Predicting judged fluency of consecutive interpreting from acoustic measures

Potential for automatic assessment and pedagogic implications

Wenting Yu and Vincent J. van Heuven
Shanghai International Studies University / University of Pannonia/Leiden University

This experimental study of consecutive interpreting investigates whether: (1) there is any correlation between assessments of its fluency and accuracy; (2) judged fluency can be predicted from computer-based measurements like articulation rate. Ten raters judged six criteria of accuracy and fluency in two consecutive interpretations of the same recorded source speech, from Chinese A into English B, by 12 trainee interpreters (seven undergraduates, five MA students). The recorded interpretations were examined with the speech analysis tool PRAAT. From a computerized count of the pauses thus detected, together with disfluencies identified by raters, 12 acoustic measures of fluency were calculated. The advanced students were more fluent than the beginners; both groups were less fluent in the initial interpretation. Statistical analysis shows: (1) a strong positive correlation between judged accuracy and judged fluency; (2) strong correlations between judged fluency and objective fluency variables; (3) the usefulness of effective speech rate (number of syllables, excluding disfluencies, divided by total duration of speech production and pauses) as a predictor of judged fluency. Other important determinants of judged fluency were the number of filled pauses, articulation rate, and mean length of pause. Potential for developing automatic fluency assessment in consecutive interpreting is discussed, as are possible training implications.

Keywords: fluency, accuracy, acoustic measures, consecutive interpreting, automatic assessment

Interpreting 19:1 (2017), **47–68.** DOI 10.1075/intp.19.1.03yu
ISSN 1384–6647 / E-ISSN 1569–982X

1. Introduction

Ever since conference interpreting became established as an internationally recognized profession in the mid-twentieth century, the concept of quality has been a major concern in professional practice and training (Pöchhacker 2012). The 1980s witnessed the burgeoning of systematic investigations and studies of interpreting quality, with growing recognition of the role played by fluency as an important criterion for assessing the interpreter's delivery. It is nevertheless only recently that interpreting scholars have begun to focus on fluency as a specific component of quality, with the aim of defining it in practical terms and determining to what extent it might affect intelligibility and user perception (Rennert 2010). So far, a number of empirical studies have indicated that: (a) there is a contrast between user expectations and user perception regarding the importance of fluency as a quality criterion in interpreting; (b) fluency is actually such an important element in perceived quality that its experimental manipulation affects user perception of other criteria (Collados Aís et al. 2007; Pradas Macías 2003, 2007; Rennert 2010). These studies argued that the importance of fluency had long been underestimated, with surveys of quality perception or expectations among interpreters and habitual listeners showing a tendency to prioritize accuracy or faithfulness over form-related features like fluency (Bühler 1986; Chiaro & Nocella 2004; Kurz 1993, 2001, 2003; Moser 1996).

While almost all the above-mentioned research on the relations between judged fluency and judged quality was carried out in the context of *simultaneous* interpreting (SI), where fluency is in any case largely subject to source speech rhythm, this study investigates the relations between judged fluency and judged quality in *consecutive* interpreting (CI). Here, a fundamental consideration with a view to training, testing and assessing interpreters' fluency is the importance of identifying what specific component(s) might affect perceived fluency in consecutive interpreting.

Studies on L1 and L2 speech have attempted to define fluency in terms of objective speech properties (Cucchiarini et al. 2000, 2002; Freed 1995; Lennon 1990; Riggenbach 1991). According to Cucchiarini et al. (2002), these studies have adopted a dual approach in which listeners' evaluations of fluency are related to objective temporal measures. This type of approach, particularly useful for gaining insight into the acoustic features underlying listeners' evaluations, has a long tradition in phonetic research (Cucchiarini et al. 2002). It has been observed that, for speech tasks entailing different degrees of cognitive effort, there will be corresponding differences in fluency rating (Bortfeld et al. 1999; Grosjean 1980) and in the most powerful objective predictors of fluency (Cucchiarini et al. 2002). It is therefore reasonable to expect that, for the cognitively demanding speech task of interpreting, the relations between auditory fluency rating and acoustic measures will differ from those in everyday L1 and L2 speech. In other words, the acoustic measures that have robust predictive power for the judged oral fluency of unconstrained L1 and L2 production might not apply to interpreting.

This study has three goals. First, we aim to investigate whether there is a connection between judged fluency (a major feature of an interpretation's form or presentation) and judged accuracy of delivery (usually considered the most important feature of content) in consecutive interpreting. Second, we attempt to investigate the relations between the subjective perception of fluency and objective acoustic measures in consecutive interpreting (here, from Chinese 'A' into English 'B'). Third, we intend to explore which acoustic measure(s) contribute(s) most to judged fluency in CI.

If the present approach proves feasible, the implication is that expert fluency ratings of CI can, in principle, be predicted on the basis of temporal fluency measures. Currently, with an increasing number of interpreting programs worldwide and various national and regional interpreting accreditation tests in operation, assessment of interpreting competence entails enormous costs in time and know-how. These costs might be at least partly saved by developing a tool for automatic assessment of fluency, though this is obviously only one parameter of overall quality. Identifying the acoustic measure(s) with the greatest predictive value for judged fluency in consecutive interpreting will also enable trainee interpreters to improve specific aspects of their production accordingly.

2. Fluency

Fluency is a multifaceted concept, for which there is no consensus on the various attempts to define it in different contexts (e.g., Brumfit 1984; Chambers 1997; Fillmore 1979; Leeson 1975; Lennon 1990, 2000; Schmidt 1992). Lennon (1990, 2000) distinguishes between a broad sense and a narrow sense of fluency. In its broad sense, it is often considered a synonym of "overall language proficiency" (Chambers 1997; Lennon 1990); fluency in its narrow sense refers to one aspect of oral proficiency, as distinct from other components like grammatical knowledge and vocabulary size (Pinget et al. 2014). Cucchiarini et al. (2000) consider their discussion of fluency in the broad sense as relevant primarily to foreign language teaching and testing, pointing out that fluent speech as such tends not to be formally used as a criterion of evaluation for native speakers. In such a perspective, native-speaker-like performance does not necessarily constitute the target to be achieved (Brumfit 1984). Some definitions of fluency refer to the temporal aspect of oral proficiency (Freed 1995; Lennon 1990; Nation 1989; Riggenbach 1991;

Schmidt 1992; Towell et al. 1996), in line with Lennon's (1990) assumption that the goal in foreign language learning consists in producing "speech at the tempo of native speakers, unimpeded by silent pauses and hesitations, filled pauses, […] self-corrections, repetitions, false starts and the like."

