



(Para)linguistic Correlates of Perceived Fluency in English-to-Chinese Simultaneous Interpretation

Chao Han

Department of Linguistics
Macquarie University, Australia

Received: 06-08-2015

Accepted: 14-09- 2015

Published: 31-10- 2015

doi:10.7575/aiac.ijclts.v.3n.4p.32

URL: <http://dx.doi.org/10.7575/aiac.ijclts.v.3n.4p.32>

Abstract

(Para)linguistic parameters (e.g., un/filled pauses, speech rate) are believed to underlie perceived fluency of simultaneous interpretation (SI). Little research, however, is available to ascertain whether and to what extent these (para)linguistic measures of SI fluency correlate with fluency ratings provided by human raters. This exploratory study investigates three questions: a) how nine selected (para)linguistic parameters correlate with each other, b) how the parameters correlate with rater-generated fluency ratings, and c) which parameter or a combination of parameters could best discriminate an interpreter into pre-determined groups of interpretation fluency. The major results are: a) three underlying dimensions of the perceived fluency emerged, including breakdown, speed, and repair fluency, b) speech rate, phonation/time ratio, and mean length of a run had higher correlation with the fluency ratings, and c) speech rate and phonation/time ratio were the best possible predictors of the interpreters' group affiliation. Implications of the results are discussed regarding fluency assessment in SI.

Keywords: Correlation, Disfluency, Fluency, (Para)linguistic Parameters, Simultaneous Interpretation

1. Introduction

Fluency is regarded as one of the most important quality criteria of simultaneous interpretation (SI). Altman (1994) even states that fluency could be the one single aspect of an interpretation that distinguishes students' performance from that of professional interpreters. The importance of fluency is also supported by a number of surveys in which users of interpretation services are asked to comment on which aspect(s) of interpretation is most valued by them (Bühler, 1986; Kurz, 1993). Fluency is consistently ranked by different user groups as an important criterion, after accuracy and fidelity. In actuality, when across-cultural communication hinges on interpretation, users of interpreting services can only "evaluate the simultaneously interpreted discourse by its form" (Yagi, 2000, p. 522). Usually, they tend to assess interpreting performance based on fluency, pronunciation, voice quality, etc. in target language discourse.

While fluency is an important quality of SI performance, assessing it could be challenging, as fluency is believed to be an elusive concept without broadly accepted definitions (Pradas Macías, 2006). In practice, two approaches have been used to assess fluency of SI: a) human assessor-mediated assessment using a fluency rating scale, and b) atomistic analysis of a wide range of (dis)fluencies or (para)linguistic parameters such as (un)filled pauses, repairs, and self-corrections in a given interpretation sample. Although there have been empirical studies using either of these approaches, as can be seen from the literature reviewed below, little research has been conducted to examine the relationship between different fluency-related (para)linguistic parameters on the one hand, and to ascertain the relationship between (para)linguistic parameters and actual ratings of perceived fluency, on the other. Against this background, the present study contributes some empirical data to provide a preliminary investigation into these two types of relationship.

2. Literature review

Traditionally, there have been two approaches to assessing SI fluency in interpreting literature. One approach is to ask human assessors/raters to provide a rating on SI fluency using a rating scale (Cheung, 2007; Hamidi & Pöchhacker, 2007; Lee, 2008; Liu et al., 2013; Pradas Macías, 2006). The rating scales used could be based on meaningful and pre-defined descriptors for each scale band (Lee, 2008; Liu et al., 2013). Descriptors that relate to SI fluency may include speech deviations such as "inarticulate speech, pauses, hesitation, false starts, fillers, irritating noise, repetition, excessive repairs or self-correction, unconvincing voice quality and monotonous intonation, & irritatingly slow speech rate" (Lee, 2008, p. 173), or "instances of hesitation, repetition, self-correction, and redundancy" (Liu et al., 2013, p. 177). Assessors are asked to read the descriptors for each scale band, listen to actual interpretations, and assign a rating that can best represent fluency features in a given interpretation. The fluency rating scale could also be based on loosely defined or an unspecified concept of SI fluency (Cheung, 2007; Hamidi & Pöchhacker, 2007). Pradas Macías (2006), for example, asked interpretation users to rate the overall fluency of SI using a five-point rating scale.

