Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20). 2008. Volume 1.
Edited by Marjorie K.M. Chan and Hana Kang. Columbus, Ohio: The Ohio State University. Pages 213-227.

Filled Pauses in L2 Chinese: A Comparison of Native and Non-Native Speakers

Chen-huei Wu
University of Illinois at Urbana-Champaign

The aim of this paper is to determine whether native and non-native speech can be predicted on the basis of temporal measurements of filled pauses by training a Classification and Regression Tree (Breiman et al. 1984). On the basis of the present results, several conclusions can be drawn: First, distinguishing between native and non-native speech can increase in accuracy based on temporal measurements of FPs. Among these variables, the rate of speech appears to be the best predictor. Second, this study suggests that information from the FPs 'uh' and 'um' is a useful predictor of fluency in further differentiating native/non-native speakers. Third, the classification can be accurately predicted with a small set of variables.

1. Introduction

"Fluency" is frequently used to describe the distinction between native and non-native language performance. The native speaker is capable of speaking smoothly and effectively, while the non-native speaker may not. The notion of fluency consists of multiple dimensions, including good oral command of phonetics and phonology, vocabulary and grammar, and the ability to talk at length with little or no disfluencies, such as filled pauses, silences, repairs, repetitions, etc.

In natural production of speech, fluency is not the opposite of disfluency. Some disfluencies may impede the flow of speech. In the natural speech produced by native speakers, disfluencies affect up to 10% of the words and an overall 1/3 of the utterances (Shriberg 2001). It is normal to have disfluencies in native speech. In Zhao and Jurafsky's (2005) study, native Mandarin speakers use more demonstrative pronouns e.g., *zhege* 'this' and *nage* 'that' as lexical fillers (hereafter, LFs) than filled pauses (hereafter, FPs) 'uh' and 'um'. In second language acquisition (hereafter, SLA), researchers have shown that the inaccurate L2 production of suprasegmentals, such as stress timing, peak alignment, speech rate, pause frequency and pause duration might contribute to foreign accents more than inaccurate L2 segmental production (Anderson-Hsieh et al. 1992; Koster and Koet 1993; Munro, 1995; Trofovimich and Baker 2006).

This observation raises the question about what actually distinguishes native speakers from non-native speakers. Different levels of fluency are not categorical distinctions, but continuous scales. The distinction between native and non-native speakers lies in the different status along the line of continuous degrees of fluency. In this

paper, three research questions will be addressed: 1) What are the similarities and differences of FPs and LFs between heritage and L2 learners? 2) Is the behavior of heritage learners more like native speakers or L2 learners? 3) Are there any specific patterns for heritage or L2 learners?

Clark and Fox Tree (2002) have proposed that the filled pauses 'uh' and 'um' have semantic distinctions on speech delay, that is, 'uh' and 'um' indicate short and long delays in speech, respectively. Therefore, filled pauses are treated as linguistic events, not like coughing, laughter in speech, which can not reflect the fluency of speech. The aim of this research is to determine whether fluency can be predicted on the basis of temporal measurements of filled pauses in speech, that is, the normalized frequency of FPs, mean length of FPs, normalized duration of FPs and rate of speech.

The goal of this study will be pursued by training a Classification and Regression Tree (Breiman et al. 1984. abbreviated as CART, hereafter) to distinguish native speakers from non-native speakers of Mandarin with a set of quantitative measurements of FP in their spontaneous speech in a classroom setting. In addition, the analysis will try to determine the contribution of the various quantitative variables to fluency and the thresholds of these variables to separate native and non-native speakers.

This paper is organized as follows. Section 2 briefly discusses the literature related to the definition of filled pauses, and the assessment of second language learners' fluency. Section 3 presents the data collection and the methodology conducted in this study. Section 4 shows the results and discussion. Section 5 will present a discussion.

2. Literature Review

2.1 Heritage speakers & Fluency in SLA studies

Until recently, researchers in SLA have investigated language learning development in heritage speakers (Polinsky 1997; Montrul 2002, 2004, forthcoming; Au et al. 2002). Heritage speakers are adult early bilinguals of minority languages. They might be the children of first generation immigrants or had moved to a L2 country at various ages during childhood. Under these conditions, the heritage language might not be completely acquired due to the fact that children of first generation immigrants have strong desires to fit into the new society. Therefore, the heritage language used at home might gradually be dominated by the majority language of the new society.

The competence and performance of heritage speakers varies to diverse degrees due to incomplete L1 acquisition (Polinsky 1997; Montrul 2002). Generally speaking, they have good speaking and listening abilities, and native-like levels of pronunciation and fluency. Due to the age effect, many characteristics of adult L2 acquisition have been identified, such as L1 transfer, incomplete attainment and fossilization. An interesting question is raised here as to whether heritage speakers perform more like native speakers or adult L2 learners.

