Research Note: The processing and evaluation of fluency in native and non-native speech

Hans Rutger Bosker
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
HansRutger.Bosker@mpi.nl

June 2014

1. Introduction

Disfluency is a pervasive feature of spontaneously produced spoken language, be it native (L1) or non-native (L2) speech. The present research note reports on empirical studies that investigated the consequences of disfluency in native and non-native speech for the listener. It is commonly assumed that speech comprehension is hindered by the disfluent nature of spontaneous speech. This assumption is also found in the language learning community. Many language learners strive hard to speak a language fluently thus hoping to improve their comprehensibility. Indeed, empirical research provides evidence that disfluent non-native speech negatively affects the impression that listeners have of the non-native speaker. Studies investigating the effect of disfluency on perceived fluency judgments (e.g., subjective ratings) will be referred to as the *evaluative* approach to the study of fluency perception (e.g., Cucchiarini, Strik, & Boves, 2002; Derwing, Rossiter, Munro, & Thomson, 2004; Kormos & Dénes, 2004; Rossiter, 2009). Studies adopting this approach investigate fluency as a global property of the spoken discourse as a whole and have primarily focused on non-native speech.

However, in contrast to the negative effects of disfluency on listeners' impressions of the speaker's fluency, there are also indications in the psycholinguistic literature that disfluencies may in fact help, rather than hinder, the listener in speech comprehension. For instance, disfluencies may help the listener avoid erroneous syntactic parsing (Brennan & Schober, 2001; Fox Tree, 2001), attenuate context-driven expectations about upcoming words (Corley, MacGregor, & Donaldson, 2007; MacGregor, Corley, & Donaldson, 2010), improve recognition memory (Collard, Corley, MacGregor, & Donaldson, 2008; Corley, MacGregor, & Donaldson, 2007) or predict particular referents (e.g., Arnold, Hudson Kam, & Tanenhaus, 2007; Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Barr & Seyfeddinipur, 2010). These studies will be referred to as the *cognitive* approach to fluency. Scholars adopting this approach study disfluency as a local property of a particular utterance and have primarily focused on native speech.

Review of the evaluative and cognitive approach reveals an apparent contradiction between the *negative* effects of non-native disfluencies on subjective fluency ratings, on the one hand, and *positive* effects of native disfluencies on speech perception, on the other. This research note summarizes some empirical work that aimed to combine the two approaches by studying the processing and evaluation of fluency in native and non-native speech. Furthermore, it proposes an integrative account of how native and non-native fluency characteristics affect both (i) the impression that listeners have of the speaker's fluency level and (ii) the processes involved in speech comprehension, such as prediction, memory, and attention. Finally, potential steps for future research are outlined and theoretical and practical implications are introduced.

2. Empirical studies

2.1 Results of the evaluative approach to fluency

Study 1 (published as Bosker, Pinget, Quené, Sanders, & De Jong, 2013) and Study 2 (published as Bosker, Quené, Sanders, & De Jong, in press) adopted the evaluative approach to fluency. Both studies investigated the effect of native and non-native fluency characteristics on perceived fluency assessment. Study 1 asked the question what it is that makes L2 speech sound fluent. Study 2 compared the way listeners assess native and non-native fluency levels.

2.1.1. What makes speech sound fluent? Study 1: Bosker, Pinget, Quené, Sanders, and De Jong (2013)

Study 1 investigated how raters assess the fluency levels of non-native speakers. For Experiment 1, subjective fluency judgments were collected from naïve listeners who were asked to base their judgments on the use of silent and filled pauses, the speed of delivery of the speech and the use of hesitations and/or corrections (Cronbach's α = 0.97). These subjective ratings were related to objective acoustic measurements of the non-native speech materials. Acoustic measurements were categorized into three fluency dimensions (Skehan, 2003, 2009; Tavakoli & Skehan, 2005): *breakdown fluency* (number of filled pauses, number of silent pauses, and silent pause durations), *speed fluency* (mean syllable duration), and *repair fluency* (number of corrections and number of repetitions). The acoustic measures were selected for their low intercollinearity: cross-correlations between the speech measures demonstrated that both within and across fluency aspects our speech measures were largely independent. This low intercollinearity aided the interpretation of the following analyses, in that the contribution of one fluency dimension (e.g., speed) could be separated from that of another dimension (e.g., pauses). In this fashion, we aimed to answer the first research question of Study 1:

RQ 1A: What are the independent contributions of three fluency dimensions (breakdown, speed, and repair fluency) to perceived fluency judgments?

Our results showed, first of all, that assessment of L2 fluency is largely dependent on the fluency characteristics of the speech signal: the six combined acoustic measures could account for 84% of the variance of the subjective fluency judgments. Secondly, breakdown fluency and speed fluency were found to contribute most to perceived fluency ratings (59% and 54% explained variance, respectively). In contrast, repair characteristics of the speech signal were observed to have only a weak relationship with fluency perception (16% explained variance).

The second aim of this study was to seek for possible cognitive factors that underlie fluency perception. We hypothesized that differences in sensitivity to the different fluency dimensions might account for differences in correlations between acoustic measures and fluency ratings. For instance, if listeners would be, in general, more sensitive to pause phenomena (than to speed or repair phenomena) then this could explain the large contribution of pause characteristics to the perception of fluency. A series of experiments was designed to establish the relative sensitivity of listeners to pause phenomena, to the speed of delivery, and to repair features in speech, thus hoping to formulate an answer to the following research question:

RQ 1B: How well can listeners evaluate the pause, speed, and repair characteristics of non-native speech?

Three new experiments made use of the same speech materials as used before. The participants in these three new experiments received instructions to either assess the speaker's pausing behavior, the speaker's speed of speaking, or the speaker's use of repair strategies. Subsequently, these subjective ratings (all Cronbach's α's > 0.94) were related to the objectively measured acoustic characteristics of the speech. The extent to which the objective measures accounted for the subjective judgments was taken as an indication of listeners' sensitivity to different speech characteristics. Our statistical models showed that the ratings of pausing behavior were best predicted by the objective acoustic measures (70% explained variance). Listeners' sensitivity to the speed and repair characteristics of speech was inferior to their sensitivity to pause phenomena (53% and 55% explained variance, respectively). This suggests that listeners are most sensitive to the pausing characteristics of speech.