The other commonly used approach to assessing SI fluency is based on atomistic analysis of (para)linguistic parameters

such as (un)filled pauses, false starts, etc. (Bakti, 2009; Cecot, 2001; Mead, 2000, 2005; Pio, 2003; Rennert, 2010; Tissi, 2000; Yagi, 2000). In other words, each (para)linguistic parameter that is believed to influence the perceived fluency is identified in a given interpretation sample and analyzed separately. Yagi (2000), for instance, observes that fluency can be studied quantitatively, if elements that contribute to a seemingly effortless, fluid, and smooth interpretation are identified. A diverse array of (para)linguistic parameters has been proposed as potential contributors to perceived fluency, including vowel or consonant lengthening, glottal click, (un)filled pauses, duration of pauses, speech rate, phonation/time ratio, articulation rate, mean length of a run, repetitions, false starts, and corrections/repairs (Mead, 2005; Pio, 2003; Rennert, 2010; Tissi, 2000).

Fluency-related parameters have been examined in numerous empirical studies that fall into three categories. In the first category, descriptive studies are conducted to account for disfluencies occurring in SI (Bakti, 2009; Mead, 2005). Studies of this type generally draw upon a corpus of interpretation and provide a descriptive analysis of a wide range of disfluency parameters such as (un)filled pauses, false starts, repetitions, etc. in speech samples. The second category features those studies that investigate how the features of source-language (SL) texts affect fluency in target-language (TL) output (Cecot, 2001; Mead, 2000; Pio, 2003; Tissi, 2000). Both Cecot (2001) and Pio (2003), for example, examined how a change of delivery speed in SL texts would impact on fluency of TL output measured by (un)filled pauses, corrections, and false starts. It was found that an increase of delivery rate of SL input generally contributed to more disfluency features in TL output. In another example, Mead (2000) found that in Italian/English bi-directional SI filled pauses and total pauses occurred more frequently in English than Italian interpretations, and that this trend achieved statistical significance. These results suggest that output was more fluent in the A language (i.e., Italian) than the B language (i.e., English). The third category of research pertains to the relationship between SI fluency and other variables (Pradas Macías, 2006; Rennert, 2010). Rennert (2010), for instance, reported a potential link between the perceived fluency of an interpretation and users' assessment of the interpreter's accuracy.

Although both rating scale-based assessment and atomistic analysis are used to assess SI fluency, there has been little research that investigates whether meaningful relationships exist between (para)linguistic measures of fluency and rater-generated ratings of perceived fluency. The present study thus draws upon an empirical dataset to provide tentative answers to this question.

3. Background to the present study

This study represents a follow-up of a factorial experiment (Han & Riazi, 2015) which investigated the effects of speakers' delivery speed and accent on three quality measures of English-to-Chinese SI. In the experiment, two independent variables (IVs) were manipulated, namely speech rate and accent, to vary on two levels, respectively. Specifically, there were a fast speech rate of about 155 wpm and a slow rate of approximately 105 wpm; there were also two native English speakers: one non-accented speaker from Australia, and one strongly accented speaker from India. These two IVs were fully crossed with each other, producing four treatment conditions: slow and non-accented (SN), slow and accented (SA), fast and non-accented (FN), and fast and accented (FA). Four SL texts were subsequently developed to comply with the four conditions. The texts were based on four authentic speeches on the general topic of Australia-China relation, which are delivered by Australian government officials. The texts were also made comparable regarding length (i.e., about 1250 words each text), lexical complexity, and propositional density, except for the IVs. Thirty-two Beijing-based interpreters were recruited to perform SI in the four tasks (i.e., Task_{SN}, Task_{SA}, Task_{FN}, and Task_{FA}), producing a total of 128 interpretations. Each interpretation was then assessed by nine trained raters using eight-point descriptor-based rating scales on three dimensions: a) information completeness (InfoCom), b) fluency of delivery (FluDel), and target language quality (TLQual).