Au et al. (2002) investigated voice onset time (hereafter, VOT) of stops /p, t, k, b, d, g/ and the lenition of voiced stops /b, d, g/ between two vowels in Spanish heritage

speakers defined as overhearing the language in childhood and minimal speaking experience. The result shows that the Spanish heritage speakers produced native-like VOT that are shorter in duration than the late L2 learners. Also, they produced lenited voiced stops more often than late L2 learners.

Most SLA phonology research focuses on segmental accuracy (vowels and consonants) as a function of the onset age of L2 acquisition or the length of residence. Few studies have investigated the suprasegmentals, such as stress assignment, peak alignment, frequency and length of pauses, and speech rate (Munro 1995; Trofovimich and Baker 2006). Trofovimich and Baker (2006) examined five suprasegmentals (stress timing, peak alignment, speech rate, pause frequency, and pause duration) and their affects on the production of adult Korean L2 learners of English and how each contributed to fluency and foreign accents. The findings show that the amount of L2 experience influenced the production of stress timing while the onset age of intensive L2 exposure influenced the others (peak alignment, speech rate, pause frequency, and pause duration). Moreover, pause duration and speech rate contributed more than others (stress timing, peak alignment) to foreign accents.

Hence, the hypothesis of the current study is that the heritage learners will behave more like native speakers, which means they use fewer FPs and more LPs than L2 learners and pause at appropriate phrase boundaries. Late L2 learners are expected to produce the most FPs among the three groups, seldom use LFs and pause at inappropriate positions and break the phrase boundaries.

2.2 Disfluencies: FPs and LFs

'Uh' and 'um' are called *filled pauses*, as opposed to silent pauses. Also, 'uh' and 'um' have been viewed as a symptom when trouble is detected in the process of producing speech (Levelt 1989) or as a non-linguistic signal when speakers are unable to produce the required words (Goffman 1981). In this perspective of 'filled pauses as non-linguistic elements', 'uh' and 'um' have been grouped with unfilled pauses or non-linguistic events such as laughter and coughing.

In other studies, filled pauses have been grouped as "fillers" with linguistic events, such as discourse markers 'you know' and 'well'. Clark and Fox Tree (2002) have proposed that 'uh' and 'um' are English words, which are planned for, formulated, and produced as parts of utterances just as other words are. In their 'filler-as-word hypothesis', 'uh' and 'um' are interjections with basic meanings (Clark and Fox Tree 2002: 79):

(a) Uh: "used to announce the initiation, at *t*('uh'), of what is expected to be a minor delay in speaking"
(b) Um: "used to announce the initiation, at *t*('um'), of what is expected to be a major delay in speaking"

In other words, producing *uh* itself indicates a brief delay, and *um*, a longer delay in the flow of speech. They are not non-linguistic events, which can reflect the fluency of speech.

In general, oral fluency is viewed as an important criterion for distinguishing between native and non-native language performance. Mohle (1984) suggested that the difference between natives and non-natives tends to be the frequency of hesitation rather than the presence of hesitation in speech. Cucchiarini et al. (1999) have demonstrated that it is possible to predict the fluency rating of L2 read speech on the basis of automatically calculated temporal measurements of speech quality such as speech rate, articulation rate, number and length of pauses, number of disfluencies, mean length of runs, and the phonation/time ratio. In their experiment, read speech of 20 native and 60 non-native Dutch speakers was evaluated by expert raters, including phoneticians, teachers of Dutch as a second language, and speech therapists. Based on the literature on temporal variables in L2 (Grosjean 1980; Towell et al. 1996), the following measures were investigated (Cucchiarini et at. 1999: 993-994):

(a) *ros* = rate of speech = # phonemes/total duration of speech including sentence-internal pauses
(b) *ptr* = phonation/time ratio = 100% x total duration of speech without pauses/total duration of speech including sentence-internal pauses
(c) *art* = articulation rate = # phonemes/total duration of speech without pauses
(d) #*p* = # of silent pauses = # of sentence-internal pauses of no less than 0.2 s
(e) *tdp* = total duration of all sentence-internal pauses of no less than 0.2 s
(f) *mlp* = mean length of pauses = mean length of all sentence-internal pauses of no less than 0.2 s
(g) *mlr* = mean length of runs = average number of phonemes occurring between unfilled pauses of no less than 0.2 s
(h) #*fp* = # filled pauses = # of uh, er, mm, etc.
(i) #*dy* = # dysfluencies = # of repetitions, restarts, repairs

These variables can be divided into three groups: speech rate, frequency effect, and duration. The results of their study suggest that: (i) all the variables are strongly related with the fluency ratings, with the exception of *mlp*; (ii) for fluency, the frequency of pauses is more relevant than their length. In other words, natives and non-natives differ more from one another in terms of pause frequency as opposed to pause length; (iii) rate of speech appears to be the best predictor for fluency rating because it incorporates the rate at which the speakers articulate the sounds and the number of pauses they make.