Against this background, it is the identification of temporal features of speech that enables quantitative studies of fluency in different contexts: previous research shows that perceived fluency is correlated with different quantitative measures, according to the language and speech task concerned – e.g., L1 speech vs. L2 speech; read speech vs. spontaneous speech (Cucchiarini et al. 2002; Kormos & Dénes 2004; Möhle 1984; Towell et al. 1996).

Studies of temporal features in L1 and L2 speech have identified a number of these quantitative variables that appear to be related to perceived fluency (Freed 1995; Goldman-Eisler 1968; Grosjean 1980; Grosjean & Deschamps 1975; Lennon 1990; Nation 1989; Riggenbach 1991; Towell et al. 1996). The clearest taxonomy is provided by Grosjean (1980), who distinguishes between primary and secondary variables. Primary variables are those "that are always present in language output" (Grosjean 1980: 40): articulation rate (number of syllables, including those that constitute disfluencies, divided by the total duration of speech apart from silent and filled pauses); speech rate, also known as rate of speech or speaking rate (number of words or syllables, including disfluencies, divided by the total duration of speech complete with all pauses); phonation/time ratio (total duration of speech without pauses, divided by total duration of speech including pauses to obtain a percentage); mean length of run (mean number of syllables produced between silent pauses); mean length of silent pauses; duration of silent pauses; and number of silent pauses (Cucchiarini et al. 2002). Secondary variables are related to hesitation phenomena such as filled pauses, repetitions, repairs, and restarts. These variables are not necessarily present in all speech, particularly if it is read speech (Grosjean 1980: 42). Tavakoli & Skehan (2005) divide temporal properties of fluency into three components: (1) speed fluency (i.e., the speed at which speech is delivered); (2) breakdown fluency (i.e., the number and length of pauses); and (3) repair fluency (i.e., the number of false starts, corrections and repetitions).

The focus of the present study is to investigate the relationship between auditory rating and acoustic analysis of fluency in the cognitively demanding speech task of CI from Chinese 'A' into English 'B' (in language teaching terms, from L1 to L2). Since this is the first study of its kind, Section 2.1 presents an overview of the existing dual studies on fluency in two types of L2 speech tasks (i.e., read speech and spontaneous speech), in which listeners' evaluations of speech are examined in relation to objective temporal measures calculated for the same speech samples.

2.1 Fluency in spontaneous and read L2 speech

Research on fluency in L2 speech is mostly concerned with one of two aims: to gain insight into the factors which determine listeners' evaluations (Cucchiarini et al. 2002); and to help develop objective tests of second language fluency that might lend themselves to automatic assessment (Cucchiarini et al. 2002; Townshend et al. 1998). Lennon (1990), Riggenbach (1991), Freed (1995), Kormos and Dénes (2004), Cucchiarini et al. (2002) and Pinget et al. (2014) carried out studies in which samples of spontaneous speech produced by non-native speakers of English were judged by experts on fluency and were then analyzed in terms of quantitative variables such as speech rate, phonation/time ratio (the percentage of speaking time used for actual speech production), mean length of run, and number and length of pauses. These studies show that: (1) fluency ratings are mainly affected by quantitative variables related to speed fluency and breakdown fluency; (2) examining the relationship between fluency ratings and temporal variables in spontaneous speech may be rather complex, since the former are affected by non-temporal language features such as grammar, vocabulary and accent (Freed 1995; Lennon 1990; Riggenbach 1991). Cucchiarini et al. (2000) conducted an experiment with L2 Dutch read speech and found that expert ratings of fluency could be predicted on the basis of quantitative measures, the best predictor being speech rate, followed by articulation rate and the number of pauses. In a later study by Cucchiarini et al. (2002), the relationship between objective properties of speech and perceived fluency in L2 Dutch read and spontaneous speech was investigated in two separate experiments: fluency ratings in both cases were closely related to speech rate, phonation/time ratio, number of silent pauses per minute, duration of silent pauses per minute, and mean length of run. While articulation rate showed almost no relationship with perceived fluency ratings in spontaneous L2 speech, the two were closely related in read L2 speech production: the authors' tentative explanation for this finding was that, since pauses tend to be much more frequent in spontaneous speech, articulation rate (which takes no account of pauses) may in practice be relegated to a position of irrelevance. Kormos and Dénes (2004) conducted a dual study on L2 Hungarian spontaneous speech fluency ratings and temporal measurements: speech rate, mean length of utterance, phonation/time ratio and the number of stressed words produced per time unit were the best predictors of fluency scores. Like Cucchiarini et al. (2002), they did not find that articulation rate, the number of filled and unfilled pauses, or other disfluency phenomena were good predictors of fluency ratings. A recent study by Pinget et al. (2014) investigated which acoustic measures of fluency can predict perceived fluency in L2 Dutch spontaneous speech: although their acoustic measures (calculated on the basis of syllable length and pause length/frequency) differed from those used in previous research, these parameters showed high predictive value for much of the variance in fluency ratings, while two measures of repair fluency (number of corrections and number of repetitions) showed a certain – albeit limited – degree of predictive value compared to other studies (cf. Cucchiarini et al. 2002; Kormos & Dénes 2004).

To sum up, a number of studies have shown that objective measures are predictive of subjective fluency ratings in L2 speech. The general consensus is that objective measures related to speed fluency and breakdown fluency are far more relevant than repair fluency in this respect. However, the predictive power of objective measures differs for L2 speech tasks in relation to the cognitive effort involved.

Section 2.2 will briefly review a number of studies on fluency in the cognitively more complex task of interpreting.

2.2 Fluency as a quality criterion of interpreting

Interpreting is cognitively more demanding than L2 oral tasks, which account for only part of the interpreter's overall activity. What distinguishes CI from everyday spoken language activity is readily appreciated by basing comparison of the two on models often used to analyse them: Levelt's (1989) speech production model and Gile's (1995) Effort Models. The main difference between the two is that the speech production model has an initial conceptualization stage, whereas CI starts with perception and comprehension of the source language, with parallel storage, processing, and retrieval of information through note-taking, memory functions, and coordination of all these efforts. As a result, more attentional resources are almost certainly required in CI than in spontaneous speech production.

