

To be published in *TESOL Quarterly* (Wiley-Blackwell)



Exploring the dynamic nature of second language listeners' perceived fluency: A mixed-methods approach.

Viktoria Magne, Shungo Suzuki, Yui Suzukida, Meltem Ilkan, Mai Tran, & Kazuya Saito

Keywords: fluency, pronunciation, perception, mixed methods, L2 listeners

Despite the status of English as the Lingua Franca of the world (Seidlhofer, 2011) and the growing number of L2 users (Pennycook, 2017), researchers have mostly investigated L2-accented English based on native speakers' perceptions. In particular, a number of previous studies have looked at native English speakers' perceptions of fluency (e.g., Bosker, Pinget, Quené, Sanders, & de Jong, 2013). Only a limited number of studies have examined L2 speakers' perceptions of the same phenomenon (for a rare exception, see Rossiter, 2009). Therefore, we know very little about how L2 users conceptualize fluency, and what their subjective conceptualizations of fluency are. Thus, the current study took the first step to investigate the factors affecting L2 users' intuitive perceptions of L2 fluency using an explanatory sequential mixed-methods design.

Background

Disentangling Fluency

The importance of fluency has been recognized in successful L2 communication and language assessment contexts. Thus, previous studies have investigated which utterance features can predict fluency judgements by native speakers of target languages. Prior research has commonly reported that NSs' perception is associated largely with speed and breakdown fluency and secondarily with repair fluency (e.g., Bosker et al., 2013). However, prior research has exclusively employed native speakers as fluency judges while a number of studies have reported that L2 English communication takes place more frequently between L2 users than between L1 and L2 users. For the sake of the ecological validity of research findings, fluency research should be extended by examining L2 users' fluency judgements (Rossiter, 2009).

Qualitative Approach to L2 learners

Despite being limited in number, qualitative methods have uncovered affective and perceptual aspects of L2 learners. For example, Derwing, Munro and Thomson (2008) conducted individual interviews with 32 L2 English speakers in Canada (L1 Chinese and Russian/Ukrainian). Their results show that affective-cognitive, social, and motivational aspects play a role in L2 speakers' self-reported willingness to communicate in English, suggesting the importance of various aspects of L2 speech production. Regarding L2 learners' perception of fluency (the focus of this study), Rossiter's (2009) qualitative findings point towards temporal features of speech such as pausing and self-repetitions rather than their social perceptions. The raters in Rossiter (2009) also negatively reacted to non-temporal features of speech during their fluency judgements including pronunciation, vocabulary, and perceived confidence. Notably, both novice and expert raters gave a lot of negative comments on lexical errors perhaps due to their strong awareness of lexical choices and appropriateness than that of L2 speakers. Similarly, Tavakoli and Hunter (2018) explored teachers' understanding of fluency using a mixed-methods approach. The results demonstrated that teachers define fluency as an overall speaking ability; teachers conflated the distinction between fluency and speaking ability. Hence, their qualitative data showed that classroom teachers may address fluency in an incompatible way with findings from fluency research which commonly specifies fluency as temporal performance of speech.

Our precursor research (Saito, Ilkan, Magne, Tran, & Suzuki, 2018) quantitatively explored a set of utterance fluency measures that differentiated between low, mid, and high levels of fluency judged by native speakers of British English in the context of picture narrative speech produced by 90 Japanese learners of English and 10 native speakers of Canadian English. The results showed that articulation rate differentiated high and nativelike

fluency, and that mid-clause pause ratio differentiated between mid and high levels while final-clause pause ratio distinguished low and mid levels of fluency. In this extension study, we recruited 10 L2 users of English in London and asked them to rate the same speech samples in Saito et al. (2018). As such, we aimed to explore how utterance fluency features are associated with L2 listeners' perceived fluency of L2 picture description speech. We also added a qualitative dimension to elicit L2 listeners' subjective perspectives on fluency. Therefore, the following research questions were formulated:

1. How are utterance fluency features associated with L2 listeners' fluency ratings?
2. Which aspects of perceived fluency do L2 listeners name as influencing their fluency ratings?

Methodology

Speech Samples

A total of 100 samples came from Saito et al. (2018). Out of 100, 90 samples were provided by L1 Japanese speakers with varied proficiency levels from inexperienced learners (Length of Residence [LOR] = 0 years), experienced learners (LOR < 5 years) to attainers (LOR > 6 years), following Trofimovich and Baker's (2006) categories. Additional 10 samples were provided by L1 Canadian English speakers recruited in Vancouver to provide a native-level baseline for the ratings.

