Studies in Second Language Acquisition, 2014, page 1 of 34.
doi:10.1017/S0272263114000114

COMMUNICATIVE FOCUS ON FORM AND SECOND LANGUAGE SUPRASEGMENTAL LEARNING

Teaching Cantonese Learners to Perceive Mandarin Tones

Kazuya Saito
Waseda University

Xianghua Wu
University of California, Davis

The current study examined how form-focused instruction (FFI) with and without corrective feedback (CF) as output enhancement facilitated second language (L2) perception of Mandarin tones at both the phonetic and phonological levels by 41 Cantonese learners of Mandarin. Two experimental groups, FFI only and FFI-CF, received a 90-min FFI treatment designed to encourage them to notice and practice the categorical distinctions of Mandarin tones through a range of communicative input and output activities. During these activities, the instructors provided CF only to students in the FFI-CF group by recasting and pushing them to repair their mispronunciations of the target features (i.e., output enhancement). The control group received comparable meaning-oriented instruction without any FFI.

Xianghua Wu is now at the University of California, Berkeley.

This research was supported by the *Language Learning* Research Grant awarded to the first author. We are grateful to Roy Lyster and four *SSLA* reviewers for providing constructive comments on drafts of this article. We gratefully acknowledge Irina Presnyakova and George Smith, who helped with data collection and manuscript preparation; the two teachers, Haisheng Jiang and Yujia Jiang; and all the volunteer students for the project.

Correspondence concerning this article should be addressed to Kazuya Saito, School of Commerce, Waseda University, 1-6-1 Nishi Waseda, Shinjuku, Tokyo 169-8050, Japan. E-mail: kazuya.saito@waseda.jp

The effectiveness of FFI was assessed via a forced-choice identification task with both trained and untrained items for a variety of tonal contrasts in Mandarin (high-level Tone 1 vs. mid-rising Tone 2 vs. high-falling Tone 4). According to statistical comparisons, the FFI-only group attained significant improvement in all lexical and tonal contexts, and such effectiveness was evident particularly in the acquisition of Tone 1 and Tone 4—supposedly the most difficult instances due to their identical phonological status in the learners' first language, Cantonese. The FFI-CF group, however, demonstrated marginally significant gains only under the trained lexical conditions. The results suggest that FFI promotes learners' attentional shift from vocabulary to sound learning (generalizable gains in trained and untrained items) and facilitates their access to new phonetic and phonological categories. Yet the relative advantage of adding CF to FFI as output enhancement remains unclear, especially with respect to the less experienced L2 learners in the current study.

Over the past 25 years, second language (L2) education studies have extensively examined the acquisitional value of form-focused instruction (FFI) and corrective feedback (CF) on production errors as ways of enhancing the rate and ultimate attainment of L2 acquisition. Whereas previous FFI studies have been exclusively concerned with the grammatical, lexical, and pragmatic domains of language, recent studies have begun to examine the pedagogical potential of FFI in the area of L2 speech acquisition (e.g., Saito & Lyster, 2012a, 2012b). Drawing on recent theoretical accounts of L2 phonetic and phonological learning (Best & Tyler, 2007; Bundgaard-Nielsen, Best, & Tyler, 2011), the current study investigated the differential effectiveness of two types of FFI options—FFI with and without CF as output enhancement—designed to help Cantonese learners perceive Mandarin tones in a tutored setting.

BACKGROUND

FFI in SLA

Form-focused instruction is defined as "any pedagogical effort which is used to draw the learners' attention to language form either implicitly or explicitly" (Spada, 1997, p. 73) and is considered most effective when integrated into communicatively oriented and content-based classrooms, in which conveying a meaningful message is a priority (Spada, 2011). The notion of FFI echoes that of focus-on-form instruction, whereby "meaning and use must be evident to the learner at the time that attention

is drawn to the linguistic apparatus needed to get the meaning across" (Doughty & Williams, 1998, p. 6). Unlike traditional teaching methods, which typically introduce language forms in a decontextualized manner (e.g., grammar-translation and audio-lingual methods), communicative focus-on-form is hypothesized to facilitate form-meaning mappings (e.g., VanPatten, 2002) as well as the proceduralization of declarative knowledge in meaningful contexts (e.g., DeKeyser, 2007), both of which may, in turn, enable L2 learners to transfer what they have learned to other communicative contexts outside the classroom (Lightbown, 2008). Specifically, several models concerning the pedagogical sequencing of FFI activities have been proposed in relation to differential levels of students' interlanguage development (e.g., Ellis, 1997; Lyster, 2007; Ranta & Lyster, 2007; VanPatten, 2002).[1]

In Ellis's (1997) computation model, L2 development starts to take place when learners (a) consciously notice linguistic information from L2 input, (b) convert input into intake, and (c) restructure and change their interlanguage representations. Such a developing L2 system is responsible for their output performance at a later stage of L2 acquisition processes (see also VanPatten, 2002). Therefore, for beginning and intermediate students with emerging metalinguistic representations and limited processing abilities in a L2, teachers first need to draw learners' attention to problematic target features via a range of preplanned tasks that manipulate the saliency or frequency of these features in oral and written input (i.e., noticing phase; Lyster, 2007). Certain researchers have emphasized that instruction at the initial stage of L2 acquisition should be mainly operationalized in a receptive mode, because placing too much pressure on students for the production of output might arguably lead these inexperienced learners to exhaust their limited cognitive resources, which are needed to attend to linguistic (instead of semantic) aspects of language (e.g., VanPatten, 2002). Such receptive FFI techniques include structured input (i.e., learners are required to process linguistic form only through input for meaning; VanPatten & Cadierno, 1993) and typographically enhanced input (i.e., target structures are highlighted by means of visual changes such as italics to induce learners to notice the forms in written L2 input; Sharwood Smith, 1993).

As learners' experience and proficiency levels increase, research has generally shown that a simple (though intensive) exposure to enhanced forms in the input alone may not be sufficient for development. To further advance L2 development, teachers should guide learners to engage in deeper and more elaborate uses of target linguistic structures (i.e., awareness phase; Lyster, 2007). In some studies (e.g., Spada, Lightbown, & White, 2005), metalinguistic information was provided before FFI lessons, especially with respect to the acquisition of nonsalient grammatical features with less communicative value (see Spada & Lightbown, 2008, for a list of relatively difficult grammatical structures that may require isolated

explicit instruction). Many other FFI studies have devised several FFI tasks that require students to not only comprehend but also produce the target structure in a linguistically accurate manner to successfully complete the communicative purposes of the task (see Swain, 2005, for her discussion of the role of comprehensible output in L2 development). It has also been claimed that the provision of output-prompting CF, such as explicit feedback (Ellis, 2001) and prompts (Lyster, Saito, & Sato, 2013), toward students' production errors in a seemingly less planned fashion during online interaction pushes learners to repetitively practice producing the target feature in communicatively authentic contexts (i.e., practice phase; Lyster, 2007). This ultimately leads to the development of more robust representations of the target feature (i.e., qualitative change) as well as various levels of processing ability (i.e., quantitative change) to produce the feature in a more nativelike and automatic manner (Gatbonton & Segalowitz, 2005).

In recent years, relevant empirical studies of L2 morphosyntax development have been summarized using narrative review methods (e.g., Ellis, 2001; Spada, 2011) as well as meta-analytic review methods (e.g., Norris & Ortega, 2000; Spada & Tomita, 2010). These reviews have reached a consensus in that (a) the effects of decontextualized teaching methods (e.g., grammar-translation methods) tend to be limited to learners' performance assessed under highly controlled task conditions, and (b) contextualized teaching methods (e.g., FFI and focus-on-form instruction) lead to substantial improvement not only at a controlled but also at a spontaneous level. Similar research findings have been observed in the context of vocabulary teaching (e.g., Schmitt, 2008) as well as pragmatic teaching (e.g., Nguyen, Pham, & Pham, 2012). These studies have shown that not only language-focused learning (targeting analytic use of language) but also meaningful input, output activities, and fluency development (targeting the functional use of language) are needed for students to attain well-balanced L2 vocabulary and pragmatic development (Nation, 2001). What is intriguing in this vein of L2 education research is the number of emerging studies that have tested the amenability of FFI to L2 speech acquisition (e.g., Saito & Lyster, 2012a, 2012b).

