

Prosodic features, self-monitoring, and dysfluency in native and non-native Mandarin speech

Bei Yang

University of Wisconsin – Madison

This study explores the relationship between the prosodic features for time delay, self-monitoring in speech production, and perceived dysfluency. In this study, twenty native and non-native speakers of Chinese took a speech test. Each speech was transcribed, prosodic features were assigned symbols, and the coding system traced self-monitoring. An additional twenty-eight native speakers assessed the fluency of the speech samples, and then the researcher matched assessment results with symbols and coding, and analyzed them. The results indicate that uh/um and self-monitoring influence perceived dysfluency in most cases while other prosodic features do not; that the filled pause in non-native speech is a salient feature of perceived dysfluency; and how a dysfluency is perceived. The study also finds the native speakers' perception bias.

Keywords: fluency, self-monitoring, pause, prosodic feature, production and perception

1. Introduction

1.1 What is fluency?

Lennon (1990) distinguished the narrow sense of fluency from the broad sense. The broad sense of fluency refers to a global oral proficiency including various components, such as accuracy and appropriateness. The narrow sense of fluency refers to only one component. However, fluency also indicates listeners' assessment of how smooth and rapid the speech is (Isaacs & Trofimovich, 2011). "Fluency is an 'automatic procedural skill' (Schmidt, 1992) on the part of the speaker and a perceptual phenomenon on the part of the listener" (Guillot, 1999) (Derwing, Thomson, & Munro, 2006: 85). In terms of production, fluency is dependent on

automaticity (Schmidt, 1992) which was first proposed as a part of speech planning by Levelt (1989). According to Raupach (1980), grammatical encoding becomes automatic gradually with the development of fluency, since “activities of planning and uttering can be executed nearly simultaneously by the speaker of the language” (Rehbein, 1987: 104). In terms of perception, fluency is a listener’s impression of psycholinguistic processing (Lennon, 1990). Therefore, fluency reflects a speaker’s speech plan, yet from the point of view of a listener’s decoding.

In the current study, the working definition of fluency follows and extends from the previous definitions and descriptions, including Lennon (1990), Schmidt (1992), Guillot (1999), Kormos & Denes (2004), Derwing et al. (2006), and Isaacs & Trofimovich (2011). In oral speech production, whether declarative knowledge is used correctly and appropriately can be determined by the correctness of words, sentences, and clauses produced. This is accuracy. Whether declarative knowledge is processed without time delay or repair, such as using an appropriate amount of time to select a word, or having no restructuring chunk, is treated as fluency. On the surface form, fluency is usually determined by the qualitative and quantitative properties of the elements that cause the time delay, such as prosodic features including silent pauses, filled pauses, and prolongations; and the elements that cause speech re-planning because of the production of an inappropriate chunk, such as self-monitoring, including repairing incorrect words. However, not all prosodic features and self-monitoring affect fluency of oral speech, since a listener also needs time to comprehend speech locally, and the delayed time during production is sometimes helpful for comprehension.

1.2 What is studied?

According to our working definition, two common approaches have been used to observe and analyze fluency in previous studies. From the production perspective, temporal features and self-monitoring are investigated. From the perception perspective raters are hired for assessment. Some features in speech production have been shown to be correlated with perception assessment, such as speech rate and average run of the speech (e.g. Kormos & Denes, 2004; De Jong & Perfetti, 2011; Derwing et al., 2006). However, some features show no correlation; for example, silent pauses and filled pauses do not influence perception of fluency (Kormos & Denes, 2004), although they could be indicators of a delay in speech planning. This shows a gap between production processing and perception assessment.

Some studies of perception of dysfluency in second language speech have been somewhat inconsistent. A few studies (Voss, 1979) have indicated that pauses and hesitation phenomena provide perceptual barriers, while others (Blau, 1991; Fox Tree & Schrock, 1999; Watanabe et al., 2008) found that filled pauses provide more

time for listeners to comprehend a passage. The working definition in each case could explain the contradictory results. Basically, prosodic properties, especially duration, represent the psychological process of speech planning at the production level. During perception, prosodic features could be interpreted from the decoding perspective. For example, if a listener decodes a sentence more slowly than a speaker's encoding, that listener may use the time that is filled by prosodic features to decode information, and then dysfluency will not be perceived. Instead, it could facilitate comprehension. On the other hand, if a listener decodes a sentence faster than a speaker's encoding, then a dysfluency will be perceived. However, the physical features, including pauses, could also be independent from the grammatical decoding process. For example, a foreign language learner, who has low proficiency and cannot understand a native speaker's speech, can still judge whether the speech is fluent or not. This could be an extreme example indicating that a listener simply uses physical features, such as silent pauses, to make an assessment. This kind of extreme example will not be considered. We aim to explore in which cases prosodic features and self-monitoring influence perceived fluency when a listener is decoding speech production, and how they accomplish this. Conversely, in which cases do they not influence fluency?

In previous studies (e.g. Derwing et al., 2004; Derwing et al., 2006; Kormos & Denes, 2004; Lennon, 1990; Towell et al., 1996), researchers have explored global fluency based on perception, and its relationship to different prosodic features. The current study further explores in which local contexts prosodic features (such as pauses) and self-monitoring are perceived as local dysfluency, in Chinese as a second language. Both native and non-native speakers' speech production samples were used and rated by native speakers of Mandarin Chinese. These are the research questions:

1. What is the relationship between global perception and local perceived dysfluency among native speakers and non-native speakers?
2. What is the extent of the relationship between prosodic features/self-monitoring and local perceived dysfluency? In which contexts are prosodic features or self-monitoring produced when dysfluencies are perceived?
3. To what extent do native speakers and non-native speakers differ on prosodic features and self-monitoring for perceived dysfluency?

1.3 Assessment of fluency

Generally, two kinds of methods are employed to assess fluency variables. One is to investigate information from oral speech production, such as pauses and self-repair. The other is to employ raters to assess the global fluency of a speech sample,

for example, with a global fluency score. The current study will follow the convention and use these two methods, since it mainly explores the relationship between prosodic features in oral speech production and perception assessment. However, it also makes use of some new methods since it assesses local dysfluency as well as global assessment.

1.3.1 *Analysis of production*

From the perspective of production, various methods have been used to explore which variables are related to dysfluency, such as speech rate (e.g. Griffiths, 1991; Riggensbach, 1991; Trofimovich & Baker, 2006; O'Brien, 2014), pause frequency and duration (e.g. Albrechtsen et al., 1980; Cenoz, 2000; Trofimovich & Baker, 2006; Derwing et al., 2008), articulation rate (e.g. Riggensbach, 1991; Kormos & Denes, 2004; O'Brien, 2014), phonation-time ratio (e.g. Towell et al., 1996; Kormos & Denes, 2004), number of words (Trenchs-Parera, 2009), and pace (e.g. Vanderplank, 1993).