Interestingly, listeners were approximately as sensitive to speed features as they were to repairs. Nevertheless, the fluency ratings had revealed that repair phenomena only contribute very little to fluency judgments. The combined results from all the experiments suggest that, despite listeners' sensitivity to repair phenomena, they do not base their fluency judgments on these repair phenomena. If the perceptual sensitivity of listeners were the only factor determining the relative contributions of fluency dimensions to fluency perception, then we would expect to have found a larger contribution of repair measures to the perception of fluency. Apparently, there is no direct link between listeners' perceptual sensitivity and listeners' fluency evaluation. This suggests that listeners first perceive the acoustic characteristics of a speaker's speech but then subsequently also weigh the importance of the perceived speech characteristics for fluency.

2.1.2. Native and non-native fluency perception
Study 2: Bosker, Quené, Sanders, and De Jong (in press)

Study 2 looked further into the weighing of acoustic fluency characteristics by comparing native and non-native fluency perception. Much of the literature on fluency assessment has focused on non-native speech (e.g., Cucchiarini, Strik, & Boves, 2002; Derwing, Rossiter, Munro, & Thomson, 2004; Kormos & Dénes, 2004; Rossiter, 2009); presumably, because native speech is supposedly perceived as fluent by default. However, the psycholinguistic literature indicates that there is considerable variation in the production of disfluencies by native speakers as well (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fox Tree, 1995). This raises the question:

RQ 2: Do listeners evaluate fluency characteristics in the same way in native and non-native speech?

Because native and non-native speech differ in a large range of linguistic aspects, correlational analyses are unsuitable for comparing the perception of L1 and L2 fluency. Therefore, we applied phonetic manipulations to speech from native and non-native speakers (each speaking about three different topics) that had been matched for one particular acoustic property. If different fluency ratings are given to two items differing in a single manipulated acoustic property, then this perceptual difference may be reliably attributed to this single manipulated acoustic property. And because the native and non-native speech has been matched, it is possible to compare the contribution of one acoustic factor across native and non-native fluency perception. Moreover, this experimental method has the additional advantage that the separate contributions of multiple acoustic factors can be investigated. Thus, it is possible to study the effect of one acoustic property on fluency judgments (e.g., the duration of pauses) whilst keeping all other possibly interacting factors (e.g., the number of pauses) constant.

Phonetic manipulations were first applied to the number and duration of silent pauses (Experiment 1), having matched the native and non-native speech materials for the number of silent pauses per 100 syllables. Three conditions were created: NoPauses - all original pauses of >250 ms had been removed; ShortPauses - all original pauses of >250 ms were altered to have a duration within the range of 250-500 ms; and LongPauses - all original pauses of >250 ms were altered to have a duration within the range of 750-1000 ms. Subjective ratings of these manipulated speech fragments from native and non-native speakers were collected in a rating experiment. Results showed that (1) native speakers were perceived to be more fluent than non-native speakers; (2) both an increase in the number of silent pauses and an increase in the duration of silent pauses negatively affected perceived fluency judgments; and (3) these manipulation effects were similar across native and non-native speech.

A similar approach was adopted in Experiment 2. Here, the speed of the speech was manipulated to compare the contribution of articulation rate and speech rate to perceived fluency. Non-native speech was increased in speed (both Articulation Rate Manipulations, ARM, and Speech Rate Manipulations, SRM) to match the mean speaking rate of the native speakers. Also, native speech was slowed down (both ARM and SRM) to match the mean speaking rate of the non-native speakers, thus making comparisons across native and non-native speakers possible. The results from Experiment 2 mirrored those from Experiment 1. Again, (1) native speech was perceived to be more fluent than non-native speech; (2) both manipulation conditions (ARM and SRM) contributed to perceived fluency judgments: slowed down native speech resulted in lower fluency judgments, and faster non-native speech resulted in higher fluency judgments; and (3) the increase in fluency ratings of the non-native speech, and the decrease in fluency ratings of native speech, were of a similar magnitude. Based on the findings from Experiment 1 and Experiment 2, we conclude that there is no difference in the way listeners weigh the fluency characteristics of native and non-native speech. When raters are instructed to evaluate fluency in the narrow sense, native and non-native fluency perception are comparable. Therefore, there is no reason to believe that listeners make a qualitative distinction between native and non-native speakers when evaluating fluency.

2.2 Results of the cognitive approach to fluency

Studies 1 and 2 focused on listeners' assessment of fluency. These studies demonstrated that (i) listeners weigh the perceived speech characteristics (breakdown, speed, and repair fluency) on their relevance for fluency perception, and (ii) that this weighing of acoustic factors is similar for native and non-native fluency assessment. These observations do not necessarily warrant the conclusion that native and non-native disfluencies are perceptually equivalent, because Study 1 and 2 have only investigated the effects of disfluencies on listeners' subjective impressions of the speaker. Therefore, Study 3 (published as Bosker, Quené, Sanders, & De Jong, 2014) and Study 4 adopted the cognitive approach to fluency to test whether native and non-native disfluencies have different effects on the cognitive processes involved in speech comprehension. Study 3 asked the question whether native and non-native *uhm*'s may guide prediction of low-frequency referents to the same extent. Study 4 compared the effects of native and non-native disfluencies on attention.

2.2.1. Disfluency and prediction

Study 3: Bosker, Quené, Sanders, and De Jong (2014)

The psycholinguistic literature on disfluencies in native speech seems to converge on the conclusion that native disfluencies may aid the listener in comprehension. Listeners use their experience with the regularities in the distribution of disfluencies to anticipate the linguistic content following a disfluency. The literature on disfluency production indicates that disfluencies tend to occur before open-class words (Maclay & Osgood, 1959), unpredictable lexical items (Beattie & Butterworth, 1979), low-frequency color names (Levelt, 1983), or names of low-codability images (Hartsuiker & Notebaert, 2010). Therefore, disfluencies cue dispreferred or more complex content.