L2 Pronunciation and Listening in Instructed Settings

Although foreign accent is a normal aspect of L2 speech production (Flege, Munro, & MacKay, 1995), students still need to fulfill the minimal phonetic and phonological requirements to be comprehensible to achieve the goal of successful communication. Furthermore, some may even wish to go beyond this (Derwing & Munro, 2005). Yet, some pronunciation teaching methods (e.g., audio-lingual methods) have

been notorious for their overdependence on controlled exercise activities, such as mechanical drills and the repetition of minimal pairs (see Trofimovich & Gatbonton, 2006, for a discussion). Whereas accurate perception abilities play a fundamental role in successful word recognition (Cutler & Broersma, 2005) and thus make an ultimate impact on the development of global listening skills (Field, 2008), previous research has exclusively been conducted within the paradigm of intensive auditory training, such as high-variability phonetic training (HVPT; see Thomson, 2012, for a review). In this training method, L2 learners are intensively exposed to minimally paired stimuli produced by a number of native speakers in a strictly controlled laboratory setting. The results have shown that adult L2 learners can demonstrate a 10–20% increase in their identification scores and can generalize their instructional gain from trained to untrained lexical contexts at both segmental (e.g., Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991) and suprasegmental levels (e.g., Wang, Spence, Jongman, & Sereno, 1999). Additionally, such perceptual improvement resulting from HVPT has been shown to be transferable to production domains without any explicit articulatory instruction (e.g., Hardison, 2003; Wang, Jongman, & Sereno, 2003).

Although we acknowledge a great deal of theoretical implications stemming from laboratory training studies, this vein of L2 speech research has generated little pedagogical relevance. In her review, Fraser (2011, p. 12) described the method as "the deprecated 'drill and kill' training," whereby L2 learners listen to only a small number of phonological targets without any contextualized use of the language for many hours (e.g., a 40-min session × eight times in Wang et al., 1999), repeating the same simple task over and over again. Not only is this method monotonous and demotivating, it is also unrealistic for most L2 classrooms, which are notorious for their minimal devotion of time and attention toward pronunciation and listening teaching (see Derwing & Munro, 2005, p. 382). Given that HVPT is one of the only methods proven to effectively improve learners' ability to acquire new sounds, the current study will test a combination of communicative input and output activities (e.g., FFI and CF) in L2 suprasegmental learning within 1.5 hr of instruction by interfacing L2 education and L2 phonology perspectives in an integrative way.

L2 Speech Learning

Recently, two crosslanguage speech perception models—namely, the perceptual assimilation model (PAM) by Best (Best & Tyler, 2007) and the speech learning model (SLM) by Flege (1995)—have been widely used to account for L2 learners' speech learning outcomes. According

to these two models, phonetic similarities and phonological correspondence between first language (L1) and L2 phonological categories may be predictive of the difficulty in discriminating L2 phonological categories. For instance, the PAM proposes that two L2 categories will be very difficult to distinguish if they are assimilated equally to the same L1 category. In contrast, if two L2 categories correspond to two different L1 categories, discrimination between the L2 categories will be relatively easy. The SLM, in contrast, proposes that perceptual success is determined by the perceived phonetic similarity between L1 and L2 phonetic categories. For example, if L1 and L2 phones are acoustically identical, there will be no difficulty in L2 acquisition. If L1 and L2 phones are phonetically dissimilar, the L2 phone will be perceived as a new phone, and thus the level of perceptual difficulty will also be low. However, if L1 and L2 phones share partial acoustic similarity, L2 acquisition will be most difficult for L2 learners.

Mandarin and Cantonese Tones

In Mandarin and Cantonese, lexical tones are realized on monosyllables to signal word meanings (Chao, 1948; Zee, 1999). For example, the syllable /pa/ with the four Mandarin tones superimposed stands for *eight* (/pa^{1}/, 八), *to pull* (/pa^{2}/, 拔), *handle* (/pa^{3}/, 把), and *father* (/pa^{4}/, 爸). The most important acoustic correlates of lexical tones are fundamental frequency (F0) height and contour (Abramson, 1975; Gandour, 1983; Lin, 1988).[2] On a 5-point scale with 1 corresponding to the lowest F0 and 5 to the highest, Mandarin Tone 1 (T1) is represented as [55], which means that it is a high-level tone that starts high and stays high; T2 is realized as a rising tone that starts low and ends high, represented as [35]. In contrast, T4 is represented as [51], representing a high-falling tone that starts high and ends low. Likewise, in Cantonese, T1 is also realized as a high-level tone [55], and T2, as a rising tone [25/35] (So & Best, 2010). As shown in Table 1, Cantonese T1 and T2 correspond to Mandarin T1 and T2, respectively, due to similar acoustic realizations of F0 height and contour. Additionally, Cantonese T1 has a high-falling allophonic variation [53] (Hashimoto, 1972; Yu, 2007), which corresponds to Mandarin high-falling T4 ([51]) due to their similar acoustic properties (i.e., initial pitch height and falling tonal trajectory). In Cantonese, the high-level and high-falling tones are collapsed for native speakers. However, only the high-level tone emerges when the target tone is followed by a high-level tone or a high-falling tone (Bauer & Benedict, 1997; Wong, 1982). This phenomenon is also known as allotone (Gandour, 1981; So & Best, 2010), a phonological change occurring in tonal languages.

Table 1. Description of tonal contrasts in Mandarin and Cantonese

Language	Tone number	Description	Pitch
Mandarin	1	high level	55
	2	high rising	35
	3	low falling rising	214
	4	high falling	51
Cantonese	1	high level	55
	2	high rising	25/35
	3	mid-level	33
	4	low falling	21
	5	low rising	23
	6	low level	22

So and Best (2010) examined the perception of the four Mandarin lexical tones by Cantonese speakers without any previous Mandarin experience. On the one hand, Cantonese listeners had less difficulty distinguishing the T1-T2 contrast by assimilating Mandarin T1 [55] and T2 [35] to the Cantonese T1 [55] and T2 [25/35], respectively. According to the PAM, this could be categorized as a two-category (TC Type) assimilation, in which two nonnative phones are assimilated to two separate native phonemes. On the other hand, Cantonese listeners demonstrated greater difficulty in identifying and discriminating the T1-T4 contrast, because Cantonese listeners perceived both Mandarin T1 [55] and T4 [51] as the Cantonese T1 [55], which has two allophonic variations—high level [55] and high falling [53]. According to the PAM, this could be categorized as a single-category (SC Type) assimilation, in which two nonnative phones assimilate equally well or poorly to a single native phoneme. In contrast, So and Best (2010) remained neutral about the relative difficulty of the acquisition of the Mandarin T2 [35] and T4 [51] contrast, because there arguably exist various possible scenarios: The learners may assimilate the Mandarin tone contrast to the Cantonese T2 [25/35] and T4 [21] contrast or to a combination of T2 [25/35] and the allophonic variation of T1 [53].

MOTIVATION FOR THE CURRENT STUDY

Given the general lack of relevant studies on instructed L2 speech learning, we previously took a first step toward conducting a series of intervention studies to investigate whether and to what degree a combination of communicative input and output activities (i.e., FFI) and CF could facilitate L2 production development of word-initial English /ɹ/ (Saito, 2013a, 2013b, 2013c; Saito & Lyster, 2012a) and English high front

vowel /æ/ (Saito & Lyster, 2012b) by Japanese learners. The results showed that FFI led the learners to attain significant pronunciation improvement not only at a controlled but also at a spontaneous speech level, especially when it was combined with CF. The facilitative role of CF can be attributed to its potentially dual pedagogical function: Pronunciation-focused CF provides students with teachers' pronunciation models while, at the same time, eliciting modified output.[3]

Due to the exploratory nature of these previous studies, they generated a number of questions worthy of future investigation for the purpose of obtaining a well-defined picture of the complex relationship between FFI, CF, and L2 phonetic and phonological development. First, these studies measured instructional gain mainly through trained lexical items that had been used during the FFI lessons. They did not reveal whether FFI enables learners to generalize what they learned during FFI lessons to novel lexical contexts beyond instructional materials (for a relevant discussion, see Logan et al., 1991). Second, these studies exclusively concentrated on L2 segmentals, targeting the acquisition of English consonant /ɹ/ and vowel /æ/. According to L2 speech theories, each of these sounds possibly has an equal amount of learning difficulty because both of the segmental sounds are dissimilar to any L1 counterparts at both phonetic and phonological levels (e.g., Flege, 1995); it still remains unclear how FFI can facilitate the acquisition of nonnative speech contrasts that entail a differential amount of learning difficulty at the suprasegmental level. Because the acquisition of Mandarin tones has provided an ideal testing ground for two different phonetic-phonological learning processes (So & Best, 2010; Wu, Munro, & Wang, 2011), investigating this topic should reveal in detail how FFI and CF can be facilitative of the relatively easy (e.g., T1-T2 for TC Type) and difficult (e.g., T1-T4 for SC Type) suprasegmental contrasts in the perception of Mandarin tones by Cantonese learners.

Third, our previous research findings were limited to the production domain. Second language pronunciation research has shown that adult L2 learners can carefully monitor their accurate production forms, drawing on explicit articulatory knowledge normally stored in general memory, regardless of the present state of their perceptual representation (Sheldon & Strange, 1982). Thus, future studies need to determine whether FFI can impact the development of new sounds in the long-term memory—where phonetic and phonological categories have perceptual bases (Flege, 1995)—or learners' surface monitoring skills, without producing much fundamental and qualitative change in their phonetic and phonological representational system (Sheldon & Strange, 1982). Given that many L2 speech researchers agree that learners' processing of new L2 sounds on the perception level is relatively automatic and that perception test scores are thought to mirror their mental representations (see Flege, 1993, p. 1605),

the current study is a first attempt to illustrate how FFI and CF can facilitate, in particular, L2 perception development.