Of particular interest in the present study are the FluDel ratings. Specifically, the raters were asked to evaluate the FluDel of the interpretations, based on the extent to which speech disfluencies (i.e., un/filled pauses, long silence, self-corrections) are present in a given interpretation sample. A generalizability (G) analysis revealed that the G coefficient (ρ^2) for the FluDel ratings was 0.89, lower than the other two dimensions: InfoCom ($\rho^2 = 0.92$), and TLQual ($\rho^2 = 0.90$) (Han, 2014). In addition, the FluDel ratings were used to categorize the interpreters into two groups: a) a more fluent (i.e., the total FluDel ratings in the four tasks ≥ 20), and b) a less fluent group (i.e., < 20).

Given the availability of the rater-generated FluDel ratings, the author decided to examine how the FluDel ratings correlate with (para)linguistic parameters. By doing so, the author could ascertain the relationship between the perceived fluency ratings and atomistically analyzed (para)linguistic parameters, and gain an insight into whether the raters appropriately used the scale descriptors. Such is the background to the present study.

4. Research questions

Given the literature review and the background provided above, a total of nine (para)linguistic parameters were selected for analysis in the study: a) the number of unfilled pauses, b) mean length of unfilled pauses, c) the total number of pauses, d) speech rate, e) phonation/time ratio, f) mean length of a run, g) the number of false starts, h) the number of reformulations, and i) the number of replacements. A detailed definition for each parameter is provided in the *Method* section below. The statistical analyses conducted in the study aim to provide preliminary answers to the following three research questions (RQ):

RQ1: Across the four tasks, is there any relationship between the selected (para)linguistic parameters?

RQ2: How and to what extent do the (para)linguistic parameters correlate with the perceived FluDel ratings across the tasks?

RQ3: Which (para)linguistic parameter or a combination of parameters could best discriminate an interpreter into the two fluency groups determined *a priori*?

5. Method

5.1 Operational definitions of the selected (para)linguistic parameters

In the present study, an unfilled pause in SI was defined as a silence of 0.5 second or greater, which falls within the range of 0.25 to 2 seconds of the cut-off criterion used in interpreting literature (e.g., Mead, 2005; Rennert, 2010; Tissi, 2000). The number of unfilled pauses (NUP) was calculated per minute. The mean length of unfilled pauses (MLUP) was the average duration (in second) of all unfilled pauses identified in an interpretation sample. The total number of pauses (TNP) was the total frequency count of both unfilled and filled (e.g., fillers such as “em”, “ah”) pauses. Speech rate (SR) was the total number of Chinese characters/syllables per minute. Phonation/time ratio (PTR) referred to the percentage of time spent speaking as a percentage proportion of the time taken to produce a speech sample. Mean length of a run (MLR) was defined as the mean number of Chinese characters/syllables between pauses. Drawing upon Skehan and Foster (1999), a false start was utterances that are abandoned before completion; in a reformulation, part of phrases or clauses was repeated with some modification either to syntax or word order; and a replacement was a complete substitution of lexical items for another. As such, the number of false starts (NFS), the number of reformulations (NRef) and the number of replacements (NRep) referred to the total number of occurrences of false start, reformulation, and replacement per minute, respectively.

5.2 Analysis of recorded interpretations

The analysis of the interpretation recordings was conducted using the software of *Cool Edit Pro 2.0*. This software can convert an acoustic signal/input into an oscillogram, providing a visual display of input sounds as a continuous wave pattern. At a sampling frequency of 44kHz, duration of different speech features (e.g., silence, pauses) can be measured in hundredths of a second.