These conclusions have been drawn based on studies of disfluent native speech. It is unknown how disfluencies in non-native speech may affect listeners' predictive strategies. Therefore, Study 3 compared the way native and non-native disfluencies affect listeners' predictive strategies. We hypothesized that, due to the fact that there are less regularities in the distribution of non-native disfluencies, non-native disfluencies would be worse predictors of the word to follow (as compared to native disfluencies).

Previous literature investigating disfluency effects on prediction have reported that listeners may interpret native disfluency as a symptom of speaker difficulty in conceptualization. For instance, listeners can attribute disfluency to trouble in recognizing unknown objects (e.g., Arnold et al., 2007) or to trouble with the object's discourse status (e.g., Barr & Seyfeddinipur, 2010). In order to compare native and non-native disfluency, we targeted listeners' attribution of disfluency to difficulty in formulation (i.e., trouble in lexical access, rather than in conceptualization). We argued that it is at this particular stage in speech planning that native and non-native speakers diverge.

Therefore, the first research question of Study 3 read:

RQ 3A: Do listeners anticipate low-frequency referents upon encountering a disfluency?

This question was addressed by means of eye-tracking experiments. Participants were presented with pictures of high-frequency (e.g., a hand) and low-frequency objects (e.g., a sewing machine). Simultaneously, fluent and disfluent spoken instructions were played (e.g., 'Click on the [target]' vs. 'Click on *uh* the [target]'). It was hypothesized that listeners might attribute the presence of the disfluency to speaker difficulty in formulating the label for the low-frequency object (rather than the high-frequency object). This would result in more looks to the low-frequency object, prior to target onset, when listeners heard an *uh*.

Results from Experiment 1 indeed indicated that listeners had a preference, prior to target onset, for looking towards the low-frequency object (e.g., sewing machine) as opposed to the high-frequency object (e.g., the hand) in the disfluent condition - not in the fluent condition. These results suggest that listeners attribute the presence of disfluency to speaker difficulty in formulation.

Experiment 2 was designed to allow for a comparison of the effects of native and non-native disfluencies on prediction:

RQ 3B: Do native and non-native disfluencies elicit anticipation of low-frequency referents to the same extent?

Experiment 2 was identical to Experiment 1, but the participants in Experiment 2 listened to non-native speech. The results from Experiment 1 and 2 were combined to test for an interaction between the type of speaker (native vs. non-native) and the presence of a disfluency bias for low-frequency referents. It was found that, where native disfluencies elicited anticipation of low-frequency referents, non-native disfluencies *did not*. We argue that listeners reduced their use of disfluencies for prediction when listening to an L2 speaker, because non-native disfluencies are worse predictors of the linguistic content to follow. These results suggest that knowledge of the non-native identity of a speaker, as evidenced by a foreign accent, influences the way listeners use performance aspects of the speech signal (i.e., disfluency) to guide prediction.

2.2.2. Disfluency and attention

Study 4

Where in Study 3 disfluency effects on prediction were targeted, Study 4 studied how native and non-native disfluencies affect listeners' attention. Earlier work on perception effects of disfluencies showed that native disfluencies may trigger listeners' attention. This raises the question whether there are differential effects of native and non-native disfluencies on attention, as addressed by the research question of Study 4:

RQ 4: Do native and non-native disfluencies trigger heightened attention to the same extent?

Disfluency effects on listeners' attention could be the result of the non-arbitrary distribution of native disfluencies: disfluencies cue relatively more complex information and, therefore, listeners may benefit from raised attention levels in order to ensure timely comprehension of the complex content. As explained above, the distribution of disfluencies in non-native speech is, from the native listener's point of view, more irregular than the disfluency distribution in native speech. Therefore, non-native disfluencies are worse cues of upcoming, relatively more complex information. We hypothesized that the effect of non-native disfluencies on listeners' attention might therefore be attenuated (relative to that of native disfluencies; cf. Study 3).

Alternatively, disfluency effects on listeners' attention could also be the result of more automatic cognitive processes in response to delay. The Temporal Delay Hypothesis (Corley & Hartsuiker, 2011) argues that the temporal delay that is inherent to disfluency facilitates listeners' recognition and listeners' retention of words. Thus, both native and non-native disfluencies would trigger heightened attention levels.

We investigated attentional effects of native and non-native disfluencies by means of the Change Detection Paradigm (CDP; see Figure 1). Participants were instructed to remember a spoken passage of three sentences. One of the words (the 'target') in the passage was either presented in a fluent or disfluent context ("... that the patient with the [*uh*] *wound* ..."). After listening to the spoken passage, participants were presented with a textual representation of the memorized spoken passage. This text sometimes contained a substitution of the target word. Participants were asked to indicate whether the text was a correct representation of the spoken passage or whether the text contained a substitution. We hypothesized that, if disfluencies trigger listeners' attention, then participants should be better in detecting a change to a target word that had been presented in a disfluent context (e.g., the *uh* wound) relative to the same target word in a fluent context (e.g., the wound).

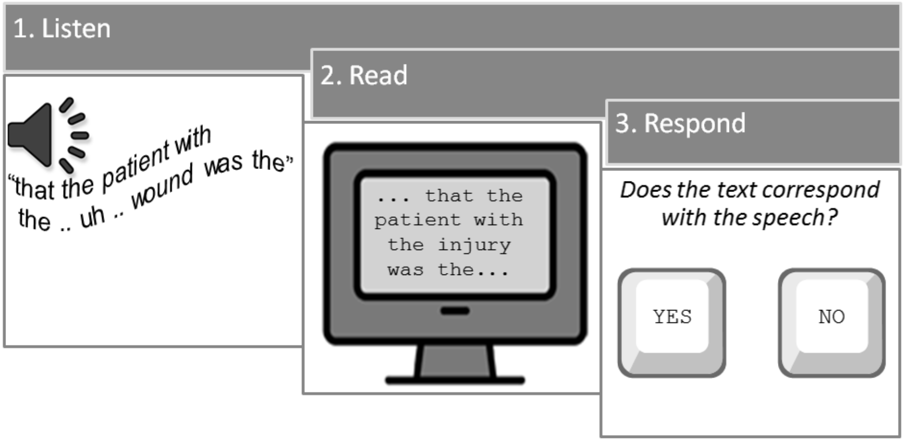

Figure 1: Schematical representation of the Change Detection Paradigm.