Finally, our previous studies included L2 learners who had a wide range of L2 experience (e.g., $M_{\text{length of residence}}$ = 15.5 months, SD = 31.8, for Saito & Lyster, 2012a) as well as learners who regularly engaged in potentially different amounts of input and interaction with native speakers outside of classrooms in a L2 learning setting. The tremendous individual differences among the participants might have interacted to conflate the effectiveness of FFI and CF. To isolate and reexamine the actual effect of instruction, it may be necessary to replicate previous studies, especially with L2 learners with homogeneous proficiency levels (i.e., beginner-to-intermediate Cantonese learners of Mandarin) and comparable learning backgrounds (i.e., the minimum input and output conditions outside foreign language classrooms; for similar discussion on the differential effectiveness of phonetic training in second vs. foreign language learning settings, see Lively et al., 1994).

CURRENT STUDY

The primary goal of the current study was to examine whether FFI with or without CF better promotes the perception development of L2 suprasegmentals in a tutored setting (two students to one instructor). This goal was achieved by conducting a quasi-experimental study with a pre- and posttest design. Form-focused instruction in the current study was defined as preplanned input- and output-based activities designed to encourage students to notice and practice Mandarin tones (the target of instruction) through meaningful teacher-student interaction. Corrective feedback in the current study comprised recasts, which were defined as a teacher's reformulation of all or part of a student's utterance, minus the error (Lyster, 2007), and which were operationalized as partial pronunciation-focused recasts (for the details of FFI and CF treatment, see the Method section). Given that recasts were directed only to production (but not to perception) errors and that pronunciation-focused recasts have been reported to elicit a great deal of modified output relative to morphosyntax-focused ones (e.g., Sheen, 2006), the CF treatment in the current study was hypothesized to function as output enhancement by pushing students to analyze target linguistic forms through their repetition of the teacher's model pronunciation (Saito & Lyster, 2012a). In this regard, one primary goal of the project was to investigate the pedagogical potential of FFI in either a more receptive (FFI-only) or a more productive (FFI-CF) mode.[4] The effects of such instructional gain on learners' integrated L2 knowledge in their phonetic and phonological representational systems (instead of their surface monitoring skills) was assessed via a forced-choice identification task of various lexical

items (trained vs. untrained items) and tonal contrasts (T1-T2, T2-T4, and T1-T4). The study's research questions are thus formulated as follows:

1. To what degree is FFI facilitative of L2 speech perception development of Mandarin tones?
2. To what degree does adding CF to FFI as output enhancement increase the size of instructional gain?

METHOD

The current project took place at a Canadian university in Vancouver, with 48 Cantonese learners of Mandarin initially recruited as participating students. Participants took a pretest in which they identified the Mandarin tones,[5] and, a few days later, they were randomly paired and divided into three groups: FFI-CF, FFI only, and control. The two experimental groups, FFI-CF and FFI only, received one 90-min communication lesson in Mandarin in which the FFI treatment was embedded to encourage them to notice and practice the categorical distinctions between the Mandarin high-level T1, mid-rising T2, and high-falling T4 in a meaningful and interactive manner. During these activities, the instructors provided CF only to students in the FFI-CF group by recasting and pushing them to repair their mispronunciations of the target features (i.e., output enhancement). Students in the control group received comparable communicative language instruction in terms of duration, but the content was taught with no focus on either the target tones or words. The researcher and a research assistant observed the tutoring sessions from the back of the classroom to ensure the consistency of treatments. All classes were video recorded. One week after the end of the lessons, all students completed posttests and participated in a final interview.[6]

Participants

Students. Unlike our previous studies (e.g., Saito & Lyster, 2012a), efforts were made to recruit only L2 learners with homogeneous (beginner-to-intermediate) proficiency levels achieved in foreign language learning settings. The project was advertised as "free 90-min Mandarin conversation tutoring sessions" via flyers distributed to students who were enrolled in first- and second-level Mandarin heritage classes open to students whose L1 was Cantonese or another Chinese dialect. According to the institute, the first-level class only allows those who have no Mandarin background but speak Cantonese as their L1 to enroll, and the

second level accepts those who have already completed the previous level or have an equal proficiency level. Because both Mandarin and Cantonese are spoken in Guangdong Province in mainland China, participants were selected such that none of them spoke Guangdong Cantonese as their L1 but rather Hong Kong Cantonese. Instead, the majority of the participants were born and grew up in Hong Kong and a small number of them were born in North America to Hong Kong families. Additionally, advertisements were posted on student community websites at the Canadian university. The flyer clearly specified that eligible participants should be beginner-to-intermediate Cantonese learners of Mandarin who had taken Mandarin courses for 3–8 months in North America. The main goal of the project (improving L2 perception and production abilities of Mandarin tones) was not divulged until all students had completed the study.[7]

Interested participants contacted the researchers to set up appointments for pretests. The 48 students who initially participated in the current study were first randomly assigned to 24 pairs of two students each, and then each of the pairs was randomly assigned to one of three groups, with each group consisting of eight pairs. Each pair received a one-time, 90-min tutoring treatment at the researcher's office. The three groups were (a) the FFI-CF group (eight pairs, $n = 16$), (b) the FFI-only group (eight pairs, $n = 16$), and (c) the control group (eight pairs, $n = 16$).

Two participants failed to complete the project for personal reasons, and five turned out to be too advanced on the basis of on their pretest performance, despite our efforts to exclude such advanced learners, and thus did not meet our proficiency criteria of beginner-to-intermediate learners (for details, see the Results section). In the end, the total number of students used in the final analyses was 41 (20 males and 21 females): $n = 13$ for FFI-CF, $n = 15$ for FFI only, and $n = 13$ for the control. These participants ranged in age from 18 to 26 years ($M_{age} = 20.5$ years, $SD = 1.9$)[8] and had a few years of Mandarin learning experience at college-level schools ($M_{years\ of\ instruction} = 1.2$ years, $SD = 0.7$). Their self-reported daily use of Mandarin was 2.4 ($SD = 1.2$) on a 6-point scale (1 = *very infrequent*, 6 = *very frequent*), indicating that their Mandarin speaking opportunities were restricted outside of classrooms, where Mandarin is a foreign language. All learners reported normal hearing and passed a pure-tone screening at octave frequencies between 250 and 4,000 Hz.

Native Speaker Controls. Six native speakers of Mandarin Chinese (three males and three females) participated voluntarily as controls ($M_{age} = 35.2$ years, $SD = 6.9$). All of them were either graduate students or Mandarin instructors at the Canadian university. They completed the same perception tests, and their scores served as baseline data.

Teachers. Two experienced Mandarin instructors (one male and one female) who worked at the Canadian university at the time of the project participated as instructors. They taught an equal number of tutoring sessions for each of the three groups (four sessions for FFI-CF, FFI only, and control, respectively).

FFI Treatment

Target Words. The Appendix presents the target words used in the pre- and posttests as well as the training tasks. In total, 10 minimal triplets of 30 words (words 1–30 in the Appendix, which differ in meaning only due to T1-T2, T2-T4, and T1-T4 distinctions) were presented in traditional Chinese characters and were embedded in a range of FFI activities to encourage students to increase their phonetic and phonological awareness of Mandarin tones. Traditional characters were selected as they were used not only in the students' L1 but also in Mandarin classes. We consulted with the instructors and examined the textbooks that most of the participating students used in their classes (first- and second-level Mandarin) to ensure that the target words were all familiar to the students. These words were italicized and highlighted in red to trigger students' noticing throughout the meaning-oriented lessons and to help them become aware of the target features at a lexical level (i.e., typographically enhanced input; Sharwood Smith, 1993).

Training Sessions. After some self-introduction and ice-breaking conversation among the participating students and teachers at the beginning of sessions, the participants engaged in a series of FFI activities. Following Lyster and Ranta's FFI model (Lyster, 2007; Ranta & Lyster, 2007) and our previous work on instructed L2 speech learning (e.g., Saito & Lyster, 2012a), we implemented a range of tasks that were designed to promote students' accurate and fluent use of Mandarin tones, first in a receptive and subsequently in a productive mode.[9] The tasks are described in the next several paragraphs.

Listening game. Thirty cards were placed on a table. Each card represented one lexical item and had a relevant picture of the Chinese character of a word. The teacher proceeded to call out words, at which point students tried to find and pick up the card corresponding to the word called as soon as possible. To get as many cards as possible, the students had to pay attention to the perceptual differences between Mandarin tones (T1 vs. T2 vs. T4) in a receptive mode (i.e., structured

input; VanPatten & Cadierno, 1993). Students were first engaged in this FFI activity with the assumption that phonetic-level noticing is a crucial first step toward developing new phonetic and phonological categories (Best & Tyler, 2007; Flege, 1995).

