Development of Second Language French Oral Skills in an Instructed Setting: A Focus on Speech Ratings

Pavel Trofimovich
Concordia University

Sara Kennedy
Concordia University

Josée Blanchet
Université du Québec à Montréal

Abstract

This study examined the relationship between targeted pronunciation instruction in French as a second language (L2) and listener-based ratings of accent, comprehensibility, and fluency. The ratings by 20 French listeners evaluating the speech of 30 adult L2 French learners enrolled in a 15-week listening and speaking course targeting segments, prosody, fluency, and connected speech processes (e.g., liaison, enchainment) were compared before and after instruction in read-aloud and extemporaneous (picture description) speaking tasks. Results showed that the learners improved in all speech ratings, especially in extemporaneous speaking. Results also revealed that accent ratings were linked to prosody (intonation accuracy, pitch range), while fluency and comprehensibility ratings were additionally linked to fluency phenomena (length of fluent run, hesitation rate). These findings are discussed in terms of their implications for L2 pronunciation learning and links between instruction, listener-rated dimensions of speech, and performance in different tasks.

Résumé

Cette étude examine la relation entre l'enseignement explicite de la prononciation en français langue seconde (L2) et l'évaluation de l'accent, de la compréhensibilité et de la fluidité par des locuteurs du français. L'étude compare les évaluations des productions orales de 30 apprenants adultes du français L2 avant et après un cours de phonétique de 15 semaines ciblant les segments, la prosodie, la fluidité et les phénomènes de resyllabification tels l'enchaînement. Les productions provenant d'une tâche de lecture à haute voix et d'une tâche de narration spontanée d'une histoire illustrée ont donc été soumises à l'évaluation (sur échelles de type Likert) de 20 juges naïfs. Les résultats montrent que les apprenants se sont améliorés dans toutes les dimensions évaluées de leur production, particulièrement en narration spontanée. On note par ailleurs une corrélation entre les jugements liés à l'accent et ceux liés à la prosodie (précision des contours intonatifs, étendue de la variation tonale). On remarque aussi que les jugements de fluidité et de compréhensibilité sont en lien avec certains phénomènes de fluidité (moyenne du temps de parole entre les pauses, hésitations). On discutera de la pertinence des résultats pour l'apprentissage de la prononciation L2 et on abordera en parallèle la relation entre l'enseignement, les dimensions de l'oral appréciées par des juges naïfs et la performance à l'oral dans diverses tâches.

Development of Second Language French Oral Skills in an Instructed Setting: A Focus on Speech Ratings

Although second language (L2) pronunciation can no longer be described as a neglected aspect of teaching and research, many questions still remain for learners of languages other than English about the links between instruction and development. For example, many textbooks for L2 French cover pronunciation (e.g., Abry & Veldeman-Abry, 2007), yet little is known about the development of French pronunciation in instructed learners. Teachers and learners thus rely on intuition, course materials, and past experience to guide their teaching and learning. This study aimed to fill this gap by investigating pronunciation development of adult L2 French learners over time, with the overall goal of understanding which specific linguistic dimensions of L2 French learners' oral performance are related to listener-rated constructs of accent (nativelikeness), comprehensibility (ease of understanding), and fluency (smoothness of speech delivery) of L2 speech. Our overarching objective was to contribute to the knowledge base about how L2 French learners' pronunciation development is linked to pronunciation instruction.

L2 French Pronunciation Teaching

The teaching and learning of L2 French oral skills, particularly the development of L2 French learners' pronunciation in instructed settings, remains a largely underexplored area of research, with existing research featuring one-time, often descriptive measures of performance as opposed to longitudinal data (e.g., Birdsong, 2003; Thomas, 2002). Existing evidence of pronunciation development comes from a handful of studies featuring longitudinal datasets. In one such study, for instance, Harnois-Delpiano, Cavalla, and Chevrot (2012) tracked the use of liaison over 1 year by Korean university learners in a weekly 3-hour French language and literature course, showing significant increases in learner production of obligatory and optional liaison in word pairs over a year (see also Howard, 2013). In another university-based study, Champagne-Muzar, Schneiderman, and Bourdages (1993) examined the effects of 12 weeks of instruction targeting French intonation, rhythm, and segments on university students' L2 French pronunciation. Pretest and posttest recordings evaluated by listeners for segments, intonation, rhythm, and global impression showed that students in the instructed group, compared to uninstructed students, made significant improvement in all four rated measures (see also Knoerr, 2000). In a recent study, Liakin, Cardoso, and Liakina (2015) evaluated the effectiveness of using mobile technology featuring automatic speech recognition software to teach university-level L2 French students the pronunciation of the French vowel /y/. After five weekly 20-minute pronunciation activities requiring students to read aloud target words from a personal tablet or phone and receive feedback on them, the students significantly improved in the accuracy of their production (but not perception) of French /y/, compared to the students who engaged in similar reading activities but received feedback from the teacher or the students who only engaged in 20-minute conversation activities with an instructor. To sum up, instructional research on French pronunciation presently provides limited evidence of instruction-pronunciation links.

Another underexplored aspect of L2 French pronunciation learning concerns the relationship between specific dimensions of learners' oral production (e.g., accuracy of liaison production, appropriateness of intonation) and such listener-based constructs of

speech as accent, comprehensibility, and fluency. The assumption here, based on prior research with L2 speakers of English (see Trofimovich & Isaacs, 2012), is that not all linguistic dimensions of L2 speech are equally relevant to these constructs. Briefly, listeners' ratings of accent, which are often used as a measure of nativeness, often capture how well an L2 speaker can approximate the speech patterns of the target language (Jesney, 2004). Ratings of comprehensibility, conceptualized as a measure of understanding used in a broad sense, target listeners' perceived ease of understanding (Isaacs & Trofimovich, 2012). And perceived fluency, as evaluated through scalar ratings by listeners (Segalowitz, 2010), refers to the degree to which speech sounds fluid (i.e., spoken without undue pauses, filled pauses, hesitations, or dysfluencies such as false starts and repetitions). The most relevant question for L2 French pronunciation instruction is to determine which linguistic dimensions of L2 learners' oral production contribute to listener perception of comprehensibility and fluency (i.e., constructs that are ostensibly most relevant to learners' communicative success) and which dimensions are merely linked to learners' sounding accented (i.e., nonnative).