We designed two experiments differing only with respect to the speaker: participants in Experiment 1 were presented with fluent and disfluent native speech, and participants in Experiment 2 heard non-native speech. The results indicated that disfluency had a beneficial effect on participants' recall: listeners were more likely to detect a substitution of a word that had been preceded by a disfluency than substitutions of words in fluent context. This disfluency effect was present in both experiments: both native and non-native disfluencies triggered heightened listeners' attention. No attenuation of the disfluency effect was observed when participants listened to non-native speech.

These findings suggest that listeners do not modulate the effect of disfluency on attention based on knowledge about the non-native identity of the speaker. This could indicate that listeners, upon encountering a disfluency, raise their attention in an automatic fashion without taking the speaker identity into account (supporting the Temporal Delay Hypothesis of Corley & Hartsuiker, 2011). However, there are several concerns about the methodology of Study 4: for instance, the non-native speech materials in Experiment 2 did not contain any grammatical errors and only occasionally a disfluency. This may have influenced the perceived L2 proficiency of the non-native speaker in Study 4. Therefore, we should be careful in drawing conclusions about attentional effects of non-native disfluencies on the basis of a null result (i.e., no interaction between the disfluency effect and the type of speaker [native vs. non-native]). Despite the fact that we cannot draw definitive conclusions about how non-native disfluencies affect listeners' perceptual mechanisms, our results, nonetheless, emphasize the role of attention in an account of disfluency processing. Hesitations trigger listeners' attention with consequences for the retention of words following the hesitation.

3. An integrative account of fluency perception

The empirical studies summarized above studied how fluency characteristics affect the perception in native and non-native speech, motivated by an apparent contradiction between, on the one hand, the *negative* effects of non-native disfluencies on subjective fluency ratings, and, on the other hand, the *positive* effects of native disfluencies on speech perception. The combined results from the empirical studies contribute to our understanding of the beneficial, and of the disadvantageous effects of native and non-native disfluency on listeners. In the following paragraphs, an attempt will be made to demonstrate that the results from these studies can resolve the apparent contradiction.

3.1 Listeners' subjective impressions

We argue that the fluency characteristics of a spoken utterance follow from the speaker's cognitive fluency: disfluency is a symptom of inefficiency of the processes involved in speech planning and production (Segalowitz, 2010). This inefficiency may arise at any of the stages in speech production: in finding out what to say (conceptualization), in finding the right words (formulation), or in generating a phonetic plan (articulation; Levelt, 1989). Both native and non-native speakers suffer from disfluency, because both types of speakers experience the time pressure under which natural conversations take place.

This does not mean that native and non-native disfluency production are identical. There are considerable quantitative and qualitative differences between native and non-native disfluency production. Regarding the quantitative differences, non-native speakers produce more disfluency: L2 cognitive fluency is less efficient than L1 cognitive fluency. With respect to the qualitative differences, non-native speakers produce disfluencies at different points in the utterance: inefficiency in L2 cognitive fluency has different origins compared to L1 cognitive fluency (De Bot, 1992; Segalowitz, 2010). Insufficient declarative (knowledge) and procedural (skill) mastery of the L2 have been identified as two sources of L2-specific inefficiency (cf. De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012).

The quantitative differences between native and non-native cognitive fluency (which surfaces in the utterance as a difference in the number of disfluencies) means that, in practice, non-native speakers are generally perceived to be less fluent than native speakers. The qualitative difference between native and non-native cognitive fluency (surfacing in a different disfluency distribution) does not seem to affect the way native and non-native fluency characteristics are weighed (e.g., it suggests that a native speaker pausing before a low-frequency word is 'just as bad' as a non-native speaker pausing before a high-frequency word). Apparently, both native and non-native fluency characteristics are perceived to be symptoms of reduced cognitive fluency and, therefore, listeners weigh the fluency characteristics of native and non-native speech in a similar way (cf. Study 2).

3.2 Listeners' predictive strategies

However, listeners are not insensitive to the qualitative differences between native and non-native cognitive fluency. Study 3 has demonstrated that listeners can make use of symptoms of cognitive inefficiency by using disfluency as a cue to upcoming, relatively more complex information. Listeners were only observed to use disfluency to predict reference to low-frequency objects when listening to a native speaker. When listeners heard a non-native speaker produce similar spoken instructions, they did not use disfluency to guide anticipation of low-frequency referents. This suggests that the non-native identity of the speaker can modulate the effect that disfluencies have on prediction.

These findings from Study 3 suggest that listeners are sensitive to the qualitative differences between native and non-native cognitive fluency. Listeners are familiar with the regularities in native disfluency production and, therefore, can use disfluency to anticipate relatively more complex information. They are also familiar with the more irregular distribution of non-native disfluencies and, therefore, attenuate the effect of non-native disfluencies on prediction. As such, listeners can, in a very clever way, make use of symptoms of inefficiency for prediction.

3.3 Listeners' attentional resources

Study 4 reported that both native and non-native disfluencies have beneficial effects on speech comprehension because they were observed to trigger listeners' attention. Thus, the positive effect of native disfluencies on attention, reported in the literature, is extended to non-native disfluency. Despite several concerns about the methodology of the experiments in Study 4, we may yet speculate as to the reasons why native and non-native disfluencies heighten listeners' attention to the same extent.