Reading game. The same cards used in the listening game were divided between the two students in each pair (15 cards per student). Instead of the teacher, each of the students took a turn reading aloud a list of words while the other student picked up the relevant cards. If a student produced a target word with a wrong tone, which might have caused the teammate to pick up a different card, the student reading the list had to say "no" and try to read it again so that the teammate could get the intended card. At this stage, students were guided to pay attention to accurately perceiving their own speech to allow their partners to get the correct cards. In this way, the students were encouraged to tune their production to model pronunciation forms on the basis of the perception of one's own and others' speech (Baker & Trofimovich, 2006).

Guessing game. Each card had one vocabulary item written in traditional Chinese characters on the upper right-hand corner. One learner paraphrased the word, and the other guessed the vocabulary item that his or her partner was trying to describe. At this stage, students were encouraged to use target words with more elaboration in an extemporaneous manner (Lyster, 2007; Ranta & Lyster, 2007).

Story-creating game. Students randomly picked up 5 out of the 30 cards labeled with the target lexical items. When they felt ready after sufficient planning time, they created and read aloud a coherent story in a few sentences using all five words. During the activity, students were pushed to practice L2 features with equal attention paid not only to phonetic accuracy but also to lexical, morphosyntactic, semantic, pragmatic, and discoursal accuracy.

Fluency development. Whereas students spent approximately 10–15 min on each game described in the preceding paragraphs, they did the same games for a second round, but this time as quickly as possible within a time limit (5 min). The notion of fluency development in the current study follows the view that L2 learners need to practice the L2 in relatively familiar and easy tasks repetitively under communicative pressure to increase their fluency and control over what they have already learned. For similar fluency teaching frameworks, see research on the automatization in communicative contexts of essential speech segments (Trofimovich & Gatbonton, 2006) and the four learning strands (Nation, 2001).

In total, the whole session—including the self-introduction, ice-breaking conversation among students at the beginning of lessons and the first and second rounds of the four FFI activities—lasted 90 min.

CF Treatment

For the FFI-CF group, the instructors consistently provided CF in the form of recasts in response to learners' mispronunciation or unclear pronunciation of the target features (Mandarin tones) during the FFI activities. Given that some L2 learners have difficulty perceiving recasts as teachers' correction of linguistic errors or confirmation of message comprehensibility (Lyster et al., 2013), the teachers in the current study were specifically asked to recast with a stressed voice only individual words that had been mispronounced (i.e., pronunciation-focused recasts). Such recast techniques are categorized as *partial recasts* and are thought to be the most perceptually salient type of recast, eliciting a great deal of modified output (e.g., Sheen, 2006).

To measure how many recasts students received and repeated, the researcher carefully watched 12 hr of FFI-CF treatment (8 sessions × 1.5 hr each) and recorded the number of recasts the instructors provided and to what degree they elicited students' repetitions. Previous classroom studies found that the amount of learners' repetition of pronunciation-focused recasts tended to be relatively high, especially compared to that of grammar-focused ones (e.g., Lyster, 1998), and previous simulated recall interview studies have confirmed that learners tend to notice the corrective purpose of pronunciation-focused recasts more often relative to morphosyntax-focused ones (e.g., Mackey, Gass, & McDonough, 2000).

Unlike grammatical and lexical errors, which can easily and clearly be transcribed for subsequent analyses, we did not further examine precisely to what degree the participants successfully self-corrected pronunciation errors, because all recordings were filled with unwanted noise during teacher-student interactions (e.g., some repeated the teacher's recasts while the other student was talking) and were thus not appropriate for refined acoustic analyses or listener judgment (for a similar procedure and decision, see Mackey et al., 2000; Sheen, 2006). In examples (1) and (2), "S" is used for students, "T" for teacher, and "*" for mispronunciation or unclear pronunciation.

(1) Story-creating game
S: 我有一只很可爱的/mau^{4}/*
"I have a very cute hat."

(According to the classifier只, the target word should be *kitten* /mau^{1}/.)
T: /mau^{1}/
"kitten" (← RECAST)
S: 我有一只很可爱的/mau^{1}/
"I have a very cute kitten." (← REPETITION)

(2) Guessing game
S1: 你妈妈的丈夫是谁?
"Who is the husband of your mother?"
S2: /pa^{1}pa/*
"father" (but pronounced as "eight")
T: /pa^{4}pa/
"father" (← RECAST)
S2: 对!
"Yes!" (← NO REPETITION)

For the FFI-only and control groups, the instructors were told not to provide any CF, even when their students made errors in their production of Mandarin tones.

Control Group

The 14 learners in the control group received 1.5 hr of meaning-oriented activities (e.g., small talk, communicative games, discussion, and presentations on topics related to Mandarin culture) but with no emphasis on pronunciation or listening practice. Efforts were made not to expose the students to any of the target words. The purpose of including the control group was to check if there existed any test-retest (instead of FFI) effects the participants might have benefitted from by taking the same test twice.

Teacher Training

The researcher provided the instructors with 3 hr of training over 2 days before the intervention commenced. First, the instructors were given a package of instructional materials with a list of the target words. Next, they were told the content and purpose of each activity as well as the way to provide CF following students' pronunciation errors. Finally,

they practiced with the researcher until they fully understood and felt comfortable with the procedure.

Perception Test

Target Words. Among the 30 words that appeared during the FFI, nine (words 22–30 in the Appendix), together with nine other words that were not used in the training session (words 31–39), were used in the tone-identification tests, specifically the pre- and posttests.

Material Preparation. A female native speaker of Mandarin read a list of 18 words three times at a normal speed. All stimuli were recorded by means of a Roland-05 Wave recorder in a soundproof booth at the Canadian university. The tokens were digitized at a sampling rate of 44.1 kHz and were normalized for peak intensity at an average level of 70 dB by means of the speech-analysis software Praat (Version 5.3; Boersma & Weenink, 2011). The best exemplars for each of the 18 items were carefully selected for the testing session.

Procedure. The participating students took a forced-choice tone-identification test right before and a week after the training sessions in a soundproof booth. All the stimuli were presented binaurally via Logitech acoustical headsets; the signal was sent using the Praat software installed on a Macintosh computer. Instructions were given orally by a trained research assistant in English. The students responded by circling one correct answer from three choices (in traditional Chinese characters) printed on a sheet. The 18 items were randomly played three times, and the total score of the test was 54. The available response time was 3 s. Before the actual test, a practice session with six tokens was provided to familiarize the students with the test. The whole test lasted about 15 min (for a similar procedure, see Wang et al., 1999).

RESULTS

Due to ceiling effects, five learners (three from FFI-CF, two from the control) whose total listening scores were more than 95% on the pretests ($x \geq 52$ out of 54 points) were eliminated from the final analysis (for a similar decision, see Iverson, Hazan, & Bannister, 2005, and other auditory training studies). For all following inferential statistical analyses, the alpha level was set at $p < .05$. Cohen's d was also calculated to measure the magnitude of instructional effectiveness between two contrast groups of means. Assumptions of ANOVAs were checked

and met in terms of equal variance (via Levine tests) and sphericity (via Mauchly's test), $p > .05$. The reliability (Cronbach's alpha) for the pretest (consisting of 54 items) among the remaining 41 learners was .775, which ensures the validity of the test format.

Baseline Data

The six native speakers of Mandarin completed the perception test to ascertain the construct validity of the listening stimuli. They demonstrated perfect scores for all test materials ($M = 54$ points).

Initial Performance

Lexical Conditions. The students' identification scores according to group (FFI-CF, FFI only, and control) and lexical (trained vs. untrained) conditions are displayed in Figure 1. To find out if there were any preexisting differences in the students' perception performance according to group and lexis, all students' identification scores were submitted to a two-way Group × Lexis ANOVA. There was no significant main or interaction effect of group, $F(1, 38) = 0.22$, $p = .644$, or lexis, $F(2, 38) = 0.20$, $p = .816$, indicating that students' perception performance was comparable between trained and untrained lexical items across the three groups ($M = 73.8$–81.5%) before receiving instruction.

Tonal Conditions. To examine whether the initial performance varied according to three Mandarin tonal contexts (supposedly entailing different learning difficulties), two subanalyses—individual tone and tone-pair analyses—were implemented.