Cucchiarini et al. (2002) have further explored the relationship between temporal measures and perceived fluency in spontaneous speech. Due to the fact that pauses are more frequent in spontaneous speech than in read speech, the variables that do not

contain information about the frequency of pauses have almost no relationship with fluency.

Although much work related to FPs in English has been done, seldom have empirical studies in Mandarin spontaneous speech been carried out. Tseng (2003, 2006) focuses on repairs and repetitions in Mandarin spontaneous speech. Zhao and Jurafsky (2005) have reported a descriptive study of Mandarin FPs. Their research was based on the data from the LDC 98-HUM5 Mandarin corpus of telephone conversations, in which the FPs 'uh' and 'um' are hand-labeled. The result shows that Mandarin speakers intensively use demonstrative *zhege* 'this' and *nage* 'that' as major types of FPs. The difference between demonstratives *zhege*/*nage* and 'uh'/ 'um' as FPs lies in their distribution in different syntactic contexts. Demonstratives are more frequently used in a nominal-searching environment, while 'uh' and 'um' are more likely to be used at clause-initial positions.

In Zhao and Jurafsky's (2005) study, native Mandarin speakers use more demonstrative pronouns as fillers than the FPs 'uh' and 'um'. This suggests that the occurrences of FPs 'uh' and 'um' are not frequent in Mandarin spontaneous speech. Compared with the occurrence of 'um' and 'uh' in the CallHome English corpus, 'um' occurs 7.15 times per 1000 words; 'uh', appears 7.1 times per 1000 words. In the 98HUB Mandarin corpus, 'um' appears 1.46 times per 1000 words; 'uh' appears 2.55 times per 1000 words.

In sum, FPs 'uh' and 'um' as linguistic events in speech can be an indicator of fluency. The current study trains CART based on temporal measurements, normalized frequency of non-lexicalized FPs 'uh' and 'um', mean length of FPs, normalized duration of FPs and rate of speech to differentiate between native and non-native speakers whose language performance differs on the scale of fluency.[1] In addition, we try to use minimum variables to reach high accuracy in the classification.

3. Methodology

Speech data in a Chinese as a second-language corpus has been videotaped in third-year and fourth-year Chinese classes at UIUC from 2004 till the present. There are two speech formats in the corpus: one is called "Variety Show" and the other is called "Debate" (Shih 2006).

In the "Variety Show" format, there are 4 main sessions: opening, talk show, formal speech and comments. Students are asked to play roles, such as a chair of the whole show, the talk show host, or the formal speech speakers. Basically, the main part is the talk show: the host prepares several topics in advance and chooses students from the audience to come to the stage and answer the questions. The answers from students are spontaneous speech, while speech from the host may be prepared. Therefore, the speech style is casual, like a dialog or general conversation.

[1] 'uh' and 'um' were put in the same category to calculate variables..

In the "Debate" format, students are divided into two sides, the "positive" and "negative" sides. A specific topic is given in advance. Some of them prepare a formal speech to express their positions on this topic; others prepare questions to ask the opposing side. Since students can prepare their speeches beforehand, some of them just read their draft out aloud. In this case, disfluencies in this style speech may be reduced. Overall, the speech style of Debate is more formal than that of Variety Show.

Based on the background information of students, they can be grouped into four types as seen in the table:

TABLE 1. BACKGROUND INFORMATION OF STUDENTS

	BACKGROUND INFORMATION	NUMBERS	LABELS
1	Students whose native language is English or Korean	9	Non-native
2	Students whose native language is Chinese, but who are born in the United States.	4	
3	Students whose native language is Chinese, but who arrived in the states for the first time during middle school or high school (early immersion students of English).	12	Native
4	Student whose native language is Chinese, but who arrived in the states for the first time after the age of 18 (late immersion students of English)		

In this study, the data produced by three Chinese instructors and twelve heritage students in group (3) and (4) and nine non-native speakers in group (1) are used. In the four hours of speech material, "Variety Show" takes up one hour, and "Debate" takes up three hours.

Four hours of speech material was orthographically transcribed in traditional Chinese characters by transcribers in Taiwan and checked by the author. In the transcription, special indications were used to indicate disfluencies or non-speech events:

(a) FPs: 'uh' is transcribed as <呃>, <哦> and 'um' as <嗯>.
(b) speaker noise: laughter, clapping, coughing, and other loud noise, transcribed as <LAUGH>, <CLAP>, <COUGH>, and <NOISE>, respectively.
(c) unclear or overlapped speech is transcribed as <SKIP>.