In the study, the middle part (i.e., approximately three minutes) of each interpretation recording was subjected to analysis, leaving the first and the last one-third of each recording unanalyzed. This arbitrary decision was made, largely because the middle part may better represent the interpreters' performance, considering that the interpreters may need some warm-up practice in the beginning, and suffer fatigue by the end of the interpretation. The length of the speech sample (i.e., about three minutes) is acceptable, given the exploratory nature of the study. As a result of this initial analysis of the speech samples, frequency counts and relevant measures were generated for each interpreter in each task. Table 1 presents the descriptive statistics for the (para)linguistic parameters and the FluDel ratings.

Table 1. Descriptive statistics for the (para)linguistic parameters and the FluDel ratings

Parameters	Descriptive statistics $M(SD)$			
	Task _{SN}	Task _{SA}	Task _{FN}	Task _{FA}
NUP (p/m)	10.93 (3.2)	13.26 (3.10)	10.01 (3.30)	11.05 (3.32)
MLUP (s)	1.14 (0.28)	1.37 (0.39)	1.13 (0.36)	1.23 (0.39)
TNP (p/m)	13.4 (3.6)	15.73 (3.63)	12.54 (3.67)	13.53 (3.22)
SR (sps)	3.47 (0.59)	3.05 (0.61)	3.97 (0.68)	3.83 (0.73)
PTR	0.77 (0.10)	0.68 (0.11)	0.79 (0.11)	0.75 (0.11)
MLR (w)	17.1 (8.44)	12.34 (4.31)	20.38 (9.65)	17.53 (6.54)
NFS (p/m)	2.67 (1.53)	2.30 (1.47)	3.41 (1.55)	2.80 (1.72)
NRef (p/m)	1.79 (1.37)	1.55 (1.18)	2.20 (1.33)	1.59 (1.13)
NRep (p/m)	0.88 (0.66)	0.76 (0.71)	1.21 (0.92)	1.22 (0.98)
FluDel ratings	4.74 (1.05)	4.07 (0.99)	5.10 (0.95)	4.34 (0.98)

5.3 Statistical analysis

To answer RQ1, bi-variate correlational analysis was conducted to compute Pearson's r for each possible pair of (para)linguistic parameters across the four tasks. Correlation matrix was then examined to detect potential patterns. To answer RQ2, correlational analysis was again carried out to produce correlation coefficients between each (para)linguistic parameter and the FluDel ratings for each task. To answer RQ3, a stepwise discriminant analysis was performed to explore which parameter or a set of parameters could best predict a fluency group an interpreter belongs to. A stepwise analysis was used, primarily because of the exploratory nature of the study. All statistical analyses were run using IBM/SPSS 21. Statistical significance at three alpha levels (i.e., $\rho < 0.1$, $\rho < 0.05$, and $\rho < 0.01$) was reported.

6. Results

6.1 Relationship between (para)linguistics parameters across the tasks

Table 2 shows the bivariate correlation coefficients between the nine (para)linguistic parameters in each task. It is noted here that only part of the correlation coefficients was presented in order to accentuate prominent patterns. As can be seen in Table 2, a consistent pattern of correlation emerged across the tasks. In general, there was positive correlation within the three sets of (para)linguistic parameters: a) NUP, MLUP and TNP, b) SR, PTR and MLR, and c) NFS, NRef and NRep, with each set encapsulated by dotted lines. In each set, most of the correlations were moderate or relatively

strong, ranging from about 0.50 to 0.90, and achieved statistical significance at different p levels. These results indicate that there was a positively linear relationship between the (para)linguistic parameters in each set. In other words, any two measures within the set of a), b) or c) vary in an approximately same manner.