Many of the indexes above are not consistently related to fluency. Several indexes are related to silent pauses, such as speech rate. However, the duration of silent pauses that could cause dysfluency has had different definitions in the previous studies. In 1.3.3, the definition of the silent pause is discussed. Since local perceptual information is the focus, the current study utilizes original prosodic features, including silent pauses, filled pauses and prolongations, and counts their frequency to perform the analysis (see 2.3.2). The mean length of runs is not used since complexity is not observed in this study. Pace (the number of stressed words in a minute) is not employed either, since stress does not work for Chinese at the lexical level; therefore it is also not considered in the current study.

1.3.2 *Perception assessment*

In previous studies, perception assessment focused on global scores; that is, raters provided a fluency score for each speech sample while ignoring dysfluent details that are generally provided by the prosodic and psycholinguistic information locally in oral speech production. Only Riggensbach (1991) used micro-analysis to observe such details of production information. In the current study, the perception assessment includes both global scores and local information, so that perception assessment results can be connected to production details.

As for perception assessments, scholars have employed numbers of raters varying from 1 to 33 to judge global fluency (Reich, 1980; Brennan & Schober, 2001; Arnold et al., 2004; Trofimovich & Baker, 2006; Derwing et al., 2004; Derwing et al., 2009). Following these parameters, 28 raters were recruited in the current study.

1.3.3 *Some terms used in the production analysis*

In the analyses of dysfluency, many terms may be used to describe prosodic features, linguistic features, and psychological features. However, the definitions of some terms vary from author to author; below are three examples salient for this study.

First, a silent pause is an important feature of dysfluency. But the definition of the term varies. It has been defined as any break in the speech stream longer than 100ms in length, a measure commonly used in pausological research (Riazantseva, 2001). In the field of second language acquisition, pauses are treated as a feature of hesitation phenomena. Then how long of a pause is regarded as a hesitation phenomenon? Some scholars believe that pauses over 200ms should be regarded as hesitation phenomena (Kormos & Denes, 2004). Following common practice, some scholars think that pauses longer than 400ms indicate dysfluency (Derwing et al., 2004; Derwing et al., 2009). In other studies, silent pauses are defined as pauses lasting over 250ms (Towell et al., 1996). Most studies cite Riggensbach's (1991) definition of the silent pause. Riggensbach categorized three kinds of pauses: a "micropause" which is less than 200ms, a "hesitation pause" which is between 300ms and 400ms, and an "unfilled pause" that is 500ms or greater (Riggensbach, 1991: 426). Since previous studies also found that short pauses of 400ms or less occur frequently in native speakers speech (Fillmore, 1979; Deese, 1984), a silent pause that is used to observe dysfluency is defined as a pause over 400ms in the current study.

Second, in most studies, filled pauses were analyzed. Filled pauses might be frequently-used interjections such as *uh* or *um*. However, filled pauses could be other things as well, such as content words. Additionally, each language has its own language-specific filled pauses. For example, *ránhòu* is a commonly-used filled pause in Mandarin Chinese. The current study mainly focuses on two kinds of filled pauses: one is non-word filler, such as *uh/um* and *ah*, and the other is content-word filler, such as *ránhòu*.

Third, many studies have explored self-monitoring. Repair and repetition are analyzed in the current study. Levelt (1989) distinguishes among three kinds of repairs. However, there are four kinds of self-monitoring types in the current study: repetition (a speaker's repeating the same unit, such as a word, a phrase, or a clause), instant repair (a speaker's retracing back to a single troublesome word, and replacing it with a correct word), anticipatory retracing (a speaker's retracing back to some point prior to the error), and restart (a speaker's dropping the original syntactic structure and just starting over).

2. Methodology

2.1 Participants

A total of 10 American learners of standard Chinese, i.e., *Putonghua* (mean age: 20.4; range: 19–21; 5 males, 5 females) and 10 native speakers of Chinese (mean age: 25.50; range: 22–29; 5 males, 5 females) participated in the production test of the study. These participants were recruited from a Midwest university in the United States. The non-native speakers had all taken Chinese for two years in the U.S. and had not studied in China. According to the short-form proficiency test they took, their Chinese oral proficiency was at the intermediate level, except for one learner whose proficiency was novice-high.

In addition, a total of 28 native speakers of *Putonghua* (mean age: 22.3; range: 18–27; 12 males, 16 females) were recruited from a university in Southeast China to assess the speech samples. The major of these participants was Chinese language and literature. All native and non-native participants were paid hourly.

2.2 Procedure

The research instrument was a 20-minute short-form speaking proficiency test. In the test, four questions were adopted from the Simulated Oral Proficiency Interview (SOPI) developed by the Center for Applied Linguistics (1995). SOPI, an alternative format of the Oral Proficiency Interview (OPI), is a semi-direct speaking test, which is tape-mediated instead of face-to-face (Ke, 2005). The adapted test in the current study covered topics of picture description, favorite reading, preference of eating at home or in a restaurant, and how to hunt for jobs in the U.S. Each participant's oral proficiency was rated based on the 20-minute speaking test. Ten learners of Chinese and 10 native speakers of Chinese took the proficiency test individually. The participants were given a test sheet while instructions were delivered through headphones attached to a computer. Participants' speech was recorded with an Olympus LS-10 Linear PCM Recorder and a Sony ECMMS907 Digital Recording Microphone.

In the current study, only the speech on the third topic was selected for the perception assessment on fluency. Topic Three required a speaker to state a preference for either eating at home or eating in a restaurant, and provide some reasons for the preference. Each participant had 20 seconds to plan the speech, and then was asked to speak for 75 seconds.

Previous studies indicated that task design and type affect speech fluency (Derwing et al., 2004; Tavakoli, 2009). In the current study, there were two reasons to select the third topic. First, more syllables on this topic were produced by

non-native speakers compared with the other topics. Second, some vocabulary and relevant patterns used by non-native speakers were learned in the second-year Chinese textbook, so non-native speakers in the current study produced some similar words and patterns; thus it was easier to compare their speech samples, because there were fewer variations in linguistic structure and discourse contents. In terms of this topic, their proficiency was the same, i.e., at the intermediate level.

The speech produced by the 10 non-native speakers and 10 native speakers was transcribed, and prosodic features and self-monitoring were identified in the transcriptions. Then the researcher removed punctuation and symbols indicating silent pause, filled pause, prolongation, and self-monitoring, so that only characters transcribed from the oral speech were retained. The adapted transcription of the 20 speech samples was copied into a file.