For L2 English speech, ratings for accent, comprehensibility, and fluency have been linked to elements such as speech rate, use of discourse markers, lexical richness, grammatical accuracy, segmental and word stress accuracy, and pitch range (Isaacs & Trofimovich, 2012; Kang, 2013; Munro & Derwing, 1998; Trofimovich & Isaacs, 2012; Williams, 1992). However, little empirical work has been conducted with other L2s, including L2 French, particularly in relation to listeners' perception of L2 accent, comprehensibility, and fluency. For instance, Caspers (2010) found that for L2 Dutch, comprehensibility ratings for L1 Chinese speakers were linked to segmental errors, but accent ratings were linked to stress errors (for further examples of work in languages other than English and French, see Abrahamsson & Hyltenstam, 2009; Szpyra-Kozlowska, 2014). In L2 French, Prefontaine (2013) showed that acoustic measures of fluency, such as speech and articulation rate, mean length of runs, and pause frequency, were often significantly related to L2 French learners' self-assessment of their fluency. And for L2 German, comprehensibility of L2 German speakers was predicted by phonetic, lexical, morphological, and fluency measures and by stress placement accuracy, but not by syntactic and phonological measures (O'Brien, 2014). The only relevant research to date comes from Kennedy, Guénette, Murphy, and Allard (2015) who recently showed that during task-based interactions between pairs of L2 French speakers, pronunciation accounted for 18% of the comprehension problems between the pairs (see also Bergeron & Trofimovich, 2017). The elements most frequently linked to comprehension problems were segments, particularly consonant production. However, that is apparently the only published research on specific pronunciation elements linked to communication difficulties, and that research featured L2 French use among fellow learners (not the perception of L2 speech by advanced or nativelike French speakers) and did not focus on the dimensions of accent and fluency. Put differently, there is little research in which learners' development of specific pronunciation aspects of L2 French is linked to listeners' judgments of these same learners' accent, comprehensibility, and fluency. Thus, teachers have little guidance regarding which specific dimensions of their learners' speech are most relevant to learners' sounding fluent in their L2 speech, as well as easy to understand.

The Current Study

To address the lack of longitudinal research on the development of L2 French pronunciation, in a previous study, we explored the effectiveness of phonetics teaching for 30 adult learners of L2 French in a 15-week listening and speaking course targeting segments, prosody, fluency, and such connected speech processes as liaison (Kennedy, Blanchet, & Trofimovich, 2014). We analyzed the learners' speech in read-aloud and picture description tasks before and after instruction, using seven measures encompassing the dimensions of segmental and suprasegmental (prosody) production as well as fluency, with all measures based on the coding by linguistically trained coders. That is, the results were not based on listeners' ratings but on pre-established speech measures. We found improvements in learners' segmental and intonation accuracy, use of enchainment, pitch range, and number of hesitations.

While encouraging, these findings of significant improvement in L2 French learners' pronunciation following extensive instruction have one important shortcoming. These results do not show whether learners improved across time in listener-based measures of L2 speech, such as accent, comprehensibility, and fluency. These findings also do not indicate the extent to which the reported gains in segmental accuracy, prosody, and fluency were linked to L2 speech characteristics that are perceptible to listeners. Indeed, what is likely most relevant to listeners are those aspects of pronunciation and fluency that feed into global perceptions of L2 speech. In other words, it is possible that results could show statistically significant improvement in segmental and suprasegmental production as well as in fluency (as coded by trained coders), yet this improvement might not be sufficiently perceptible to listeners and/or might not influence their judgments of L2 French speech. Therefore, in this study, we revisited the data from our original research to examine the impact of phonetics instruction on listener-based ratings of accent, comprehensibility, and fluency in L2 French speech before and after instruction. The specific research questions were:

1. Do L2 French learners improve in listener-based ratings of accent, comprehensibility, and fluency following a 15-week listening and speaking course?
2. Which segment, prosody, and fluency aspects of learner speech are associated with these listener-based ratings?

Method

Participants

The L2 French learners included 30 adult speakers of French (23 women, seven men) who at the time of the study resided in Montreal, Quebec, and were enrolled in an intermediate-level listening and speaking course at a French-medium university. The learners (M_{age} = 35.8 years, *range* = 27–52 years), who had resided in Quebec for an average of 3.2 years (*range* = 0.3–10 years), represented various language backgrounds, including Mandarin (11), Russian (7), Farsi (3), Spanish, Portuguese, Cantonese (2 each), as well as Korean, Malay, and Romanian (1 each). At the outset of the study, the learners self-rated their French ability at a mean of 4.1 (*range* = 2–7) on a 9-point scale (1 = *very poor*, 9 = *excellent*), suggesting that their general level of French ability was intermediate.

The Course

The 15-week listening and speaking course took place once per week for 3 hours, with about 1 hour devoted to practice in a multimedia lab. The instructor was a native speaker of Quebec French with a graduate degree in applied linguistics and 12 years of teaching experience. The instruction focused on segmental and suprasegmental aspects of spoken French. The main focus was on connected speech processes, which included enchaînement and liaison, and on developing fluency and prosody through work on phrasal stress (rhythmic groups) and intonation. *Enchaînement* refers to a link between a word-final consonant and the initial syllable of the following word (e.g., *elle aime* [*she likes*] becomes [E-lEm]) or the link between word-final and word initial vowels (e.g., *Hugo aime* [*Hugo likes*] becomes [u-γo~Em]). *Liaison* stands for the overt production of a word-final consonant that is typically silent, and its resyllabification into the initial syllable of the following word (e.g., *ils aiment* [*they like*] becomes [il-zEm]). For both enchaînement and liaison, the emphasis was on comprehension but learners were encouraged to produce them through practice. For phrasal stress and intonation, the emphasis was on fluid delivery of speech, with practice involving both controlled output recorded in the lab and guided tasks (e.g., practicing a scene from a play). In a typical pedagogic sequence, each topic was covered in one class meeting and reviewed during the following class. Each meeting started with a discovery activity, followed by the teacher's explanation of the targeted aspect, then by controlled practice. The learners then practiced the targeted aspect through communicative and fluency tasks (e.g., role plays, shadowing). Lab-based dictation or production tasks involved short sentences illustrating the targeted aspects.