The perception and expectation of interpretation quality is generally agreed to involve not only content (mainly accuracy and completeness of information), but also form (fluency of delivery, accent, intonation, and voice quality) (e.g., Bühler 1986; Zwischenberger & Pöchhacker 2010). Fluency is among the most important formal criteria contributing to the overall quality of interpreting, though it has received very little systematic attention in the teaching and training of interpreting. In Kurz's (1993) survey of how different user groups and interpreters rate various features of conference interpreting, fluency is placed fifth out of eight criteria in the overall ranking – ahead of correct grammatical usage, native accent, and a pleasant voice. Bühler (1986) and Chiaro and Nocella (2004) find that fluency ranks fourth on the list of interpreting quality criteria, following content-related criteria such as sense consistency and completeness of information. In a recent large-scale global survey on conference interpreting quality by Pöchhacker (2012), involving 704 AIIC interpreters worldwide, fluency was perceived as very important by 71% of the participants and ranked third out of eleven quality criteria (behind sense consistency and logical cohesion). In addition, a few studies have examined how the relative importance of fluency differs between users' expectations and their actual assessment of interpretations: results suggest that limited fluency may impact negatively on the overall judged quality of an interpretation (e.g., Collados Aís 1998; Pradas Macías 2003; Rennert 2010).

Research on fluency in interpretation, though still in its infancy, is an exciting area of study (Mead 2005). It offers insights into what features of an interpretation actually constitute and affect judged fluency, providing a starting point from which to help trainee interpreters negotiate the many difficulties of the learning process and perhaps also to develop an automatic assessment tool for fluency in interpreting.

Fluency ratings have been much less widely studied in interpreting than in L2 production as a whole (see Section 2.1). There is scarcely any study that has attempted to explore the underlying temporal parameters that constitute fluency, except Mead's (2005) pioneering work on elaborating a conceptual approach to quantitative assessment of fluency in interpreting. Based on his analysis of five temporal parameters of fluency, he suggested that speech rate, pause duration, and length of fluent run can be taken as the most relevant parameters in assessing fluency. However, these three objective measures were not checked against subjective perception of fluency, thus ultimately offering no empirical evidence of whether they are powerful predictors of judged fluency in actual practice. Another limitation of Mead's (2005) initial exploration of interpreters' fluency is that he did not examine repair disfluencies such as false starts, restarts, corrections and repetitions.

To sum up, no interpreting research has emulated studies of L1 and L2 speech in exploring whether (and which) objectively quantifiable speech parameters can be predictive of subjective fluency ratings. Such an approach is potentially of considerable significance, as a basis for developing an automatic assessment tool that may help make evaluation of interpreting less labour-intensive.

2.3 Implications of this study for interpreter assessment and training

Most interpreters with Chinese A and English B are required to interpret in both directions. For trainee interpreters who are not early bilinguals, the disparity in command of the A and B languages always poses a major challenge. Accordingly, oral proficiency in B often receives a great deal of attention throughout the entire interpreting syllabus. Fluency, as an important aspect of oral proficiency, should therefore be given due attention in interpreter training and related research.

The results of this study are expected to shed light on the relations between subjective ratings and objective measures of fluency in consecutive interpreting. If objective measures correlate well with subjective ratings on fluency in interpreting, the most powerful predictor(s) might ultimately lend themselves to development of automatic assessment for trainee interpreters' fluency in an examination setting. While quantitative fluency parameters are certainly not bound to reflect the overall quality of interpreting, the distinct practical advantage of a quantitative assessment tool is the scope it offers for a more clear-cut evaluation than content-related parameters like completeness or correctness: the latter may ultimately prove more difficult to pin down, and their evaluation may differ from one assessor to another (Mead 2005). Another benefit of a quantitative assessment tool is that it can help avoid the influence of content-related judgment on the rating of fluency, since judgments of content accuracy and fluency are sometimes hard for raters to disentangle. In sum, an automatic assessment tool focusing on quantitative aspects of fluency may enhance the efficiency of assessment.

In Section 3, we present our experiment to explore the correlations between raters' subjective assessments of consecutive interpreters' fluency and: (1) content-related features such as accuracy; (2) objective measurements of fluency parameters.

3. **Methods**

3.1 Participants

Twelve students from Shanghai International Studies University participated in this study: seven third-year BA translation and interpretation majors, with a mean age of 20; and five second-year MA students, with a mean age of 25, from the Graduate Institute of Interpretation and Translation. The BA students were still working on development of basic interpreting skills, while the MA students were already working part-time as conference interpreters and were oriented towards obtaining their professional qualification as such. By the time of the experiment, the BA students had done three basic one-semester CI training courses; the MA students had completed three semesters of intensive and advanced interpreter training (at least three hours a day, covering both CI and SI). All participants had Chinese A and English B.

3.2 Material

A source audio clip in Chinese (3.5 minutes in duration, with a total of 501 Chinese characters comprising six paragraphs) was prepared from recordings of the press conference (2.5 hours) held by the former Chinese Premier Wen Jiabao during the National People's Congress in 2009. The audio clip was played to the student participants and they interpreted it consecutively into English. Of the six interpreted paragraphs, two (paragraphs 4 and 5) were selected for perceptual rating and acoustic analysis. The reason for focusing on an extended extract, rather than the whole interpretation, was that the rating task had to be performed within manageable time so as to avoid rater fatigue.

3.3 Procedures

3.3.1 *Experiment*

The experiment, originally designed and run to test an earlier hypothesis regarding improvement of judged fluency when exactly the same speech is interpreted a number of times in quick succession (Yu & van Heuven 2013), took place in conference rooms equipped with booths for simultaneous interpreting. The source stimulus material and the participants' rendition were digitally recorded on separate tracks maintaining time differences. One of the authors monitored the interpreters over headphones and ensured that all of them had finished interpreting one paragraph before the next was played to them. Participants were instructed to interpret the same source speech three times (deliveries 1, 2 and 3), paragraph by paragraph, with a break of two minutes between deliveries. Delivery 1 and delivery 3 were selected for both auditory rating and acoustic analysis. Delivery 2 was excluded, because previous studies (e.g., Zhou 2006) suggest that the third delivery is often the most proficient during oral task repetition.

3.3.2 *Fluency ratings*

The online survey software Qualtrics was used for the rating procedure. Twenty-four clips (12 interpreters × 2 interpretations each) were rated on six measures related to accuracy and fluency: (i) accuracy of information; (ii) grammatical correctness; (iii) speed of delivery; (iv) control of pauses (both silent and filled); (v) control of other disfluencies (unnecessary repetitions, false starts, inappropriate lengthening of syllables, self-corrections); and (vi) overall fluency, on a scale between a minimum of 1 and a maximum of 10. The presentation of the 24 clips was randomized for each rater individually, to prevent a potential order effect in the ratings.