Another group of correlations that showed relatively strong relationship concerned the two sets of parameters: a) NUP, MLUP and TNP versus b) SR, PTR and MLR. Despite statistically significant correlation in all cases, the linear relationship was negative, suggesting that an increase in NUP, MLUP, or TNP would be accompanied by a drop in SR, PTR or MLR. The last group of correlations of interest pertained to a) NUP, MLUP and TNP versus c) NFS, NRef, and NRep. It appears that most of the correlations were weak, and they did not display a meaningful pattern, as some of them were positively correlated, and others negatively related.

Table 2. Patterns of correlation between the (para)linguistic parameters

Task ID	NUP	MLUP	TNP	SR	PTR	MLR	NFS	NRef	NRep
Task_{SN}									
NUP	1	.56***	.67***	-.60***	-.83***	-.762***			
MLUP		1	.34*	-.65***	-.88***	-.52***			
TNP			1	-.52***	-.60***	-.88***			
SR				1	.73***	.74***	.18	.17	.08
PTR					1	.73***	.24	.25	.05
MLR						1	.18	.22	-.04
NFS							1	.90***	.45***
NRef								1	.03
NRep									1
Task_{SA}									
NUP	1	.08	.71***	-.38**	-.66***	-.67***			
MLUP		1	-.07	-.66***	-.77***	-.41**			
TNP			1		-.47***	-.74***			
SR				1	.73***	.79***	.36**	.57***	-.21
PTR					1	.75***	.22	.34	-.09
MLR						1	.20	.42**	-.27
NFS							1	.88***	.62***
NRef								1	.017
NRep									1
Task_{FN}									
NUP	1	.59***	.82***	-.65***	-.84***	-.80***			
MLUP		1	.36**	-.56***	-.90***	-.48***			
TNP			1	-.54***	-.67***	-.87***			
SR				1	.70***	.75***	.39**	.41**	0.06
PTR					1	.71***	.45**	.43**	.13
MLR						1	0.09	.19	-.13
NFS							1	.81***	.52***
NRef								1	.03
NRep									1
Task_{FA}									
NUP	1	.39**	.76***	-.66***	-.79***	-.84***			
MLUP		1	.13	-.69***	-.85***	-.49***			
TNP			1	-.40**	-.54***	-.82***			
SR				1	.80***	.81***	.35	.42**	.14
PTR					1	.79***	.11	.17	.02
MLR						1	.06	.13	-.01
NFS							1	.84***	.78***
NRef								1	.31*
NRep									1

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

6.2 (Para)linguistic correlates with perceived fluency across the tasks

Table 3 summarizes the correlational results between each of the nine (para)linguistic parameters and the FluDel ratings in each task. As can be seen in Table 3, the correlation between the first three parameters (i.e., NUP, MLUP and TNP) and the FluDel ratings were negative, which makes conceptual sense, given that these parameters pertain to disfluency features that hinder fluency of SI. These correlations were also weak overall. The next three parameters (i.e., SR, PTR,

and MLR) displayed a weak to moderate positive relationship with the FluDel ratings across the tasks. Except for TaskSA, these correlations were statistically significant at different alpha levels. The last three parameters (i.e., NFS, NRef, and NRep) were negatively related to the FluDel ratings overall. But the strength of these correlations was very weak across the tasks. Taken together, it seems that the second set of parameters, namely SR, PTR, and MLR, could be better predictors of the FluDel ratings.

Table 3. Correlation between the (para)linguistic parameters with the fluency ratings

Parameters	Ratings of FluDel			
	Task _{SN}	Task _{SA}	Task _{FN}	Task _{FA}
NUP	-.30*	-.01	-.36**	-.25
MLUP	-.29*	-.36**	-.22	-.62***
TNP	-.38**	.10	-.37**	-.26
SR	.49***	.23	.46***	.55***
PTR	.42**	.23	.32*	.55***
MLR	.29*	.10	.40**	.49***
NFS	-.001	-.07	-.21	-.23
NRef	.09	.06	-.07	-.09
NRep	-.20	-.27	-.25	-.29