Task Procedure

To elicit their spontaneous speech, all speakers performed a timed picture description (Saito et al., 2018 for details of the task format and procedure). Following previous research in L2 fluency (Derwing, Munro, Thomson, & Rossiter, 2009), participants described seven pictures with five seconds of planning. Out of seven pictures, the last three were used for the rating and linguistic analysis. Each picture was presented with three keywords to facilitate language production for less proficient speakers. The first 10 seconds of the last three picture descriptions were extracted using Praat (Boersma & Weenink, 2012) and combined into a single WAV file totaling 30 seconds.

Listeners

A total of 10 L2 users of English were voluntarily recruited. The listeners came from different European language backgrounds (Russian = 1; Spanish = 4; Hungarian = 2; French = 2; Ukrainian = 1). Their age ranged from 30 to 38 ($M_{\text{age}} = 32.9$). They were all skilled professionals working in London at the time of data collection. They all came to the UK after the age of 16 but demonstrated a high level of fluency in English as evidenced from their self-reports in the interview and from the first author's assessment. They were born outside of English-speaking countries and learned English as a foreign language in a classroom setting. None of them had studied Japanese prior to participating in the project. Their reported relatively low familiarity with Japanese-accented English ($M = 1.5$, $Range = 1-4$) on a 6-point scale (1 = *not at all*, 6 = *very much*).

Rating procedure

Following prior research (e.g., Derwing et al., 2009), the listeners first were given a brief definition of fluency (i.e., the flow and smoothness of speech). They then proceeded with the practice rating of three samples to familiarize themselves with the range of the dataset and rating procedure. After they completed the practice task and verbally confirmed their understanding of the rating procedure with the researcher, they moved on to evaluate the

100 samples on a 9-point scale (1 = *not fluent*, 9 = *very fluent*). The speech samples were presented in a randomized order using Praat.

Inter-rater agreement. The results of the Cronbach alpha analyses confirmed the high inter-rater agreement of our 10 raters' fluency judgements ($\alpha = .97$) in line with previous fluency studies (e.g., Bosker et al., 2013 for $\alpha = 0.97$). Therefore, their judgements were averaged to compute fluency scores for each speech sample.

Interview Procedure

After completing the rating session, the listeners were invited to the follow-up semi-structured interviews. All participants agreed to being recorded and cited for the purposes of this study. The interview questions covered their perceptions of fluency during the rating task and, more specifically, the features they were attending to while differentiating the level of fluency of speech. To facilitate the interview process, the participants were given terminologies that could help them articulate their experiences: speed (not too fast/slow), breakdown (pause frequency within/between sentences or clauses), and repair (repetitions). The first author conducted the interviews, using the above areas as guidelines with some flexibility and spontaneity. All responses were audio-recorded. The interviews varied in length from 8 to 25 minutes ($M = 16.5$).

Data Analysis

Utterance fluency analysis. In line with previous research, all speech samples were transcribed into Analysis of Speech Units (Foster, Tonkyn, & Wigglesworth, 2000). Building on Kormos's (2006) utterance and cognitive fluency model, the speech samples were then coded for three dimensions of utterance fluency by three trained researchers. For breakdown fluency, the number of filled and unfilled pauses (> 250 ms: Bosker et al., 2013) in the middle and the end of clauses were divided by the total number of words. For speed fluency, the mean number of syllables per second was computed over the total phonation time. Repair fluency was obtained through the mean number of repetitions and self-corrections over the total number of words (for descriptive statistics, see Table 1).

Interpretative phenomenological analysis. The interview data were transcribed and analyzed by the first author using interpretative phenomenological analysis (IPA; Smith, Flower, & Larkin, 2009). At the initial stage of data analysis, exploratory comments were put in the left margin. Next, initial themes were identified and recorded in a separate document. Afterwards, the connections between the initial themes were identified and grouped to form superordinate themes. A table of themes with superordinate and subordinate themes was then produced. The same process was repeated for the remaining nine interviews. Following Isaacs and Trofimovich (2012), the subordinate themes were then transformed into quantitative data by calculating word frequency for each theme. By combining the quantitative analysis of speech and the listeners' fluency scores with qualitative comments, we aimed to capture the dynamic nature of L2 listeners' perceived fluency.

Results

Quantitative Results

Initially, a set of Pearson correlational analyses was performed to explore the relationship between L2 users' fluency judgements (perceived fluency) and utterance fluency features. As indicated in Table 1, perceived fluency was correlated strongly with articulation rate (speed) and mid-clause pause ratio (breakdown) but weakly with final-clause pause ratio (breakdown) and repetition ratio (repair). To further investigate the relative weights of

utterance fluency measures in perceived fluency, we ran a stepwise multiple regression analysis. As summarized in Table 2, the regression model included only articulation rate and mid-clause pause ratio, accounting for 50.9% of the total variance. The model shows that L2 listeners' fluency perception was predicted primarily by speed of delivery and secondarily by clause-internal pausing. According to Plonsky and Ghanbar's (2018) field-specific benchmark, the explained variance could be considered relatively large ($R^2 > .50$), suggesting that our L2 listeners greatly relied on temporal information (speed, breakdown) in their fluency judgements.