During the perception assessment section, 28 native speakers were seated in a phonetics lab, and provided with a portfolio including the adapted transcription.

Before the perception assessment, the raters were instructed on how to evaluate fluency. They were requested to assess fluency while comprehending the speech. When a rater found that the speaker was thinking about how to organize or produce the speech, for example, using time to plan speech, the rater could note it as a local dysfluency.

Raters were required to listen to each sample three times. The first time, they were asked to provide a global score for each speech sample, scaled from 1 to 10. The 10-scale system, in which 10 refers to the highest fluency and 1 refers to the lowest fluency, is similar to the general grading system in which 10 refers to the best performance. Therefore, it was easy for the raters who were not familiar with the methods of fluency study to assess the speech samples. On the second listen, the raters were required to use underlines and vertical lines to indicate places where they heard dysfluency. Underlines were used to mark dysfluent chunks, while vertical lines were used for boundaries. Finally, the raters were allowed to listen to the speech samples for a third time, to double-check their markings. The raters could write some comments at the end of each sample. Two one-hour assessment sessions were scheduled. There was a 15-minute break between the two sessions.

2.3 Analysis

2.3.1 *Transcription*

Two Chinese native speakers transcribed 10 non-native speakers' and 10 native speakers' proficiency test recordings into Chinese characters with the following symbols to signal prosodic features:

- (0.25) a number in single brackets indicates an interval of exact time (two digits after the decimal).
- (.) a full stop in single brackets indicates an interval of a tenth of a second or less in the stream of speech.
- oh: a colon indicates an extension of the sound or syllable it follows (more colons prolong the stretch)

Filled pauses, e.g., “uh” and “um”, were identified and transcribed accordingly. The computer program Audacity was used to measure exact silent pause duration. Audacity is open-source software that is used to record and edit sounds. Prolongation was determined by listening to the speaker’s general speech rate and comparing the duration of the syllable with other syllables nearby using Audacity. Self-monitoring items were marked on the righthand column of the transcription sheet. The researcher double-checked the transcription.

2.3.2 *Counting and coding system*

The means and standard deviations of the global scores of the native speakers group and the non-native speakers group were calculated, and then the researcher counted the number of local dysfluencies. If over 68% of raters perceived and marked a dysfluency, the dysfluency was considered a local perceived dysfluency. This criterion was based on the central limit theorem. A local dysfluency was marked if more than 19 raters marked a chunk or a clause boundary as a dysfluency within two syllables’ difference.

Three kinds of prosodic features were coded: silent pauses that are longer than 400ms; filled pauses including “uh”, “um”, “ah”, *ránhòu* (‘after that’), and prolongation. Four types of self-monitoring were coded: repetition, instant repair, anticipatory retracing and restart.

After marking local dysfluencies, the researcher counted the number of silent pauses, filled pauses, prolongations, and self-monitorings in the speech production transcription according to the location of the dysfluency. Generally, the prosodic features or self-monitoring coded occurred within a marked dysfluent chunk, or no earlier or later than two syllables of a marked dysfluency located at a clause boundary. The researcher also counted the total number of silent pauses, filled pauses, prolongations, and self-monitorings occurring in each speech sample.

3. Results

3.1 Global score and local dysfluency number: highly correlated

An independent-samples *t*-test was conducted to compare global scores of the NS group and the NNS group. The NS global fluency score ($M = 7.75$, $SD = 0.84$) was significantly higher than that of the NNS ($M = 4.49$, $SD = 1.04$); $t(18) = 7.70$, $p = 0.000$. The NS local dysfluency number ($M = 5.1$, $SD = 2.47$) was significantly lower than that of the NNS ($M = 10.4$, $SD = 4.09$); $t(18) = -3.51$, $p = 0.003$.

A Pearson correlation coefficient was computed to assess the correlation between the global score and the number of local dysfluencies. A scatterplot summarizes the results (Figure 1). Overall, there was a strong negative correlation between the global score and the local number for both NS ($r = -0.755$, $n = 10$, $p = 0.012$) and NNS ($r = -0.834$, $n = 10$, $p = 0.003$). If the global score was higher, the number of local dysfluencies was smaller. This shows that the global perception and local perception of fluency were consistent.

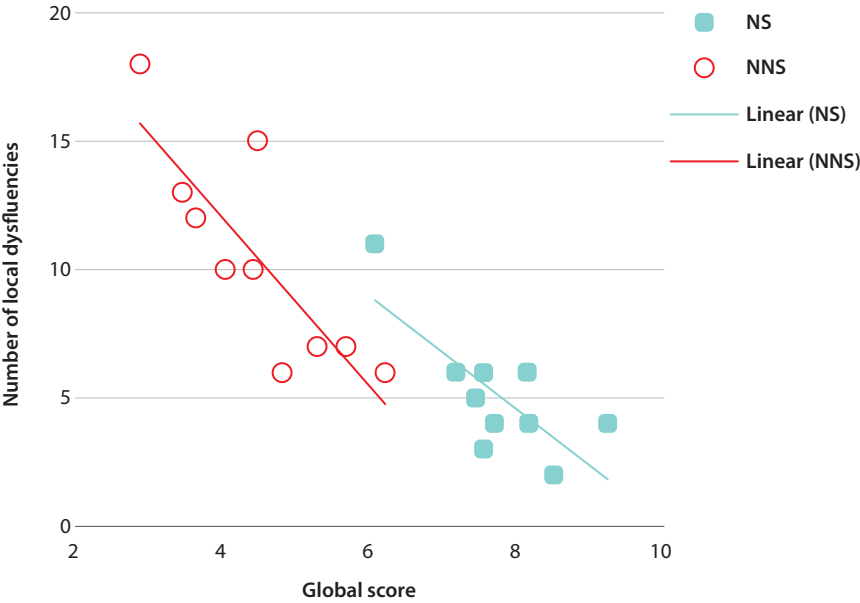


Figure 1. Correlation between global score and local dysfluency number

Figure 1 indicates that NS global scores were higher than those for NNS, and NS number of local dysfluencies was lower than that for NNS. It also reveals that the global scores for NS were higher than those for NNS when both produced a similar number of local dysfluencies. For example, when the number of local dysfluencies was about 6, NS global scores were around 8 while NNS were around 6.