Ten raters (five men, five women), with a background of studying or teaching at Leiden University, participated in the online rating: three native English speakers (two UK, one US), and seven with near-native English proficiency (six Dutch L1 and one Portuguese L1, all members of the academic staff in the English section, or PhD candidates in linguistics). The raters were informed that the entire rating session would last an hour and advised to take a ten-minute break after rating twelve clips, so as to avoid fatigue. They were then asked to complete a background survey, after which they carefully read through an English translation of the two paragraphs so that they understood the messages to be interpreted. Subsequently, audio clips of two specimen interpretations were played to them: one very good, the other less so. These were recordings of interpreters who were not actually included in the experimental sample.

The ten raters scored all the twelve subjects across two deliveries, as explained above (3.3.2). Means were calculated, to obtain one score for each single delivery on each rating measure. A total of 24 ratings was thus obtained for each of the six rating measures.

3.3.3 *Acoustic correlates of fluency*

The following measures were selected for investigation:

(1) articulation rate = number of syllables, including disfluencies, divided by total duration of speech apart from all (silent and filled) pauses longer than 0.25 seconds;
(2) speech rate = as for articulation rate, but including all pauses in the total speech duration;
(3) effective speech rate = as for speech rate, but excluding disfluencies from the syllable count;
(4) number of silent pauses above 0.25 seconds in duration;
(5) mean length of silent pauses longer than 0. 25 seconds;
(6) number of filled pauses (*uh*, *er*, *mm*, etc.);
(7) mean length of all filled pauses;
(8) number of pauses = sum of (4) and (6);
(9) mean length of pauses = mean of (5) and (7), weighted by their respective frequencies (items 4 and 6);
(10) number of other disfluencies (repetitions, restarts, false starts, corrections);
(11) mean length of fluent runs = mean number of syllables produced between silent pauses longer than 0.25 seconds;
(12) phonation/time ratio, calculated on the basis of items 4 and 6, as a percentage of overall speech time = (total duration of speech without pauses, divided by total duration of speech including pauses) × 100

The threshold of 0.25 seconds for silent pauses is often chosen to distinguish hesitation in speech (Towell, Hawkins & Bazergui 1996) from pauses that are part of normal articulation for some combinations of sounds or may be classified as micro-pauses (Pinget et al. 2014; Riggenbach 1991).

According to Tavakoli and Skehan's (2005) perspective on the temporal properties of fluency, the above acoustic measures are predictive of speed fluency (1, 2), breakdown fluency (2, 4, 5, 6, 7, 8, 9, 11 and 12), repair fluency (10), or all three categories (3).

In the present study, speech length is measured in syllables. This is in line with Pöchhacker's (1993) observation that use of the syllable as a standard international unit of measurement obviates the practical drawback caused by the sometimes considerable variability in word length across different languages. Calculation of some temporal measures (e.g., phonation/time ratio) requires a count of the transcribed syllables: in our study, this was done manually by the first author and checked by a student assistant.

For all 24 clips, the transcription was made by a graduate student assistant and checked by the first author. The transcriber was instructed to listen very carefully, noting any apparently unpronounced syllables for the purpose of the subsequent syllable count. No syllables had actually been omitted. Irrespective of this, detailed phonetic transcription and a manual syllable count would be too labour-intensive for longer speech samples. An automatic syllable count can be done by the PRAAT speech analysis software (De Jong & Wempe 2009), with the prospect of further improvements for future use in research such as this. Automatic phonetic transcription is also very likely to be facilitated as automatic speech recognition technology evolves. Currently, however, we consider that manual calculation offers the best guarantee of accuracy.

The transcription of the 24 clips included filled pauses and all types of disfluencies. Silent pauses were detected by running the MarkInterval script (developed by Jos Pacilly) on the PRAAT software: this makes it possible to convert an acoustic signal into an oscillogram and/or spectrogram, visualizing sounds as a continuous wave pattern in which any segment can be matched with the corresponding recording. At a sampling frequency of 44.1KHz, duration of different speech features can be measured in milliseconds. Together with the oscillogram, two annotation levels ('tiers' in a 'textgrid') were created for the transcribed texts and the labelled disfluencies (see Figure 1). The length of each silent and filled pause detected, the total duration and number of all pauses, and the number of disfluencies were automatically computed by routines implemented in AWK.

Silences at the very beginning and end of every delivery were discarded by the editing function in PRAAT. The selection of the variables in this study is slightly different from Cucchiarini et al.'s (2002) choice of nine temporal variables related to fluency in L2 spontaneous speech. First, a distinction was made between effective speech rate (a variable proposed by us) and speech rate in general. Effective speech rate is arithmetically calculated by excluding the syllables identified as disfluencies (e.g., involuntary repetitions), because these are very likely to be more frequent in the cognitively demanding speech task of interpreting than in unconstrained speech. Second, syllables were used as the units of measurement. Third, mean length of filled pauses was added. The rationale for this choice was that interpreting is probably more conducive than spontaneous speech to hesitation pauses, as the interpreter takes time to analyze incoming information while also planning and retrieving the components of target language production. Finally, the number and mean length of pauses were also added so as to measure overall pausing.

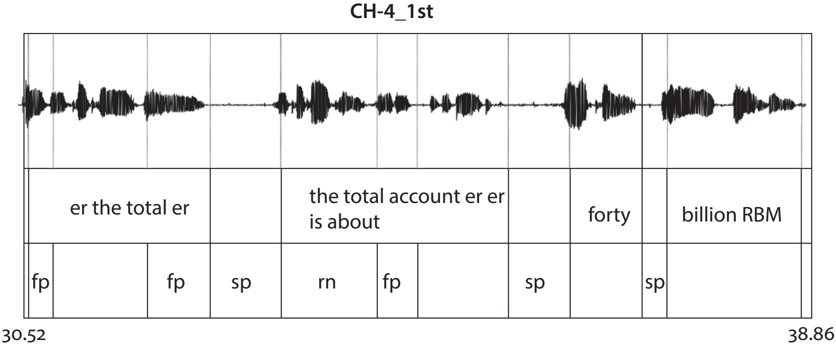

Figure 1. Visualization of a selected sound wave, with two annotation tiers in a text grid, in PRAAT

The objective fluency measures described above were calculated separately for every single interpretation (i.e., two values per subject for each parameter). A total of 24 values was thus obtained for each objective measure. Correlations between these values and the judged fluency ratings were then analyzed.