The speech material was also segmented with Praat and the duration of FPs was extracted from the manual label. Previous studies (Grosjean. 1980; Towell et al. 1996;

Cucchiarini et al. 1999, 2002) have shown a number of quantitative variables related to Fluency. Four variables with slightly different definitions were used in this study[2]:

TABLE 2. DEFINITION OF QUANTITATIVE MEASUREMENTS

NAMES	DEFINITIONS
Normalized frequency of FPs	total # of FPs / total # of syllables
Mean length of FPs	sum (duration) of FPs / total # of FPs
Normalized duration of FPs	sum (duration of FPs) / total speech time
Rate of Speech	total # of syllables / total speech time (seconds)

These definitions were employed to train CART using the method of cross validation. The frequency of FPs was normalized by the total number of syllables. Also, the total duration of FPs produced by each speaker was normalized relative to his/her total speech time. Rate of speech was calculated using the total number of syllables against total speech time. The total syllable and total speech time include utterance of internal FPs.

4. Results and Discussion

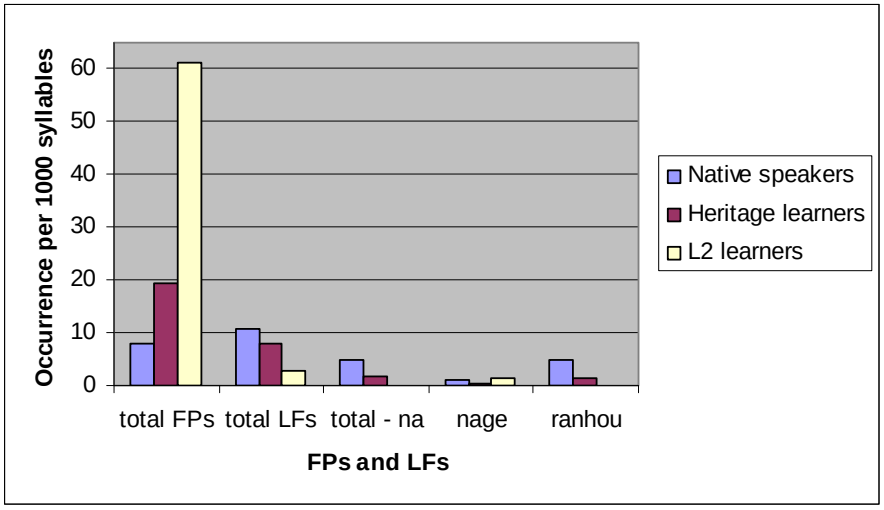
Table 3 shows the total FPs and each LF used out of 1000 syllables in each group's speech production. Figure 1 demonstrates that L2 learners use the most FPs; the native speakers used the fewest FPs and heritage learners are somewhere between native speakers and L2 learners. In contrast to the production of FPs, native speakers used the most LFs, L2 learners used the fewest LFs and the heritage speakers are somewhere between these two groups. For LFs, native speakers used three of them, *na* 'then', *nage* 'that' and *ranhou* 'and then' while L2 learners only used *nage* 'that' in their speech when they had trouble retrieving lexical items. Heritage learners used LFs *na* 'then' and *ranhou* 'and then' in their speech, but less than native speakers did.

TABLE 3. OCCURRENCE OF FPS AND LFS PER 1000 SYLLABLES

	Total FPs	*na*	*nage*	*ranhou*	Total LFs	Total syllables
Native speakers	7.94	4.69	1.08	4.87	10.65	5539
Heritage learners	19.25	1.86	0.51	1.35	7.94	11843
L2 learners	61.22	0	1.52	0	2.71	7906