3.2 The relationship between prosodic features/self-monitoring and perceived dysfluency

There were 51 local dysfluencies in NS speech and 104 in NNS speech according to the perception assessment results. A dysfluency could be involved in one or more prosodic features and/or self-monitoring. In this section, the relationship between prosodic features/self-monitoring and perceived dysfluency is analyzed from four perspectives. 3.2.1 analyzes the numbers and percentages of prosodic features and self-monitoring produced in NS and NNS speech when dysfluencies were perceived. 3.2.2 analyzes the numbers and the percentages of perceived dysfluencies when prosodic features/self-monitoring were produced in NS and NNS speech. Then, in 3.2.3, combinations of prosodic features and self-monitoring are analyzed. 3.2.4 analyzes some examples to explain how prosodic features and self-monitoring in NS and NNS speech influenced perceived dysfluencies.

3.2.1 *How many prosodic features or self-monitorings are produced when dysfluencies are perceived?*

Table 1 provides the number of silent pauses, filled pauses, prolongations, and self-monitorings in NS and NNS speech, and their numbers and percentages produced when dysfluencies are perceived. The results indicate that over 50% of filled pauses and self-monitoring in NS speech and over 80% of filled pauses and self-monitoring in NNS speech occur when dysfluencies are perceived.

Independent-samples *t*-tests were conducted to compare the total numbers of silent pauses, filled pauses, prolongations, and self-monitorings in NS and NNS speech, respectively; and to compare their numbers in NS and NNS speech when dysfluencies are perceived.

There is a significant difference in the total number of filled pauses in NS speech ($M = 6.2$, $SD = 3.58$) and in NNS speech ($M = 12.2$, $SD = 7.53$): $t(18) = -2.28$, $p = 0.035$; and there is a significant difference between the number of filled pauses in NS speech co-occurring with perceived dysfluencies ($M = 3.2$, $SD = 2.04$) and that in NNS speech ($M = 10.1$, $SD = 6.23$): $t(18) = -3.38$, $p = 0.003$. This shows that non-native speakers use significantly more filled pauses in their speech production. However, there is no significant difference between NS and NNS speech with regard to the total number of self-monitorings or the amount of self-monitoring co-occurring with perceived dysfluencies. This could be because only a small amount of self-monitoring occurs in both NS and NNS speech.

The *t*-test results also indicate that native speakers use significantly more prolongations ($M = 18.7$, $SD = 8.78$) than non-native speakers ($M = 7.2$, $SD = 4.16$): $t(18) = 3.74$, $p = 0.001$. Yet there is no significant difference between NS and NNS with regard to the number of prolongations co-occurring with perceived

Table 1. Percentages of silent pause, filled pause, prolongation, and self-monitoring produced in NS and NNS speech when dysfluencies are perceived

Factors		NS	NNS
silent pause (SP)	Number of SP in dys.	44	104
	Total number	164	222
	% of SP in dys.	27%	47%
filled pause (FP)	Number of FP in dys.	32	100
	Total number	62	122
	% of FP in dys.	52%	82%
prolongation (PL)	Number of PL in dys.	27	31
	Total number	187	72
	% of PL in dys.	14%	43%
self-monitoring (SM)	Number of SM in dys.	17	27
	Total number	26	34
	% of SM in dys.	65%	80%

Note: More than one prosodic feature or self-monitoring could occur when a dysfluency is perceived.

dysfluencies, since only a small number of prolongations are involved in dysfluencies. The total numbers of silent pauses in NS and NNS speech are not significantly different. However, significantly fewer silent pauses in NS speech co-occur with perceived dysfluencies ($M = 4.4$, $SD = 3.24$) than in NNS speech ($M = 10.5$, $SD = 5.4$): $t(18) = -3.06$, $p = 0.007$.

3.2.2 *How many dysfluencies are perceived when a specific feature is produced?*

Table 2 lists the numbers and the percentages of the dysfluencies that are perceived when prosodic features or self-monitorings are produced in NS and NNS speech. It reveals several important points.

In NS speech, 71% of dysfluencies are perceived where silent pauses are produced; about 50% of dysfluencies are perceived where filled pauses or prolongations are produced; and only 31% of dysfluencies are perceived where self-monitorings are produced.

In NNS speech, over 72% of dysfluencies are perceived where silent pauses or filled pauses are produced. However, only 26%–30% of dysfluencies are perceived where self-monitorings or prolongations are produced.

Table 2. Percentages of perceived dysfluencies co-occurring with produced features

Percentage of dysfluencies		NS	NNS
co-occurring with silent pause	No. of dysfluencies	36	75
	% of dysfluencies	71%	72%
co-occurring with filled pause	No. of dysfluencies	29	86
	% of dysfluencies	57%	83%
co-occurring with prolongation	No. of dysfluencies	29	31
	% of dysfluencies	57%	30%
co-occurring with self-monitoring	No. of dysfluencies	17	27
	% of dysfluencies	31%	26%
Local dysfluency number		51	104

3.2.3 *Feature combinations and perceived dysfluencies*

The researcher coded all the prosodic features and self-monitoring for each local dysfluency perceived. Figure 2 shows the feature combinations for those local dysfluencies, and reveals that most features produced that co-occur with perceived dysfluency are multiple-feature combinations in both NS and NNS speech. In NNS speech, a 2-feature combination is salient.

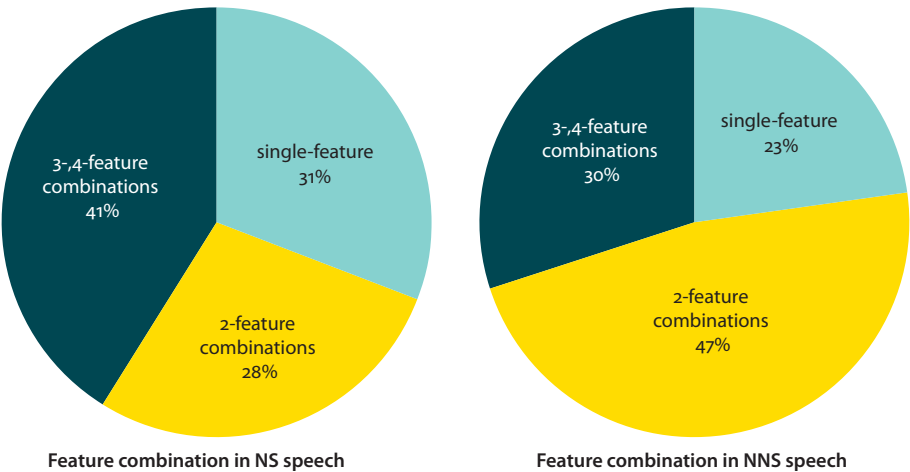


Figure 2. Feature combinations and perceived dysfluencies

A combination could contain more than one prosodic feature or instance of self-monitoring. In (1), the dysfluent chunk contains two prolongations, two silent pauses, and two filled pauses.

