音声研究　第 21 巻第 3 号
2017（平成 29）年 12 月
41–52 頁

Journal of the Phonetic Society
of Japan, Vol. 21 No. 3
December 2017, pp. 41–52

特集論文

Chinese Demonstratives and Their Spoken Forms in a Conversational Corpus

Shu-Chuan Tseng*

会話コーパスにみる中国語指示詞とその発音

SUMMARY: Demonstratives are a particular group of linguistic expressions in Chinese, categorized as pronouns, determiners, adverbs, connectives, or fillers. Utilizing a phonetically labeled conversational corpus, we examined the spoken form of Chinese demonstratives in this study. Specific selection preferences for phonological variants were identified in the corpus. Filler demonstratives are generally longer than their lexical counterparts. Disyllabic lexical and filler demonstratives show seemingly contrasting duration patterns. Deviating from the falling tone carried by lexical originals, filler demonstratives tend to be flat with a mid-height onset, as illustrated by representative tonal contours obtained by a proposed computational tone model.

Key words: Mandarin Chinese, conversational corpus, usage-based analysis, computational modeling, duration and tonal pattern

1. Introduction

Human speech, as a linguistic system, evolves in a communicative, socio-pragmatic context in accordance with users' natures and needs. By this usage-based nature, different types of shaping and transforming processes simultaneously proceed in the use of language. These processes include phonetic variation, meaning and structure generalization as well as functional innovation. However, a stabilized variant, regardless of the linguistic level, must fit the fundamental linguistic system *per se*. Chinese demonstratives form an illustrative group of linguistic expressions that transform a physical, spatial relationship between a speaker and an object into a deictic, referential context in language use (Hockett 1963). The spatial distance can be far away or nearby, differentiating between distal and proximal references in linguistic expressions. The need of developing a deictic reference is considered to evolve in a face-to-face social context (Lyons 1977, Levinson 1983), as indicated by Wu (2004, p. 4): "*Spatial demonstratives provide evidence for the immediate situational and displaced modes of language use.*" A close examination of the various linguistic forms of Chinese demonstratives reveal them be a language process that pinpoints the development from a lexical word. Each Chinese demonstrative is an expression conveying manner and time; it can also serve as a connective linking a proposition with the previous ones in an interactional discourse. *The symbolic referential use* (e.g., *this man*) is immediately generated from the spatial relationship in a face-to-face interaction. *The adverbial use* (e.g., *this far*) is derived from the referential indication.

Chinese demonstratives can also be used as fillers, e.g. *Nǐ NAGE NAGE bù huílái ma*? (You *NAGE NAGE* will not come back?). The filler *NAGE* that appears twice in the above example serves the same function as simple filled pauses, such as *uh* and *uhn* in English. However, despite the wide usage of demonstratives in human speech communication, research toward the use and form of demonstrative fillers in Chinese started only recently (Zhao and Jurafsky 2005, Tseng 2015, Maekawa, Nishikawa and Tseng 2017). With regard to the prosodic properties of simple filled pauses, they often differ from the neighboring context in terms of the duration, fundamental frequency (F0), and spectra (Gabrea and O'Shaughnessy 2000). Batliner, Kießling, Burger and Nöth (1995) concluded that F0 is lower in filled pauses in German and its contour tends to be falling, suggesting that filled pauses are similar to a type of parenthetical chunks in the speech flow. Furthermore, Shriberg (2001) suggested F0 values and slopes in English have the same tendencies. Related to discourse boundaries, Swerts proposed that the use

* Institute of Linguistics, Academia Sinica, Taiwan（中央研究院・語言学研究所（台湾））

of simple filled pauses was correlated with discourse boundaries in Dutch (Swerts 1998), and simple filled clauses can also be used as cues to the complexity of upcoming phrases in Japanese (Watanabe, Hirose, Den and Minematsu 2008).

This paper closely examines the use of Chinese demonstratives in a corpus of face-to-face conversations, with a schema proposed to group Chinese demonstratives into lexical and filler categories. In addition to the analysis of duration patterns in lexical and filler demonstratives, we adopt a computational methodology, which differs from earlier works by focusing on only F0 values, to compute and visualize the tonal contour in order to closely examine the presence of any difference among demonstratives.

2. Chinese Demonstratives

2.1 Word Categories

In language use, demonstratives perform a two-by-two contrast: *entity referring* versus *place referring* and *proximity* versus *nonproximity* distinctions. In English, demonstrative tokens are this/these (entity referring, proximity), that/those (entity referring, nonproximity), here (place referring, proximity), and there (place referring, nonproximity) (Wu 2004). Compared with English, Chinese has a wider variety of morphosyntactic variants. Stem components in the Chinese demonstrative system include *nà* (literally, *that*) and *zhè* (*this*)[1]. *Nà* is used for the distal reference of a previously named entity or place, and *zhè* is used for the proximal reference (Li and Thompson 1981, pp. 130–131). *Nà/zhè* can be used in plural forms with *xiē*, a bound morpheme for plurality. Similar to English, *nà*/*zhè* can be used to indicate a location and time (Wu 2004, Wang 1987); for instance, *nàlǐ/zhèlǐ* (there/here) and *nàshíhòu/zhètiān* (that moment/this day). Apart from these uses, Chinese demonstratives are often discussed in the context of noun phrases (NPs) or determiner phrases, focusing mostly on syntactic theories, such as whether they are the head of a definite NP or whether the meaning of the definiteness comes from demonstratives (Tang 1990, Lin 1997, Cheng and Sybesma 1999, Li 1999, Simpson 2001). Regarding the word category, *nà*/*zhè* can be used as demonstrative **pronouns**, such as ***nà*** *shì wǒ de* (that is mine) and ***zhè*** *shì tā de* (this is his) (Chao 1968, p. 649). They are also used as **determiners** in the form of *nà* classifier (CLS; e.g., ***nà ge*** *shì wǒ de* [that particular one is mine] and *zhè* CLS (e.g., ***zhè ge*** *shì tā de* [this particular one is his]). In the context of Chinese discourse, the emergence of the category "definite article" in the use of *nà ge* was primarily explored by Huang (1999). Alternatively, demonstratives also form a full-fledged *nà* CLS NP (e.g., ***nà ge rén*** *shì wǒ bàbà* [that CLS man is my father][2] and ***zhè bù chē*** *hěnhǎo* [this CLS car is good]) or an absent classifier (e.g., ***nà rén*** *wǒ bú rènshì* [I don't know that man] and ***zhè chē*** *bú shì wǒ de* [that car does not belong to me]) (Chao 1968, Li and Thompson 1981). **Adverbs** are specifically used for expressing manner and degree (e.g., *nàme/zhèyàng* [that way/like this]).