Qualitative results

After carrying out IPA, the overarching theme of equating fluency with proficiency in English emerged with three superordinate themes, which the participants believed affected their judgements of fluency: (a) temporal factors; (b) non-temporal factors; and (c) social factors.

The first theme, similar to Rossiter (2009), included subcategories of speed and breakdown fluency, which divided speech samples into two broad categories: speech with the optimal speed of delivery and with excessive pausing. For instance, Participant 1 described 'that was difficult to follow because the gaps were too long.' However, this theme did not account for the majority of the comments related to fluency ($N = 96$ out of 339) while the participants received a definition of fluency as a temporal construct prior to the rating. Notably, there were few comments related to repair fluency as a factor influencing L2 raters' fluency judgements whereas it was found to be significantly correlated with their numerical ratings.

Theme 2 and 3 received more comments from the participants ($N = 113, 130$, respectively). The second theme looked at lexical richness, grammatical accuracy and pronunciation in general. Participants' narratives indicated that standard grammar and varied or sophisticated vocabulary were indicative of high fluency whereas 'struggling to get the right words' (Participant 3), non-standard grammar and not having 'good pronunciation' (Participant 2) signalled low-level fluency and by extension low proficiency in English.

Participants' comments revealed strong emphasis on social factors and their effect on their perceptions of speech. The third theme encompassed L2 users' experiences of Standard English, evoking the notions of high socio-economic status, gender stereotypes, L1 English speaker superiority, and the subsequent inferiority of L2 accents. Participants expressed their internal points of reference to native speakers of English as the benchmark for fluency judgements in a number of comments. As Participant 10 summarized, 'some of them are native speakers but speak very slowly, and some of them speak very very fast. And both of them, I would consider them as being fluent, obviously.' Consequently, participants would assign higher fluency ratings to speech samples that sounded natively like to them.

To further explore the qualitative findings, a 13-category coding scheme was developed under the three main themes identified using IPA. Frequency of coded categories from L2 raters' transcribed interviews are summarized in Table 3. The main challenge was to disentangle the categories of 'Natively like/Not natively like' and 'Pronunciation'. For example, 'then it's pronunciation that's the how you pronounce consonants and vowels' was classed under non-temporal factors whereas 'fluent means um as close as to native, with the accent or obviously, it would be the best to speak to native speakers' were classed under Social factors

DYNAMIC NATURE OF L2 FLUENCY

Table 1

Descriptive Statistics and Correlational Coefficients of Utterance Fluency Measures and Perceived Fluency

	<i>M</i>	<i>SD</i>	2		3		4		5		6	
			<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Perceived fluency												
1. Perceived fluency	5.04	2.10	0.665**	<.001	-0.339**	<.001	-0.607**	<.001	0.218*	0.029	0.135	0.182
Utterance fluency												
2. Articulation rate	3.03	7.02	–	–	-0.399**	<.001	-0.574**	<.001	0.192	0.056	0.076	0.451
3. Final-clause pause ratio	0.14	0.10			–	–	0.089	0.381	-0.276**	0.005	-0.092	0.364
4. Mid-clause pause ratio	0.18	0.15					–	–	-0.29**	0.003	-0.094	0.352
5. Repetition ratio	0.98	0.04							–	–	0.425**	<.001
6. Self-correction ratio	0.98	0.04									–	–

Note. * indicates $p < .05$, ** indicates $p < .01$

Table 2.

Results of Stepwise Multiple Regression Analysis Using Utterance Fluency Measures as Predictors of Perceived Fluency

Outcome variable	Predictor variables	<i>Adj. R²</i>	<i>R² change</i>	β	<i>F</i>	<i>p</i>	VIF
Perceived fluency	Articulation rate	0.437	0.437	0.472	77.82	<.001	1.49
	Mid-clause pause ratio	0.509	0.072	-0.337	52.26	<.001	1.49

as they were contextually linked to the notion of native speakerism. These results point to the multifaceted nature of perceived oral fluency, which is seen as a combination of several intertwined factors.

Table 3.

Frequency of Coded Categories from L2 Raters' Transcribed Interviews.