(1) NNS

不要: (0.68) 呃 (0.27) 嗯 (0.62) 打扫: (.)
bú yào: (0.68) *uh* (0.27) *um* (0.62) *dǎsǎo:* (.)
 not need uh um clean
 'do not need to clean'

In Table 3, only 4 categories comprise over 10% of the total numbers of perceived dysfluencies in NS speech: single silent pause, single self-monitoring, filled pause + prolongation, and silent pause + filled pause + prolongation; and only 3 in NNS speech: single filled pause, silent pause + filled pause, and silent pause + filled pause + prolongation.

Table 3. Feature combinations and perceived dysfluencies

Combination category		NS	NNS
Single feature	silent pause	9	6
	filled pause	0	13
	prolongation	1	1
	self-monitoring	6	4
2-feature combination	silent pause + filled pause	4	36
	silent pause + prolongation	2	1
	silent pause + self-monitoring	1	3
	filled pause + prolongation	6	4
	filled pause + self-monitoring	1	4
	prolongation + self-monitoring	0	1
3,4-feature combination	silent pause + filled pause + prolongation	12	16
	silent pause + filled pause + self-monitoring	1	7
	silent pause + prolongation + self-monitoring	3	2
	filled pause + prolongation + self-monitoring	1	2
	silent pause + filled pause + prolongation + self-monitoring	4	4
Total		51	104

Note: a combination could contain more than one filled pause, silent pause, prolongation, or instance of self-monitoring.

The data indicate that the numbers of the combination of silent pause + filled pause + prolongation in both NS and NNS speech are more than those of other 3- or 4-feature combinations. Much of the data supports this, such as the silent pause, filled pause, and prolongation in (1). The reason is that other combinations containing three or four features must contain self-monitoring, while the total

number of self-monitorings is much smaller than that of prosodic features. This means that the number of other combinations cannot easily exceed 10%.

Meanwhile, the filled pause is a salient feature for NNS dysfluency. It not only occurs independently, but also co-occurs with silent pauses in NNS speech. The number of silent pause + filled pause in NNS speech is much greater than that of any of the other two-feature combinations. However, the filled pause is not salient for NS dysfluency.

3.2.4 *In which cases do prosodic features and self-monitoring influence fluency?*

In this section, some examples are analyzed to demonstrate how major prosodic features and self-monitoring influence NS and NNS fluency. Prolongations will not be analyzed, since few prolongations occur in NNS speech or co-occur with perceived dysfluency.

3.2.4.1 Analysis of silent pause. The current study finds that speech rate is moderately correlated to global score: 0.72 for NS and 0.66 for NNS. Meanwhile, dysfluency is not perceived when some single individual silent pauses occur. In NS speech (2) and NNS speech (3), (1.05) and (1.39) are located at the clause boundaries when speakers plan the following clauses. However, no dysfluency is perceived in either case.

(2) NS

... 就 没有 餐厅 的 (.) 那么 好 吃 了 (1.05) 所以 ...
 ... *jiù méiyǒu cāntīng de(.) nàme hǎo chī le* (1.05) *suǒyǐ...*
 ... then no restaurant REL such good eat CRS so
 ‘... then, (it is) not so delicious as what is in a restaurant, so ...’

(3) NNS

不如: (1.60) 呃 (0.48) 餐厅 的 饭 (0.27) 好 吃 (1.39) 那 ...
bùrú: (1.60) eh (0.48) cāntīng de fàn (0.27) hǎo chī (1.39) nà ...
 not as uh restaurant REL rice good eat that ...
 ‘not so delicious as restaurant foods, that ...’

Some single individual silent pauses occurring within a clause in NNS speech are not perceived as dysfluency, even if they are long silent pauses. For example, in (4), few native speakers perceived (0.52), (1.38) or (0.82) as a dysfluency. All these silent pauses are located within a clause. The speaker was re-planning the chunk *shì hěn hǎochī yīdiǎnr guì* (*shì* was used to repair *de* instantly) when he produced the silent pause (0.52); from the listener’s perspective, he was decoding *fàn hǎochī... de* before (0.52), so no dysfluency was perceived near this silent pause. When the

speaker was planning *hěn yǒuyìsi* during the silent pause (1.38), the listener was decoding *hěn hǎochī yīdiǎnr guì* which was not native-like, so it took some time for the listener to decode it. Therefore, no dysfluency was perceived. When the speaker was planning *yě shì hěn fāngbiàn* during the silent pause (0.82), the listener was decoding *hěn yǒuyìsi* which did not make sense in context, so it took some time for the listener to comprehend it, although the listener failed in the end. Therefore, no dysfluency was perceived near (0.82). All of these examples indicate that non-native speakers need more time to plan speech; however, when native speakers perceive NNS speech in which meaning or grammar is not so transparent or native-like, they also need time to comprehend it, and therefore may not perceive a pause as a dysfluency.

(4) NNS

饭 好 吃 (0.42) 呃 的 [instant repair] (0.52) 是 (0.38) 很 好 吃
fàn hǎo chī (0.42) *eh de* [instant repair] (0.52) *shì* (0.38) *hěn hǎo chī*
 rice good eat uh REL is very good eat
 一点儿 贵
yīdiǎnr guì
 a little expensive
 ‘foods are delicious is indeed very delicious, a little expensive,’
 可是 (1.38) 很 有意思 (0.82) 也 是 很 方便
kěshì (1.38) *hěn yǒuyìsi* (0.82) *yě shì hěn fāngbiàn*
 but very interesting also is very convenient
 ‘but it is very interesting and also very convenient.’

Some single individual silent pauses occur within a clause in NNS speech when a dysfluency is perceived. For example, in (5), there was a silent pause (0.74) at the word boundary when a dysfluency was perceived. From the speaker’s perspective, he was planning the chunk *měiwèi yīdiǎn* when the silent pause (0.74) was produced. From the listener’s perspective, he was decoding *yībān dōu bǐjiào* which had a clear meaning and structure, so the listener did not need much time to decode it. Therefore, a dysfluency was perceived at this point.

(5) NNS

一般 都 (.) 比较 (0.74) 美味 一点
yībān dōu (.) *bǐjiào* (0.74) *měiwèi yīdiǎn*
 generally all relatively delicious a bit
 ‘Generally, they are a little more delicious.’

3.2.4.2 Analysis of filled pause. Table 4 shows the numbers of filled pauses produced in NS and NNS speech and the numbers of filled pauses involving dysfluency.