Speech Samples

Learner production was analyzed in two tasks, which differed in degree of formality (controlled reading vs. spontaneous speaking in response to a picture prompt), both administered before and after instruction. In the original study, a focus on two speech tasks was motivated on the finding that L2 learners differ in accuracy and fluency of speech by task, such that read-aloud tasks often elicit more accurate production of segments and prosody than more spontaneous tasks, such as storytelling and interviews (Rau, Chang, & Tarone, 2009). The first task was a read-aloud story (163 words) featuring a three-sentence narrative followed by a dialogue between a woman standing in a ticket line and a man who wanted to cut into the line (five turns, nine sentences). All sentences were about 10-15 words long (M = 11 words), and 90% of all vocabulary were among the first 1,000 most frequent words in French (Cobb, 2000). The second task was an oral picture description based on an eight-panel image sequence depicting a woman and a man who accidentally exchanged identical suitcases after colliding with each other on a busy street corner (e.g., Derwing, Munro, & Thomson, 2008). The tasks were administered twice, in Week 3 as a pretest and in Week 15 as a posttest, using the same equipment, instructions, and procedure. The learners recorded their speech in a multimedia lab using interactive software CAN-8 Virtual Lab (1990). For the read-aloud task, they received a copy of the text and had 2 minutes to review the text, after which they were given 2.5 minutes to record the text. For picture description, the learners received a copy of the picture story entitled *Erreur sur la*

valise [*Suitcase Mix-Up*], to contextualize the story's central element, and then had 2 minutes to review the images and 5 minutes to record their narrative.

Coded Measures of Speech

The audio recordings from both tasks, considered along with transcripts of the recordings, were coded for accuracy by two trained coders (native French speakers). The coders were first trained by one of the researchers (henceforth, Researcher 1), who coded 10% of all data together with the coders. The coders then analyzed an additional 5% of the data independently, with the opportunity to clarify any remaining issues with this researcher. The rest of the data were coded independently, such that each coder took responsibility for one task, with the possibility of consulting with each other. Interceded agreement reached 98–100% for all speech judgments after another researcher (henceforth, Researcher 2) reanalyzed 10% of the data from the picture description (spontaneous) task, where the likelihood for inconsistent judgments was arguably greater than in the read-aloud (controlled) task. No reanalysis of the read-aloud task was deemed necessary because the original coder of the read-aloud task worked in close collaboration with Researcher 1, who contacted the coder frequently and checked consistency of coding decisions on multiple occasions as the coding proceeded. For each learner, the data from both tasks were coded for the following seven measures:

1. Segmental errors: single segment additions, deletions, or substitutions (e.g., *temps* [*time*] spoken as *dent* [*tooth*]), and spelling-based mistakes (e.g., *loup* [*wolf*] spoken as *loupe* [*magnifying glass*]). The measure was a ratio of all segment errors over all words produced.
2. Intonation errors: inappropriate pitch moves, with rising or flat contours signalling closure; and falling or flat contours used for an expected signal of continuity (e.g., *Peut-être que vous ne savez pas lire, monsieur!* [*Perhaps you don't know how to read, sir!*] with *lire* spoken with a falling pitch). This measure was a ratio of the total number of inappropriate intonation contours produced over the total number of expected contours.
3. Enchaînement use: a successful consonant-to-vowel or vowel-to-vowel link (e.g., *elle aime* [*she likes*] becomes [E-lEm] and *Hugo aime* [*Hugo likes*] becomes [u-ɣo~Em]). Enchaînement required an expected rise-fall pattern signalling a word boundary (i.e., *Elle ↑a~un↓ rhume* [*She has a cold*]) and had to involve a continued, unbroken phonation (absence of perceptible pausing). This measure was a ratio of the total number of successfully produced enchaînement over the total number of contexts where such links could potentially occur.
4. Liaison use: categorical, near-categorical, and frequent variable liaison between two words (e.g., between personal pronouns and verbs, determiners and nouns). This measure of liaison use matched the content of instruction, as students were taught to produce liaison in contexts corresponding to categorical, near-categorical, and very frequent variable liaisons (Durand & Lyche, 2008). Thus, students were expected to produce liaison in its most commonly heard contexts. Erratic (or prohibited) liaisons were neither taught nor tallied. Liaison had to involve a proper grammatical context (e.g., *vous allez* [*you go*] becomes [vu-za-le]), accurate phonetic realization of the

linked consonant (e.g., [s] produced as [z] in *gros arbres* [*big trees*], spoken as [γ{o-za{b{]), and an expected rise-fall pattern signalling a word boundary, with no perceptible pausing between words. This measure was a ratio of the total number of successfully realized liaisons out of the total number of contexts for liaison in each learner's production.

5. Fundamental frequency (F0) range: difference between highest and lowest F0 values, extracted from a pitch tracker display (Boersma & Weenink, 2010). This measure was to capture the degree of pitch variation for each learner, in absolute terms, on the assumption that narrower pitch ranges characterize flat, monotonous delivery and wider ranges describe lively, animated speech (see Wennerstrom, 2001). Although F0 range might encompass only extreme (high and low) values that might not apply to extensive discourse content or might be subject to pitch tracking inaccuracies, it was nevertheless considered a simple, broad measure of acoustic correlates of pitch that also allowed for direct comparisons of the current data with comparable L2 English datasets (Isaacs & Trofimovich, 2012; Trofimovich & Isaacs, 2012).
6. Mean length of run (MLR): mean number of syllables produced between two adjacent filled or unfilled pauses of 400 milliseconds or longer, following Riggenbach (1991).
7. Speech hesitations: all dysfluencies such as filled and unfilled pauses (e.g., *i ils marchent . . . puis [euh . . .] ils tournent* [*they walk . . . and (hmm . . .) they turn*], where " . . . "designates an unfilled pause), and pauses inside a rhythmic group (e.g., *ça fait une . . . demi-heure maintenant* [*It has been a . . . half hour now*]). This measure was a ratio of the total number of hesitations over all syllables in each sample.