Coded category	Total comments
Temporal	96
Speed	62
Pauses	29
Rhythm	5
Repair	5
Non-temporal	113
Vocabulary	10
Grammar	21
Pronunciation (general comments)	44
Sentence patterns	38
Social	130
Education	3
Socio economic status	3
Gender	4
Nativeness	45
L2 accent/Not nativelike	75

Discussion

It has been suggested that native speakers' L2 fluency judgements have been mostly influenced by temporal factors such as speed and breakdown fluency. However, given the lack of studies on L2 speakers' perceptions of fluency, the present study attempted to identify key factors underlying L2 users' perceived fluency using an explanatory sequential design. It appears that fluency, similar to comprehensibility (Isaacs and Trofimovich, 2012), tends to be conceptualized as a broad concept (for similar findings with teacher see Tavakoli and Hunter, 2018). Our quantitative findings indicate that L2 speakers' perceived fluency is mainly associated with the speed of delivery and pausing behaviors within clauses. The magnitude of the link between acoustic information and L2 fluency judgements is relatively large ($R^2 > .50$) (Plonsky & Ghanbar, 2018). The results here serve as additional supportive evidence for the significant role of temporal features in L2 fluency judgements.

In the context of native speakers' judgements, pausing behaviors between clauses is also found to be a significant predictor (Saito et al., 2018). However, our regression model in the case of L2 speakers' perceived fluency excluded final-clause pause ratio which is relatively independent of L2 proficiency (Kormos, 2006). This might suggest that L2 speakers are more aware of different speech processing in response to pause locations than monolingual native speakers. Our L2 listeners focused more on mid-clause pauses which is assumed to reflect linguistic encoding processes (i.e., formulation) than on final-clause pauses which is expected to mirror conceptualization. This finding may suggest L2 speakers' sensitivity to linguistic processing underlying speech production probably due to their experience of L2 learning, subsequently supporting the advantages of employing a diverse

range of listeners in speech judgement tasks (e.g., Munro, Derwing, & Morton, 2006 for L2 users with varied experience backgrounds).

These quantitative findings have been further substantiated by our qualitative findings based on the interview data. First, as with the regression model, theme 1 pointed to the importance of optimal speed and the right placement of pauses in L2 speakers' notion of 'being fluent' in L2. More importantly, the qualitative findings uncover factors explaining the dynamicity of L2 listeners' judgements of fluency. Our qualitative results are consistent with Rossiter's (2009) claim that non-temporal variables such as pronunciation, grammar, and vocabulary have an effect on fluency judgments. Whilst participants recognized the importance of these variables in assigning fluency ratings, the speech samples triggered a wealth of comments related to speaking English fluently which were classified under the umbrella term 'social factors'. The negative comments associated with low-level fluency were contrasted in the data against nativelikeness which, in turn, is associated with the notion of high social status and advanced proficiency in the target language. Based on this, it is possible to hypothesize that L2 speakers might be more sensitive to monolingual views of bilingualism rooted in essentialist ontologies of language that relegate non-native speakers to a deficit model (Ortega, 2019). Our L2 listeners might have experienced linguistic insecurity, which is still extant among multilinguals. They appeared to evaluate their L2 English oral proficiency vis-a-vis monolingual native speakers of English and also to self-judge their L2 fluency with native-like fluency or accent as a referential point. Therefore, they might have invoked such social factors during their judgements of speech produced by L2 speakers from a different L1 background (Japanese).

Social aspects of perceived fluency are substantively difficult to operationalize and measure in a quantitative manner. However, the present study suggests that it is crucial for further research to acknowledge that linguistic studies need to focus more on language as a social construct that is never neutral or easily quantifiable. While the findings reported here provide insight into the dynamic nature of L2 listeners' perceived fluency, it should be noted that LOR could only give a rough index for the level of L2 proficiency.

References

- Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer software]. Retrieved from www.praat.org/
- Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & de Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30(2), 159–175. <http://doi.org/10.1177/0265532212455394>
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29(3), 359–380. <http://doi.org/10.1093/applin/amm041>
- Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31(4), 533–557. <http://doi.org/10.1017/S0272263109990015>
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354–375. <http://doi.org/10.1093/applin/21.3.354>
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34(3), 475–505. <http://doi.org/10.1017/S0272263112000150>
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28(1), 111–131. <http://doi.org/10.1017/S0272263106060049>
- Ortega, L. (2019). SLA and the study of equitable multilingualism. *The Modern Language Journal*, 103(S1), 23–38. <http://doi.org/10.1111/modl.12525>
- Pennycook, A. (2017). *The cultural politics of English as an international language*. Routledge.
- Plonsky, L., & Ghanbar, H. (2018). Multiple Regression in L2 Research: A Methodological Synthesis and Guide to Interpreting R² Values. *The Modern Language Journal*. <http://doi.org/10.1111/modl.12509>
- Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review*, 65(3), 395–412. <http://doi.org/10.3138/cmlr.65.3.395>
- Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics*, 39(3), 593–617. <http://doi.org/10.1017/S0142716417000571>
- Seidlhofer, B. (2011). *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.
- Smith, J. A., Flower, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method and research*. London: SAGE.
- Tavakoli, P., & Hunter, A.-M. (2018). Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*, 22(3), 330–349. <http://doi.org/10.1177/1362168817708462>
- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28(1), 1–30. <http://doi.org/10.1017/S0272263106060013>