In both NS and NNS speech, “uh/um” occurs most, and most of the occurrences are involved in dysfluency. Meanwhile, “uh/um” in NNS speech occurs much more frequently than that in NS speech, and is much more highly correlated with dysfluency.

Table 4. Occurrences of filled pauses and numbers of filled pauses involved in dysfluency

	NS		NNS	
	occurrence	dysfluency	occurrence	dysfluency
uh/um	40	27	121	99
ah	7	4	1	1
ránhòu	15	1	0	0
Total	62	32	122	100

The filled pause *ah* is used in NS speech infrequently, but half of the instances are involved in dysfluency. Only one *ah* occurs in NNS speech.

The content word *ránhòu* (‘after that’) is a filled pause unique to Chinese, so it only occurs in NS speech. In Table 4, only one *ránhòu* is involved in dysfluency. This indicates that *ránhòu* has functions besides speech planning, such as discourse functions as in (6). It is located at the clause boundary, and no dysfluency is perceived.

(6) NS

..... 你 就 要 必须做 非常 (.) 多 的 饭 菜 (0.46)

..... nǐ jiù yào bìxū zuò fēicháng (.) duō de fàn cài (0.46)

..... you then need must make very more REL rice dish

然后: (.) 这 (0.11)

ránhòu: (.) zhè (0.11)

after that this

‘(If friends come to your home,) you have to cook many dishes, (so)’

工作量 就 非常 (.) 大 (0.55)

gōngzuòliàng jiù fēicháng (.) dà (0.55)

work load then very large

‘the work load will be very large.’

3.2.4.3 Analysis of self-monitoring. In 3.2.1, Table 1 shows that 65–80% of self-monitorings occurred when dysfluencies are perceived. Self-monitoring is used for time delay (repetition) or to correct or re-plan speech (repair, anticipatory retracting, and restart), causing a change in meaning. When self-monitoring occurs in speech, a listener generally can smoothly comprehend the just-produced chunk, so they do not need to spend more time on decoding, but the meaning change may

cause incoherence. Therefore, dysfluency is easily perceivable. Table 5 shows the total numbers of self-monitorings in NS and NNS speech and the numbers of self-monitorings perceived as dysfluency.

Table 5. Occurrence of self-monitoring and numbers of self-monitorings involved in dysfluency

	NS		NNS	
	occurrence	dysfluency	occurrence	dysfluency
repetition	13	8	18	16
instant repair	3	1	8	6
anticipatory retracting	4	2	5	3
restart	6	6	3	2
Total	26	17	34	27

In Table 5, restart occurs three times more in NS dysfluency than in NNS, while NNS use more “repetition” and “instant repair” than NS. Because of the low cognitive load, the chunks that non-native speakers plan are shorter than those that native speakers plan. Repetition and instant repair occur in short chunks, such as a word, while restarts only occur in a longer unit. Therefore, there are more restarts in NS speech since it requires more cognitive load. Native speakers and non-native speakers use similar numbers of anticipatory retracing when dysfluencies are perceived.

Table 6. Feature combinations involving self-monitoring in perceived dysfluency

	NS	NNS
single self-monitoring	6	4
multi-feature combination including self-monitoring	11	23

According to Table 6, there were more single self-monitorings in NS dysfluency than in NNS dysfluency. In (7), there is no prosodic feature when restart occurs. The restart begins at *hěnduō* (‘many’). NS perceived incoherence in the meaning when *hěnduō* occurred, since the meaning cannot be connected to the previous structure smoothly. However, listeners could hear coherent structures of chunks both before and after *hěnduō*. Therefore, at the restart point, a dysfluency was perceived.

(7) NS

可以 吃 到 我 非常 想 [restart] 很多 我 想 吃 到 的
kěyǐ chī dào wǒ fēicháng xiǎng [restart] hěnduō wǒ xiǎng chīdào de
 can eat to I unusually think many I think eat arrive REL
 东西
dōngxi
 stuff
 'can eat I want very much [restart] many foods that I want to eat'

According to Table 6, when self-monitoring is perceived as dysfluency in NNS speech, most of it is accompanied by one or more prosodic features. The phrase in (8) was produced by a non-native speaker, and perceived as a dysfluency by native speakers. This is a combination of silent pause + filled pause + prolongation + self-monitoring. In other words, there is a feature combination including a silent pause, a filled pause, and a prolongation between the original word and the repeated word. Obviously, the NNS needed more time for planning. However, NS did not need more time to comprehend the word *yīnwèi* whose meaning was clear, and there was no meaning change here. Therefore, the time delay represented by the combination of silent pause, filled pause, and prolongation was obvious, and a dysfluency was perceived.

(8) NNS

因为: (.) 呃: [repeat] (2.26) 因为
yīnwèi: (.) eh: [repeat] (2.26) yīnwèi
 because eh because
 'because because'

4. Discussion

4.1 NS perceptual bias

The results show that the global perception score is strongly correlated to the number of local perceived dysfluencies. Native speakers have significantly higher global scores than non-native speakers do. Native speakers rate NS speech samples as more fluent at the global level than NNS speech samples even if both samples have the same number of local dysfluencies. This might be caused by the non-native accent in articulating speech segments and/or tones. These results reveal that NS perceptual bias exists, and this bias is worth further study.

4.2 How do prosodic features and self-monitoring influence perceived fluency?

The current study finds that most silent pauses do not influence perceived fluency; this finding is similar to that of Kormos & Denes (2004), especially when silent pauses are located at a clause boundary. Only small numbers (27% for NS and 47% for NNS) of silent pauses influenced perceived dysfluencies. However, the results indicate that silent pauses occur in both NS and NNS speech in 71–72% of perceived dysfluencies (Table 2). This explains why the speech rate is moderately correlated to the global score: 0.72 for NS and 0.66 for NNS, which confirms previous results (e.g. Kormos & Denes, 2004; De Jong & Perfetti, 2011; Derwing et al., 2006). As some studies (Riggenbach, 1991; Yang, 2012) have indicated, silent pauses occur more frequently at the clause boundary. The reason is that speakers produce a portion, pause to plan the next portion, then produce again, and pause again, often repeatedly (Beatti, 1983). In other words, speakers plan and produce speech in cycles (Henderson, et al., 1966). At the same time, the silent pause causes time delay so it is sometimes perceived as dysfluency when it occurs in speech production.

The examples in 3.2.4.1 show the contexts in which silent pauses do not influence the perception of fluency and those in which they do. In NNS speech, when a single silent pause occurs within a clause and the meaning or the linguistic structure before the silent pause is not transparent or native-like, a dysfluency generally will not be perceived. In this case, the speaker uses the time of the silent pause to plan speech; yet the listener also needs that time to decode the chunk before the silent pause. However, if the meaning or linguistic structure is transparent and easy to comprehend, a dysfluency will be perceived. Moreover, when a dysfluency is perceived, in most cases, silent pauses co-occur with other prosodic features in speech.