Speech Ratings

Pretest and posttest samples from both tasks were saved individually and then shortened to 20 seconds of speech, excluding initial hesitations and dysfluencies. The samples were organized by task and then presented to listeners for rating, using two randomized orders within each task and opposite sequences of task presentation (Task 1–2 vs. Task 2–1), with an equal number of listeners assigned to each combination of task and sample orders. The listeners were 20 French speakers (13 females, seven males) who had all grown up in Francophone households and all but one was educated in French. The listeners (M_{age} = 28.2 years) were university students in the fields of linguistics, education, and psychology, with no formal training in L2 phonetics or pronunciation. The majority of listeners (17) reported high familiarity with accented French speech, most often citing English, Spanish, Arabic, Chinese, and German as familiar language backgrounds of L2 French speakers. Most listeners (15) reported having work experience with speakers of L2 French in a French environment, and all reported knowledge of at least one additional language (e.g., English, Spanish).

The listeners rated the speech samples individually in a self-paced task on a computer, using three 9-point rating scales in the following order: accent (1 = *accent marqué* [*heavy accent*], 9 = *pas d'accent* [*not accented*]); comprehensibility (1 = *difficile à comprendre* [*hard to understand*], 9 = *facile à comprendre* [*easy to understand*]); and

fluency. For fluency, the third construct, the scale endpoints were labelled differently depending on the task, to reflect the controlled and spontaneous nature of fluency in reading aloud (1 = *la lecture n'est pas du tout fluide* [*reading is completely dysfluent*], 9 = *la lecture est très fluide* [*reading is very fluent*]) versus extemporaneous speaking (1 = *ne parle pas du tout couramment* [*does not speak fluidly at all*], 9 = *parle très couramment* [*speaks very fluidly*]). The listeners were first given definitions of each construct, then rated five practice files illustrating speech samples by speakers from different language backgrounds featuring various combinations of speech patterns (e.g., speech that is accented yet easy to understand, easy to understand but dysfluent, non-accented and fluent). They worked at their own pace in approximately 1-hour sessions, playing each consecutive file and recording their ratings in the booklet, with replays permitted. The listeners showed high rating consistency (Cronbach's alpha) for accent (a = .89–.93), comprehensibility (a = .93–.96), and fluency (a = .94–.97), so mean scores were computed per learner by averaging across 20 listeners' ratings for each rated construct, separately for each task at each testing time.

Results

Speech Ratings Across Time and Task

To address the first research question, which asked whether L2 French learners improved in listener-based ratings of accent, comprehensibility, and fluency following 12 weeks of targeted instruction (Weeks 3–15), we first focused on the learners' accent, comprehensibility, and fluency scores across time and task (see Table 1 for descriptive information for all measures).

Table 1
Means and Standard Deviations for All Speech Ratings by Task and Time

Measure	Read-Aloud Task				Picture Description Task			
	Pretest		Posttest		Pretest		Posttest	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Accent	3.17	0.96	3.41	1.07	3.16	0.78	3.35	0.73
Comprehensibility	4.82	1.11	4.72	1.13	3.64	1.34	4.20	1.09
Fluency	4.36	1.38	4.47	1.34	3.26	1.35	3.77	1.12

We compared these three sets of ratings through analyses of variance (ANOVAs), with time (pretest, posttest) and task (read-aloud, picture description) as repeated measures. For accent, there was only a significant effect of time, F (1, 29) = 5.84, p = .022, η_p^2 = .17, but no significant effect of task, F (1, 29) = .10, p = .749, η_p^2 = .01, and no significant time × task interaction, F (1, 29) = .07, p = .787, η_p^2 = .01. Thus, the learners improved in accent, although modestly, in both tasks (as illustrated in Figure 1).

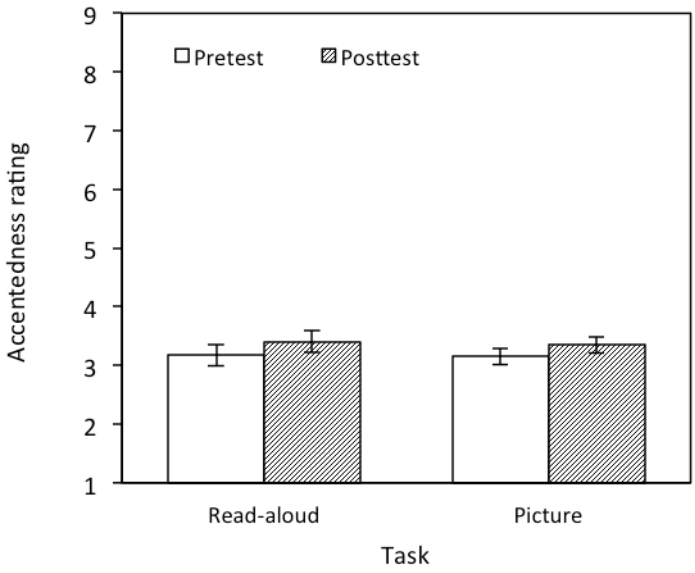

Figure 1. Accent scores across task (read-aloud, picture description) and time (pretest, posttest), with error bars enclosing two standard errors of the mean.

For comprehensibility, there was no significant effect of time, F (1, 29) = 2.40, p = .132, η_p^2 = .08, but there was a significant effect of task, F (1, 29) = 22.63, p < .001, η_p^2 = .44, and a significant time × task interaction, F (1, 29) = 5.33, p = .028, η_p^2 = .16. Follow-up tests of interaction effects (with a Bonferroni correction), carried out to explore the significant interaction, revealed that the learners improved in comprehensibility only in the picture task (p = .016), as illustrated in Figure 2.

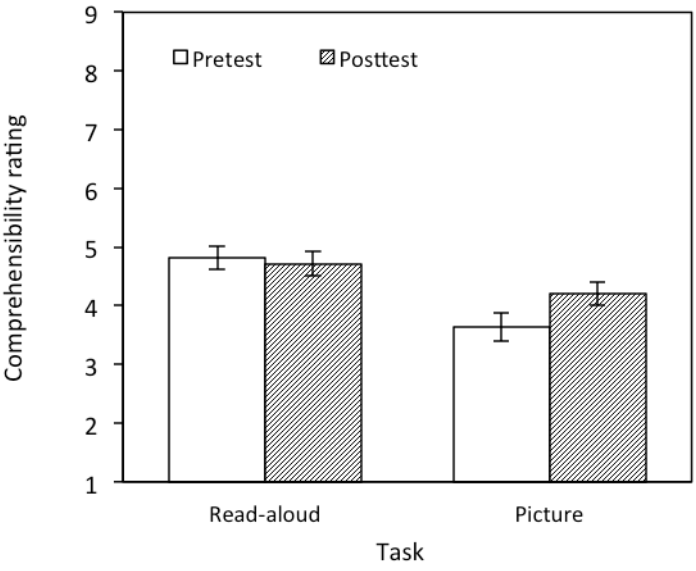

Figure 2. Comprehensibility scores across task (read-aloud, picture description) and time (pretest, posttest), with error bars enclosing two standard errors of the mean.