3.4 Statistical analysis

The ten raters were highly consistent in their ratings on the six measures related to accuracy and fluency, with a mean Cronbach's alpha of 0.96. The highest Cronbach's alpha was 0.97, for grammatical correctness and speed of delivery; the lowest Cronbach's alpha was 0.94, for accuracy of information. Table 1 shows the inter-rater reliability for the six fluency rating items.

Table 1. Inter-rater reliability coefficients (Cronbach's alpha) for the ten raters

Rating item	Cronbach's alpha
Accuracy of information	0.94
Grammatical correctness	0.97
Speed of delivery	0.97
Control of pauses	0.95
Control of disfluencies	0.96
Overall fluency	0.95

For the statistical analyses of the rating results and acoustic measures, Pearson's *r* and (multiple) linear regression were performed.

4. Results

The results of the accuracy and fluency ratings assigned by the ten raters are presented first, followed by those for the objective acoustic measures of fluency. Finally, the relationship between the two is studied.

4.1 Auditory ratings of fluency and accuracy

Before calculating the correlation coefficients between accuracy-related and fluency-related ratings, we first present the mean scores on the six ratings for accuracy and fluency parameters (see Table 2). Ratings differ for the beginners and the advanced students, as well as for deliveries 1 and 3. The analysis in our earlier study (Yu & van Heuven 2013) shows significant main effects of both repetition (delivery) and proficiency on perceived accuracy and fluency. This means that the advanced students are judged as being significantly more accurate and fluent than the beginners, while all subjects are judged as being significantly more accurate and fluent in delivery 3 than in delivery 1.

The correlations between accuracy-related and fluency-related ratings are shown in Table 3: both the former are closely correlated with the ratings on the four fluency-related items. All correlations are positive and highly significant ($p < 0.01$). This indicates that, for CI, ratings of fluency and accuracy are likely to influence each other.

4.2 Acoustic measures of fluency

In this section, the twelve acoustic fluency variables are calculated (see Table 4). Table 4 also shows values of the different acoustic fluency variables for delivery 1 *vs* delivery 3. The D3/D1 ratio is the mean of each acoustic variable for delivery 3, divided by that for delivery 1.

For most of the acoustic variables, scores are better in delivery 3. The only exceptions are mean length of filled pauses and phonation/time ratio, which hardly change. The number of filled pauses and number of disfluencies are halved in delivery 3. Overall, the acoustic parameters of fluency are consistent with the trends for fluency ratings in our earlier study, where both beginners and advanced students achieved significantly higher scores in delivery 3 (Yu & van Heuven 2013).

Table 2. Mean ratings for the six measures related to accuracy and fluency

		Delivery 1						**Delivery 3**					
		Accuracy		**Fluency**				**Accuracy**		**Fluency**			
Proficiency	ID	AI	GC	SD	CP	CD	OF	AI	GC	SD	CP	CD	OF
BA	1.	5.7	5.7	5.7	4.8	4.8	5.3	6.1	6.4	6.7	6.1	5.4	6.1
BA	2.	5.6	5.2	5.5	4.9	5.0	5.6	7.5	6.6	6.5	5.5	5.6	6.3
BA	3.	4.5	4.6	5.0	4.7	4.3	4.5	6.4	6.0	6.2	5.9	5.9	6.2
BA	4.	5.2	4.9	4.4	4.0	3.6	4.1	5.1	5.2	5.9	5.6	5.2	5.7
BA	5.	5.6	5.3	5.2	5.0	4.8	5.1	6.8	6.6	6.1	6.4	6.0	6.2
BA	6.	4.7	5.2	4.9	4.6	4.4	4.8	6.7	5.9	5.5	5.6	6.0	5.9
BA	7.	5.5	5.7	5.0	3.9	4.4	4.4	5.7	6.0	6.3	5.7	5.1	5.7
Mean (BA)		5.3	5.2	5.1	4.6	4.5	4.8	6.3	6.1	6.2	5.8	5.6	6.0
MA	8.	7.0	6.6	7.0	6.4	6.1	6.8	7.5	6.5	7.3	6.9	6.5	7.2
MA	9.	7.6	6.8	7.7	6.7	6.4	7.1	8.1	7.6	7.9	7.6	7.4	7.6
MA	10.	7.8	6.9	7.6	6.9	6.8	7.3	8.2	7.7	8.0	7.2	7.2	7.8
MA	11.	7.1	5.7	6.3	5.7	5.7	6.1	7.4	6.0	6.8	6.6	6.2	6.6
MA	12.	5.6	5.5	5.6	4.8	4.7	5.3	6.2	6.7	6.7	5.8	5.2	5.8
Mean (MA)		7.0	6.3	6.8	6.1	5.9	6.5	7.5	6.9	7.3	6.8	6.5	7.0
Mean (grand)		6.0	5.7	5.8	5.2	5.1	5.5	6.8	6.4	6.7	6.2	6.0	6.4

Notes. ID = identification number of participants, AI = accuracy of information, GC = Grammatical correctness, SD = speed of delivery, CP = control of pauses, CD = control of other disfluencies, OF = overall fluency

Table 3. Pearson's *r* correlations between accuracy-related and fluency-related ratings

	Speed of delivery	**Control of pause**	**Control of disfluencies**	**Overall fluency**
Accuracy of information	0.88*	0.86*	0.92*	0.92*
Grammatical correctness	0.90*	0.85*	0.86*	0.88*

Notes.
* $p < .01$ (two-tailed)

In Section 4.3, the relations between acoustic fluency measures and fluency ratings will be explored in greater detail.

