13			01:01	79	12/01/2013	T
29	S13			01:28	103	01/09/2014	T

Since the duration of each conversation in the table refers to the total duration when each conversation lasts between the interlocutors, the number of Chinese character produced by the sampled speaker is given in the sixth column to measure the length of the speech produced by the sampled speaker only. That explains why a longer conversation appears to consist of fewer Chinese characters as Conversation No.2 displays. The column of *Time* indicates the exact time at which each conversation was occurring. In the column of *Medium*, *F* indicates a face-to-face conversation, and *T*, a conversation on telephone.

3.3 Procedures and measures

3.3.1 Procedures

In consideration of the fact that each information unit has multiple values, and of the requirement that the FSR parameters must have a high frequency of occurrence in relevant materials (Rose, 2002: 51), to begin with, the values wanting to be investigated have to be determined.

Once the values to be investigated are determined, the sampled conversations are to be analyzed in terms of these values with DIA to extract potentially speaker-specific features. Next, Experiments 1 and 2 are to be conducted to test the potential of the extracted features to discriminate speakers and the extent to which they represent a speaker's individual speaking style.

3.3.2 Measures

Firstly, the distribution of 15 types of information knots was displayed in the form of the percentage of conversations containing each type in the sampled data set（the first row in Table 4）and of the occurrence percentage of each type in all conversations as a whole (the second row in Table 4).

Table 4. Distribution of 15 Types of Information Knots

	WT	*WB*	*WF*	*WI*	*WP*	*WO*	*WN*	*WR*	*HW*	*WY*	*WE*	*WC*	*WA*	*WG*	*WJ*
(%)	100	13.8	62.1	24.1	27.6	17.2	24.1	17.2	6.9	6.6	24.1	17.2	51.7	6.9	0
(%)	51.8	0.8	13.2	3.3	3.0	1.7	3.6	1.9	0.8	6.6	2.2	1.7	8.3	0.3	0

Table 4 illustrates that the information knot of *WT* presents in all conversations (100%), and meanwhile occurs well above other types of information knot (51.8%). Therefore, the information unit at the knot of *WT* will be observed so as to extract potentially speaker-specific features.

After careful observation, two features were selected as the potentially speaker-specific features. One feature concerns information unit, and another concerns information elements. The first feature is the duration of the information unit at the information knot of *WT* and is measured in millisecond, which is represented by *P1*. Different from other measures of speech tempo, here, information unit is set as the measure, which includes all kinds of pauses. *P1* was measured in the Forensic Linguistic Laboratory with CSL4500. The second feature is the ratio between the total number of information elements in each WT information unit and the total number of information elements in the conversation being observed, which is represented by *P2*. *P2* is expected to reflect a speaker's strategy to organize information elements in a conversation.

4 Experiments

4.1 Experiment 1

Experiment 1 intends to prove that the two features *P1* and *P2* are potentially speaker-specific. It means that they can distinguish speakers to some extent through working together or separately.

Usually, in FSR research speakers in dataset used to extract and

test discriminating ability of FSR parameters should be kind of homogeneous as far as certain basic information is concerned. For instance, the speakers are expected to be of the same sex, and similar in age, dialect region, and voice quality, etc. when acoustic parameters are to be compared (Morrison 2010). Hughes *et al*., (2013) exemplified that more broadly, sociolinguistically homogeneous speakers were qualified, and they considered the DyVis speakers in their research to be sociolinguistically homogeneous who "are all young (aged 18-25), male speakers of Standard Southern British English from the University of Cambridge".

A speaker's individual speaking style is dependent on his linguistic ability and cognitive ability (Guan 2014b), thus, sociolinguistically homogeneous speakers will also be appropriate as a rule to sample speakers to test FSR parameters reflecting a speaker's individual speaking style. As such, the conversations from S1 to S10 composing the dataset in Experiment 1 were produced by sociolinguistically homogeneous speakers, who are of similar age, of the same sex, and from the same school of the same university.