Although Chinese demonstratives have been widely studied, their use in spoken discourse is under-investigated. Similar to simple fillers, demonstrative **fillers** are mainly used to fill a gap during the course of conversation (Zhao and Jurafsky 2005). Three examples obtained from the MCDC8, a conversational corpus used for analysis later in this paper, illustrate the use of filler demonstratives. In this paper, we have used capital letters with no tone labels to denote filler demonstratives. In *érqiě yě NEGE bāng wǒ dìdì* (besides, it also NEGE helped my brother), *NEGE* has no concrete meaning; the same is with *NAGE* and *ZHEGE* in *tā míngtiān NAGE jiù bú qù le* (he NAGE will not join tomorrow, then) and *yīnwéi ZHEGE jìchéngchēduì táotài hěn kuài* (because *ZHEGE* taxi fleets will be replaced rather rapidly), respectively. *NEGE*, *NAGE*, and *ZHEGE* in the aforementioned examples have one thing in common. The definite meaning is lost. From the syntactic point of view, in the example of "he *NAGE* will not join tomorrow, then", *NAGE* does not appear before a noun, apparently violating the structure of a demonstrative construction. In the example "because *ZHEGE* taxi fleets will be replaced rather rapidly", although *ZHEGE* is followed by the NP "the taxi fleets," the speaker does not refer to any specific taxi fleets; thus, no definite meaning is referred to here.

In addition to the lexical use as pronouns, determiners and fillers, demonstratives are also used as **connectives**. When *nà* appears in an utterance-initial position, it can mean "as for that" or "in that case," as in Chao's example ***nà*** *wǒ méi bànfǎ le* "As for that, I don't know what to do about it" (1968, p. 650), which is similar to a discourse marker meaning "well," "then," or "so", as proposed by Schiffrin (1988) for English. They resemble connectives, rather than adverbs, as suggested by Chao (1968) or simple fillers with no explicit lexical meaning.

2.2 Variants

From the point of view of the writing system, *nà* and *zhè* both have homographs in the form of Chinese

Table 1 Chinese demonstratives.

	Lexical meaning	Definiteness	Discourse function	Examples	English translation
Pronoun	+	+	−	***Nà*** *shì wǒ de* ***Zhè*** *shì tā de*	That is mine This is his
Determiner	+	+	−	***Nà ge rén*** *shì wǒ bàbà* ***Zhè bù chē*** *hěnhǎo* ***Nà rén*** *wǒ bú rènshì* ***Zhè chē*** *bú shì wǒ de*	That CLS man is my dad This CLS car is good That man, I don't know This car is not mine
Connective	+	−	+	***Nà*** *wǒ méi bànfǎ le*	Then, I ran out of solutions
Adverb	+	+	+	*Nǐ wèishéme* ***nàme*** *bù tīnghùa* ***Nàyàngzi*** *dehùa, nǐ bù néng zùo*	Why are you so disobedient Under that circumstance, you cannot do it
Filler	−	−	+	*Érqiě yě* ***NEGE*** *bāng wǒ dìdì* *Tā míngtiān* ***NAGE*** *jiù bú qù le*	Besides, it also NAGE helped my brother. He will NAGE not go there tomorrow, then.

characters. But from the viewpoint of word pronunciation, alternatively, these homographs can be regarded as phonological variants. The standard pronunciation of *nà* and *zhè* is [na] and [tʂə], both articulated with a falling tone and written in 那 and 這, respectively. In addition to [na], 那 has three variants, [nə], [nei], and [nai], in spoken use. The variant [nai] is speculated to result from the merged form of 那一個 *nà yī ge* (that one CLS). However, additional studies on this issue are required. 這 is pronounced [tʂə] or [tʂei]. For notations used in this paper, 那／這 with a standard pronunciation is written as *nà*/*zhè*. The other variants [nə], [nei], [nai], and [tʂei] are noted as *nè*, *nèi*, *nài*, and *zhèi*, respectively, all pronounced with a falling tone. *Nè* is often regarded as a phonetically reduced alternative for *nà*, different from other variants. To distinguish them from the lexical use, filler demonstratives are noted as *NA/ZHE*, *NE*, *NEI*, *NAI*, and *ZHEI*, accordingly. This study also examines the usage distribution of the shortlisted variants in conversation. Given a classification schema of demonstratives, our corpus results may provide empirical evidence for the notion that particular selection preferences for phonological variants are imposed by users.

2.3 Taxonomy

According to the linguistic and interactional characteristics of demonstratives, three aspects, *lexical meaning*, *definiteness*, and *discourse function*, are preliminarily adopted to distinguish varying uses of Chinese demonstratives in conversation.

Table 1 shows the classification taxonomy of Chinese demonstratives with illustrative examples. Demonstrative fillers do not convey explicit lexical meaning or deictic definiteness. However, in conversation, they perform the discourse function that signals a kind of hesitation. In contrast to fillers, lexical demonstrative pronouns, determiners, and adverbs (except for demonstrative connectives) have a specific meaning and convey definiteness. When used as demonstrative connectives, the original referring definiteness is not preserved in the process of developing into the role of connecting consecutive propositions, unlike the referring definiteness of demonstrative adverbs. In spoken use, demonstrative connectives and adverbs can sometimes be produced with emphasized intonation patterns to express linguistic implications that are applicable in their particular discourse contexts. Different from the other demonstrative word categories, adverbs have a considerable number of morphological compositions (e.g., *nàme* [then] and *nàyàngzi* [in that way][3]). Because the current study focuses on the contrasting uses and forms of lexical and filler demonstratives, not all tokens involving demonstratives can be considered due to the scarcity of our data. For instance, plural uses (e.g., *nàxiē* [those] and *zhèxiē* [these] and locative uses (e.g., *nàlǐ*, *nàr*, *nàbiān*, *zhèlǐ*, *zhèr*, and *zhèbiān*[4]) are excluded from the current study. These tokens themselves form a specific group that requires independent analyses to investigate their forms and functions in

Table 2　Corpus use of demonstratives.

	nà	*nè*	*nèi*	*nài*	*zhè*	*zhèi*	TOTAL
Pronoun	88	71	6	1	117	2	285
Determiner (+N, +CLS, +CLS N)	84	430	203	16	365	94	1,192
Connective	130	12	1			1	144
Adverb	10	80	8		366	19	483
SUM (lexical use)	312	593	218	17	848	116	**2,104**
Filler	*NA* 560 *NAGE* 53	*NE* 89 *NEGE* 258	*NEI* 18 *NEIGE* 70		*ZHE* 2 *ZHEGE* 63	*ZHEI* 4	
SUM (filler use)	613	347	88		65	4	**1,117**

conversational use.