As for filled pauses, the current study finds that in most cases, “uh/um” occurs when a dysfluency is perceived. Therefore, filled pauses influence perceived fluency. This finding is different from that of Kormos & Denes (2004). In NS speech, 15 *ránhòu* (‘after that’) were produced and almost no *ránhòu* were perceived as dysfluency. However, NNS produced no *ránhòu*. The explanation for this phenomenon is that “um/uh” and *ah* are totally different from meaningful words in Chinese, while *ránhòu*, which only occurs in NS speech, is a content word, i.e., a meaningful disyllabic word in Mandarin Chinese, and it does not cause incoherence in speech meaning. Although in some cases *ránhòu* does not convey the meaning of “then”, or “after that”, native speakers still accept the word *ránhòu* as natural, indicating it is perceived as a consequence of what preceded it in the speech, or the development of the speech gradually, as in the instance of *ránhòu* in

(6). Therefore, the filled-pause *ránhòu* does not cause any time delay in the speech, and a dysfluency is not perceived.

In most cases, self-monitoring in both NS and NNS speech influences perceived fluency. Self-monitoring generally causes time delay or incoherence because the speech is replanned. However, the number of self-monitorings is much smaller than those of silent pauses and filled pauses, so there are not many perceived dysfluencies involving self-monitoring.

Based on the results and discussion above, “um/uh” and self-monitoring are proposed as heterogeneous elements when a speaker plans or replans speech, because the sounds of “um” and “uh” are different from normal speech sounds, and the meaning of the chunks that cause self-monitoring is different from that of normal coherent speech. Therefore, when they occur in speech, dysfluency tends to be perceived. However, silent pauses and prolongation do not create unusual sound production. In other words, a silent pause is silent, and prolongation just follows the previous pronunciation; i.e., both of them are homogeneous elements. Therefore, when silent pauses and prolongations occur in speech, in most cases, dysfluency will not be perceived. This could also explain why feature combinations occur when most dysfluencies are perceived. Table 3 shows that almost all feature combinations contain filled pauses and/or self-monitoring, except 2 combinations in NS speech and 1 in NNS speech. Since filled pauses and self-monitoring are heterogeneous elements, when they are combined with other homogeneous elements, such as silent pauses or prolongation, the nature of the combinations is heterogeneous and dysfluencies are perceived.

4.3 Comparison between NS and NNS

With regard to prosodic features and self-monitoring in perceived dysfluency, the current study indicates some similarities between NS and NNS speech. First, the results indicate that NS generally perceived a local dysfluency when a filled pause or self-monitoring was produced in both NS and NNS speech. Second, when NS perceived a local dysfluency, in 71–72% cases, there was a silent pause produced in NS and NNS speech. However, a much smaller number of silent pauses (fewer than 47%) were produced in perceived dysfluencies. Third, when a dysfluency was perceived, in over 69% cases, multi-feature combinations were produced in NS and NNS speech.

The results also reveal some differences. First, the filled pause in NNS speech production is a salient feature for perceived dysfluency. The results show that the total number of filled pauses in NNS speech was significantly more than that in NS. This could be caused by NNS lack of sense of the language or language proficiency. Meanwhile, the number and the percentage of filled pauses for NNS dysfluency

were also significantly more than those for NS; many more dysfluencies were perceived when there were filled pauses produced in NNS speech, and the number of single-feature and multi-feature combinations containing filled pauses in NNS speech with perceived dysfluency is obviously more than those in NS speech. One reason for this is that the native speakers produced 15 *ránhòu* while the non-native speakers produced none, and very few *ránhòu* were involved in dysfluency. A second reason is that the total number of filled pauses in NS speech was much lower than that in NNS speech, and “um”/“uh” was almost the only type of filled pause in NNS speech, except for one *ah*. Another reason could be the heterogeneous nature of the filled pause.

The second difference between NS and NNS is self-monitoring. Self-monitoring in NS speech is used more to replan long chunks, while that in NNS speech is used for short chunks, such as words or syllables. This is determined by NS and NNS cognitive loads which influence NS and NNS speech planning; i.e., NS plan and self-monitor longer chunks while NNS (re-)plan short chunks. Third, the data reveals that fewer silent pauses occur in NS speech (27%) than in NNS speech (47%) when dysfluencies are perceived. This might cause native listeners' bias.

5. Conclusion

The current study indicates that the global perception score is strongly correlated to the number of local perceived dysfluencies. However, NS have a perceptual bias on NNS dysfluency.

The results indicate that silent pauses, prolongations, and *ránhòu* do not influence perceived fluency in most cases. In NNS speech, if the meaning or the linguistic structure is transparent and easily comprehended, a dysfluency is perceived when a silent pause is produced. However, in most cases, the filled pauses “uh”/“um” and self-monitoring influence perceived fluency. To explain the results, the current study proposes that some filled pauses (“uh”/“um”) and self-monitoring are heterogeneous elements while silent pauses, prolongations, and other filled pauses (content words) are homogeneous elements.

With regard to the differences between NS and NNS, the filled pause is a salient feature for NNS dysfluency; self-monitoring in NS speech is used more to replan long chunks while that in NNS is used for short chunks because of the cognitive load; and many more silent pauses occur in NNS speech when dysfluencies are perceived.

6. Pedagogical implications

In the classroom, instructors need to be attuned to the various aspects of fluency development that their students demonstrate, and that mark their development along the way. For example, this study's results show that filled pauses are salient in NNS speech, since NNS do not produce word-like filled pauses while NS do. Instructors of Chinese could be mindful of students' filled pauses when their students' proficiency reaches advanced level, and guide them to avoid filled pauses, such as "um"/"uh", which are easily perceived as dysfluency. They could teach students to use some native Chinese content words instead, such as *ránhòu*, which is close to NS usage and not generally perceived as dysfluency by native speakers. In this way, students could have more time to plan their speech, while listeners would not perceive the time delay; at the same time, students' speech would be closer to native-like speech. Textbooks could also provide the kinds of words that can be used as filled pauses.

Moreover, teachers could encourage students not to be afraid to self-monitor their speech; that is, to practice self-monitoring to replan their speech and in doing so, improve accuracy; on the other hand, as the researcher concluded above, native speakers generally need more time to understand non-native speakers' speech when non-native speakers replan their speech, as in the instant repair in (4). The inherent chunk caused by self-monitoring is not perceptually salient to native listeners if not many prosodic features involve time delay; therefore dysfluency based on self-monitoring is not easily perceived.