As reviewed, the common way of selecting potentially useful parameters is to inspect the ratio of between-speaker to within-speaker variation with the Analysis of Variance. Thus a one-way between-subject multivariate analysis of variance was conducted in SPSS19 where *P1* and *P2* were the dependent variables. If there is statistically significant difference among the 10 speakers in terms of the two features jointly or separately, the potential speaker-discriminating power of these two features will have been tested and the second experiment will be activated.

4.2 Experiment 2

Experiment 2 depends on the positive conclusion of Experiment 1, in which the potential of the extracted features to discriminate speakers has been tested, and it intends to prove that the two features *P1* and *P2* can reflect a speaker's individual speaking style. It means that they would stay consistent among a speaker's conversations across speech situations and time.

The conversations from S11, S12, and S13 were used as the experimental materials. Each speaker's sampled conversations

occurred at different time and there is no overlap among the persons with whom the sampled speaker was talking. Given this, the conversations produced by each speaker can be considered as conversations across different speech situations and time. In addition, both female and male speakers were sampled to improve the reliability of the examination of individual speaking style.

One way to test the consistency of the extracted features among a speaker's different conversations is to test that statistically there is no significant difference among the sampled conversations from a speaker in terms of the two features separately. Thus three one-way between-subject multivariate analysis of variance were conducted in SPSS19 separately, where *P1* and *P2* were the dependent variables.

5 Results of experiments and discussion

5.1 Results of Experiment 1

For Experiment 1, $p = 0.000$ in Bartlett's Test of Sphericity, and $p > 0.05$ for *P1* and *P2* in Levene's Test of Equality of Error Variances, which shows that the data are qualified for a one-way between-subject multivariate analysis of variance. Box's $M = 42.422$, $p = 0.155$ in Box's M test, which is bigger than the significant level 0.05, thus Willk's Lambda was used to assess the multivariate effect, where Willk's Lambda = 0.456, and $p = 0.001$. The lower p value indicates that the 10 conversations were significantly different in terms of the two features jointly.

Furthermore, the univariate ANOVAs conducted in terms of either feature separately produced p values lower than the significant level 0.05, where p for *P1* is equal to 0.038, and p for *P2* is equal to 0.003. It demonstrates that either of the two features can distinguish speakers.

To sum up, the results of the one-way between-subject multivariate analysis of variance have showed that in Experiment 1 the 10 conversations from the 10 speakers can be predicted not to belong to one speaker in terms of the two features jointly or in terms of either of them separately. In other words, either of the extracted features with DIA has been tested to be statistically significant between speakers. Such results exemplified that DIA did work to extract discourse information features that can discriminate speakers

to some extent. As a consequence, Experiment 2 was activated and conducted to test whether these potentially speaker-specific features reflect a speaker's individual speaking style.

5.2 Results of Experiment 2

For Experiment 2, the Levene's Test of Equality of Error Variances and Bartlett's Test of Sphericity in three sub-experiments, see Table 5 and Table 6, indicate that the data are qualified for a one-way between-subject multivariate analysis of variance.

In the three sub-experiments, $p > 0.01$, the significant level, in Box's Test of Equality of Covariance Matricesa, and thus Wilks' Lambda was used to assess the multivariate effect, where for S11, Wilks' Lambda = 0.545, $p = 0.134$; for S12, Wilks' Lambda = 0.700, $p = 0.196$; for S13, Wilks' Lambda = 0.690, $p = 0.591$. the p values larger than the significant level of 0.01 indicate that the conversations in each sub-experiment are predicted to be from one speaker in terms of the two features jointly.