3. Corpus-based Inspection of Spoken Demonstratives

3.1 Corpus Overview

The Sinica MCDC8 consists of eight free conversations between sixteen native Taiwan Mandarin speakers, with each conversation lasting approximately one hour (Tseng 2013)[5]. Concerning the corpus size, the Sinica MCDC8 contains 93,533 words, equivalent to 136,229 syllables or 310,625 phonemes. For text processing, transcripts were segmented into words by using the CKIP automatic word segmentation and POS tagging system (Chen, Huang, Chang and Hsu 1996). Manual post-editing and -correction were required because the CKIP word segmentation system was mainly trained on written data. For signal processing, long speech stretches were first segmented into interpause units based on the occurrence of silent pauses and paralinguistic sounds, such as laughter and inhalation. With word-segmented transcripts and the corresponding sound files, a phone aligner, which was specifically trained on manually labeled training data of spontaneous Chinese speech, was used to perform the forced alignment to obtain initial timestamps for word, syllable, and phoneme boundaries (Liu, Tseng and Jang 2014). Extremely reduced phonemes were excluded from the present study, because the quality of the extracted acoustic features and the result of our proposed models could be affected due to inadequate acoustic information in the unstable physical signal.

3.2 Data Inspection

Table 2 lists the corpus use of Chinese demonstratives. The MCDC8 corpus contains 2,104 lexical uses and 1,117 filler uses of demonstratives. The tokens listed in Table 2 alone account for 3% of the entire corpus. In our calculation, the cases of determiners include those followed by a noun, a classifier, or a classifier plus a noun. In addition, to analyze duration and tone patterns later, word-internal syllable boundary locations must be available in phonetic forms. Thus, disyllabic demonstrative tokens that are syllable mergers are excluded from the acoustic analysis.

3.3 Variant Selection Preferences

According to the usage in our conversational corpus, specific preferences exist for pronunciation variants for both lexical and filler demonstratives (Figure 1a and 1b).

For pronoun use, *nà* and *nè* occur similarly often. However, for determiner use, which mostly appears in polysyllabic forms, *nè*, the phonetically reduced form, is preferred over *nà* and *nèi*. For adverb use, which always appears in a polysyllabic form, *nè* is also preferred. However, for monosyllabic connective use, *nà* is the preferred form. When lexical units involve more than the demonstrative token alone, the phonetically reduced variant *nè* is preferred by speakers. The variant *nài* was rarely used in our corpus, with only one token as a pronoun and 16 tokens as determiners. This may be attributed to our earlier speculation that *nài* is a merged form of "that one CLS," carrying a concrete referring meaning. Thus, the usage as an adverb or a connective in which a discourse function is imposed is hindered. For monosyllabic filler use, *NA* is the most preferred form, but in the case of disyllabic fillers, *NEGE* is preferred. The filler variant *NEI* is seldom used in our corpus data, and *NAI* was not used at all. For *zhè* cases, *zhè* is clearly preferred over *zhèi* in both lexical and filler uses. Despite individual preferences for selecting

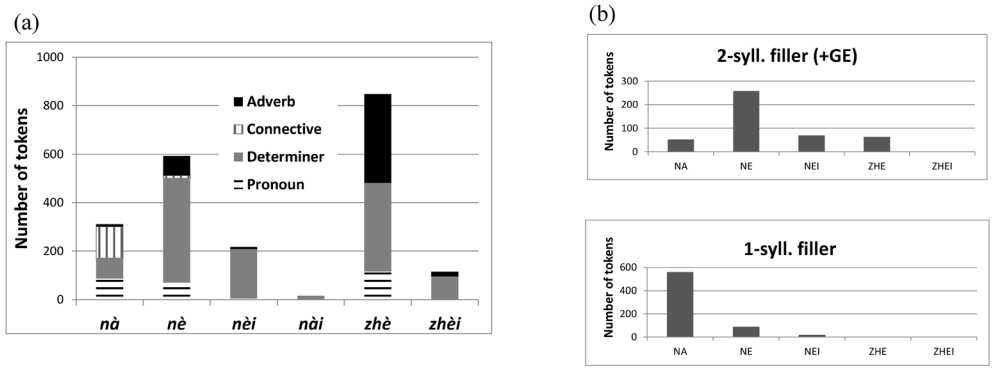

Figure 1 a: Distribution of lexical use, b: Distribution of filler use.

Table 3 Temporal properties of monosyllabic demonstrative use.

	nà/NA			*nè/NE*		
	MEAN	VARIANCE	MEDIAN	MEAN	VARIANCE	MEDIAN
Pronoun	0.198	0.016	0.153	0.105	0.005	0.078
Connective	0.188	0.031	0.151	0.057	0.001	0.040
Filler	0.261	0.046	0.186	0.203	0.028	0.150

spoken word forms, a collective behavior of variant selection preferences was observed in our 16-speaker corpus. To explicitly identify whether it is a stabilized preference in the speaker community, further experimental analysis is necessary.

4. Temporal Property: Lexical versus Filler Demonstratives

Demonstratives are produced with different preferences of phonological variants in spoken use, regarding the segmental properties. What about the suprasegmental ones? The existence of any prosodic markedness that distinguishes lexical from filler demonstratives and the existence of such markedness among the observed phonological variants remain unclear. In order to explore the suprasemental properties of demonstratives, we examined duration patterns first. Considering the data size, we only took into account the monosyllabic variant pairs of *nà/NA* and *nè/NE* and the disyllabic *nàge/NAGE* pair. In the acoustic–prosodic analysis, adverb cases were not included.