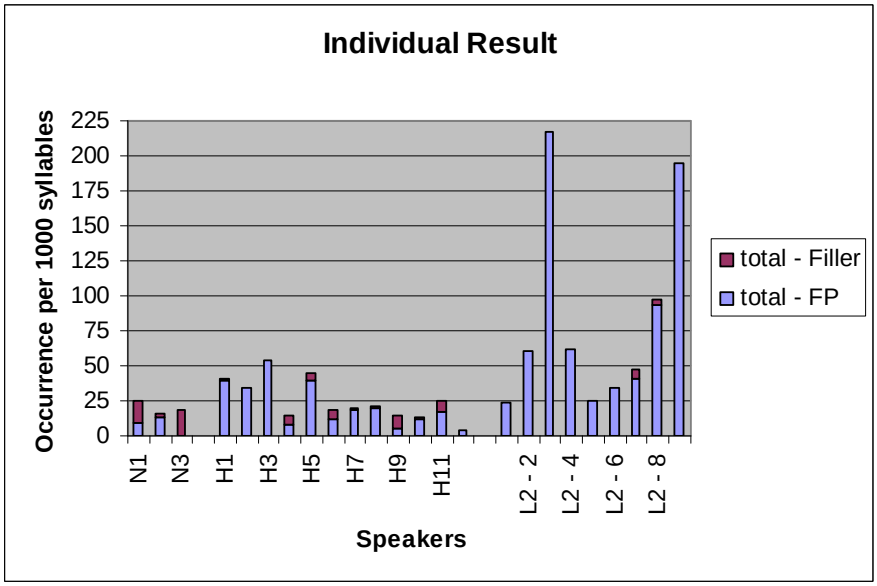
[2] Grosjean (1980) used syllables to calculate variables, while Cucchiarini *et at.* (1999, 2002) used phonemes instead.

FIGURE 1. OCCURRENCE OF FPS AND LFS PER 1000 SYLLABLES



In Figure 2, we can see the trend that heritage learners produced more FPs and LFs than native speakers, and L2 learners produced more than heritage learners. Some patterns of heritage learners are more like L2 learners, such as H1/H2 – L5, H3 – L2/L4 (they only produced FPs and had similar frequency) and H5 – L7 (they produced fewer LFs and more FPs and had similar frequency as well). Others are more like native speakers, such as H4/H6/H9 – N2 and H11 – N1.

FIGURE 2. OCCURRENCE OF FPS AND LFS PER 1000 SYLLABLES



N = native speaker; H = heritage speaker; L2 = L2 learner

The result here confirmed the findings in Zhao and Jurafsky (2005), which reported that native Mandarin speakers tend to use demonstrative *zhege* 'this' and *nage* 'that' as major types of FPs. In addition, we found that the performance of heritage learners is more native-like and L2 learners are less fluent than heritage speakers. The level of fluency of heritage learners is somewhere between native speakers and L2 learners. The findings in the current study are in line with previous literature (Au et. al. 2002; Trofovimich and Baker, 2006).

Based on the result above, in the training of CART, the heritage learners and native speaker were grouped into the label of native; L2 learners were labeled as non-native.

Table 4 shows the occurrence of FPs per 1000 words produced by native and non-native speakers. As we can see, the occurrence of the FP 'uh' produced by non-natives is five times more frequent than that by natives, while 'um' of non-native speakers is 1.5 times more frequent than that of native speakers. Both natives and non-natives produce more 'uh' than 'um'.

TABLE 4. OCCURRENCE OF FILLED PAUSES PER 1000 WORDS

FPs	Native	Non-native
uh	10.88	54.27
um	4.02	6.76
Total	14.90	61.04

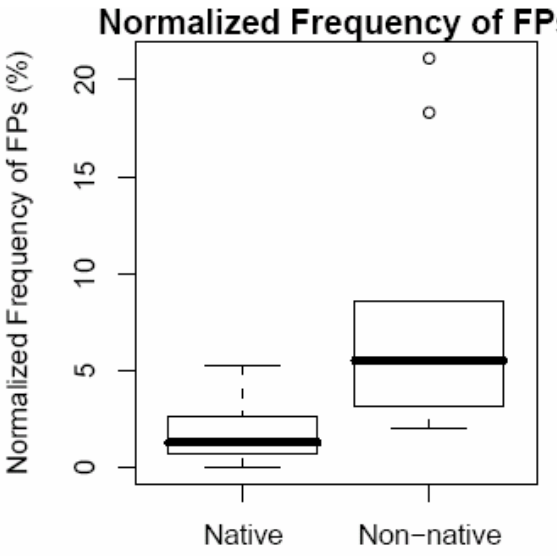
Table 5 shows the average of each variable, and boxplots in Figure 3-6 present the distribution of each variable. Normalized frequency and normalized duration of FPs show that non-native speakers have more FPs in their speech. Also, non-native speakers have longer lengths of FPs than native speakers do. For rates of speech, the native speaks faster (3.77 syllables per second) than the non-native (2.74 syllables per second).