Table 5. The Results of Levene's Test of Equality of Error Variances in Experiment 2

	P1(S11)	*P2(S11)*	*P1(S12)*	*P2(S12)*	*P1(S13)*	*P2(S13)*
F	.391	1.873	.937	.241	1.418	1.530
df1	4	4	5	5	7	7
df2	20	20	37	37	31	31
Sig.	**.819**	**.155**	**.469**	**.942**	**.234**	**.194**

Note: α = .01

Table 6. The Results of Bartlett's Test of Sphericitya in Experiment 2

	S11	*S12*	*S13*
Likelihood Ratio	.000	.000	.000
Approx. Chi-Square	42.287	76.411	201.670
df	5	5	5
Sig.	**.001**	**.000**	**.000**

Note: α = .01

The univariate ANOVAs conducted in terms of both features separately in every sub-experiment all gave the p values much larger

than the significant level of 0.01, see Table 7. The results indicate that both features show consistency within a speaker's conversations across speech situations and time.

Table 7. The Results of Univariate ANOVAs in Experiment 2

Source	*Dependent variable*	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
S11	P1	.261	4	.065	1.874	**.155**
	P2	.068	4	.017	.367	**.829**
S12	P1	.250	5	.050	1.614	**.181**
	P2	.176	5	.035	1.167	**.344**
S13	P1	.495	7	.071	.162	**.991**
	P2	.005	7	.001	1.069	**.406**

Note: α = .01

5.3 Discussion

This experimental study was conducted to verify the potential speaker-discriminating power of individual speaking style. DIA is considered to be an appropriate approach to analyze natural conversations and find out non-phonetic features at the level of discourse information that reflect individual speaking style.

Different from the prior experiments in FSR research, the experimental materials used in this study are natural conversations instead of lab-recorded audio materials. All these natural conversations occurred in real-world conditions and were being recorded with nothing controlled.

Firstly, based upon the distribution of 15 types of information knots in all sampled conversations, the information unit at the information knot of *WT* has been determined to be the object of investigation. That ensures that the two extracted features meet one of the six criteria for ideal FSR parameters that they should have a high frequency of occurrence in relevant materials.

One of the explored features is the duration of WT information unit. It can be easily measured with computerized speech lab, or with voice analysis software like Praat. Another is the ratio between the total number of information elements in each WT information unit and the total number of information elements in the conversation

being observed. It can be extracted through simply counting the number of information elements in each WT information unit and a conversation and then computing the ratio. Therefore, the two explored features meet one more of the six criteria for ideal FSR parameters that they should be relatively easy to extract and measure.

Then, in Experiment 1, the two explored features have been tested to have enough higher *F*-ratio to discriminate the sampled speakers both jointly and separately. In other words, the two features tend to meet the most important criterion for ideal FSR parameters that they should show high between-speaker variability and low within-speaker variability.

In Experiment 2, the two explored features have been tested to stay consistent within 3 different speakers respectively. Because all involved conversations from each speaker occurred in different speech situations and at different time, the tested within-speaker consistency illustrates that the two features tend to meet another two criteria for ideal FSR parameters that they should be resistant to attempted disguise or mimicry and be robust in transmission.

In addition, *P1* and *P2* has been tested in SPSS19 to be uncorrelated with $p = .524$. That is to say, they meet the sixth criterion for ideal FSR parameters that each parameter should be maximally independent of other parameters.

In summary, the results of the experiments demonstrate that the two explored quantitative features tend to meet all six criteria for ideal FSR parameters. It provides evidence for the individual speaking style as well as their potential speaker-discriminating power. Put another way, supposing the consistency was accidental due to small datasets, and then the extracted speaker-specific features might reflect the common sense *style* in sociolinguistics at the level of language instead of *the individual speaking style* at the underlying level. However, “there are no single style speaker” (Labov, 1984) in that style reflects the interaction between a speaker and contexts (Eskénazi, 1993). Then as a consequence, these features would not have demonstrated consistency between any two of the conversations occurring under different conditions or on different occasions. As such, consistency across more than five speech situations and time in three cases from both female and male speakers is convincing to

some extent.

Most importantly, the experimental materials in this study are natural conversations. It means that they could be put into practice directly if the explored features were further tested to represent individual speaking style to a great extent by large datasets of natural conversations.

6 Conclusion

The experiments were designed in order to test that individual speaking style has potential speaker-discriminating power as predicted and is potentially qualified non-phonetic FSR parameters.