4.1 Method: Normalization

To resolve the issue of cross-speaker discrepancies, we used a scaled method for normalizing raw measurements of duration patterns, defined by (x−xmin)/(xmax−xmin), where x is the syllable duration and xmax and xmin are the maximum and minimum values measured from the tokens of a given syllable spoken by a given speaker in the MCDC8. That is, the maximum and minimum duration measurements are first computed based on the tokens of *nà*, *NA* and all the other Chinese characters that have the syllable structure [na] produced in the entire MCDC8 corpus. Subsequently, the original duration is scaled to the interval between 0 and 1, representing the minimum and the maximum values, respectively, for the purpose of inter-speaker comparison (Lobanov 1971).

4.2 Monosyllabic *nà/NA* and *nè/NE* Comparison

The temporal properties of *nà/nè* and *NA/ZHE* are shown in Table 3 and Figure 2.

First of all, *nè* is in general shorter than *nà* in both lexical and filler categories. This supports the notion that is usually accepted from the phonetic point of view, suggesting that *nè* is a reduced form of *nà*. Furthermore, *NA/NE* as fillers, are longer than their lexical counterparts *nà/nè*. The variance among fillers is larger than that among lexical uses. This may be due to the fact that in conversation, the context and phonological environment in which demonstrative fillers are used

vary considerably. For instance, the monosyllabic demonstrative fillers *NA/NE* tend to be used in a prosodically initial position, accounting for 67% of *NA* and 71 % of *NE* tokens. Disyllabic tokens of *NAGE* (87%) and *NEGE* (100%) are produced mostly in a medial position. Concerning the discourse function, demonstrative fillers can be used to indicate a clear hesitation with a long duration or to just fill the gap, which is often spoken in a very short duration in a prosodically initial position. As a result, the duration difference between lexical and filler demonstratives is rather marginal. To account for these two factors (discourse function and prosodic position) that play crucial roles in their prosodic realization, further studies are necessary.

4.3 Disyllabic *nàge/NAGE* Comparison: A Contrast

Because of our imbalanced data, we only observed the general duration tendency for disyllabic uses (Figure 3), offering no detailed statistical analysis. Similar to monosyllabic demonstratives, the lexical uses *nàge/nège/nèige/zhège* are generally shorter than the filler uses *NAGE/NEGE/NEIGE/ZHEGE*. In detail, the classifier *ge* is shorter than the determiner or is similarly long in the case of the distal demonstratives *nà/nè/nèi*.

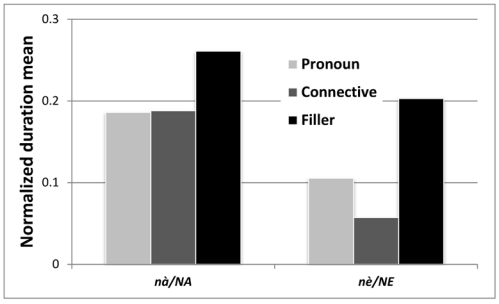

Figure 2 Duration pattern: Monosyllabic demonstratives.

For the proximal demonstrative *zhè*, the classifier is clearly longer than the determiner. This result may be due to the distribution of the classifier types that follow demonstratives. In the case of *zhè*, it is mostly *ge* that follows (82%). For the filler use, the tendency is clearer. *GE* is consistently longer than *NA/ZHE*, showing an iambic rhythmic pattern instead. This can be regarded as a cue suggesting that the discourse function of hesitation has an effect on the realization of the duration pattern, as *GE* is part of a filler, instead of a classifier that is usually pronounced reduced as a function word.

5. Spoken Demonstratives: Tonal Patterns

5.1 Computational Model for Deriving Representative Tonal Contours

Chinese has lexical tone distinction. *Nà/zhè* and all of their phonological variants are pronounced with a falling tone. In the literature, simple filled pauses tend to have a falling intonation contour (Batliner et al. 1995, Shriberg 2001); thus, for Chinese, it is interesting to examine whether the filler use preserves the falling tonal contour from the original lexical demonstratives. Because the type of our spoken data is spontaneous speech, a computational tone modeling algorithm was proposed to objectively derive representative tonal contours from the overall tokens in the corpus. Taking Chinese phonology into consideration, the algorithm must fulfill two theoretical assumptions. First, the abstraction must be phonologically adequate, that is, the acoustic features used for modeling the abstracted contour should accurately reflect the prosodic organization of Chinese tones (e.g., the number of computing sections along the course of a pronounced tone). Second, the abstraction should be computationally derivable with no human intervention. In other words, the representative contours should be automatically selected by mathematically defined algorithms. Below are the concrete procedures for conducting our tone

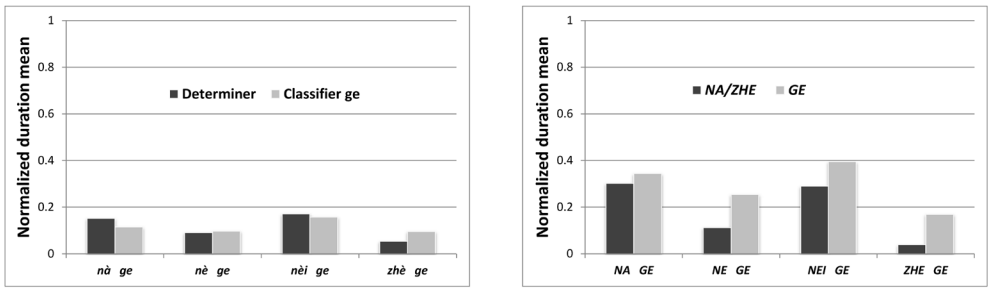

Figure 3 Duration pattern: Disyllabic demonstratives.