Table 4. Values of 12 acoustic fluency measures for 12 subjects, in two consecutive interpretations

Variable		BA student ID						MA student ID						Mean	D3/D1
	Delivery	1	2	3	4	5	6	7	8	9	10	11	12		
Articulation rate (syll./sec.)	1	5.2	4.5	5.7	3.6	3.6	3.2	4.4	5.1	4.8	4.8	5.4	4.3	4.6	1.1
	3	5.1	3.8	4.7	4.4	4.5	4.0	5.0	5.1	5.0	5.0	5.0	6.2	4.8	
Speech rate (syll./sec.)	1	3.2	2.3	3.1	2.3	2.3	2.1	2.0	3.7	3.7	4.0	3.7	2.9	2.9	1.2
	3	3.8	2.3	3.3	3.1	3.5	2.8	3.2	4.1	3.8	3.9	3.7	3.8	3.4	
Effective speech rate (syll./sec.)	1	2.8	1.8	2.8	1.7	2.2	1.9	1.7	3.7	3.7	3.9	3.5	2.6	2.7	1.2
	3	3.3	2.2	3.3	2.9	3.4	2.8	2.9	4.1	3.7	3.8	3.6	3.4	3.3	
No. silent pauses	1	42	95	54	61	60	43	63	36	34	28	54	6	53	0.7
	3	41	59	35	32	38	30	45	27	33	29	42	54	39	
Length of silent pauses (sec.)	1	0.7	0.5	0.7	0.6	0.7	1	1.1	0.5	0.4	0.4	0.4	0.5	0.6	0.8
	3	0.4	0.6	0.7	0.5	0.4	0.8	0.7	0.4	0.5	0.5	0.4	0.5	0.5	
No. filled pauses	1	15	61	6	35	9	11	25	5	6	0	3	14	16	0.5
	3	13	40	0	9	2	3	7	5	3	0	6	11	8	
Length of filled pauses (sec.)	1	0.3	0.5	0.2	0.4	0.4	0.3	0.4	0.3	0.2	0.0	0.4	0.3	0.3	1
	3	0.3	0.4	0.0	0.5	0.4	0.3	0.2	0.3	0.3	0.0	0.3	0.3	0.3	
No. pauses	1	57	156	60	96	69	54	88	41	40	28	57	82	69	0.7
	3	54	99	35	41	40	33	52	32	36	29	48	65	47	
Length of pauses (sec.)	1	0.6	0.5	0.7	0.5	0.6	0.8	0.9	0.5	0.4	0.4	0.4	0.5	0.6	0.8
	3	0.4	0.5	0.7	0.5	0.4	0.8	0.6	0.4	0.5	0.5	0.4	0.5	0.5	
No. disfluencies	1	12	17	7	28	8	8	16	2	3	2	15	9	11	0.5
	3	14	4	3	9	3	2	7	0	3	3	4	11	5	
Length of fluent runs (in syllables)	1	4.8	2.2	4.7	3.0	4.1	4.9	3.2	6.9	6.6	9.5	5.3	4.2	5.0	1.3
	3	5.8	2.9	7.3	5.7	6.7	7.1	5.2	8.2	7.6	9.9	5.5	4.8	6.4	
Phonation/time ratio (%)	1	0.6	0.5	0.6	0.6	0.7	0.7	0.5	0.7	0.8	0.8	0.7	0.7	0.7	1
	3	0.7	0.6	0.7	0.7	0.8	0.7	0.6	0.8	0.8	0.8	0.7	0.6	0.7	

4.3 Correlations between acoustic fluency measures and fluency ratings

In this section, acoustic fluency measures are compared with fluency ratings so as to determine how, and to what extent, the two are related. Pearson's *r* results are shown in Table 5. Eleven out of the twelve acoustic measures are closely correlated with judged fluency. Effective speech rate is the most closely correlated ($r = 0.84^{**}$), followed by mean length of fluent runs ($r = 0.78^{**}$) and phonation/time ratio ($r = 0.78^{**}$).

Table 5. Pearson's *r* correlations between fluency ratings and acoustic fluency measures for 12 subjects, in two consecutive interpretations

	Accuracy of information	**Grammatical correctness**	**Speed of delivery**	**Control of pauses**	**Control of disfluencies**	**Overall fluency**
Articulation rate (syll./sec.)	0.27	0.34	0.50*	0.42*	0.33	0.36
Speech rate (syll./sec.)	0.65**	0.66**	0.83**	0.84**	0.74**	0.77**
Effective speech rate (syll./sec.)	0.72**	0.70**	0.86**	0.90**	0.82**	0.84**
No. silent pauses	−0.50*	−0.54**	−0.58**	−0.68**	−0.64**	−0.60**
Length of silent pauses (sec.)	−0.56**	−0.42*	−0.62**	−0.64**	−0.51*	−0.62**
No. filled pauses	−0.33	−0.37	−0.43*	−0.55**	−0.50*	−0.42*
Length of filled pauses (sec.)	−0.44*	−0.52**	−0.56**	−0.50*	−0.52**	−0.52**
No. pauses	−0.44*	−0.49*	−0.53**	−0.65**	−0.60**	−0.54**
Length of pauses (sec.)	−0.54**	−0.43*	−0.62**	−0.59**	−0.46*	−0.58**
No. disfluencies	−0.57**	−0.56**	−0.62**	−0.71**	−0.73**	−0.69**
Length of fluent runs (in syllables)	0.66**	0.68**	0.73**	0.82**	0.82**	0.78**
Phonation/time ratio (%)	0.68**	0.63**	0.72**	0.83**	0.77**	0.78**

Notes.
* $p < .05$
** $p < .01$ (two-tailed)

4.4 Acoustic measures as indicators of judged fluency

Several linear regression models were built in SPSS, to investigate to what extent the twelve acoustic measures of fluency could explain the variance in fluency ratings. Tables 6, 7, 8 and 9 show the adjusted proportion of variance explained (R^2) for these models, and thus the predictive power of each acoustic parameter.

First, model 1 evaluates all twelve acoustic fluency measures as predictors of fluency ratings: the adjusted R^2 shows that 78.9% of the variance in the ratings on speed of delivery may be explained on the basis of two acoustic measures – i.e., effective speech rate, with $R^2 = 72.1\%$; and number of filled pauses, with $R^2 = 6.8\%$ (Table 6). Effective speech rate appears to be the best indicator of the ratings on speed of delivery. Second, 88.2% of the variance in the ratings for control of pauses may be explained on the basis of three acoustic measures – i.e., effective speech rate, with $R^2 = 79.6\%$; articulation rate, with $R^2 = 4.5\%$; and number of filled pauses, with $R^2 = 4.1\%$ (Table 7). Again, effective speech rate appears to be the best indicator here. Third, model 3 shows that 87.6% of the variance in the ratings on control of disfluencies may be explained on the basis of four acoustic measures – i.e., effective speech rate, with $R^2 = 66.1\%$; speech rate, with $R^2 = 9.1\%$); number of filled pauses, with $R^2 = 6.6\%$; and mean length of fluent runs, with $R^2 = 5.8\%$ (Table 8). Here too, effective speech rate appears to be the best indicator. Finally, model 4 shows that 90.2% of the variance in the overall fluency ratings may be explained on the basis of four acoustic measures – i.e., effective speech rate, with $R^2 = 68.7\%$; number of filled pause, with $R^2 = 6\%$; articulation rate, with $R^2 = 11.8\%$; and mean length of pause, with $R^2 = 3.7\%$ (Table 9). Once again, effective speech rate appears to be the best indicator of the ratings on overall fluency.