The results of Experiment 1 demonstrate that the two features concerning discourse information and extracted with DIA did show high between-speaker variability and low within-speaker variability. On the basis of Experiment 1, Experiment 2 further verified that the two features extracted in Experiment 1 did stay consistent among the same speaker's conversations across different speech situations and time. It proves that they represent a speaker's individual speaking style to some extent.

Moreover, the two features are quantitative, which makes it easier to evaluate them with likelihood-ratio approach as the new paradigm shift requires. Further, the potential speaker-discriminating power of the two quantitative features was tested with natural conversations that occurred in real-world conditions, and they remained consistent under different conditions and on different occasions. It indicates that individual speaking style features tend to be immune to within-speaker variability and the gap between FSR research and practice is to be bridged if the forensic significance of individual speaking style features can be further verified and evaluated.

To sum up, the results of Experiment 1 and Experiment 2 together have provided support to the following predictions. First of all, individual speaking style is potentially speaker-specific. Next, individual speaking style parameters tend to resistant to within-speaker variability in voice and the effects of forensically realistic conditions and meet all criteria for ideal FSR parameters. More importantly, natural conversations have been introduced into

FSR research through this study, which plays a key role in bridging the gap between FSR research and practice.

Certainly, the potential speaker-discriminating power of the two features, as well as the extent to which they represent a speaker's individual speaking style, expects to be tested and evaluated with large datasets. Furthermore, it is hoped that more non-phonetic features are to be explored and tested inspired by this experimental study.

References

Alexander, A., & J. K. McElveen. (2007). Approaches to speaker recognition: A primer. *Clarifying Technologies*.

Beritelli, F. & A. Spadaccini. (2012). Performance evaluation of automatic speaker recognition techniques for forensic applications. In Yang, J. & S. J. Xie (eds.), *New Trends and Developments in Biometrics*. InTech, 129-148.

Bijhold J., Ruifrok A, M. Jessen, Z. Geradts, S. Ehrhardt, & I. Alberink. (2007). Forensic audio and visual evidence 2004-2007: A review. *15th INTERPOL Forensic Science Symposium*.

Broeders, A. P. A. (2001). Forensic speech and audio analysis forensic linguistics. *Proc. 13th INTERPOL Forensic Science Symposium*, 16-19.

Cambier-Langeveld, T. (2007). Current methods in forensic speaker identification: Results of a collaborative exercise. *International Journal of Speech Language and the Law* 14(2): 223-243.

Coulthard, M. & A. Johnson (eds.). (2007). *An Introduction to Forensic Linguistics-Language in Evidence*. London & New York: Taylor & Francis Group.

Du, J. B. (2004). *Forensic Linguistics.* Shanghai: Shanghai Foreign Language Education Press.

Du, J. B. (2007). A study of the tree information structure of legal discourse. *Modern Foreign Language* 30(1): 40-50.

Du, J. B. (2008). A study on the theories and methodologies of discourse analysis. *Foreign Language Research* 140(1): 92-98.

Du, J. B. (2011). Discourse information analysis: a new research perspective in forensic linguistics. *Chinese Social Sciences Weekly* 5.24 (015).

Du, J. B. (ed.). (2013). *The Course of Discourse Analysis*. Wuhan: Wuhan University Press.

Elliott, J. R. (2001). Auditory and F-pattern variation in Australian Okay: a forensic investigation. *Acoustic Australia* 29(1): 37-41.

Eskénazi, M. (1993). Trends in speaking styles research. *Third European Conference on Speech Communication and Technology*, 501-509.

Gold, E., & P. French. (2011). An international investigation of forensic speaker comparison practices. *Proceedings of the 17th International Congress of Phonetic Sciences*, 751-754.

Gonzalez-Rodriguez, J., P. Rose, D. Ramos, D. T. Toledano, J. Ortega-Garcia. (2007). Emulating DAN: Rigorous quantification of evidential weight in transparent and testable forensic speaker recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 15: 2104-2115.