Individual tone analysis. The participants' initial performance for each individual tone (T1 vs. T2 vs. T4) is illustrated in Figure 2. Whereas a two-way Group × Tone ANOVA demonstrated no significant main effect for group, $F(2, 38) = 0.47$, $p = .459$, and no interaction effect, $F(4, 76) = 0.09$, $p = .598$, a significant main effect of tone was obtained, $F(2, 76) = 21.68$, $p < .001$. According to Bonferroni multiple comparisons, their performance of T2 ($M = 86.4\%$) was significantly higher than that of T1 ($M = 74.7\%$, $p < .001$, $d = 0.70$) and T4 ($M = 70.5\%$, $p < .001$, $d = 1.02$). Yet no significant difference was found between the difficulty in T1 and T4 acquisition ($p = .167$). The results suggest that although the participating students generally showed 75–80% accuracy on the pretests, their seemingly good performance could

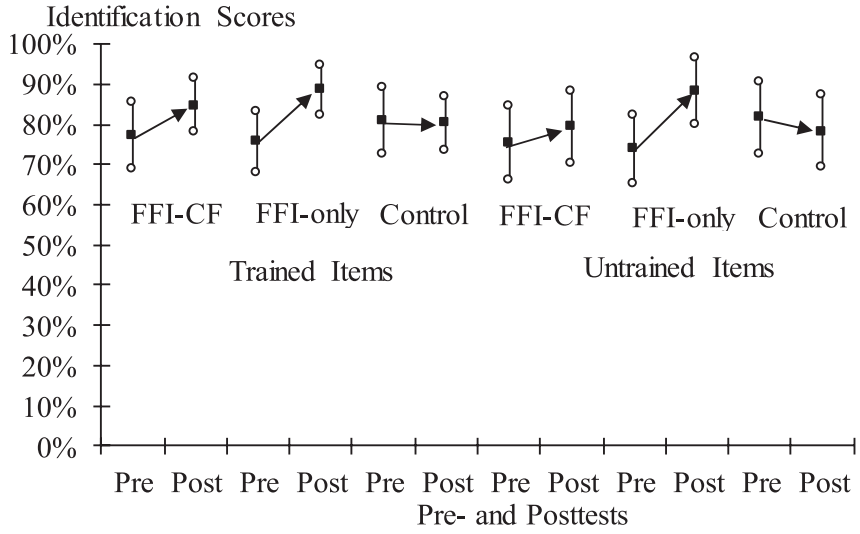

Figure 1. 95% confidence intervals and mean values of the students' pre- and posttest scores. Whereas the FFI-CF group noted a marginally significant improvement only in trained items after instruction, the members of the FFI-only group significantly enhanced their identification scores in both trained and untrained lexical contexts.

be attributed to the relatively easy T2 acquisition (both Cantonese and Mandarin have a similar mid-rising tone) as opposed to the relatively difficult T1 and T4 acquisition (high-level and high-falling tones, which constitute the same phonological category as an allophonic variation in Cantonese).

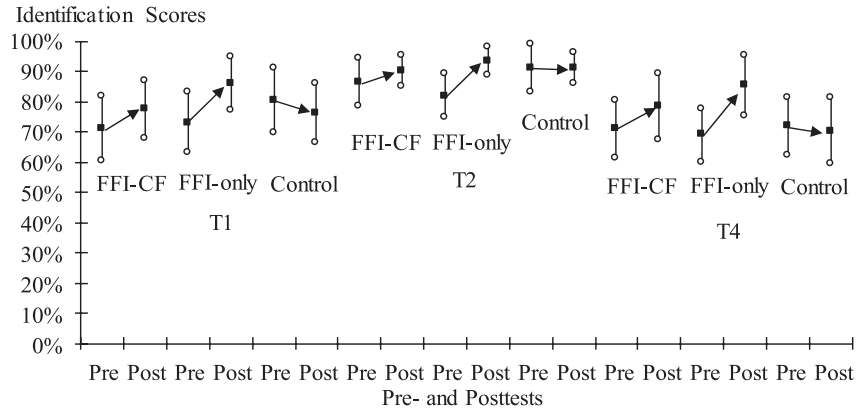

Figure 2. 95% confidence intervals and mean values of the students' pre- and posttest scores according to tonal conditions.

Tone-pair analysis. The initial performance according to three phonological contrasts (T1-T2, T2-T4, and T1-T4) was also analyzed via a confusion matrix, which summarizes the number of errors the trainees made for each tone pair out of a total of 54 instances (6 syllables × 3 tones × 3 repetitions) per participant. The descriptive results of the number of errors are plotted in relation to the three tonal learning contrasts in Figure 3. For instance, the number of errors for T1 and T2 is the sum of misperceptions of both T1 as T2 and T2 as T1. A two-way Group × Tone ANOVA yielded significant main effects of tone, $F(2, 76) = 67.18$, $p < .001$, but no main or interaction effects of group were found: main effects of group: $F(2, 38) = 0.84$, $p = .439$; interaction effects of group: $F(4, 76) = 1.98$, $p = .109$. According to Bonferroni multiple comparisons, the students made a significantly higher number of errors in the T1-T4 contrast ($M = 7.1$ errors) than in the T1-T2 contrast ($M = 2.5$ errors, $p < .001$, $d = 1.52$) and the T2-T4 contrast ($M = 2.7$ errors, $p < .001$, $d = 1.35$). The results suggest that all groups of Cantonese learners likely exhibited limited proficiency in identifying Mandarin T1-T4 (SC Type) relative to T1-T2 and T2-T4 contrasts (TC Type) before the treatment, as predicted by the PAM (So & Best, 2010).

Effects of Instruction

Lexical Conditions. According to the descriptive results of the students' pre- and posttest scores (see Figure 1), whereas the control group displayed little change over time in any of the trained and untrained

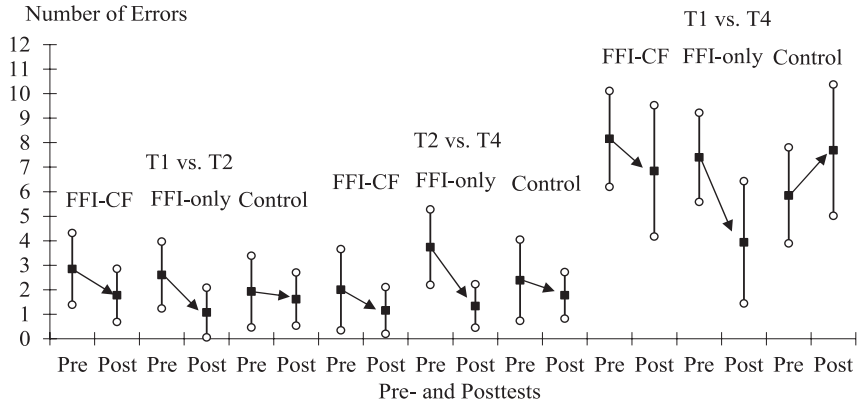

Figure 3. 95% confidence intervals and mean values of the learners' pre- and posttest scores. The FFI-only group significantly reduced the number of errors, especially in the difficult learning instance (T1-T4 for TC Type).

lexical contexts, both of the experimental groups seemed to enhance their perception scores by roughly 10–15% as a result of the training. To examine the statistical significance of the instructional gain, their identification scores were submitted to two-way Group × Time ANOVAs with group as a between-groups factor and time (pre- vs. posttests) as a within-groups factor under trained and untrained lexical conditions, respectively.

With respect to the trained items, the two-way ANOVA exhibited significant Group × Time interaction effects, $F(2, 38) = 3.30$, $p = .048$. Whereas the results of Bonferroni multiple comparisons demonstrated no significant improvement for the control group ($p = .883$), they identified a significant gain for the FFI-only group ($M = 75.6 \rightarrow 88.4\%$, $p = .001$, $d = 0.83$) as well as a gain for the FFI-CF group that was approaching significance ($M = 76.9 \rightarrow 84.6\%$, $p = .052$, $d = 0.64$). With respect to the untrained items, there was also a significant Group × Time interaction effect, $F(2, 38) = 7.89$, $p = .001$. According to Bonferroni multiple comparisons, although neither the FFI-CF group nor the control group attained significant improvement over time ($p = .234$ and .306, respectively), the members of the FFI-only group substantially improved their perception performance ($M = 73.6 \rightarrow 87.9\%$, $p < .001$, $d = 0.94$).

Taken together, the results revealed the following patterns: (a) the FFI-only treatment enabled the students not only to significantly enhance their perception abilities but also to generalize their improvement from trained to untrained items; and (b) the FFI-CF treatment led the students to attain marginally significant gains, but only in the trained lexical items.

Tonal Conditions. Next, we examined whether the degree of FFI effectiveness differed with regard to relatively easy and difficult instances of phonetic and phonological acquisition. The results of the descriptive statistics in Figures 2 and 3 indicate that the experimental groups—FFI only, in particular—substantially increased their individual T4 tone identification and also experienced a reduced number of errors in the T1-T4 contrast. Their performance was analyzed via a set of two-way Group × Time ANOVAs according to the phonetic and phonological categories (via individual tone analysis) and contrasts (via tone-pair analysis).

Individual tone analysis. For T1, there were significant effects of Group × Time interaction, $F(2, 38) = 3.98$, $p = .027$. The results of Bonferroni multiple comparisons identified a significant improvement for the FFI-only group in terms of their correct identification of T1 ($M = 73.0 \rightarrow 85.9\%$, $p = .004$, $d = 0.72$). For T2, there were only significant effects of time, $F(1, 38) = 6.78$, $p = .013$. This indicates that all of the groups (FFI-CF, FFI only, and control) equally improved their perception scores

(M = 86.4 → 91.6%, d = 0.44). For T4, there were significant effects of Group × Time interaction, $F(2, 38) = 6.18$, $p = .005$. The results of Bonferroni multiple comparisons identified a significant improvement for the FFI-only group in terms of their correct identification of T4 (M = 68.9 → 85.2%, $p < .001$, d = 1.06).