One possible explanation of the attentional effects of disfluency is related to the Temporal Delay Hypothesis (proposed by Corley & Hartsuiker, 2011). This hypothesis argues that temporal delay in the speech signal improves speech comprehension. The temporal delay may provide the listener with additional time to orient to the upcoming information (disengagement from previous linguistic content and shift to new information). This would suggest that disfluency effects on attention are an automatic consequence of intrinsic temporal delay, independent of knowledge about the identity of the speaker. Both native and non-native disfluencies inherently introduce temporal delay and, therefore, both types of disfluencies trigger an orienting response.

There are, however, some empirical findings that challenge the Temporal Delay Hypothesis. One of its conjectures is that any kind of delay in the speech signal should improve speech comprehension: disfluencies, such as filled pauses and silent pauses, but also coughs, beeps, or barks. However, the evidence from previous studies for beneficial effects of delays (that are not disfluencies) on speech comprehension is equivocal (cf. Bailey & Ferreira, 2003; Barr & Seyfeddinipur, 2010; Corley & Hartsuiker, 2011). For instance, Fraundorf and Watson (2011) found that filled pauses did improve listeners' recall of previously remembered stories, but coughs, matched in duration to the filled pauses, did not.

Alternatively, the null result in Study 4 may be explained by listener strategies in response to anticipated comprehension effort. Both native and non-native disfluencies arise through relatively high cognitive load in speech production. This cognitive load experienced by the speaker may also carry consequences for listener effort in speech comprehension. For instance, finding the right label for a low-frequency object may present a native speaker with additional cognitive load, as evidenced by a higher probability of disfluency. At the same time, low-frequency words are also more cognitively demanding (for the listener) to comprehend, as evidenced by slower responses in word recognition (e.g., Marslen-Wilson, 1987) and lower recognition accuracy (e.g., Goldinger, Luce, & Pisoni, 1989). Listeners may anticipate this increased effort upon encountering a disfluency and, therefore, adopt precautionary comprehension strategies, such as the raising of attention.

The strategy of heightened attention levels in response to disfluency may also apply to the processes involved in comprehension of non-native speech. For instance, the difficulty experienced by a non-native speaker in planning and producing L2 speech may result in semantic inaccuracy, grammatical errors, or poor pronunciation. All of these challenge the listener in comprehension. Therefore, listeners may benefit from strategically raising their attention levels when encountering disfluency - both when listening to native and non-native speech - to ensure timely comprehension of cognitively demanding linguistic input.

Based on the data from Study 4, we cannot yet discriminate between these two explanations of the attentional effects of native and non-native disfluencies. New investigations will have to determine the value of either of the two explanations (see next section). Nevertheless, the combined findings from the different studies do resolve the apparent contradiction between, on the one hand, the *negative* effects of non-native disfluencies on subjective fluency ratings, and, on the other hand, the *positive* effects of native disfluencies on speech perception.

We argue that negative and positive effects of disfluency are the result of different listener considerations. Native and non-native disfluencies have negative effects on listeners' judgments about the speaker's fluency level, because listeners are assumed to consider both types of disfluencies to be symptoms of speech production difficulty. Despite these negative effects, listeners are capable of using these symptoms of speaker difficulty to anticipate reference to relatively more complex linguistic content (e.g., low-frequency words). However, listeners only adopt this predictive strategy when listening to native speech, because of the regularities in the distribution of native disfluencies. The distribution of non-native disfluencies is, from the native listener's point of view, much more irregular, leading to an attenuation of the effects of non-native disfluency on prediction. With respect to attention, both native and non-native disfluencies may heighten listeners' attention to the following information, either because of the delay intrinsic to native and non-native disfluency, or because of listeners taking precautionary measures to reduce anticipated cognitive effort in comprehension.

4. Future research

The sketched account of the perception of fluency may motivate future studies to test the account's conjectures and/or to expose its limitations. For instance, our conclusions about the evaluation of fluency were based on experiments (Studies 1 and 2) in which we presented raters with specific instructions on how to assess fluency. Current language tests commonly provide their raters with explicit instructions about how to assess oral fluency by reference to fluency characteristics of the utterance, such as speed of delivery, pauses, and hesitations. This tendency is found both within language testing practice and within the literature on fluency perception (e.g., Derwing et al., 2004; Kormos & Dénes, 2004; Rossiter, 2009). Because we also adopted this procedure in Studies 1 and 2, our findings could be directly applied to language testing practice where similar methods are used.

However, exactly because of the prevalence of very specific fluency instructions, the relationship between ratings of fluency in the broad sense (i.e., overall speaking proficiency) and ratings of fluency in the narrow sense (i.e., a component of overall speaking proficiency) has not (yet) received much attention. Some studies have investigated the componential nature of overall speaking proficiency (Adams, 1980; Higgs & Clifford, 1982; McNamara, 1990), or have targeted the factors that contribute to oral proficiency (Ginther, Dimova, & Yang, 2010; Iwashita, Brown, McNamara, & O'Hagan, 2008; Kang, Rubin, and Pickering, 2010). However, an investigation of how fluency characteristics in the speech signal contribute to ratings of fluency in the narrow and in the broad sense has not yet been undertaken.

This carries consequences for our conclusions in Study 2. The conclusion that native and non-native fluency perception are comparable, was drawn on the basis of data collected through evaluations of fluency in the narrow sense. This raises the question whether the similarity of native and non-native fluency perception also applies when listeners assess fluency in the broad sense (for instance, in fluency assessment without any instructions on what comprises fluency). This question is very much relevant for everyday situations in which interlocutors in a conversation draw inferences about the other (native or non-native) speaker's social status (Brown, Strong, & Rencher, 1975), emotion (Scherer, 2003), physical properties (Krauss, Freyberg, & Morsella, 2002), metacognitive state (Brennan & Williams, 1995), fluency level (e.g., Studies 1 and 2), etc. Listeners' considerations in these spontaneous, uncontrolled situations have been under-investigated in the literature and future studies may find ways of tapping listeners' underlying deliberations in these situations. Until that time, it is uncertain whether our conclusions about native and non-native fluency in the narrow sense generalize to situations without clearly formulated fluency assessment instructions.