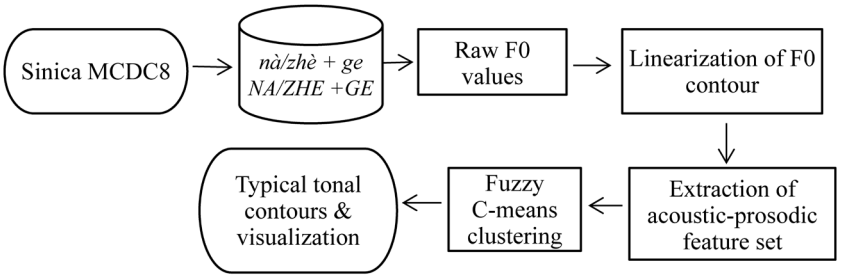

Figure 4 Procedure flowchart of computational tone modeling.

modeling experiment. First of all, we obtained raw F0 measurements from timestamps in the annotated corpus by using PRAAT for pitch tracking (Figure 4) (Boesma and Weenink 2013). After the F0 values and duration had been normalized across syllables and speakers, acoustic–prosodic features derived from the results of a linearization process were used for an automatic clustering calculation (Hartigan and Wong 1979, Bezdek 1981). The resulting major clusters were regarded as typical tonal contours. Later in the paper, the data point that is the closest to the cluster centroid is illustrated for the purpose of visualization.

5.2 F0 Contour Linearization

Onset and offset F0 values identified from the voiced region of each demonstrative token were used to anchor the time span for F0 measurement extraction. To overcome the limitations of conventional linear stylization, we also made use of the maximum and minimum F0 values within the voiced region. Together with the onset and offset F0 values, at most three fitted lines can be obtained if all four values are distinct. However, for a phonologically adequate model of Chinese tones, the model allows, at most, only two fitted lines, because Chinese dialects rarely have a three-section contour tone in the tonal system. To implement the two-section principle, the less representative line was merged with the immediately adjacent major line.

5.3 F0 and Time Normalization

Speakers in conversation may from time to time produce extremely high- or low-pitch height in situations where they make use of their pitch height to express or respond to interactional needs. Therefore, if we naively take the difference between the minimum and maximum F0 values as a speaker's pitch range, the result would lack discriminatory effects. Below is our proposed procedure for normalizing F0. For all syllables in a data set produced by a given speaker, representative F0 values measured at the position with the maximum intensity value within a syllable are collected. Most of these representative values should be stable because the most sonorous position within a syllable is likely to be around the place with the maximum intensity, normally where the vowel is located. Figure 5 shows the F0 histograms of all sixteen speakers in the MCDC8 used for conducting the computation of the speakers' pitch ranges[6].

Subsequently, the logF0 values between the 25% and 75% levels of speakers' syllable data are used for fitting a normal distribution. The F0 range of the speaker is then determined by taking the 0.1-th and 99.9-th percentiles as the lower and upper bounds, respectively. Later in the tone contour clustering experiment, based on the individually computed pitch ranges, the logF0 values of each speaker are rescaled to 0–1 by using (x−min)/(max−min), denoted as **NormF0**. F0 values outside the lower and upper bounds are discarded. Similar to F0 normalization, the time scale is rescaled to 0–1 for each syllable, denoted as **NormT**.

5.4 Pitch Range Computation

Applying the F0 normalization procedure introduced in the aforementioned section, we present the normal probability of our speakers in Figure 6, with information about the pitch range, respectively. Compared with simplex differences between the F0 maximum and minimum values, the pitch ranges calculated using our method seem to be able to accurately reflect individual peculiarity among the speakers.

5.5 Acoustic–prosodic Features

The Fuzzy C-means clustering algorithm is used to calculate the degree to which a data point belongs to a cluster in a multidimensional vector space (Bezdek 1981). The major clusters separated by a defined distance after 100 iterations were regarded as typical variant groups of tonal contours in this study. The

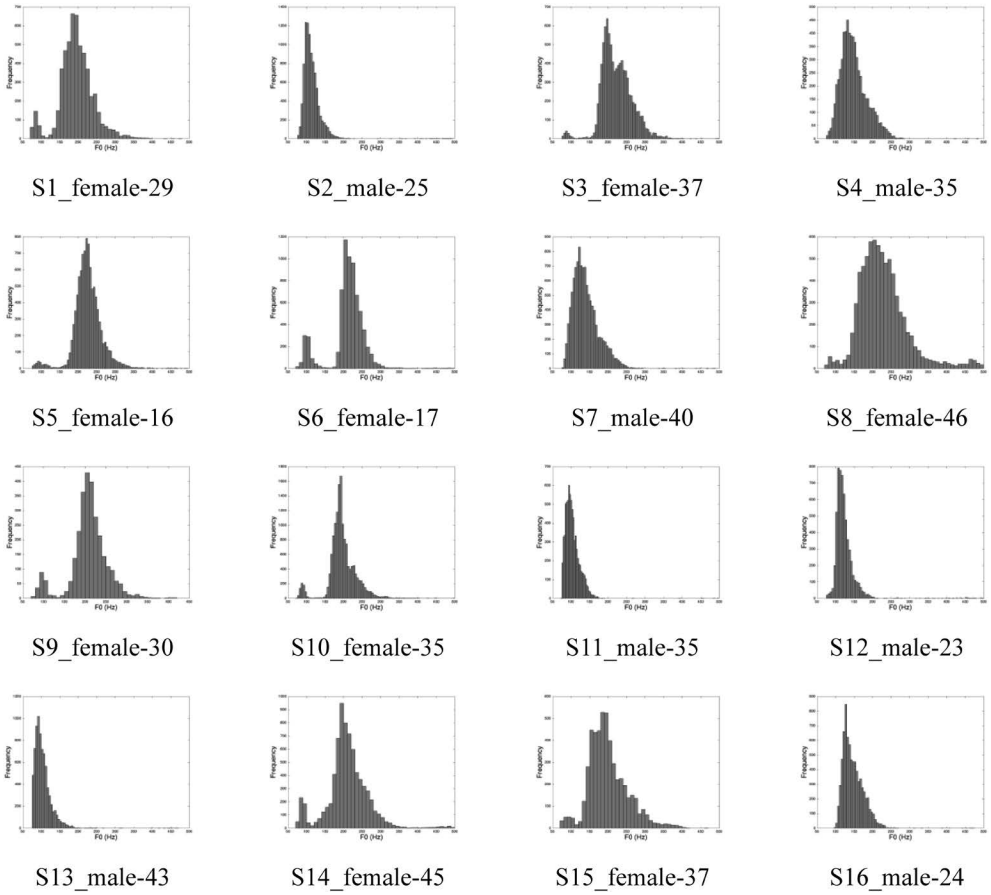

Figure 5　F0 Histogram (gender-age).

major clusters were those that contained the most data points. MATLAB Fuzzy Logic Toolbox was used for the clustering experiment. A set of features must be primarily defined for executing the experiment. On the basis of linguistic considerations, we used the following five acoustic–prosodic features that are relevant for tone production: (1) **OnsetNormF0**: the normalized onset F0 value; (2) **SlopeL1**: the slope of the first fitted line; (3) **SlopeL2**: the slope of the second fitted line, if any. In case of only one fitted line, SlopeL2 was set to 0; (4) **TurningTime**: because the turning point is crucial to approximate the turn between two opposite trends, it was used as a feature. It was defined to be the (normalized) time of the turning point in case of two fitted lines. If only one fitted line existed, then TurningTime was set to 1; and (5) **OffsetNormF0**: the normalized offset F0 value.