For fluency, there were significant effects of time, F (1, 29) = 5.97, p = .021, η_p^2 = .17, and task, F (1, 29) = 20.33, p < .001, η_p^2 = .41, but no significant time × task interaction, F (1, 29) = 2.32, p = .138, η_p^2 = .07. The two significant main effects indicated that the learners showed greater fluency in the read-aloud than picture description task and that they overall improved in fluency over time, although (as shown in Figure 3) mostly in the picture task.

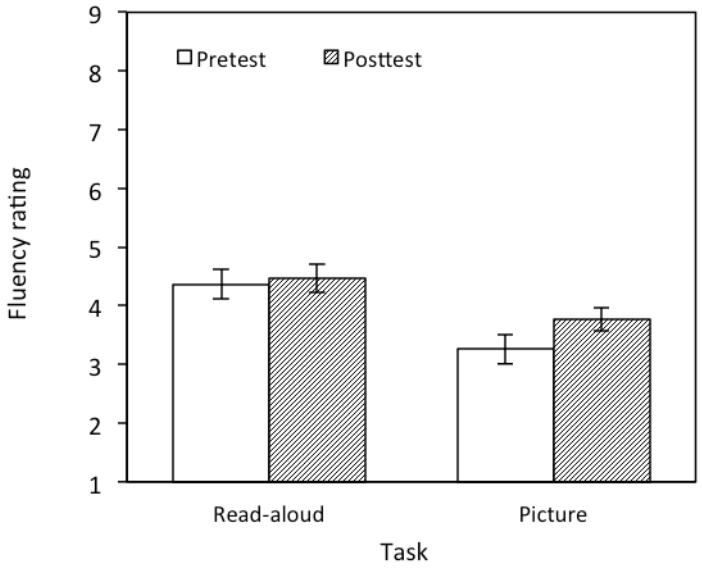

Figure 3. Fluency scores across task (read-aloud, picture description) and time (pretest, posttest), with error bars enclosing two standard errors of the mean.

Speech Ratings and Linguistic Aspects of Learner Speech

To address the second research question, which asked which segment, prosody, and fluency aspects of learner speech were associated with the listener-based ratings of accent, comprehensibility, and fluency; we then explored contributions of the seven speech measures to listener ratings via partial correlations. These correlations were carried out between each rating set (accent, fluency, comprehensibility) and each speech measure at posttest, with the relevant pretest measure partialled out. For example, for the correlation between the learners' segmental errors at the posttest and their posttest comprehensibility ratings, we partialled out the learners' segmental error rates from the pretest. By controlling initial performance, we examined the extent to which each speech measure was related to listener ratings at the end of the course. The results of partial correlations are summarized in Table 2.

Table 2
Partial Correlations Between Listener-Based Ratings and Individual Speech Measures from the Posttest, with Relevant Pretest

	Accent		Comprehensibility		Fluency	
Speech measure	Read-aloud	Picture	Read-aloud	Picture	Read-aloud	Picture
Segmental errors	−.25	−.19	−.23	−.03	−.14	.05
Intonation errors	−.42*	−.43*	−.37*	−.36*	−.38*	−.37*
Enchaînement	−.06	−.04	−.29	−.01	.02	−.03
Liaison	.13	−.07	.13	.10	.03	.21
F0 range	.26	−.36*	.10	−.31*	−.01	−.36*
MLR	.18	.22	.31*	.08	.49**	.15
Hesitations	−.11	−.14	−.04	−.45**	−.24	−.40*

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

As shown in Table 2, for accent, less accented L2 speech was linked to fewer intonation errors (both tasks) and a narrower F0 range (picture task). For comprehensibility and fluency, more comprehensible and fluent speech was linked to fewer intonation errors (both tasks), longer fluent speech runs (read-aloud), narrower F0 range, and fewer hesitations (picture task).

Discussion

The current project was conceptualized as a follow-up analysis to our previous study, which showed significant improvement in L2 French learners' pronunciation following 15 weeks of instruction (Kennedy et al., 2014). Our objective was to determine if previously reported significant gains in L2 French learners' segmental and intonation accuracy, use of enchaînement, F0 range, and number of hesitations were associated with listeners' judgments of learner speech, in terms of its accentedness (nativelikeness), comprehensibility (ease of understanding), and fluency (smoothness of speech delivery). The current findings showed that, following instruction, learners received significantly better ratings for accent, comprehensibility, and fluency, usually across both tasks (read-aloud, picture description). In addition, learners' gains in two aspects of prosody (intonation accuracy, F0 range) and in one aspect of fluency (hesitation rate) reported in our earlier study were also associated with posttest ratings of accent, comprehensibility, and fluency. This finding complements previous research on L2 French pronunciation (e.g., Harnois-Delpiano et al., 2012; Howard, 2013; Liakin et al., 2015) by showing that focused pronunciation instruction can have measurable benefits for the listener (see also Champagne-Muzar et al., 1993). Moreover, this finding builds on research on learners of

other L2s, indicating that global improvements in listener-based measures (e.g., comprehensibility, fluency) following targeted instruction (e.g., Derwing, Munro, & Wiebe, 1998; Galante & Thomson, 2017; Kennedy et al., 2014) are linked to learner gains in specific dimensions of speech, such as prosody and fluency measures of hesitations and pausing. All in all, the current results are noteworthy as they imply that focused phonetics instruction has an impact beyond specific aspects of L2 speech, contributing to listeners' global judgments of L2 French speech. In other words, the benefits of L2 instruction are evident not only in particular coded measures of learners' speech, but also in the impressionistic judgments of listeners, who are likely to interact with L2 learners in real-world contexts outside the classroom.