Note: *** $\rho < 0.01$, ** $\rho < 0.05$, * $\rho < 0.1$

6.3 Discriminant analysis: (Para)linguistic parameters

As has been explained in the *Method* section, a stepwise discriminant analysis was conducted for each task to explore which (para)linguistic parameter or a set of parameters could best discriminate the interpreters into two fluency groups (i.e., FluDel rating ≥ 20 and < 20). Table 4 summarizes the results of the discriminant analysis. As can be seen in Table 4, out of the nine potential predictors (i.e., the nine parameters), PTR and SR could best predict to which group an interpreter belongs. Particularly, PTR was selected as the best possible predictor in Task_{SN}, Task_{SA}, and Task_{FA}, and SR was a better predictor in Task_{SN}.

Table 4. The discriminant analysis results

Task ID	Selected parameter(s)	Wilks' Lambda				Exact F			
		Lambda	df ₁	df ₂	df ₃	Statistic	df ₁	df ₂	ρ
Task _{SN}	PTR	0.81	1	1	30.00	6.89	1	30.00	**
Task _{SA}	PTR	0.78	1	1	30.00	8.39	1	30.00	***
Task _{FN}	SR	0.86	1	1	30.00	4.89	1	30.00	**
Task _{FA}	PTR	0.77	1	1	30.00	9.02	1	30.00	***

Note: *** $\rho < 0.01$, ** $\rho < 0.05$, * $\rho < 0.1$

7. Discussion

The results from the analysis of relationship between (para)linguistic parameters suggest that although the perceived fluency of SI is often regarded by interpreting researchers as a unitary variable, and could be measured by different (para)linguistic parameters, it may be best represented as a multi-dimensional concept. As the correlational patterns in Table 2 show, there could be three underlying dimensions: a) *breakdown fluency*, represented by NUP, MLUP and TNP, b) *speed fluency*, relating to SR, PTR and MLR, and c) *repair fluency*, attributable to NFS, NRef, and NRep (see De Jong & Hulstijn, 2009). It was also shown that not all sub-dimensions contributed equally to the perceived fluency, as demonstrated by the correlational analysis of the (para)linguistic parameters and the fluency ratings. The results of the analysis show that the rater-generated ratings were negatively correlated with NUP, MLUP, and TNP, and positively related to SR, PTR, and PTR, in a statistically significant way. Although the raters were instructed to assess interpretation based on breakdown fluency or the disfluency features such as (un)filled pauses and corrections, there were higher absolute correlation coefficients between the FluDel ratings and the speed fluency parameters such as SR, PTR, and MLR across most of the tasks. This result suggests that the speed fluency variables (e.g., speech rate) were more likely to be the factors that influence raters' evaluation. Compared to assessing disfluency features, the raters may find it more attractive and less cognitively taxing to assess speed fluency. The dilemma in which the raters were asked to provide ratings based on breakdown fluency on the one hand, and used speed fluency parameters on the other may account for the relatively lower generalizability coefficient than the other quality dimensions. It seems that the raters did not consistently apply assessment criteria to evaluate the FluDel, thus causing more variability. The results from the discriminant analysis seem to support that a closer relationship exists between the speed fluency parameters (i.e., SR, PTR, and MLR) and the perceived fluency ratings. Particularly, phonation/time ratio was selected as the best possible

predictor of the interpreters' group affiliation in the three tasks. This result indicates that the raters may evaluate the FluDel based on how long an interpreter actually speaks in relation to the total time involved in interpreting.

These results have some implications for how SI fluency can be assessed, particularly when a descriptor-based rating scale is used by human raters. Given that the speed fluency variables such as speech rate, mean length of a run, and phonation/time ratio seem to be more appealing to raters, and can better predict interpreters' affiliation to a fluency group, scale descriptors may need to be re-constructed accordingly. Instead of anchoring scale descriptors on breakdown and/or repair fluency parameters, it could be more relevant and meaningful to align definitions of perceived SI fluency to speed fluency variables. However, it should be pointed out that the current sample of interpretation is taken from a group of practicing interpreters. As such, the results may not be applicable to the assessment of student interpreters.