Another issue that prospective studies may address is related to the differential effects of native and non-native disfluencies on prediction. Study 3 has revealed that non-native disfluencies do not guide prediction of low-frequency referents (whereas native disfluencies do). This does not necessarily imply that non-native disfluencies do not guide prediction at all. Our experiments targeted listeners' attributions of disfluency to speaker trouble in *formulation* (i.e., lexical retrieval). It is, as yet, unclear whether non-native disfluencies also have differential effects on speech comprehension (relative to native disfluencies) when listeners attribute disfluency to speaker trouble in *conceptualization*. For instance, an experiment may be proposed in which, following Arnold et al. (2007), listeners are presented with visual arrays of known vs. unknown objects (e.g., a picture of an ice-cream cone paired to a picture of an abstract symbol). When a native speaker is heard producing disfluent instructions to click on one of the objects, listeners have been shown to anticipate reference to the unknown object (Arnold et al., 2007), suggesting that listeners attribute the disfluency to trouble in conceptualization of the unknown object. A new experiment may test how listeners deal with a situation in which a non-native speaker struggles to produce these kinds of instructions.

It could be argued that non-native disfluency does not affect listeners' predictive mechanisms in any situation. Listeners may consider the non-native speaker to have equal trouble with the production of known and unknown words in their L2 (i.e., naming an ice-cream cone is more difficult for an L2 speaker as compared to an L1 speaker). In this case, the perception of non-native speech would pattern with the perception of atypical native speakers (e.g., a native speaker with object agnosia; Arnold et al., 2007). Alternatively, one could also hypothesize that non-native disfluency, just like native disfluency, may guide prediction of unknown referents. This would suggest that listeners are aware that both native and non-native speakers encounter similar troubles in conceptualizing unknown referents. As such, the non-native identity of the speaker would play a role in listeners' attributions of disfluency to speaker trouble at the phase of conceptualization. In this fashion, new studies into prediction as a component of speech comprehension may determine in which cases the non-native identity of a speaker plays a role in speech comprehension, and in which cases it does not.

A final issue that may encourage follow-up studies concerns the two introduced explanations for the attentional effects observed in Stduy 4: they are either due to the delay intrinsic to native and non-native disfluency (he Temporal Delay Hypothesis; Corley & Hartsuiker, 2011), or they are the result of listeners taking precautionary measures to reduce anticipated cognitive effort in comprehension. New experiments may be designed that test these two explanations. For instance, researchers may present listeners with speech that contains forms of delay that do not necessarily cue more complex linguistic content (e.g., coughs). If listeners are found to be more accurate in recalling words that were preceded by such delays, this would suggest that delay alone can account for heightened attention. Thus, the field of speech perception may benefit from investigations into the effects of different types of delay in various comprehension tasks (e.g., recognition, prediction, retention, syntactic parsing, etc.).

Since the definitive explanation for the observed attentional effects of disfluency is, as yet, lacking, the relationship between the attentional and prediction effects of disfluency is also unclear. If both attentional and prediction effects of disfluency are caused by listeners' experience with the regularities of disfluency production, another interesting question regards the time course of these two types of effects. Does prediction precede heightened attention? Or vice versa? And does one effect elicit the other? Do heightened attention levels trigger predictive mechanisms, or does prediction of relatively more complex information implicate the attention levels required for the processing of this information? These questions may form an agenda for future investigations into the cognitive effects of disfluency.

5. Implications

5.1 Rater instructions

The findings from Study 1 have revealed that the pausing and speed characteristics of L2 speech are the most important contributors to perceived fluency judgments. Non-native speakers' repair strategies (e.g., corrections, repetitions) also contribute to fluency perception, but these acoustic features play a much smaller role. These observations are applicable to testing procedures in language testing practice. Many language tests use speaking rubrics with explicit mention of aspects of fluency, but the way in which raters are instructed to assess fluency differs. For instance, for IELTS (IELTS, [n.d.]), raters are instructed to assess, amongst others, speakers' 'fluency and coherence' on a 9-band proficiency scale. Descriptives of speech at each of the 9 proficiency bands are provided, such as references to the length of the speech performance, pauses, hesitations, repetitions, and self-corrections. The descriptives of several bands contain reference to repair strategies: for example, speakers at band 5 use "repetition, self-correction and/or slow speech"; and speakers at band 6 are described as "willing to speak at length, though [they] may lose coherence at times due to occasional repetition, self-correction or hesitation". In contrast, hardly any descriptives contain reference to speed characteristics. The rating procedure of this language test (and also others) may be informed by our hierarchy of fluency dimensions, as described in Study 1. For instance, the contributions of pause and speed characteristics can be stressed, whereas reference to repair strategies could be minimized. This does not mean that references to repair strategies should be removed from rater instructions, but they certainly should not be emphasized either.

In a similar way, our findings about the relevance of the different fluency dimensions for fluency perception can also be applied to instruments for automatic fluency assessment. Such instruments are already being used in official language tests, such as the PTE Academic, TOEFL iBT, and the Dutch Immigration Test. Our results can guide designers of these tests to adjust the weights that are applied to automatically measured acoustic fluency characteristics. For instance, it is possible to implement a higher fluency penalty on pausing phenomena as compared to repair phenomena. In this fashion, automatic fluency assessment is expected become a better approximation of human fluency judgments.

5.2 Relevance of fluency dimensions

The first experiment from Study 1 suggested a hierarchy in the contributions of fluency dimensions to fluency perception: pause and speed characteristics of speech contribute most to perceived fluency judgments, and repair strategies contribute only very little. The following three experiments of Study 1 addressed the question why listeners adopt this hierarchy of fluency dimensions. We proposed that differences in perceptual sensitivity to the three fluency dimensions might account for this result. If pause and speed characteristics are more salient than repair phenomena, then this may explain why raters base their fluency judgments more on the speaker's pause behavior and speed of speaking rather than on the speaker's repair behavior. However, listeners were found to be as sensitive to repair characteristics as to speed characteristics. Therefore, perceptual sensitivity alone cannot account for the observed relative contributions of the different fluency dimensions to fluency perception.