Linguistic Dimensions of Speech Ratings

Although our original comparisons of learners' development across time (based on coding by trained coders) showed improvements in multiple linguistic areas, such as segmental and intonation accuracy, use of enchaînement, F0 range, and number of hesitations (Kennedy et al., 2014), only a subset of these dimensions was associated with posttest speech ratings by untrained French listeners in this study. More specifically, less accented (more nativelike) speech was linked to fewer intonation errors and a narrower F0 range, while more fluent and comprehensible L2 output was related to intonation, F0 range, and also longer fluent speech runs and fewer hesitations.

Among these improvement areas, intonation errors and F0 range (aspects of prosody) appeared to be related to all speech ratings; in particular, for both tasks, the numbers of intonation errors were consistently associated with speech ratings. These findings highlight the importance of prosody for L2 speech learning and teaching in general and more specifically for L2 French. For instance, various measures of prosody, which included intonation choice and boundary tone marking, have been shown to account for up to 50% of the variance in L2 English accent, comprehensibility, and overall proficiency judgments for L2 speakers from multiple linguistic backgrounds (e.g., Kang, Rubin, & Pickering, 2010). Prosodic factors, including intonation, have also been linked to various listener-based measures of L2 speech (Field, 2005; Mennen, 1998; Pickering, 1999; Wennerstrom, 2000), and focused instruction targeting prosody and fluency dimensions of speech appears to lead to improvements in L2 speech production, as measured through listener judgments (e.g., Derwing et al., 1998). And although a narrower F0 range was associated in this study with higher speech ratings, which at first glance appears counterintuitive, this finding has been attested in prior research. Kang et al. (2010) showed negative associations between measures of pitch (encompassing pitch height and range) and listener ratings of L2 English speakers' comprehensibility and overall proficiency (but see Trofimovich & Isaacs, 2012, where F0 range was unrelated to L2 English accent or comprehensibility). It might be that an exaggerated F0 range—although typical of lively, animated speech—might lead listeners to downgrade their evaluations. This result notwithstanding, the overall pattern of findings for prosody suggests that these aspects of L2 speech might be highly relevant to various dimensions of L2 performance, including L2 French speech by classroom learners, applying across various listener-based constructs (Bergeron & Trofimovich, 2017; Kang, 2012; Kang et al., 2010).

Fluency was another dimension of L2 French speech consistently associated with the speech ratings at the posttest. Both comprehensibility and fluency were associated with

the production of longer, unbroken stretches of speech (in the read-aloud task) and with fewer hesitations (in the picture description task). Although the specific fluency measures linked to speech ratings differed, these measures involve one common characteristic—they reflect temporal dimensions of speech output, such as frequency of pausing and duration of speaking, suggesting an important role of temporal measures of fluency for such listener-based constructs as comprehensibility and perceived fluency (e.g., Kang et al., 2010; Rossiter, 2009). For instance, L2 speakers often have a different distribution of pauses in their speech, compared to native speakers, even though the overall number of pauses might be the same (Bosker, Quené, Sanders, & de Jong, 2014), which implies that measures of pausing can and do influence listener-based judgments of L2 speech. In addition, listeners likely use "fluency" in its broader sense, as a marker of general language ability (Lennon, 1990). In fact, L2 learners' proficiency gains have been shown to be associated with enhanced L2 fluency (Freed, 2000), which would be consistent with the finding that the speech that includes longer stretches of pause-free output and fewer hesitations would be perceived as being easier to understand and more fluid.

Speaking Task Effects

Although the task variable (comparison of read-aloud vs. picture description tasks) was not a primary target of this research, the use of two speaking tasks—one eliciting controlled production while the other requiring spontaneous speech output—revealed several interesting patterns. The first pattern was that pretest-posttest improvement across the three speech ratings varied as a function of task. Whereas the learners overall improved in accent in both tasks (though this improvement was very small and thus not particularly meaningful, as shown in Figure 1), the improvement in comprehensibility and fluency was mostly restricted to the picture description task (see Figures 2 and 3). This finding aligns well with previously reported improvements in comprehensibility and fluency, particularly when evaluated through spontaneous production tasks, in both instructed L2 learners (e.g., Derwing et al., 1998; Derwing, Munro, Foote, Waugh, & Fleming, 2014; Galante & Thomson, 2017) and uninstructed L2 users (Derwing & Munro, 2013). This result is also in line with small and often nonsignificant changes in accent following instruction or long-time residence in an L2 environment (e.g., Derwing et al., 2014; Kennedy, Foote, & Buss, 2015; Kennedy & Trofimovich, 2010, 2013;). And the second pattern of findings was that different fluency measures were associated with comprehensibility and fluency ratings, depending on the speaking task, which supports prior work on task effects on comprehensibility and fluency ratings (e.g., Crowther, Trofimovich, Isaacs, & Saito, 2017; Derwing, Rossiter, Munro, & Thomson, 2004). The measure of hesitation frequency was (negatively) correlated with comprehensibility and fluency in the picture description task, while the measure of pause-free speech output (MLR) was (positively) associated with both constructs in the read-aloud task (see Table 2). It is possible that reading aloud, which was based on a text provided to the learners, along with a 2.5-minute deadline to complete the reading, may have encouraged learners to complete the reading as efficiently as possible, with fewer hesitations and disruptions to the flow of speech. In contrast, the unscripted picture descriptions may have elicited spontaneous expression, with the consequence that amount of pausing clearly distinguished those who were more fluent from those who showed less fluid performance.