5.6　Contrasting Tonal Patterns in Lexical versus Filler Demonstratives

The results of the clustering experiment are shown in Figures 7 and 8, with visualizations of tokens that were the closest to the centroids of resulting major clusters. Because the clustering calculation requires a reasonable amount of data, for our data size, we only conducted the experiment for the monosyllabic *nà/NA* and the disyllabic *nàge/NAGE* and *nège/NEGE* contrasting pairs from our corpus for inspecting the experiment results.

The filler demonstrative *NA* has a lower F0 onset, and the falling tendency is not as apparent as that of the lexical demonstrative *nà* used as a pronoun and connective. When used as a connective, *nà* connects the previous proposition with the following one; thus, it is not surprising that we obtained an uprising section in the end of the connective, whereas the pronoun *nà* shows a standard falling lexical tone contour in the

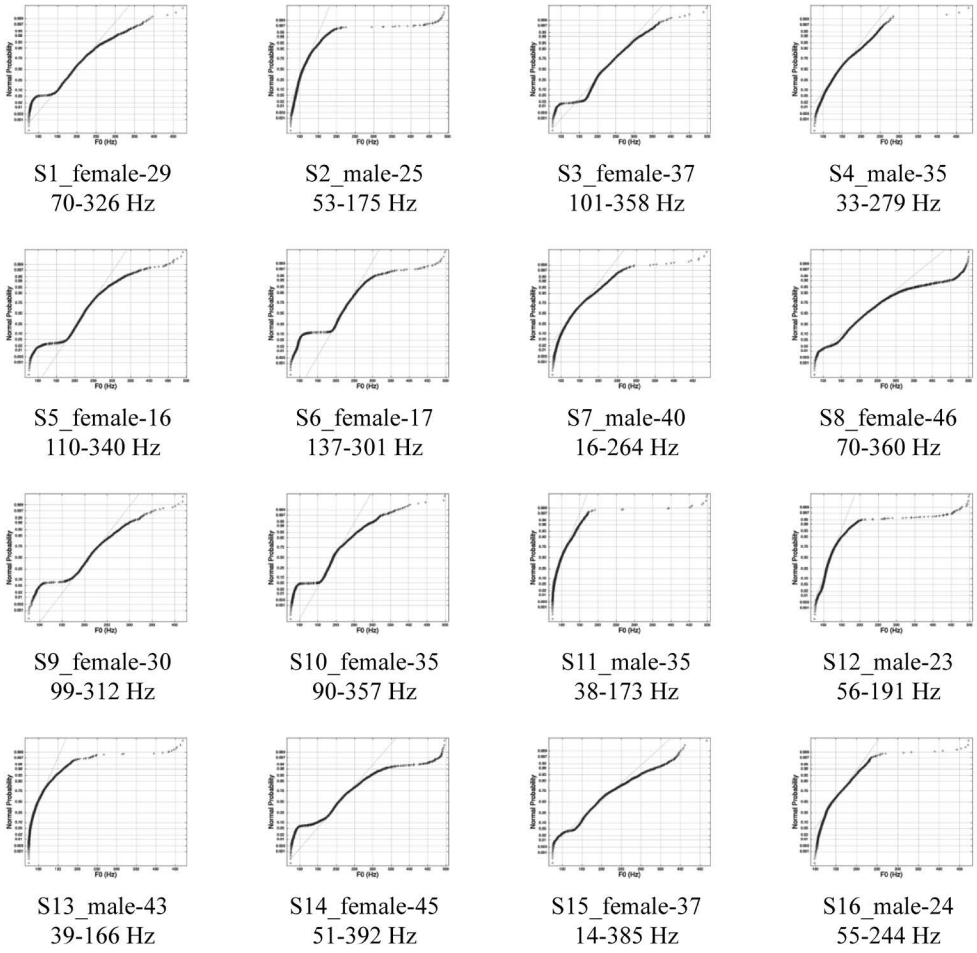

Figure 6 F0 normal probability with pitch range information.

Mandarin tone system.

In cases of disyllabic contrasting pairs, the result was complex. First, we observed that the filler *NAGE* was relatively flat with a middle onset F0 value, where the lexical *nàge* shows more contour changes, with a rising section in the beginning, then continues with falling sections. However, both *NEGE* and *nège* pairs start with similarly high onset F0 values, and *NEGE* is both slightly falling. However, *nège* first stays relatively flat, then turns low in the beginning of *ge*, and rises until the short falling section toward the end. The difference in representative tonal patterns among these four cases can be attributed to duration patterns. In the comparison of *nàge* with *nège*, *ge* is shorter than *nà*, but longer than *nè*. This implies that for *nège*, the determiner is reduced, followed by a strong classifier, probably to prepare for the head of a NP; thus, the presence of the rising contour of *ge* is not particularly surprising. While *nàge* forms a more structurally coherent duration pattern, where the classifier plays the least important role in the expression of meaning, the two consecutive falling tones preserve their tendencies in tonal patterns. The fillers *NAGE* and *NEGE*, tend to be rather flat, regardless of the duration pattern.