TABLE 5. AVERAGE VALUE OF ALL VARIABLES

Variables	Native	Non-native
Normalized Frequency FPs (%)	1.78	7.84
Mean length of FPs (ms)	245.73	298.98
Normalized duration of FPs (%)	1.66	6.07
Rate of speech (seconds)	3.77	2.74

As we can see, between the native and the non-native, the distribution of normalized frequency of FPs and normalized duration of FPs (Figure 3 and Figure 5) have similar patterns, the distribution in the non-native is more scattered than that in the native. These two variables show similar information for FPs; normalized frequency is calculated by the number of FPs relative to total syllables, and normalized duration of

FPs reflects the time information. The distribution of the mean length of FPs (Figure 5) and rates of speech (Figure 6) in the native is more scattered than that in the non-native. The mean length of FPs in the non-native group is around 300 ms, while the ranges of the native is from 180-350 ms. For rates of speech, the non-native groups average around 2.7 syllables per second, while the native ranges from 2.6 to 4.75 syllables per second.



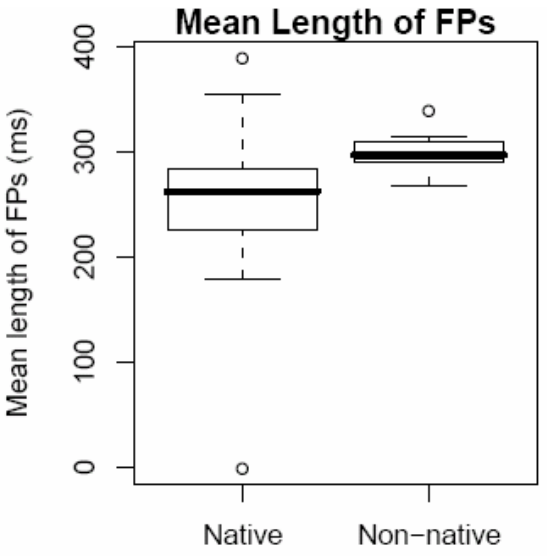
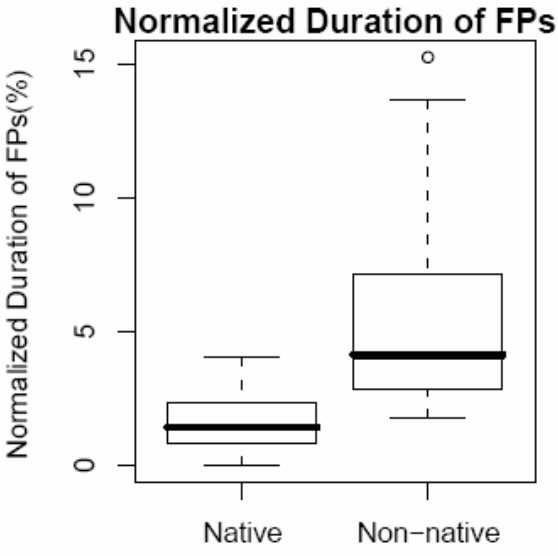


FIGURE 3. NORMALIZED FREQUENCY OF FPS **FIGURE 4. MEAN LENGTH OF FPS**



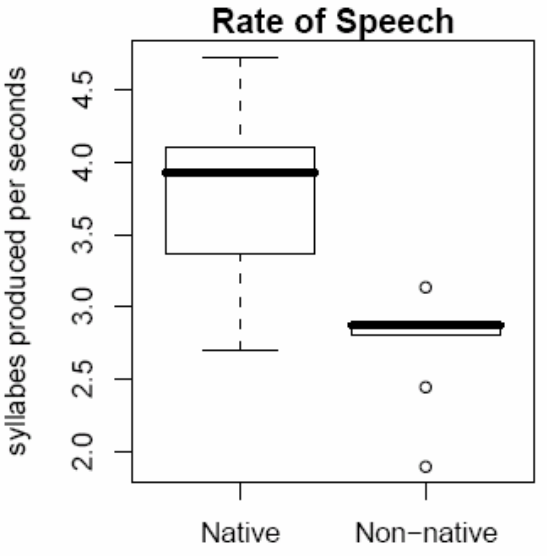


FIGURE 5. NORMALIZED DURATION OF FPS **FIGURE 6. RATE OF SPEECH**

CART were trained using variables listed in Table 2 in two ways: CART trees were created for each of these variables and the whole set of variables to compare the performance of each variable and the whole set. In Figure 7, when the speech rate is smaller than 3.14054, the speaker was identified as a non-native speaker (labeled as L2 in the tree); when the speech rate is bigger than 3.14054, the speaker is classified as a native speaker.