With regard to the silent pause, Chinese language instructors need to know that it has multiple effects, such as the perception of dysfluency and the marking of a clause boundary. Educators need to focus on functions like the clause boundary marker, and guide students to segment phrases or clauses appropriately in connected speech.

Acknowledgements

This research was funded by the project PRJ61TU of Wisconsin Alumni Research Foundation. The author is very grateful to Dr. Murray J. Munro for his valuable comments on the early version of this paper. Thanks are also due to the former editor Dr. Zheng-sheng Zhang and the anonymous reviewers for their detailed comments and suggestions.

References

- Albrechtsen, D., Henriksen, B., & Faerch, C. (1980). Native speaker reactions to learners' spoken interlanguage. *Language Learning*, 30(2), 365–396. doi: 10.1111/j.1467-1770.1980.tb00324.x
- Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new disfluency and reference resolution. *Psychological Science*, 15(9), 578–582. doi: 10.1111/j.0956-7976.2004.00723.x
- Beatti, G. (1983). *Talk: An analysis of speech and non-verbal behaviour in conversation*. Milton Keynes, England: Open University Press.
- Blau, E. K. (1991). More on comprehensible input: the effect of pauses and hesitation markers on listening comprehension. Paper presented at the Annual Meeting of the Puerto Rico Teachers of English to Speakers of Other Languages: from ERIC database.
- Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, 44(2), 274–296. doi: 10.1006/jmla.2000.2753
- Cenoz, J. (2000). Pauses and hesitation phenomena in second language production. *ITL. Instituut voor Toegepaste Linguïstiek*, 127–28, 53–69.
- Deese, J. (1984). *Thought into speech: the psychology of a language*. Englewood Cliffs, NJ: Prentice-Hall.
- De Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 61(2), 533–568. doi: 10.1111/j.1467-9922.2010.00620.x
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). L2 fluency: Judgments on different tasks. *Language Learning*, 54, 655–679. doi: 10.1111/j.1467-9922.2004.00282.x
- Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System*, 34, 183–193. doi: 10.1016/j.system.2006.01.005
- Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31(04), 533–557. doi: 10.1017/S0272263109990015
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29, 359–380. doi: 10.1093/applin/amm041
- Fillmore, C. J. (1979). On fluency. In C. J. Fillmore, D. Kempler & W. S-Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85–101). New York: Academic Press.
- Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language*, 40, 280–295. doi: 10.1006/jmla.1998.2613
- Griffiths, R. (1991). Pausological research in an L2 context: A rationale, and review of selected studies. *Applied Linguistics*, 12(4), 345–364. doi: 10.1093/applin/12.4.345
- Guillot, M. -N. (1999). *Fluency and its teaching*. Clevedon: Multilingual Matters.
- Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1966). Sequential temporal patterns in spontaneous speech. *Language and Speech*, 9, 207–216.
- Isaacs, T., & Trofimovich, P. (2011). Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of second language speech. *Applied Psycholinguistics*, 32(1), 113–140. doi: 10.1017/S0142716410000317

- Ke, C. (2005). Research on the relationship between the OPI and the SOPI and their pedagogical implications. *Journal of Chinese Language Studies*, 1, 156–168.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145–164. doi: 10.1016/j.system.2004.01.001
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387–417. doi: 10.1111/j.1467-1770.1990.tb00669.x.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning*, 64(4), 715–748. doi: 10.1111/lang.12082
- Raupach, M. (1980). Temporal variables in first and second language speech production. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 263–270). Berlin: Mouton de Gruyter. doi: 10.1515/9783110816570.263
- Rehbein, J. (1987). On fluency in second language speech. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 97–105). Praeger Pub Text.
- Reich, S. S. (1980). Significance of pauses for speech perception. *Journal of Psycholinguistic Research*, 9(4), 379–389. doi: 10.1007/BF01067450
- Riazantseva, A. (2001). Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition*, 23(04), 497–526. doi: 10.1017/S027226310100403X
- Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse processes*, 14(4), 423–441. doi: 10.1080/01638539109544795
- Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in second language acquisition*, 14(04), 357–385. doi: 10.1017/S027226310001189
- Tavakoli, P. (2009). Assessing L2 task performance: Understanding effects of task design. *System*, 37(3), 482–495. doi: 10.1016/j.system.2009.02.013
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17(1), 84–119. doi: 10.1093/applin/17.1.84
- Trenchs-Parera, M. (2009). Effects of formal instruction and a stay abroad on the acquisition of native-like oral fluency. *Canadian Modern Language Journal*, 65(3): 365–393. doi: 10.3138/cmlr.65.3.365
- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in second language acquisition*, 28(01), 1–30. doi: 10.1017/S0272263106060013
- Vanderplank, R. (1993). 'Pacing' and 'spacing' as predictors of difficulty in speaking and understanding English. *ELT Journal*, 47(2), 117–125. doi: 10.1093/elt/47.2.117
- Voss, B. (1979). Hesitation phenomena as sources of perceptual errors for non-native speakers. *Language and Speech*, 22(2), 129–144.
- Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication*, 50(2), 81–94. doi: 10.1016/j.specom.2007.06.002
- Yang, B. (2012). How do pauses reveal linguistic plans by L2 learners of Chinese? In Q. Ma, H. Ding, & D. Hirst (Eds.), *Proceedings of the 6th International Conference on Speech Prosody (SP2012)*. International Speech Communication Association online archive.

摘要

本文从言语表达和感知的角度研究造成延时的韵律特点、自我监控和不流利之间的关系。十位汉语母语者和十位汉语学习者参加了本研究项目中的口语考试。口试的内容被录音，并被转写成汉字，韵律特点用特殊的符号标明，而四种自我监控形式也被标注。同时，另有二十八位汉语母语者对这些口语语料进行评估。研究者将评估的结果与转写中的韵律标示和自我监控标示进行对比，分析他们的关系。结果表明在大多数情况下，口语表达中有声停顿呃/嗯和自我监控影响不流利的感知，而无声停顿、延长和实词作为有声停顿则不影响不流利感知；汉语学习者口语表达中的有声停顿是感知不流利的非常鲜明的特点。结果显示了在何种情况下，汉语学习者口语表达中的无声停顿会被感知为不流利。本研究还发现了母语者在感知汉语学习者不流利时的偏见。

关键词: 流利度, 自我监控, 停顿, 韵律特点, 口语表达与感知

Author's address

Bei Yang
University of Wisconsin – Madison
1266 Van Hise Hall, 1220 Linden Dr.
Madison, WI 53706
byang25@wisc.edu