6. Discussion

Applying the three operational criteria *lexical meaning*, *definiteness*, and *discourse function*, we identified five demonstrative uses in a conversational corpus: **pronoun**, **determiner**, **connective**, **adverb**, and **filler**. Accounting for phonological variants, specific preferences were found for the spoken use of Chinese demonstratives. When used as a **pronoun**, *nà* and *nè*

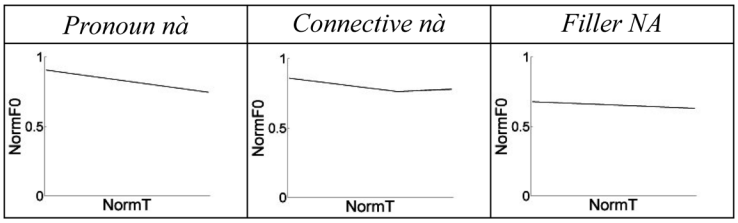

Figure 7　Tonal pattern of *nà/NA*.

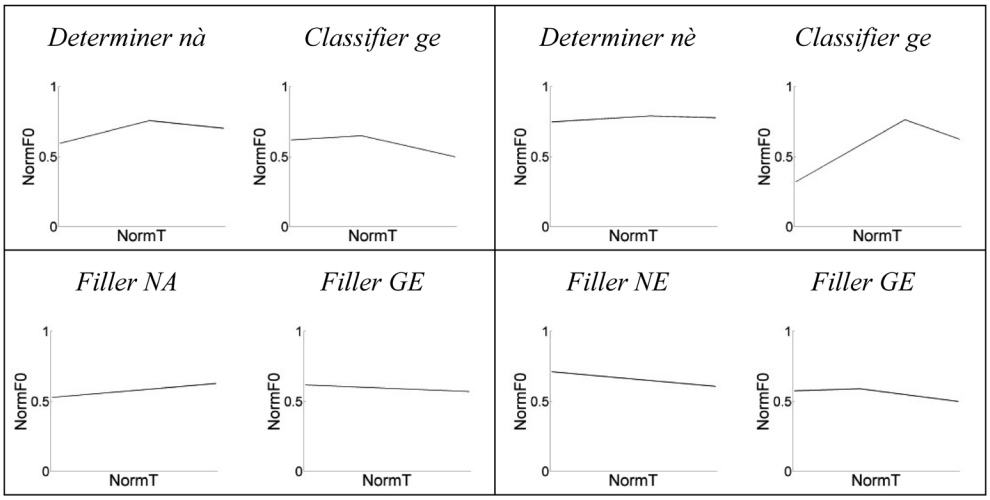

Figure 8　Tonal patterns of *nàge/NAGE* and *nège/NEGE*.

often appeared similarly often. But as a **determiner**, the phonetically reduced word form *nè* was clearly preferred, whereas *nà* was preferably used as a **connective**. Because of the syntactic property of lexical demonstratives, Chinese demonstratives usually appeared in the utterance-medial position, except when used as a **connective**. However, when used as **fillers**, the monosyllabic *NA* and disyllabic *NEGE* were preferred over *NE* and *NAGE*, respectively. *NA* mostly appeared in the initial position, whereas *NAGE* and *NEGE* were frequently produced in the medial position, suggesting that *NA* was probably involved at a sentential level, which was higher than *NAGE/NEGE* at a within-sentence level. While annotating the instances, we initially encountered issues with some of the *NA* occurrences that occurred in the initial position regarding whether to annotate them as fillers or connectives. In our annotation schema, a connective should link two propositions rather than merely serving the function of filling the gap between propositions. According to tonal contour results, the initial connective *nà* carried a falling contour, but the filler *NA* often had a flat contour with a low onset. A revised taxonomy considering the prosodic position may be required to separate demonstrative tokens in a finer way than the current system. By doing so, the results of the representative tonal contours of the automatic clustering experiment would be more closely connected with relevant linguistic implications. For the case of proximal demonstratives, the standard pronunciation *zhè* was clearly preferred over *zhèi* in both lexical and filler uses. Overall, an interesting issue related to phonological variants is whether they should be regarded as ordinary lexical entries with a distinctive pronunciation or merely pronunciation variants. This is concerned with the arrangement of lexicographical entries in dictionaries and the content of pronunciation dictionaries for automatic speech recognition systems. These tokens account for 3% of the overall corpus use. Thus, it is difficult to overlook their existence in spoken discourse. Our results indicated that selection preferences were apparent, suggesting that the dictionary should at least include two distinct lexical entries, *nà* and *nè* for automatic speech recognition.

Currently, computational modeling is commonly

applied on large-scale corpora to obtain representative regularities in language and speech in an objective manner. One-by-one description and assessment of all tokens in a corpus, irrespective of them being quantitative or qualitative, is unlikely considering the amount of the data. Representativeness must be based on mathematical calculations that exclusively rely on quantitative measurements and derived indices. Nevertheless, expert opinion and interpretation are still necessary to enhance a proper annotation scheme and to select an efficient data analysis method. The results of this study indicate that insightful observations can be derived by applying computational linguistic methodology with a well-designed taxonomy. The Chinese tone system is widely accepted in the notion of phonology. However, concerning the phonetic forms, only equivalents of canonical forms were mentioned in the literature in most cases, and the tonal description for lexical and filler demonstratives was not reported. Maekawa, Noshikawa and Tseng (2017) concluded that the acoustic features of vowels in simple fillers in Chinese can be clearly distinguished from those in ordinary words, but not from vowels involved in demonstratives. The results in our study further suggest that the difference may come from the suprasegmental level, such as duration or tonal patterns. The features we defined and used for the tone modeling experiment could be concretely helpful for classification tasks, such as those adopted by Maekawa, Nishikawa and Tseng (2017).

Finally, the development process of Chinese demonstratives can be related with the issue of the grammaticalization process. For example, the grammaticalization of directional complement constructions develops from spatial verbs, directional complements, and finally to tense markers. As the syntactic function is enhanced, the lexical meaning decreases. In the case of demonstratives, the spatial reference of demonstratives between the speaker and objects has developed into determiners, adverbs, connectives, and finally (in spoken discourse) fillers. Linguistic forms at different development stages are likely to be marked by certain linguistic characteristics. Phonetic forms, including prosodic properties, may be a part of the characteristics. We hope this paper provides motivation for more analyses on Chinese demonstratives. To investigate differences among these categories, experimental studies with robust controls over various variables are needed. The scope should cover the construction, meaning, function, and phonetic representation of Chinese demonstratives.