Tone-pair analysis. For the T1-T2 pairing, there were only significant main effects of time, $F(1, 38) = 6.23$, $p = .017$. The results, in turn, identified a significant improvement (i.e., reduction in error rate) over time for all of the students regardless of the group conditions (M = 2.5 → 1.5 errors, d = 0.44). Similarly, for T2-T4, there were only significant main effects of time, $F(1, 38) = 6.23$, $p = .006$. This suggests that all of the groups significantly reduced the number of errors after instruction (M = 2.7 → 1.4 errors, d = 0.54). For T1-T4, there was a significant interaction effect of Group × Time, $F(2, 38) = 9.81$, $p < .001$. The results of Bonferroni multiple comparisons showed that the FFI-only group significantly reduced the number of errors after instruction (M = 7.4 → 3.9 errors, $p < .001$, d = 1.04).

Tone analysis summary. To summarize, the results of the individual tone and tonal pair analyses suggest that all of the students enhanced their perception abilities of the relatively easy phonological and phonetic categories (T2) and contrasts (T1-T2, T2-T4), probably owing to the test-retest effects. The FFI effectiveness was evident especially where much learning was needed (i.e., the acquisition of the new T4 category; SC Type of T1-T4).

Fidelity of Implementation

Thirteen learners in the FFI-CF group received 252 recasts and repeated 228 of them, yielding a relatively high repetition rate (90.4%). On average, each learner in the FFI-CF group received 15.8 recasts and repeated 14.3 of them. In line with previous descriptive studies (e.g., Mackey et al., 2000; Sheen, 2006), the results indicated that the learners likely noticed the instructors' corrective intention of pronunciation-focused recasts and were thus pushed to process a great number of enhanced output opportunities during FFI activities.

DISCUSSION

The current study examined to what extent two types of FFI—with and without CF as output enhancement—could promote the perceptual

acquisition of Mandarin tones by Cantonese learners in relation to various lexical and tonal contexts. In this section, we first discuss how FFI alone can promote the development of L2 speech perception (Research Question 1 for the FFI-only approach) and then explore the extent to which adding CF to FFI (i.e., output enhancement) can increase instructional gain (Research Question 2 for the FFI-CF approach).

Effects of FFI

Whereas postpubertal L2 learners likely show tremendous resistance to learning new sounds due to the loss of neural plasticity in the brain (Abrahamsson, 2012) and the influence of L1 phonetic and phonological systems (Best & Tyler, 2007; Flege, 1995, 2003), the mechanism of instructed L2 speech acquisition has traditionally been discussed only within the paradigm of intensive auditory training studies (e.g., HVPT), in which students are involved only with computerized identification tasks for many hours without any contextualized use of the language. Building on recent trends in L2 education research, the current study was the first attempt to adapt the FFI approach (i.e., a combination of communicative input and output activities) as "a set of psycholinguistically motivated pedagogic options" (Ellis, 2001, p. 12) in the context of L2 suprasegmental learning. As we discuss in detail in the next paragraph, the nature and size of the gains resulting from FFI could be considered as comparable to previous HVPT studies in similar contexts (e.g., Wang et al., 1999, a gain of approximately 20% after approximately 6 hr of HVPT).

According to the results of the pre- and posttest scores, a brief amount of FFI treatment (1.5 hr) led students to attain a medium-to-large degree of improvement (an approximately 15% gain) in trained items, and to successfully transfer their gains to untrained items beyond the provided instructional materials. This learning pattern could be considered a transition from lexically to phonetically driven L2 speech learning in line with recent theories (e.g., Best & Tyler, 2007, for the PAM; Bundgaard-Nielsen et al., 2011, for the vocabulary model; Walley, 2007, for the restructuring model). According to these theoretical accounts, L2 learners initially attend to lexical units of L2 ambient input as a whole to extract meaning as a part of their early vocabulary learning. As they encounter and learn more L2 lexical items (i.e., vocabulary spurt), however, they will be forced to fill in phonetic and phonological details to differentiate, in particular, phonetically similar words, such as minimal pairs and lexically dense words. At this stage, learners pay simultaneous attention not only to lexical but also to sublexical units of L2 input, and their speech performance tends to be less subject

to lexical factors, such as frequency and familiarity (see also Bradlow & Pisoni, 1999; Flege, Frieda, Walley, & Randazza, 1998; Imai, Walley, & Flege, 2005; Trofimovich, Collins, Cardoso, White, & Horst, 2012). In this regard, the students' robust and generalizable gain in their identification scores in the current study suggests that the FFI-only approach effectively and efficiently induced the students to increase their phonetic and phonological awareness of Mandarin tones, thereby permitting them to attain highly consistent patterns of sound recognition, regardless of the lexical (trained vs. untrained) conditions.

Another primary goal of the current study was to reveal how FFI could be facilitative of the two phonetic and phonological learning instances entailing a differential amount of learning difficulty. The results support the facilitative effect of FFI; however, this depends on which levels of L2 speech learning (phonetic vs. phonological) students need to process. For the relatively difficult learning contexts (SC Type: T1-T4), the gains of the FFI group were evident, as the rate of confusion between the three Mandarin tones dropped significantly, particularly for the FFI-only group. This is arguably because the acquisition of the T1-T4 contrast requires the students to develop two new tone categories in their phonological systems via noticing not only the acoustic-phonetic differences but also higher phonological-level distinctions between the two L2 tones (see Wu et al., 2011), and FFI was needed in this difficult L2 learning instance in which students had much room for further improvement. For the relatively easy learning contexts (TC Type: T1-T2), however, the advantage of FFI remained unclear: All of the students in the study were better at distinguishing the two L2 tones after receiving any type of instruction, regardless of the presence of the FFI component, arguably due to phonetic- and phonological-level correspondence between these two Mandarin tones and their Cantonese counterparts.

As mentioned earlier, the efficacy of FFI in the development of the L2 speech perception abilities of Mandarin tones could be due to several factors. First, a communicative focus on form triggered by FFI treatment may simulate optimal L2 speech learning processes in naturalistic settings. In the SLM, Flege (1995) claimed that "a new phonetic category can be established for an L2 sound that differs from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds" (p. 239). In the current study, the students who received FFI were guided to notice such phonetic dissimilarity (and similarity) of new L2 sounds through meaningful teacher-student interactions, particularly by way of intensive exposure to a number of minimally paired L2 lexical items, which is hypothesized to be a crucial step toward attaining success at both the phonetic and phonological levels in speech acquisition (Best & Tyler, 2007; Bundgaard-Nielsen et al., 2011; Walley, 2007). Additionally, teacher talk during FFI may provide ideal linguistic input to help students to identify and discriminate new

sounds from their L1 counterparts, such as increased pitch, simplified prosody, and vowel and consonant hyperarticulation. It has been claimed that acoustic enhancements could be the best way for adult L2 learners to circumvent L1 interference effects and increase their awareness of new sounds (see Kuhl, 2000, p. 11855). Iverson et al. (2005) found that exposing Japanese learners to synthetically exaggerated speech stimuli, with F3 and F2 variation maximized and duration lengthened, helped them improve their perception abilities of the nonnative /ɹ/-/l/ contrast (see also McClelland, Fiez, & McCandliss, 2002).

The last variable to be discussed is the role of social interaction in L2 speech learning. In the early bilingual literature, Kuhl and her colleagues similarly stated that social interaction plays a crucial role in the computational mechanism in human language learning. For instance, Kuhl, Tsao, and Liu (2003) showed that American infants who were exposed to foreign sounds through interaction with Mandarin caretakers demonstrated more phonetic learning than those who received the same amount of input but through audiovisual or audio-only recordings. Their results indicated that "phonetic learning from complex language input relies on more than raw auditory sensory information" (p. 9100). That is, L2 speech acquisition can be facilitated more through interacting with a live person rather than simply being exposed to computerized audio-only input. This is because the former may provide motivating learning conditions in which learners can process words and sounds from ongoing speech streams through a shared perception of the communicative intentions of others (i.e., interactivity) as well as speech information via eye gazing and gestures to objects of reference (i.e., contingency; Conboy & Kuhl, 2011; see also Kuhl, 2007). To this end, the results of the current study with college-level students are consistent with the idea that processing new sounds through online teacher-student interaction via FFI could substantially impact various aspects of adult L2 phonetic and phonological development, even with a limited amount of instructional treatment (i.e., 1.5 hr).