Gruber, J. S., & F. T. Poza. (1995). *Voicegram Identification Evidence*. Lawyers Cooperative Publishing.

Guan, X. (2014a). Study on FSR cross-validation method. *Journal of Guangdong University of Foreign Studies* 5: 52-57.

Guan, X. (2014b). Forensic Speaker Recognition Based on Discourse Information (in progress) [D]. Guangdong University of Foreign Studies.

Hollien, H. (1990). *The Acoustics of Crime: The New Science of Forensic Phonetics*. New York: Plenum Press.

Hollien, H. (2002). *Forensic Voice Identification*. Academic Press.

Hughes, V., A. Brereton, & E. Gold. (2013). Reference Sample Size and the Computation of Numerical Likelihood Ratios Using Articulation Rate. *York Papers in Linguistics* 2(13): 22-46.

Johnstone, B. (1996). *The Linguistic Individual: Self-Expression in Language and Linguistics*. New York and Oxford: Oxford University Press.

Kersta, L. G. (1962).Voiceprint identification. *Nature* 196: 1253-1257.

McDermott, M. C., T. Owen, & F. M. McDermott. (1996). Voice

Identification: The Aural Spectrographic Method. *Owl Investigations Web Site*, http://www. owlinvestigations. com/forensic_artic les/aural_spectrographic/fulltext. html.

Morrison, G. S. (2009). The place of forensic voice comparison in the ongoing paradigm shift. *The 2nd international conference on evidence, law and forensic science, conference thesis (1)*, 20-34.

Morrison, G. S. (2010). Forensic voice comparison. In Freckelton, I. & H. Selby (eds.). *Expert Evidence (Ch. 99)*. Sydney, Australia: Thomson Reuters.

Morrison, G. S., P. Rose & C. Zhang. (2012). Protocol for the collection of databases of recordings for forensic-voice-comparison research and practice. *Australian Journal of Forensic Sciences* 44(2): 155-167.

Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.

Nolan, F. (1997). Speaker recognition and forensic phonetics. In Hardcastle & Laver (eds.), *The Handbook of Forensic Sciences*, 744-767.

Pruzansky, S., & M. V. Mathews. (1964). Talker-recognition procedure based on analysis of variance. *JASA* 36: 2041-2047.

Qian, G. L. (2006). Assumptions on Speechology. *Foreign Language Research* 129(2): 34-37.

Rose, P. (1996). Speaker Verification under Realistic Forensic Conditions. *Proceedings of the Sixth Australian International Conference on Speech Science and Technology*, Australian Speech Science and Technology Association, Canberra, 109-114.

Rose, P. (2002). *Forensic Speaker Identification*. London & New York: Taylor & Francis.

Rose, P. (2005). Technical forensic speaker recognition: Evaluation, types and testing of evidence. *Computer Speech and Language* 20(2): 159-191.

Rose, P. (2006). Catching criminals by their voice–combining automatic and traditional methods for optimum performance in forensic speaker identification. http://rose-morrison.forensic-voice-comparison.net/.

Sapir, E. (1927). Speech as a personality trait. *The American Journal*

of Sociology 32 (6): 892-905.
Tosi, O., H. Oyer, W. Lashbrook, C. Pedrey, J. Nicol, & E. Nash. (1972). Experiment on Voice Identification. *The Journal of the Acoustical Society of America* 51(6B): 2030-2043.
Tosi, O. I. (1979). *Voice Identification: Theory and Legal Applications*. Baltimore: University Park Press.
Wolf, J. J. (1972). Efficient acoustic parameters for speaker recognition. *JASA* 51: 2044-2056.
Zhang, C. L. (2009). *Research on Forensic Voice Technology*. Beijing: China Social Sciences Press.

Bionote

Xin Guan is a lecturer at Zhaoqing University in Guangdong and a PhD student in forensic linguistics in Guangdong University of Foreign Studies. Her research of interest focuses on forensic speaker recognition (FSR).