Acknowledgments

The author sincerely thanks two anonymous reviewers of the *Journal of the Phonetic Society of Japan* for their constructive comments as well as Yi-En Hsu, Ye-Sheng Lin, and Dr. Yi-Fen Liu for their careful work on data annotation and processing. The study presented in this article was financially supported by the Ministry of Science and Technology (MOST), under Grant 105–2410-H-001-084.

Notes

1) All Chinese examples used in this article are written in Hanyu Pinyin with an English translation. The tone is indicated by diacritics with *mā*, *má*, *mǎ*, *mà*, and *ma* denoting the high level, low rising, dipping, falling, and the neutral tone patterns. The word segmentation follows the principles used in the CKIP automatic word segmentation and POS tagging system (Chen, Huang, Chang and Hsu 1996).
2) *Ge* is a classifier that does not indicate a specific association with nouns, whereas *bù* is a classifier usually used for vehicles and machines.
3) These are *zhème*, *zhèmeyàng*, *zhèyàng*, *zhèyàngzi*, *nàme*, *nàmeyàng*, *nàyàng*, *nàyàngzi*. They all share the same meaning; they can be translated into "in this/that way" or "like this/that."
4) These words indicate a proximal/distal location, meaning "here/there" or "on this/that side."
5) It is publically distributed by the Association for Computational Linguistics and Chinese Language Processing (ACLCLP), http://www.aclclp.org.tw/.
6) The author would like to acknowledge the efforts contributed by Dr. Yi-Fen Liu for conducting the Fuzzy C-means clustering experiment, as well as Yi-En Hsu and Victor Ye-Sheng Lin for preparing and annotating the filled pause data.

References

Bezdek, J. C. (1981) *Patten recognition with fuzzy objective function algorithms*. Plenum Press: New York.

Boersma, P. and D. Weenink (2013) PRAAT: Doing phonetics by computer [Computer program], version 5.3.48. Available at: http://www.praat.org/ (accessed May, 1 2013).

Batliner, A., A. Kießling, S. Burger and E. Nöth (1995) "Filled pauses in spontaneous speech." *Proceedings of the 13th International Congress of Phonetic Sciences*, Vol. 3, 472–475, Stockholm.

Chao, Y. R. (1968) *A grammar of spoken Chinese*. University of California Press: Berkeley and Los Angeles.

Chen, K.-J., C.-R. Huang, L.-P. Chang and H.-L. Hsu (1996) "SINICA CORPUS: Design methodology for balanced corpora." *Proceedings of the 11th Conference on Language, Information and Computation (PACLIC 11)*, 167–176.

Cheng, L. L.-S. and R. Sybesma (1999) "Bare and not-so-bare nouns and the structure of NP." *Linguistic Inquiry* 30(4), 509–542.

Gabrea, M. and D. D. O'Shaughnessy (2000) "Detection of filled pauses in spontaneous conversational speech." *INTERSPEECH*, 678–681.

Hartigan, J. A. and M. A. Wong (1979) "A k-means clustering algorithm." *Journal of the Royal Statistical Society* 28, 100–108.

Hockett, C. F. (1963) "The problem of universals in language". In *Universals of Language*, J. H. Greenberg (ed.), 1–22, MIT Press.

Huang, S. (1999) "The emergence of a grammatical category definite article in spoken Chinese." *Journal of Pragmatics* 31(1), 77–94.

Levinson, S. C. (1983) *Pragmatics*. Cambridge University Press.

Li, C. N. and S. A. Thompson (1981) *Mandarin Chinese: A functional reference grammar*. University of California Press: Berkeley Los Angeles London.

Li, Y.-H. A. (1999) "Plurality in a classifier language." *Journal of East Asian Linguistics* 8, 75–99.

Lin, J.-W. (1997) "Chinese noun phrase structure: DP or NP?" In F.-F. Tsao and H. S. Wang (eds.) *Chinese language and linguistics III: Morphology and lexicon*, 401–434. Taipei: Academia Sinica.

Liu, Y.-F., S.-C. Tseng and R. J.-H. Jang (2014) "Phone boundary annotation in conversational speech." *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, 848–853. Reykjavik.

Lobanov, B. M. (1971) "Classification of Russian vowels spoken by different speakers." *Journal of the Acoustical Society of America* 49, 606–608.

Lyons, J. (1977) *Semantics I & II*. Cambridge University Press.

Maekawa, K., K. Nishikawa and S.-C. Tseng (2017) "Phonetic characteristics of filled pauses: A preliminary comparison between Japanese and Chinese." *Proceedings of the ESCA Workshop on Disfluency in Spontaneous Speech (DiSS 2017)*, 41–44. Stockholm, Sweden.

Schiffrin, D. (1988) *Discourse markers*. Cambridge University Press.

Shriberg, E. (2001) "To 'errrr' is human: Ecology and acoustics of speech disfluencies." *Journal of the International Phonetic Association. 31*(1), 153–169.

Simpson, A. (2001) "Definiteness agreement and the Chinese DP." *Language and Linguistics* 2(1), 125–156.

Swerts, M. (1998) "Filled pauses as markers of discourse structure." *Journal of Pragmatics* 30(4), 485–496.

Tang, C.-C. J. (1990) "A note on the DP analysis of the Chinese noun phrase." *Linguistics* 28, 337–354.

Tseng, S.-C. (2015) "Disfluent speech." In R. Sybesma, W. Behr, Y. Gu, Z. Handel, J. Huang, J. and Myers (eds.) *Encyclopedia of Chinese language and linguistics*, 105–108. Leiden, Netherlands: Brill.

Tseng, S.-C. (2013) "Lexical coverage in Taiwan Mandarin conversation." *International Journal for Computational Linguistics and Chinese Language Processing* 18(1), 1–18.

Wang, L. (1987) *A Grammar of Modern Chinese II* [1943]. Landeng Publisher.

Watanabe, M., K. Hirose, Y. Den and N. Minematsu (2008) "Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners." *Speech Communication 50*(2), 81–94.

Wu, Y. (2004) *Spatial Demonstratives in English and Chinese*. John Benjamins.

Zhao, Y. and D. Jurafsky (2005) "A preliminary study of mandarin filled pauses." *Proceedings of Disfluency in Spontaneous Speech (DiSS'05)*, 179–182.

(Received Jul. 23, 2017, Accepted Dec. 5, 2017)