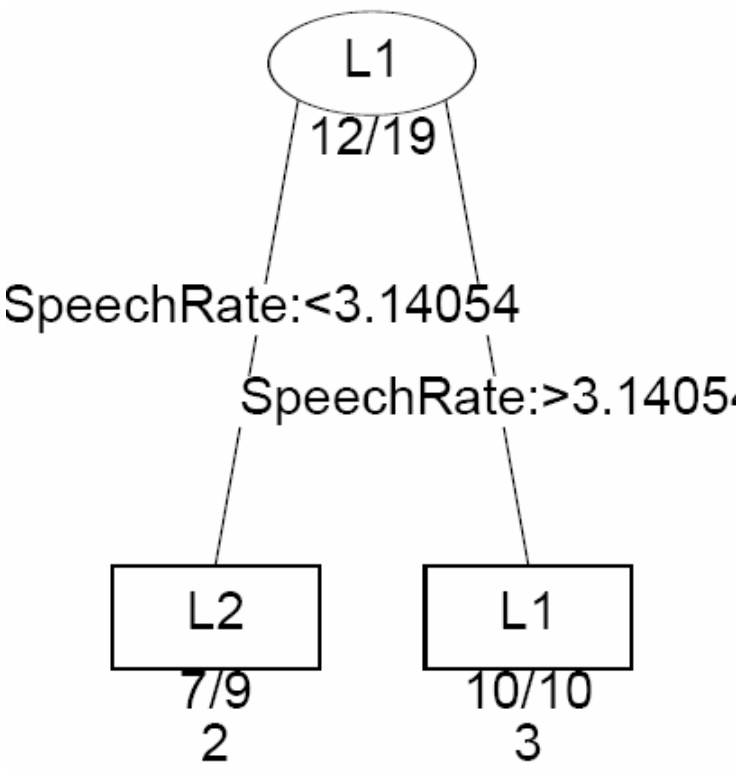


FIGURE 7. PREDICTING SPEAKER CLASSIFICATION

The system was evaluated by precision, recall, and the F-score as defined in the following way:

TABLE 6. DEFINITION OF PRECISION, RECALL AND F-SCORE

NAMES	DEFINITIONS
native speaker precision	# of correct-native prediction / # of predicted native
native speaker recall	# of correct-native prediction / # of actual native
native speaker F-score	2 x precision x recall / (precision +recall)
non-native speaker precision	# of correct-non-naïve prediction / # of predicted non-native
non-native speaker recal	# of correct-non-native prediction / # of actual non-native
F-score non-native	2 x precision x recall / (precision +recall)

The results of the performance for each variable and the whole set of variables are shown in Table 7.

TABLE 7. PRECISION, RECALL AND F-SCORE

	Native speakers			Non-native speakers		
Baseline	.625			.375		
Evaluation	precision	recall	F-score	precision	recall	F-score
Normalized Frequency of FPs	.750	.800	.754	.600	.500	.507
Mean Length of FPs	.783	.800	.779	.600	.600	.567
Normalized Duration of FPs	.600	.800	.686	.680	.380	.480
Rate of speech	.950	.867	.891	.867	.900	.853
All variables	1	.867	.920	.867	.900	.853

The baseline (.625 for the native and .375 for the non-native) represents the probability that the majority of the speakers in the data are native speakers. Among the performances of each variable, *rate of speech* has the best predictive power for identifying native (F-score: .891) and non-native speakers (F-score: .853), while normalized duration of FPs has the least ability to distinguish the native (F-score .686) and the non-native (F-score: .480). The performance of normalized frequency of FPs and mean length of FPs by themselves reaches the F-score above .750 for predicting the native, while the performance for predicting the non-native is not as good as the prediction of the native.

When all the variables are combined, the result shows excellent performance (F-score: .920) for recognizing the native, while the performance for identifying the non-native is the same as using the variable 'rate of speech' by itself.

These results show that the variable *rate of speech* is the best quantitative predictor for classifying the native and non-native. Variables *normalized frequency of FPs* and *mean duration of FPs* are also useful (to a less degree) in the task, which improves the F-score of predicting the native. In previous studies (Cucchiarini et al. 1999,

2002), they had reported that variables contain information about the frequency of FPs such as *rate of speech, phonation/time ratio, number of silent pauses per minute, duration of silent pauses per minute and mean length of runs* are particularly related to fluency in spontaneous speech, but not *articulation rate* and *mean length of silent pauses*. Our results show that the variable *rate of speech, normalized frequency of FPs* and *mean length of FPs* are good quantitative predictors, among which the latter two improve the F-score for predicting the native. Rate of speech is a complex variable which incorporates two aspects of frequency of FPs (total number of syllables including FPs) and the time information of FPs (Cucchiarini et al. 1999).

In our observation, perhaps frequency of FPs of non-native speakers is underestimated because non-natives produce fewer filled pauses in the "Debate" speech format. Some students just read their draft for the speech aloud. Therefore, the production of speech was not entirely spontaneous.