Adding CF to FFI as Output Enhancement

Next, we turn our discussion to whether and to what degree adding CF treatment to FFI can increase the overall effectiveness of meaning-oriented phonetic and phonological training. In the current study, CF was operationalized as pronunciation-focused recasts directed at production (but not perception) errors. According to the results of the fidelity of implementation analysis, the students received a great number of recasts during the 1.5 hr of instruction (M = 15.8 recasts per student), with a relatively high rate of repetition (90.4%). This indicates

that the CF treatment generated a strong illocutionary force for modified output and thus functioned as output enhancement (see also Sheen, 2006). However, compared to the FFI-only approach, which demonstrated a robust and generalizable gain in all lexical and tonal conditions, the FFI-CF approach resulted in somewhat limited improvement. The results of the pre- and posttests showed that the FFI-CF group marginally improved their perception performance only under trained lexical conditions. The lack of the generalizability indicates that the participants in the FFI-CF group improved correct pronunciation forms with their primary attention to lexical units of L2 speech; however, it remained unclear whether they became aware of the phonetic aspects of L2 speech and transferred their gain from trained to untrained lexical contexts. Thus, the learning patterns of the FFI-CF approach could thus be characterized as vocabulary rather than sound learning. Taken together, the findings of the current study suggest that, whereas FFI without CF can sufficiently promote the acquisition of Mandarin tones, it remains unclear whether adding CF can increase the overall effectiveness of FFI. The unexpected results are carefully discussed in relation to previous relevant theoretical and experimental work in the field of instructed SLA and L2 speech in an interdisciplinary manner.

The findings can be explained using the perception-first view of L2 phonological and phonetic development that many L2 speech researchers have claimed (e.g., Flege, 1995; Kuhl, 2000). According to these models, once L2 learners hear phonetic dissimilarity (or similarity) relative to their L1 counterparts, perception of new L2 sounds promotes the reorganization of their existing representational systems as well as the development of new categories. Subsequently, this perception-level restructuring activates relevant sensorimotor skills and leads to better production ability. From an educational perspective, these models suggest that only after L2 learners fully or partially restructure and develop new perceptual representations in their long-term memory via more receptive types of FFI activities should they be encouraged to practice it as production via more output-prompting types of FFI activities; this is especially true for pronunciation-focused recasts, so that learners can increase control over their retrieval abilities not only at a controlled level but also at a spontaneous level (see also Piske, 2007).

According to their Mandarin learning backgrounds, there is some evidence that the students in the current study could be considered less experienced (rather than advanced) learners at the initial stages of L2 development. First, all participating students were recruited from first- and second-level Mandarin heritage classes at the time of the project. Second, they reported relatively limited opportunities to use the L2 outside of classrooms in their foreign language learning settings (M = 2.4 for their mean use of Cantonese on a 6-point scale: 1 = *very infrequent* to 6 = *very frequent*). Finally, despite their generally good proficiency,

especially in terms of pretest scores for Mandarin T2 (M = 86.4%), the students demonstrated limited knowledge in the identification of Mandarin T4 (likely misperceived as two allophonic variations within the same category of Cantonese T1) at the beginning of the project (M = 70.5%).

Regardless of their limited access to Mandarin use on a daily basis (they infrequently used Mandarin) and insufficient preexisting knowledge of the target structures (especially for the T1-T4 contrast), the Cantonese students in the FFI-CF group were explicitly pushed to practice target L2 tones (T1 vs. T2 vs. T4) in production via recasts whenever they made pronunciation errors. It is important to note that these learners never received any type of explicit instruction regarding (a) the purpose of recasts (specifically directed to their pronunciation errors on Mandarin tones) or (b) how to repair their errors (i.e., accurate metalinguistic knowledge on Mandarin T1-T2-T3) as a way to help such less experienced learners process pronunciation-focused recasts in a complementary fashion. Given their insufficient amount of relevant L2 experience and the lack of any remedial activities such as explicit instruction before FFI, it seems reasonable to assume that the learners in the current study were developmentally unready to make the best of the CF treatment followed by a great number of modified output opportunities (FFI-CF). Rather, such learners should receive a more receptive form of FFI (FFI only) to focus on the improvement of their perception abilities, which production abilities come to reflect at later stages of L2 speech learning.

The findings indeed concur with previous instructed SLA research findings on the absence of CF advantage for less experienced L2 learners. Whereas Saito (2013b) found that FFI led beginner-to-intermediate Japanese learners in foreign language settings (with very similar backgrounds to those of the participants in the current study) to attain significant improvement on their perception and production development of English /ɹ/, the CF treatment did not result in any advantage. With respect to L2 morphosyntax development, following the well-established developmental sequence of English question forms, Mackey and Philp (1998) found that recasts were effective for developmentally ready learners at Stage 3 (using direct questions with some form of fronting) but not for developmentally unready learners at Stage 2 (using canonical word order with question intonation). Ammar and Spada (2006) found that recasts helped only advanced French learners (with pretest scores above 50%) significantly improve their performance of English possessive determiners from recast treatments. In their review on the pedagogical value of CF in various domains of SLA, Lyster et al. (2013) claimed,

> students may have target language knowledge that continues to be accessible for comprehension but that requires further activation before becoming

> readily available for accurate production. CF can be used to revisit target items . . . in ways that encourage the gradual development of a network of associations that become increasingly accessible for learners during communicative interaction. (p. 13)

In other words, CF can be intended to consolidate partially acquired knowledge rather than activate the acquisition of relatively new knowledge (see also Shintani, Li, & Ellis, 2013).[10]

Open to debate for future research is thus how and when teachers should introduce production enhancements in the form of pronunciation-focused CF as a means of enhancing adult L2 speech acquisition. In conjunction with the developmental theories in both L2 phonetics (Flege, 1995) and instructed SLA (e.g., Ellis, 1997), we suggest that teachers first ensure that their students have certain levels of integrated L2 phonetic and phonological knowledge to be able to process the intensive and repetitive output practice activities within a communicative context. For example, all learners in Saito's (2013a, 2013c) studies listened to the teachers' model pronunciation of the target feature (English /ɹ/) and received explanation on the relevant articulatory configurations of /ɹ/ before moving on to the FFI lessons. It was found that those who received CF demonstrated large and perceptible improvement in their production abilities of /ɹ/, especially compared to those who did not receive any CF. In L2 morphosyntax development, Sheen (2007) investigated the effects of metalinguistic correction: the addition of metalinguistic information while recasting adult English as a L2 learners' errors on the use of English articles (e.g., "You should use the definite article *the* because you've already mentioned fox."). The results showed that the metalinguistic correction group significantly outperformed both the recast-only and the control group. Needless to say, continued research is warranted to better understand the extent to which pronunciation-focused recasts can facilitate instructed L2 speech acquisition, by recruiting a wide range of adult learners (e.g., experienced vs. inexperienced learners) in the context of different phonetic targets, such as speech rate (Munro & Derwing, 2001), lexical stress (Field, 2005), and syllable structure (Couper, 2006). As a result, such future studies will shed some light on what kinds of L2 learners benefit most from CF treatment and how recasts boost the effects of FFI at all levels of L2 suprasegmental learning.

CONCLUSION AND FUTURE DIRECTIONS

Drawing on the notion of FFI in the field of L2 education research (e.g., Ellis, 2001; Spada, 2011), the current study took an exploratory approach toward examining the relative efficacy of FFI with and without CF as output

enhancement on the perceptual acquisition of Mandarin tones by Cantonese learners in a tutored setting. Similar to precursor studies conducted in the domain of L2 segmental production learning (e.g., Saito & Lyster, 2012a, 2012b) as well as intensive auditory training studies in similar tonal learning contexts (e.g., HVPT; Wang et al., 1999), the overall results showed that a communicative focus on form can likewise be effective for instructed L2 suprasegmental acquisition. In terms of type of instruction, the FFI-only group (receptive FFI) attained significant improvement in all lexical and tonal contexts. The FFI-CF group (productive FFI), however, demonstrated marginally significant gains only under the trained lexical condition. The results in turn suggest that FFI without CF as output enhancement may sufficiently promote learners' attentional shift from vocabulary to sound learning (generalizable gain in trained and untrained items) and may facilitate their access to new phonetic and phonological categories, especially when much learning is required (T1-T4). Drawing on the developmental theories in L2 phonetics (e.g., Flege, 1995) and the instructed SLA literature (e.g., Ellis, 1997), we claim that beginner-to-intermediate L2 learners should initially be encouraged to notice and practice the target suprasegmental feature(s) (Mandarin tones) only through FFI in a receptive mode without much pressure for modified output (recast treatment); many theorists and practitioners claim that this could be a first step toward establishing new phonetic and phonological categories (Best & Tyler, 2007; Flege, 1995; Kuhl, 2000).

Given the general lack of research on instructed L2 speech learning as well as several limitations in the current study, we call for more research of this kind that specifically examines (a) whether and to what degree any remedial techniques can promote the effectiveness of CF during FFI activities, especially in L2 perception and production development, such as explicit instruction (cf. Saito, 2013a, 2013c) and different types of CF (e.g., metalinguistic correction; Sheen, 2007), and (b) how instructors can go about maximizing adult L2 speech acquisition by combining intensive auditory training (e.g., HVPT) and a range of communicative input and output activities (i.e., FFI). Second, to investigate more realistic pedagogical implications of FFI for pronunciation and listening teaching, our laboratory-based study needs to be replicated in actual classroom contexts using not only short- but also long-delayed posttest measures (cf. Couper, 2006). Finally, we acknowledge that the forced-choice identification test (a time-pressuring and automatic but still form-oriented approach) was adopted as the only outcome measure to evaluate instructional gain in the current study. Little research attention has been given to assessing L2 perception (and production) abilities of certain sounds in the context of spontaneous speech, likely because of the difficulties in evaluating speech elicited in such a way (Flege, 1993). Thus, future

studies are needed to elaborate and validate reliable assessments to examine the impact of FFI on explicit and implicit L2 phonetic and phonological knowledge.