Implications

Besides providing evidence for the effectiveness of targeted L2 pronunciation instruction for the development of L2 French oral language skills, the current dataset also contributes to ongoing research efforts to isolate linguistic aspects of L2 speech associated with listener ratings of accent, comprehensibility, and fluency, especially across tasks (Bergeron & Trofimovich, 2017; Crowther, Trofimovich, Isaacs, & Saito, 2015; Crowther et al., 2017; Kang et al., 2010). Our results imply a distinction between ratings of accent on the one hand, and ratings of fluency and comprehensibility on the other, in that more speech measures were associated with the latter ratings (see Table 2). This result complements prior research showing that listeners' perceptions of comprehensibility are associated with a wider range of linguistic dimensions of L2 speech, spanning the domains of pronunciation (individual segments, prosody, fluency) and lexicogrammar (varied/appropriate use of words and accurate/complex grammar), compared to listeners' judgments of accent, which are mostly restricted to measures of segmental and prosodic accuracy (e.g., O'Brien, 2014; Saito, Trofimovich, & Isaacs, 2016, 2017; Trofimovich & Isaacs, 2012). The conclusion that ratings of comprehensibility and fluency were quantitatively different from judgments of accent is also supported by correlations among the three sets of ratings in the picture description task, with fluency and comprehensibility sharing 81% of variance (r = .90); in contrast, these two ratings shared only 42% of variance with accent (r = .65 in each case). To our knowledge, this is among the first studies, with a focus on French, targeting several listener-rated measures of L2 speech in relation to specific linguistic properties of learners' oral production (see also Bergeron & Trofimovich, 2017).

Limitations and Conclusion

Last but not least, it is important to acknowledge several limitations of the current research. First, in the absence of a control group in this study, it would be premature to draw unequivocal conclusions as to the effectiveness of the instructional method used in this study and the practical significance of the obtained differences. However, the improvement documented here and in the companion study (Kennedy et al., 2014), which motivated this research, is nevertheless notable. The target participants in this study had been immersed in a French-speaking environment for a mean of 3.2 years before the study. As we argued in the original report, the pronunciation development of learners with prolonged L2 exposure in a naturalistic environment is affected only minimally, if at all, by additional naturalistic exposure within a period of instruction. This is based on the finding that environment-driven (i.e., noninstructed) improvement in pronunciation is usually rapid, occurring within 6 months to 2 years of initial naturalistic exposure, such that further development often slows down or cannot be detected (Derwing & Munro, 2013; Munro & Derwing, 2008). Furthermore, the lack of a control/comparison group does not invalidate possible relationships uncovered here between linguistic dimensions of L2 French speech and listener ratings. Whether or not such relationships can be attributed to instruction, they were attested in the current dataset, suggesting links between certain linguistic properties of L2 French speech (which are likely specific to the proficiency level of learners targeted in this study) and listeners' perceptions of L2 accent, comprehensibility, and fluency.

Another set of limitations pertains to the choice of targeted listeners and speakers. Regarding listeners, it would be important to determine if the current findings generalize to other types of listeners (i.e., not necessarily native French users), given that language learners do interact with fellow nonnative speakers, often more so than with native speakers (Crowther, Trofimovich, & Isaacs, 2016). Listeners' linguistic background, their formal training and experience, as well as language and speech awareness may all contribute differently to listeners' perceptions of learner speech (Isaacs & Thomson, 2013; Isaacs & Trofimovich, 2011; Saito, Trofimovich, Isaacs, & Webb, 2017; Winke, Gass, & Myford, 2013); this implies that the same instructional intervention might be found to be more or less successful depending on the different types of listeners evaluating learners' oral language. And with respect to speakers, it would be interesting to extend this research to learners of different proficiency levels (beginner, advanced) and to speakers from specific language backgrounds, in order to tease apart those linguistic dimensions of L2 French speech specific to speakers of particular languages and proficiency levels from those dimensions that might cut across various learner profiles.

In conclusion, taken together with the findings from our original research in the same context (Kennedy et al., 2014), the current results appear to provide heartening news for instructors of L2 French in university-level settings. These results suggest that a consistent instructional emphasis on fluency, expressiveness, and intonation, and the perception and production of prosody and connected speech processes, such as liaison and enchaînement, might have positive consequences for learners' development of L2 French oral skills, in a manner that is relevant for listeners. Our findings are generally promising for both researchers and teachers as they suggest that L2 pronunciation, despite the inherent difficulty it poses for adult learners, is a skill that can be learned in classroom contexts.

Correspondence should be addressed to Pavel Trofimovich.
Email: Pavel.Trofimovich@concordia.ca

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of acquisition and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*, 249-306. Retrieved from https://doi.org/10.1111/j.1467-9922.2009. 00507.x

Abry, D., & Veldeman-Abry, J. (2007). *La phonétique : audition, prononciation, correction*. Paris, France: CLE International.

Bergeron, A., & Trofimovich, P. (2017). Linguistic dimensions of accentenedness and comprehensibility: Exploring task and listener effects in second language French. *Foreign Language Annals*, *50*, 547-566. Retrieved from https://doi.org/10.1111/flan.12285

Birdsong, D. (2003). Authenticité de prononciation en français L2 chez des apprenants tardifs anglophones. *Acquisition et Interaction en Langue Étrangère*, *18*, 17-36.

Boersma, P., & Weenink, D. (2010). *Praat: Doing phonetics by computer* [Computer program]. Retrieved from http://www. praat.org

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). The perception of fluency in native and nonnative speech. *Language Learning*, *64*, 579-614. Retrieved from https://doi.org/10.1111/lang.12067

CAN-8 Virtual Lab. (1990). [Computer software]. Toronto, ON: Sounds Virtual.

Caspers, J. (2010). The influence of erroneous stress position and segmental errors on intelligibility, comprehensibility and foreign accent in Dutch as a second language. *Linguistics in the Netherlands*, *27*, 17-29.

Champagne-Muzar, C., Schneiderman, E., & Bourdages, J. (1993). Second language accent: The role of the pedagogical environment. *International Review of Applied Linguistics*, *31*, 143-161.

Cobb, T. (2000). The compleat lexical tutor [website]. http://www.lextutor.ca

Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility: Nonnative listeners' perspectives. *Journal of Second Language Pronunciation*, *2*, 160-182. Retrieved from https://doi.org/10.1075/jslp.2.2.02cro

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does speaking task affect second language comprehensibility? *The Modern Language Journal*, *99*, 80-95. Retrieved from https://doi.org/10.1111/modl.12185

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2017). Linguistic dimensions of L2 accentedness and comprehensibility vary across speaking tasks. *Studies in Second Language Acquisition*. Published online 22 August 2017. Retrieved from https://doi.org/10.1017/S027226311700016X

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*, 163-185. https://doi.org/10.1111/lang.12000

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*, 359-380. https://doi.org/10.1093/applin/amm041

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*, 393-410. https://doi.org/10.1111/0023-8333.00047

Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning*, *64*, 526-548. https://doi.org/10.1111/lang.12053

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R I. (2004). Second language fluency: judgment on different tasks. *Language Learning*, *54*, 655-679. https://doi.org/10.1111/j.1467-9922.2004. 00282.x

Durand, J., & Lyche, C. (2008). French liaison in the light of corpus data. *Journal of French Language Studies*, *18*, 33-66.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*, 399-423. https://doi.org/10.2307/3588487

Freed, B. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243-265). Ann Arbor, MI: University of Michigan Press.