5. Conclusion

On the basis of the present results, conclusions can be drawn: first, overall, heritage learners are orally more fluent than L2 learners. Some of them used fewer FPs and LPs than L2 learners, just as native speakers did, while some of them use as many FPs as L2 learners did. Second, the behavior of heritage learners is at a degree between native speakers and L2 learners, but they are near the native level. Third, the performance for distinguishing between the native and the non-native can reach excellent accuracy based on temporal measurement, normalized frequency of FPs, mean length of FPs, and rate of speech. Among these variables, rate of speech appears to be the best predictor. Fourth, this study suggests that information about FPs 'uh' and 'um' can accurately predict fluency thus differentiating between native/non-native speakers. Fifth, the classification can be accurately predicted with a small set of variables.

A potential further study of this current paper is to train the system to recognize the fluency rating of speakers instead of the native and non-native labels using more factors pertaining to other disfluencies such as repairs, repetitions, lexicalized fillers, prolongation and silent pauses, sentence length, lexical choice and grammar.

It is possible that some non-native speakers have achieved native-like oral fluency and that it is not necessary for native speakers to produce fluent speech. Since the system was trained to recognize labels of the native and non-native instead of fluency rating, it is reasonable that the system treats native-like performance produced by non-native speakers as natives. We believe that the system will be improved with the feature sets we propose in the study if the system is trained to recognize the speaker's fluency rating.

REFERENCES

Anderson-Hsieh, J., Johnson, R., & Koehler, K. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42. 529-555.

Au, Kit-fong, Terry., Leah Knightly M., Sun-Ah Jun, and Janet S. Oh. 2002. Overhearing a language during childhood. *Psychological Science* 13. 238-243.

Breiman, Leo, Jorome H. Friedman, Richard A. Olshen, and Charles J. Stone, 1984. *Classification and Regression Tree*, Wadsworth, Belmonk, CA.

Clark, H. Herbert and Fox Tree J. E. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84. 73-111.

Cucchiarini, Catia, Helmer Strik and Lou Boves. 1999. Quantitative assessment of second language learner's fluency by means of automatic speech recognition technology. *JASA* 107(2). 989-999.

Cucchiarini, Catia, Helmer Strik and Lou Boves. 2002. Quantitative assessment of second language learner's fluency: Comparisons between read and spontaneous speech. *JASA* 111(5). 2862-2872.

Grosjean, Francois. 1980. Temporal variables within and between languages, in *Towards a Cross-Linguistic Assessment of Speech Production*, edited by H. W. Dechert and M. Paupach (Lang, Frankfurt), pp. 39-53.

Goffman, Erving. 1981. Radio talk. In E. Goffman (ed.), *Forms of talk* (pp. 197-327). Philadelphia, PA: University of Pennsylvania Press.

Koster, Cor J. and Ton Koet. 1993. The evaluation of accent in the English of Dutchmen. *Language Learning* 43.69-92.

Levelt, Willem. J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Munro, Murray J. 1995. Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition* 17.17-34.

Mohle, D. 1984. A comparison of the second language speech production of different native speakers, in *Second Language Production*, edited by H. W. Dechert, D. Mohle and M. Raupach (Narr, Tubingen), pp. 26-49.

Montrul, Silvina. 2002. Imcomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingalism: Langauge and Cognition* 5, 1. 39-68.

Montrul, Silvina, 2004. Subject and object expression in Spanish heritage speakers, A case of morpho-syntactic convergence. *Bilingalism: Langauge and Cognition* 7, 2.125-142. Special issue on *Convergence* edited by Jacqueline Toribio.

Montrul, Silvina. Forthcoming. *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor*. Amsterdam: John Benjamins: Manuscript version.

Polinsky, Maria. 1997. American Russian: Language loss meets language acquisition. Proceedings of the Annual Workshop on formal Approaches to Slavic Linguistic, 370-406. Ann Arbor: Michigan Slavic Publications.

Shriberg, Elizabeth. 2001. To 'errrr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetics Association* 31, 1. 153-169.

Shih, Chilin. 2006. The language class as a community: A task design of speaking proficiency training. *Journal of the Chinese Language Teachers Association* 41, 2. 1-22.

Towell, R., Hawkins, R., and Bazergui, N. 1996. The development of fluency in advanced learner of French. *Applied Linguistics* 1. 84-119.

Tseng, Shu-chang. 2003. Repairs and repetitions in spontaneous Mandarin. Proceeding of DISS'03. 73-76.

Tseng, Shu-chang. 2006. Repairs in Mandarin conversation. *Journal of Chinese Linguistics* 34, 1. 80-120.

Trofimovich, Pavel and Baker, Wendy. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28, 1. 1-30.

Zhao, Yuan and Dan Jurafsky. 2005. A preliminary study of Mandarin filled pauses. Proceeding of Diss'05. 179-182.