8. Limitations and conclusion

Given the exploratory nature of the study, the results reported here need to be used with caution. Specifically, there was only one coder (i.e., the author) involved in analysis of the (para)linguistic parameters. Consequently, inter-coder reliability could not be checked. In addition, the small sample used may lead to unstable results of the stepwise discriminant analysis. Despite these limitations, the study affords a tentative insight into the relationship between (para)linguistic parameters and perceived fluency ratings. It identifies three potential underlying dimensions of fluency: breakdown, speed, and repair fluency. It also reveals that the speed fluency parameters are more likely to be related to perceived fluency ratings. Future studies could use larger and heterogeneous samples to verify the identified relationships in this study.

References

- Altman, J. (1994). Error analysis in the teaching of simultaneous interpretation: A pilot study. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the Gap. Empirical Research in Simultaneous Interpretation*. Philadelphia/Amsterdam: John Benjamins, 25-38.
- Bakti, M. (2009). Speech disfluencies in simultaneous interpretation. [Online] Available <http://www.kuleuven.be/cetra/papers/papers.html> (March 6, 2014)
- Bühler, H. (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua*, 5-4, 231-235.
- Cecot, M. (2001). Pauses in simultaneous interpretation: A contrastive analysis of professional interpreters' performances. *The Interpreters' Newsletter*, 11, 63-85.
- Cheung, A. K-F. (2007). The effectiveness of summary training in consecutive interpreting (CI) delivery. *Forum*, 5 (2), 1-23
- De Jong, N. & Hilstijn, J. (2009). Relating ratings of fluency to temporal and lexical aspects of speech. [Online] http://www.ealta.eu.org/conference/2009/docs/saturday/deJong_Hulstijn.pdf (December 15, 2013)
- Han, C. (2014). Investigating score reliability in English/Chinese interpreter certification performance testing: A generalizability theory approach. Manuscript submitted for publication.
- Han, C. & Riazi, M. (2015). A partial replication study to investigate the effects of speech rate and accent on simultaneous interpretation: A mixed-methods approach. Manuscript submitted for publication.
- Hamidi, M. & Pöchhacker, F. (2007). Simultaneous consecutive interpreting: A new technique put to the test. *Meta*, 52 (2), 276-289.
- Kurz, I. (1993). Conference interpretation: Expectations of different user groups. *The Interpreters' Newsletter*, 5, 13-21.
- Lee, J. (2008). Rating scales for interpreting performance assessment. *The Interpreter and Translator Trainer*, 2 (2), 165-184.
- Liu, M. (2013). Design and analysis of Taiwan's interpretation certification examination. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting*. Frankfurt: Peter Lang, 163-178.
- 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.
- Pio, S. (2003). The relation between ST delivery rate and quality in simultaneous interpretation. *The Interpreters' Newsletter*, 12, 69-100.
- Pradas Macías, M. (2006). Probing quality criteria in simultaneous interpreting: The role of silent pauses in fluency. *Interpreting*, 8 (1), 25-43.
- Rennert, S. (2010). The impact of fluency on the subjective assessment of interpreting quality. *The Interpreters' Newsletter*, 15, 101-115.
- Skehan, P. & Foster, P. (1999). The influence of task structure and processing conditions on narrative retelling. *Language Learning*, 49 (1), 93-120.
- Tissi, B. (2000). Silent pauses and disfluencies in simultaneous interpretation: A descriptive analysis. *The Interpreters' Newsletter*, 10, 103-127.
- Yagi, S. M. (2000). Studying style in simultaneous interpretation. *Meta*, 45 (3), 520-547.