Galante, A., & Thomson, R. I. (2017). The effectiveness of drama as an instructional approach for the development of second language oral fluency, comprehensibility, and accentedness. *TESOL Quarterly*, *51*, 115-142. Retrieved from https://doi.org/10.1002/tesq.290

Harnois-Delpiano, M., Cavalla, C., & Chevrot, J.-P. (2012). L'acquisition de la liaison en L2 : étude longitudinale chez des apprenants coréens de FLE et comparaison avec

enfants francophones natifs. *SHS Web of Conferences*, *1*, 1575-1589. Retrieved from https://doi.org/10.1051/shsconf/20120100150

Howard, M. (2013). La liaison en français langue seconde : une étude longitudinale préliminaire. *Language, Interaction, and Acquisition*, *4*, 190-231. https://doi.org/10.1075/lia.4.2.04how

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*, 135-159. https://doi.org/10.1080/15434303.2013.769545

Isaacs, T., & Trofimovich, P. (2011). Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of L2 speech. *Applied Psycholinguistics*, *32*, 113-140. https://doi.org/10.1017/S0142716410000317

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34*, 475-505. Retrieved from https://doi.org/10.1017/S0272263112000150

Jesney, K. (2004). *The use of global foreign accent rating in studies of L2 acquisition.* Calgary, AB: University of Calgary Language Research Centre Reports.

Kang, O. (2012). Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly*, *9*, 249-269. https://doi.org/10.1080/15434303.2011.64263

Kang, O. (2013). Relative impact of pronunciation features on ratings of non-native speakers' oral proficiency. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 10-15). Ames, IA: Iowa State University.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of English language learner proficiency in oral English. *The Modern Language Journal*, *94*, 554-566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Kennedy, S., Blanchet, J., & Trofimovich, P. (2014). Learner pronunciation, awareness and instruction in French as a second language. *Foreign Language Annals*, *47*, 79-96. Retrieved from https://doi.org/10.1111/flan.12066

Kennedy, S., Foote, J., & Buss, L. (2015). L2 speakers at university: Longitudinal development and rater behaviour. *TESOL Quarterly*, *49,* 199-209.

Kennedy, S., Guénette, D., Murphy, J., & Allard, S. (2015). Le rôle de la prononciation dans les problèmes de compréhension entre locuteurs de français lingua franca. *Canadian Modern Language Review*, *71*, 1-25.

Kennedy, S., & Trofimovich, P. (2010). Language awareness and second language pronunciation: A classroom study. *Language Awareness*, *19*, 171-185. Retrieved from https://doi.org/10.1080/09658416.2010.486439

Kennedy, S., & Trofimovich, P. (2013). First and final-semester non-native students in an English-medium university: Judgments of their speech by university peers. *Language Learning in Higher Education*, *3*, 283-303. Retrieved from https://doi.org/10.1515/cercles-2013-0015

Knoerr, H. (2000). Pratique intonative et utilisation d'un logiciel de visualisation dans un cours de prononciation en français langue seconde : une étude descriptive. *Canadian Journal of Applied Linguistics*, *3*, 123-140.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*, 387-417. https://doi.org/10.1111/j.1467-1770. 1990.tb00669.x

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, *32*, 1-25. https://doi.org/10.1558/cj.v32i1.25962

Mennen, I. (1998). Second language acquisition of intonation: The case of peak alignment. *Chicago Linguistic Society*, *34*, 327-341.

Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning*, *48*, 159-182. https://doi.org/10.1111/1467-9922.00038

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, *58*, 479-502. https://doi.org/10.1111/j.1467-9922.2008.00448.x

O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning*, *64*, 715-748. https://doi.org/10.1111/lang.12082

Pickering, L. (1999). *An analysis of prosodic systems in the classroom discourse of native speaker and nonnative speaker teaching assistants* (Unpublished doctoral dissertation). University of Florida, Gainesville, FL.

Prefontaine, Y. (2013). Perceptions of French fluency in second language speech production. *Canadian Modern Language Review*, *69*, 324-348. https://doi.org/10.3138/cmlr.1748

Rau, D., Chang, H. H. A., & Tarone, E. E. (2009). Think or sink: Chinese learners' acquisition of the English voiceless interdental fricative. *Language Learning*, *59*, 581-621. https://doi.org/10.1111/j.1467-9922.2009.00518.x

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

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review*, *65*, 395-412. https://doi.org/10.3138/cmlr.65.3.395

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, 37, 217-240. Retrieved from https://doi.org/10.1017/S0142716414000502

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*, 439-462. Retrieved from https://doi.org/10.1093/applin/amv047

Saito, K., Trofimovich, P., Isaacs, T., & Webb, S. (2017). Re-examining phonological and lexical correlates of second language comprehensibility: The role of rater experience. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 141-156). Bristol, United Kingdom: Multilingual Matters.

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

Szpyra-Kozlowska, J. (2014). *Pronunciation in EFL instruction: A research-based approach*. Bristol, United Kingdom: Multilingual Matters.

Thomas, A. (2002). La variation phonétique en français langue seconde au niveau universitaire. *Acquisition et Interaction en Langue Étrangère, 17*, 101-121.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*, 905-916. Retrieved from https://doi.org/10.1017/S1366728912000168

Wennerstrom, A. (2000). The role of intonation in second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 102-127). Ann Arbor, MI: University of Michigan Press.

Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. Oxford, United Kingdom: Oxford University Press.

Williams, J. (1992). Planning, discourse marking, and the comprehensibility of international teaching assistants. *TESOL Quarterly*, *26*, 693-711. https://doi.org/10.2307/3586869

Winke, P., Gass, S., & Myford, C. (2013). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, *30*, 231-252. https://doi.org/10.1177/0265532212456968