Table 6. Model 1 (dependent variable: judged speed of delivery)

Predictors	**R^2**	**Adj R^2**	**increment**	**SE of estimate**
effective speech rate	.734	.721		.53479
number of filled pauses	.807	.789	.068	.46557

Table 7. Model 2 (dependent variable: judged control of pauses)

Predictors	**R^2**	**Adj R^2**	**increment**	**SE of estimate**
effective speech rate	.805	.796		.45052
articulation rate	.855	.841	.045	.39763
number of filled pauses	.897	.882	.041	.34262

Table 8. Model 3 (dependent variable: judged control of disfluencies)

Predictors	R^2	Adj R^2	increment	SE of estimate
effective speech rate	.676	.661		.55945
speech rate	.773	.752	.091	.47875
number of filled pauses	.842	.818	.066	.40961
mean length of fluent runs	.897	.876	.058	.33850

Table 9. Model 4 (dependent variable: judged overall fluency)

Predictors	R^2	Adj R^2	increment	SE of estimate
effective speech rate	.701	.687		.56339
number of filled pauses	.769	.747	.060	.50647
articulation rate	.883	.865	.118	.36983
mean length of pauses	.919	.902	.037	.31471

5. Discussion and conclusion

The study focuses on correlations between judged fluency and judged accuracy of CI, as well as between judged fluency and automatically quantifiable fluency measures. Ratings for accuracy of information and grammatical correctness were both closely correlated with the raters' judgment of four fluency-related items – i.e., speed of delivery, control of pauses, control of disfluencies, and overall fluency (Table 3). This indicates that good interpretation requires both accuracy and fluency.

With regard to the relations between judged fluency and acoustic measures of fluency, all four judged fluency criteria (speed of delivery, control of pauses, control of disfluencies, overall fluency) correlated significantly with almost all the acoustic measures of fluency: only judged control of disfluencies and judged overall fluency did not correlate significantly with articulation rate (Table 5). This suggests that there is likely to be some overlap between the four judged fluency criteria, as well as between the twelve acoustic fluency measures – e.g., it is possible that speed of delivery will be associated with both control of pauses and control of disfluencies. The overall fluency rating might be connected with the three partial criteria of fluency, while mean length of fluent runs might be related to the number of filled pauses and number of disfluencies.

The study also set out to identify which of the twelve acoustic fluency measures afforded the best prediction of fluency ratings. The results of the linear regression models show that effective speech rate (i.e., number of syllables, excluding disfluencies, divided by the total duration of speech production and pauses) appears to be the best predictor of all four judged fluency criteria in CI (Tables 6–9). The other temporal measures related (albeit less closely) to the variance of fluency ratings are number of filled pauses, articulation rate, and mean length of pause. The reason that effective speech rate appears to be the most powerful predictor of judged fluency in CI might be that effective speech rate incorporates three aspects of fluency (speed fluency, breakdown fluency and repair fluency). Given the possibility of automatically detecting syllables, pauses and disfluencies by running relevant scripts on PRAAT, there seems to be some likelihood that the labour-intensive rating of CI exams may be partly facilitated by instrumental measurement of effective speech rate as the best fluency predictor. This could be quite efficient, at least in screening out candidates who do not score satisfactorily on effective speech rate. Our study should thus serve as an initial step towards the development of an automatic quantitative assessment tool for fluency in interpreting.

One of the limitations of this study lies in the experimental setting, since no speaker or audience was present. This, of course, does not reflect the situational dynamics of CI in actual practice. Further research is needed in real-life settings, to complement the findings obtained in this study. Another limitation lies in the problem of being potentially over-categorical in labeling silent pauses and repetitions as disfluencies, since the interpreters might deliberately use these for clarity and emphasis. This needs to be taken into consideration in future studies, to distinguish those 'disfluencies' that may actually enhance communication.

Along with their practical implications for testing, the results also allow for a review of teaching practices. More specifically, what seems to emerge is that interpreting trainees need to develop a more comprehensive understanding of an integrated concept of fluency, incorporating not only speed fluency but also breakdown and repair fluency. Our experiment shows that the number of filled pauses and the number of disfluencies are halved in delivery 3 over delivery 1 (Table 4). Two aspects of fluency (i.e., breakdown fluency and repair fluency, or what Grosjean referred to as secondary fluency variables) thus seem to have a more important effect on fluency ratings in CI than was the case in previous research on L2 read and spontaneous speech (Cucchiarini et al. 2000, 2002; Kormos & Dénes 2004; Pinget et al. 2014).

It is, therefore, reasonable to see interpreting as a cognitively demanding task that entails not only analysis of the source language, but also rapid and accurate target language formulation together with close monitoring of output so as to ensure the logic and consistency of the ongoing interpretation (Mead 2000). In this regard, trainers should focus more closely on helping trainee interpreters control their breakdown fluency and repair fluency.

References

AIIC (2002). *Regulation governing admissions and language classification*. Geneva: AIIC.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F. & Brennan, S. E. (1999). Which speakers are most disfluent in conversation and when? *Proceedings ICPhS99 Satellite Meeting on Disfluency in Spontaneous Speech*, 7–10.

Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge: Cambridge University Press.

Bühler, H. (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua* 5 (4), 231–235. doi: 10.1515/mult.1986.5.4.231

Chambers, F. (1997). What do we mean by fluency? *System* 25, 535–544. doi: 10.1016/S0346-251X(97)00046-8

Chiaro, D. & Nocella, G. (2004). Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the World Wide Web. *Meta* 49 (2), 278–293. doi: 10.7202/009351ar

Collados Aís, A. (1998/2002). Quality assessment in simultaneous interpreting: The importance of nonverbal communication. In F. Pöchhacker & M. Shlesinger (Eds.), *The interpreting studies reader*. London/New York: Routledge, 327–336.

Collados Aís, A., Pradas Macías, M., Stévaux, E. & García Becerra, O. (Eds.) (2007). *La evaluación de la calidad en interpretación simultánea*: Parámetros de incidencia. Granada: Comares.