What potential other factor may account for the observed hierarchy of fluency dimensions? Why do listeners consider a speaker's pause behavior to be a better indicator of the speaker's fluency level as compared to the speaker's repair behavior? One might hypothesize that pauses are better indicators of an L2 speaker's overall speaking proficiency (relative to repairs and repetitions). De Jong, Steinel, Florijn, Schoonen, and Hulstijn (2013) investigated the relationship between measures of L2 utterance fluency and measures of L2 cognitive fluency, by collecting data from a large cohort of non-native speakers (N=179). Utterance fluency was operationalized as a set of acoustic fluency characteristics (e.g., articulation rate, number of silent pauses, number of corrections, etc.). Cognitive fluency was operationalized as the sum of the speakers' L2 linguistic knowledge and processing skills (e.g., L2 grammatical knowledge, speed of L2 lexical retrieval, etc.). Relating the utterance fluency measures to the cognitive fluency measures, the authors found that a speaker's L2 proficiency correlated most strongly with the speaker's (inverse) articulation rate: 50% of individual variance in (inverse) articulation rate was explained by the speaker's L2 cognitive fluency. In contrast, average pausing duration was only weakly related to linguistic knowledge and processing skills. This finding suggests that inefficiency in L2 speech production primarily surfaces in the non-native speaker's articulation rate, rather than his/her average pausing behavior.

Nevertheless, the literature on fluency perception has repeatedly argued that listeners do take a speaker's pausing behavior to be indicative of that person's fluency level. This presents an apparent conundrum to the study of fluency production and perception: listeners base their fluency judgments on pausing and speed characteristics, while only speed of articulation truly reflects the L2 speaker's underlying cognitive fluency. One possible solution to the puzzle may lie in the perceptual relationship between speakers' speed of speaking and speakers' pausing behavior. Both the study of De Jong et al. (2013) and the experiments in Study 1 used 'independent' acoustic measures: that is, pause and speed measures that portray low intercollinearity, such as the measure *articulation rate*) which is calculated by a speaker's speaking time, excluding silent pauses (vs. the measure *speech rate* which is calculated by a speaker's total time, including silent pauses). This approach is useful when one wants to distinguish the separate contributions of the three fluency dimensions. However, we do not know whether listeners are also capable of perceptually distinguishing speed fluency from breakdown fluency. Alternatively, one could hypothesize that these two dimensions together load onto one perceptual category: pause & speed fluency.

Further inspection of the data supports this latter suggestion. In Study 1, Experiment 2 was designed to collect perceptual judgments of L2 speakers' pausing behavior, and Experiment 3 was designed to collect perceptual judgments of L2 speakers' speed of speaking. Correlations between objective pause measures and the speed measure, as reported in Bosker et al. (2013), did not exceed r=0.4. Nevertheless, supplementary analysis of the relationship between the subjective judgments of pausing behavior (Experiment 2) and speed behavior (Experiment 3) reveals a strong correlation between the pause and speed ratings: Pearson's r=0.839, p<0.001. This indeed suggests that pause and speed characteristics, despite a relationship weak correlations on an acoustic level, do load onto one pause & speed percept in perceived fluency. Both independent parts of this percept are strong predictors of perceived fluency judgments, but only an L2 speaker's speed of articulation - not pausing - is strongly correlated with L2 knowledge and skills (i.e., oral proficiency; De Jong et al., 2013).

This raises the question which fluency construct language tests should reflect: perceived fluency or cognitive fluency? The current situation in language testing practice is that language test scores approximate perceived fluency judgments. This is inherent to many language tests because they make use of human raters. Thus, a low score on a language test correlates with low subjective fluency judgments. However, instead of reflecting perceived fluency, language tests could also reflect the underlying efficiency of the cognitive processes involved in L2 speech production, that is, cognitive fluency. Language tests evaluating cognitive fluency would provide insight into the speaker's underlying L2 knowledge and skills. As such, these tests require automatic assessment of those speech characteristics that reflect an L2 speaker's underlying L2 proficiency (e.g., articulation rate; De Jong et al., 2013), forgoing human perception.

Ultimately, designers of language tests will have to decide what fluency construct their language test is to reflect, on the basis of the particular goals of the designers. The outcomes of language tests that have been designed to reflect native listeners' impressions (perceived fluency) are applicable to social interactions with native speakers. Scores on such language tests inform non-native speakers about how native speakers will perceive their proficiency. Alternatively, if a language test is designed to reflect underlying L2 proficiency (cognitive fluency), its outcomes provide insight into speakers' underlying L2 knowledge and skills (irrespective of native speakers' subjective prejudices, stereotypes, etc.). Thus, language tests reflecting cognitive fluency constitute are expected to be more objective assessment tools, since they produce highly reliable output.

5.3 Norms and standards

On the basis of the results from Study 2, we concluded that, when raters are instructed to evaluate fluency in the narrow sense, native and non-native fluency perception are comparable. This implies that listeners do not make a qualitative distinction between native and non-native speakers in fluency assessment. Rather, the difference between native and non-native fluency is gradient: variation in fluency judgments of different native and non-native speakers can be accounted for by quantitative differences. This conclusion is good news for language learners. Our results suggest that there is no insurmountable obstacle preventing them from achieving native-like fluency levels in their second language.

The conclusion that native and non-native fluency perception are comparable also entails that, contrary to common opinion, native speech is not perceived to be fluent by default; instead, there is considerable variation in the perceived fluency of native speakers. Variation in fluency characteristics of native speech (artificially created through phonetic manipulations, Study 2, or naturally observed variation, Bortfeld et al., 2001) influences how fluent native speakers are perceived to be. This observation led us to conclude that a single ideal native fluency standard does not exist.

This carries consequences for language testing practice, where non-native fluency levels are regularly assessed on grounds of an idealized disfluency-free norm. However, the results from Studies 3 and 4 show that disfluency-free speech should not be considered the norm, because disfluencies can serve a purpose in speech comprehension. They may guide listeners to predict upcoming content and/or may trigger heightened attention levels. This suggests that, instead of penalizing all disfluent speech, language tests should rather aim to assess fluency by penalizing only those speech characteristics that hinder communication.