Cucchiarini, C., Strik, H. & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America* 107, 989–999. doi: 10.1121/1.428279

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. doi: 10.1121/1.1471894

De Jong, N. H. & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41 (2), 385–390. doi: 10.3758/BRM.41.2.385

Fillmore, C. J. (1979). On fluency. In C. J. Fillmore, D. Kempler & W. S. Wang (Eds.), *Individual differences in language ability and language behaviour*. New York: Academic Press, 85–101. doi: 10.1016/B978-0-12-255950-1.50012-3

Freed, B. F. (1995). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study-abroad context*. Amsterdam: John Benjamins, 123–148. doi: 10.1075/sibil.9.09fre

Gile, D. (1995). *Basic concepts and models for interpreter and translator training*. Amsterdam: John Benjamins. doi: 10.1075/btl.8(1st)

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Grosjean, F. (1980). Temporal variables within and between languages. In H. W. Dechert & M. Raupach (Eds.), *Towards a cross-linguistic assessment of speech production*. Frankfurt: Lang, 39–53.

Grosjean, F. & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'Anglais et du Francais: Vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica* 31, 144–184. doi: 10.1159/000259667

Kormos, J. & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32, 145–164. doi: 10.1016/j.system.2004.01.001

Kurz, I. (1993). Conference interpretation: Expectations of different user groups. *The Interpreters' Newsletter* 5, 3–16.

Kurz, I. (2001). Conference interpreting: Quality in the ears of the user. *Meta* 46 (2), 394–409. doi: 10.7202/003364ar

Kurz, I. (2003). Quality from the user perspective. In A. Collados Aís, M. Fernández Sanchez & D. Gile (Eds.), *La evaluación de la calidad en interpretación: Investigación*. Granada: Comares, 3–22.

Leeson, R. (1975). *Fluency and language teaching*. London: Longman.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning* 3, 387–417. doi: 10.1111/j.1467-1770.1990.tb00669.x

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency*. Michigan: The University of Michigan Press, 25–42.

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

Mead, P. (2000). Control of pauses by trainee interpreters in their A and B languages. *The Interpreters' Newsletter* 10, 89–102.

Mead, P. (2005). Methodological issues in the study of interpreters' fluency. *The Interpreters' Newsletter* 13, 39–63.

Möhle, D. (1984). A comparison of the second language speech production of different native speakers. In H. W. Dechert, D. Möhle & M. Raupach (Eds.), *Second language productions*. Tübingen: Gunter Narr, 26–49.

Moser, P. (1996). Expectations of users of conference interpretation. *Interpreting* 1 (2), 145–178. doi: 10.1075/intp.1.2.01mos

Nation, P. (1989). Improving speaking fluency. *System* 3, 377–384. doi: 10.1016/0346-251X(89)90010-9

Pinget, A., Bosker, H. R., Quené, H., Sanders, T. & De Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing* 31 (3), 349–365. doi: 10.1177/0265532214526177

Pradas Macías, E. M. (2003). *Repercusión del intraparámetro pausas silenciosas en la fluidez: Influencia en las expectativas y en la evaluación de la calidad en interpretación simultánea*. PhD dissertation, University of Granada.

Pradas Macías, E. M. (2007). La incidencia del parámetro fluidez. In A. Collados Aís, M. Pradas Macías, E. Stévaux & O. García Becerra (Eds.), *La evaluación de la calidad en interpretación simultánea: Parámetros de incidencia*. Granada: Comares, 53–70.

Pöchhacker, F. (1993). On the science of interpretation. *The Interpreters' Newsletter* 5, 52–59.

Pöchhacker, F. (2012). Interpreting quality: Global professional standards? In W. Ren (Ed.), *Interpreting in the age of globalization: Proceedings of the 8th National Conference and International Forum on Interpreting*. Beijing: Foreign Language Teaching and Research Press, 305–318.

Rennert, S. (2010). The impact of fluency on the subjective assessment of interpreting quality. *The Interpreters' Newsletter* 15, 101–115.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of non-native speaker conversations. *Discourse Processes* 14, 423–441. doi: 10.1080/01638539109544795

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition* 14, 357–385. doi: 10.1017/S0272263100011189

Tavakoli, P. & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language*. Amsterdam: John Benjamins, 239–273. doi: 10.1075/lllt.11.15tav

Towell, R., Hawkins, R. & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics* 1, 84–119. doi: 10.1093/applin/17.1.84

Townshend, B., Bernstein, J., Todic, O. & Warren, E. (1998). *Estimation of spoken language proficiency. Proceedings of the ESCA Workshop Speech Technology in Language Learning (STiLL 98)*, 179–182.

Yu, W. T. & van Heuven, V. J. (2013). Effects of immediate repetition at different stages of consecutive interpreting: An experimental study. In: S. Aalberse & A. Auer (Eds.), *Linguistics in the Netherlands 2013*. Amsterdam: John Benjamins, 201–213. doi: 10.1075/avt.30

Zhou, D. (2006). A study on the effects of input frequency and output frequency. *Modern Foreign Languages* 29, 154–163.

Zwischenberger, C. & Pöchhacker, F. (2010). Survey on quality and role: Conference interpreters' expectations and self-perceptions. *Communicate!* http://www.aiic.net/ViewPage.cfm/article2510.htm (accessed 21 January 2013).

Authors' addresses

Wenting Yu
Room 320, Building 1
550 Dalian Xi Road
Shanghai 200083
China

wenting_yu@163.com

Vincent J. van Heuven
Spanjaardslaan 111
8917 AP Leeuwarden
The Netherlands

Biographical notes

Wenting Yu teaches interpreting for the translation and interpretation majors at Shanghai International Studies University (SISU). She obtained her PhD degree in Linguistics from SISU in 2012. She was a visiting post-doctoral researcher at Leiden University Center for Linguistics (2012–2013). She specializes in interdisciplinary studies of interpreting, prosody, psycholinguistics and second language acquisition. Her publications include work on cognitive processing in interpreting, and interpreter training.

Vincent van Heuven is an emeritus professor of Experimental Linguistics and Phonetics at Leiden University, and a professor at the University of Pannonia (Veszprém, Hungary). He served as Director of the Holland Institute of Linguistics and the Leiden University Centre for Linguistics, Chairman of the Netherlands Graduate School in Linguistics, and Vice-President/Secretary of the Permanent Council of the International Phonetics Association. A former (associate) editor of several international book series and professional journals, he is a member of the Royal Netherlands Academy of Arts and Sciences.