6. Conclusions

The proposed account of fluency perception argues that disfluency arises from speech production difficulty. The symptoms of such speaker trouble (i.e., individual disfluencies in the speech signal) negatively affect listeners' impressions of the speaker's fluency level. Nonetheless, listeners can, in a very clever way, make use of the disfluent character of spontaneous speech for comprehension: disfluencies may elicit anticipation of - and heightened attention to - subsequent linguistic content. This account of fluency perception is the result of combining the evaluative approach and the cognitive approach to fluency. We hope to have convinced the reader of the advantages of combining these different approaches in order to arrive at a more comprehensive account of how native and non-native fluency characteristics are perceived. More specifically, because of the combination of the evaluative and the cognitive approach, the findings from one approach can inform the other.

For instance, the evaluative approach can help put the findings from the cognitive approach into perspective. Many studies into the cognitive effects of disfluency report on experiments that use monologue speaking conditions, involving artificially manipulated, one-sentence stimuli. However, spoken communication in everyday life takes place in real face-to-face conversations with free interaction between multiple interlocutors. In these natural circumstances, social factors play a major role. For instance, in natural conversations people do not take the speech signal merely as input for comprehension but also as input for social inferencing (e.g., the speech signal as an indicator of social status, physical attributes, and cognitive state). The evaluative approach to fluency reminds the cognitive approach that fluency perception, ultimately, takes place in a social setting outside of the psycholinguistic lab. In these natural communicative settings, disfluency does not only guide prediction or attract listeners' attention, but it also forms an integral part of social interaction in which people form subjective impressions of the other persons around them. Without an understanding of how listeners arrive at these socially relevant impressions, the findings from the cognitive approach are limited to very specific, highly controlled situations.

Conversely, the findings from the cognitive approach also inform the evaluative approach to fluency. Studies 3 and 4 advocate a more fine-grained perspective on fluency assessment. These chapters demonstrate that disfluencies can sometimes help the listener in comprehension. Therefore, fluency assessment should not take place along disfluency-free standards, but language tests should ultimately discriminate between those speech characteristics that help, and those that hinder communication. This calls for a large-scale research program that aims to provide a way of distinguishing communicatively 'helpful' and 'unhelpful' speech characteristics. It will have to combine methods and insights from various scientific disciplines, such as (applied) linguistics, communication sciences, psychology, and their various subfields, in order to fully understand the mechanisms involved in human spoken communication.

This research note contributes to the study of spoken communication through an investigation of fluency, combining the evaluative and the cognitive approach. Thus, it has extended our understanding of how both native and non-native fluency characteristics are weighed for fluency assessment, and how these characteristics affect speech comprehension. The described implications of the empirical studies and the call for new investigations testify to the notion that speech performance matters: communication through spoken language does not only depend on what is said, but also on how it is said and by whom. As such, these studies have put speech performance at the heart of the study of spoken communication, right where it belongs.

7. Acknowledgments

This research note is an adapted version of the concluding chapter of the author's PhD thesis (defended at Utrecht University on May 23, 2014, *cum laude*). The research reported in this PhD thesis was carried out at the Utrecht institute of Linguistics OTS (UiL OTS), Utrecht University, and it was supported by Pearson Language Testing by means of a grant awarded to Nivja H. de Jong: "Oral Fluency: Production and Perception".

References

Adams, M. L. (1980). Five co-occurring factors in speaking proficiency. In J. Frith (Ed.), *Measuring spoken language proficiency* (p. 1-6). Washington, DC: Georgetown University Press.

Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say -thee uh- you're describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 914-930.

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.

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language, 49*(2), 183-200.

Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes, 25*(4), 441-455.

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech, 22*(3), 201-211.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*, 123-147.

Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T. J. M., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30*, 157-175.

Bosker, H. R., Quené, H., Sanders, T. J. M., & de Jong, N. H. (in press). The perception of fluency in native and non-native speech. *Language Learning.*

Bosker, H. R., Quené, H., Sanders, T. J. M., & de Jong, N. H. (2014). Native 'um's elicit prediction of low-frequent referents, but non-native 'um's do not. *Journal of Memory and Language 75*, 104-116.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44*(2), 274-296.

Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (2008). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(3), 696-702.

Corley, M., & Hartsuiker, R. J. (2011). Why um helps auditory word recognition: The temporal delay hypothesis. *PLoS One, 6*(5), e19792.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, *er*, say it: Hesitations in speech affect language comprehension. *Cognition, 105*, 658-668.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America, 111*, 2862-2873.

De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics, 13*(1), 1-24.

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition, 34*, 5-34.

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics, 34* (5), 893 - 916.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54*, 655-679.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34*, 709-738.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition, 29*(2), 320-326.

Fraundorf, S. H., & Watson, D. G. (2011). The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language, 65*(2), 161-175.

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing, 27*, 379-399.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28(5), 501-518.

Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. Experimental Psychology, 57(3), 169-177.

Higgs, T. V., & Clifford, R. (1982). The push toward communication. In T. V. Higgs (Ed.), *Curriculum, competence and the foreign language teacher.* (p. 243-265). Skokie, Illinois, USA: National Textbook Company.

IELTS. [n.d.]. IELTS Speaking band descriptors. Retrieved from http://www.ielts.org/PDF/UOBDs SpeakingFinal.pdf (November 2013).

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics, 29*, 24-49.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal, 94*, 554-566.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*, 145-164.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: The MIT Press.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Investigating the consequences of disfluent silent pauses in speech for listeners. *Neuropsychologia, 48*, 3982-3992.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word,* 15, 19-44.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1), 71-102.

McNamara, T. F. (1990). Item response theory and the validation of an ESP test for health professionals. *Language Testing, 7*(1), 52-76.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and nonnative speakers of English. *Canadian Modern Language Review, 65*, 395-412.

Segalowitz, N. (2010). *Cognitive bases of second language fluency.* New York: Routledge.