Received 11 September 2012
Accepted 28 June 2013
Final Version Received 15 August 2013

NOTES

1. Because our study mainly concerns L2 phonetic and phonological development whereby the perception-production link has been empirically supported (e.g., Flege, 1995), the review concentrates on FFI models that hypothesize that comprehension and production derive from the same interlanguage system (e.g., VanPatten, 2002; but see DeKeyser, 2007, for the skill specificity of SLA). For a comprehensive review on the complex relationship between comprehension and production in instructed SLA, see Shintani, Li, and Ellis (2013).

2. Fundamental frequency (F0) is an acoustic measure that refers to the periodicity of a sound. Tonal changes are cued substantially by, though are not limited to, F0 differences.

3. This research investigated the value of recasts that serve multiple remedial functions (i.e., provision of positive and negative evidence as well as elicitation of modified output) but without additionally teasing apart precisely which components of these recast functions contributed to CF effectiveness. Future research should further pursue this topic.

4. Our conceptualization of FFI only as receptive instruction and FFI-CF as productive instruction echoes Shintani et al.'s (2013, p. 3) definitions of comprehension-based instruction (i.e., providing input "which learners have to comprehend and which induces learners' conscious attention to specific forms and the meanings these encode" without proscribing production) and production-based instruction (eliciting learners' self-modified output of the target structures especially via "feedback directed at correcting any errors that learners make in production").

5. They also took a set of production tests. The results of acoustic and listener judgment data (especially regarding the possibility and validity of measuring spontaneous tone production) will be reported in another venue.

6. This testing interval was determined on the basis of psycholinguistic research evidence that the integration of novel phoneme sequences into the mental lexicon takes around a week or so, rather than occurring immediately after the exposure (e.g., Gaskell & Dumay, 2003).

7. Given the explicit nature of the FFI activities (see the subsection on FFI treatment) and the pretest format (i.e., a forced-choice identification task) in the current study, the students were assumed to notice Mandarin tones deductively (but not inductively).

8. The total number of 41 students in the current study could be considered sufficient compared to other major HVPT studies in the field of L2 phonetics, such as Lively et al. (1994; n = 19), Wang et al. (1999; n = 16), and Thomson (2012; n = 26).

9. We used similar FFI activities to guide Japanese learners to learn English /ɹ/ and found that these tasks generally led the participating students to focus simultaneously on form and meaning, on the basis of the results of the retrospective interview after instruction (Saito, 2013c). Yet we do acknowledge that the first two games in the current study (i.e., the listening and reading games) seem to be less communicative than the other FFI activities (i.e., the guessing and story-creating games). This was done to ensure students' noticing of target sounds explicitly at the onset of the FFI lessons.

10. In their meta-analysis of 30 intervention studies, Shintani et al. (2013) provided some evidence that comprehension-based instruction relatively facilitates the initial stage of acquisition, whereas production-based instruction helps increase control over partially acquired knowledge.

REFERENCES

Abrahamsson, N. (2012). Age of onset and nativelike L2 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition*, *34*, 187–214.

Abramson, A. S. (1975). The tones of central Thai: Some perceptual experiments. In J. G. Harris & J. R. Chamberlain (Eds.), *Studies in Tai linguistics in honor of William J. Gedney* (pp. 1–16). Bangkok, Thailand: Central Institute of English Language.

Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, *28*, 543–574.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching*, *44*, 231–250.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin, Germany: Mouton de Gruyter.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 13–34). Amsterdam, the Netherlands: Benjamins.

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer (Version 5.3) [Computer software]. Retrieved from http://www.praat.org

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener- and item-related factors. *Journal of the Acoustical Society of America*, *106*, 2074–2085.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, *33*, 433–461.

Chao, Y. R. (1948). *Mandarin primer: An intensive course in spoken Chinese*. Cambridge, MA: Harvard University Press.

Conboy, B. T., & Kuhl, P. K. (2011). Impact of second-language experience in infancy: Brain measures of first- and second-language speech perception. *Developmental Science*, *14*, 242–248.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect*, *21*, 46–66.

Cutler, A., & Broersma, M. (2005). Phonetic precision in listening. In W. J. Hardcastle & J. Mackenzie Beck (Eds.), *A figure of speech: A Festschrift for John Laver* (pp. 63–91). Mahwah, NJ: Erlbaum.

DeKeyser, R. (Ed.). (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. New York, NY: Cambridge University Press.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*, 379–397.

Doughty, C., & Williams, J. (Eds.). (1998). *Focus on forms in classroom second language acquisition*. New York, NY: Cambridge University Press.

Ellis, R. (1997). *Second language acquisition*. Oxford, UK: Oxford University Press.

Ellis, R. (2001). Investigating form-focused instruction. *Language Learning*, *51*, 1–46.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*, 399–423.

Field, J. (2008). Emergent and divergent: A view of second language listening research. *System*, *36*, 2–9.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, *93*, 1589–1608.

Flege, J. E. (1995). Second language speech learning. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E. (2003) Assessing constraints on second-language segmental production and perception. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319–355). Berlin, Germany: Mouton de Gruyter.

Flege, J. E., Frieda, E. M., Walley, A. C., & Randazza, L. A. (1998). Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition*, *20*, 155–188.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125–3134.

Fraser, H. (2011). Teaching teachers to teach /r/ and /l/ to Japanese learners of English: An integrated approach. In M. Ashby (Ed.), *Proceedings of the Phonetics Teaching and Learning Conference* (pp. 11–15). Retrieved from https://www.ucl.ac.uk/psychlangsci/ptlc/ptlc2011-proceedings

Gandour, J. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics*, *9*, 20–36.

Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 149–175.

Gaskell, G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*, 105–132.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review*, *61*, 325–353.

Hardison, D. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*, 495–522.

Hashimoto, O. K. Y. (1972). *Phonology of Cantonese*. Cambridge, UK: Cambridge University Press.

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *Journal of the Acoustical Society of America*, *117*, 896–907.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /ɹ/-/l/ to Japanese adults. *Journal of the Acoustical Society of America*, *118*, 3267–3278.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Science*, *97*, 11850–11857.

Kuhl, P. K. (2007). Is speech learning "gated" by the social brain? *Developmental Science*, *10*, 110–120.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*, 9096–9101.

Lightbown, P. M. (2008). Transfer appropriate processing as a model for class second language acquisition. In Z. Han (Ed.), *Understanding second language process* (pp. 27–44). Clevedon, UK: Multilingual Matters.

Lin, M. C. (1988). Putonghua shengdiao de shengxue xingzhi yu zhijue zhengzhao [Acoustic properties and perceptual cues of Mandarin tones]. *Zhongguo Yuwen*, *204*, 437–445.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*, 2076–2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874–886.

Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, *48*, 183–218.

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam, the Netherlands: Benjamins.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, *46*, 1–40.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, *22*, 471–497.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal*, *82*, 338–356.

McClelland, J. L., Fiez, J. A., & McCandliss, B. D. (2002). Teaching the [r]-[l] discrimination to Japanese adults: Behavioral and neural aspects. *Physiology & Behavior*, *77*, 657–662.

Munro, M., & Derwing, T. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, *23*, 451–468.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. New York, NY: Cambridge University Press.
Nguyen, T., Pham, T., & Pham, M. (2012). The relative effects of explicit and implicit form focused instruction on the development of L2 pragmatic competence. *Journal of Pragmatics*, *44*, 416–434.
Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.
Piske, T. (2007). Implications of James E. Flege's research for the foreign language classroom. In O.-S. Bohn & M. J Munro (Eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 331–347). Amsterdam, the Netherlands: Benjamins.
Ranta, L., & Lyster, R. (2007). A cognitive approach to improving immersion students' oral language abilities: The Awareness-Practice-Feedback sequence. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 141–160). New York, NY: Cambridge University Press.
Saito, K. (2013a). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English /ɹ/ to adult Japanese learners. *Language Learning*, *63*, 499–529.
Saito, K. (2013b). Communicative focus on second language phonetic form: Teaching Japanese learners to perceive and produce English /ɹ/ without explicit instruction. *Applied Psycholinguistics*. Advance online publication. doi:10.1017/S0142716413000271
Saito, K. (2013c). Reexamining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition*, *35*, 1–29.
Saito, K., & Lyster, R. (2012a). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, *62*, 595–633.
Saito, K., & Lyster, R. (2012b). Investigating the pedagogical potential of recasts for L2 vowel acquisition. *TESOL Quarterly*, *46*, 387–398.
Schmitt, N. (2008). State of the art: Instructed second language vocabulary acquisition. *Language Teaching Research*, *12*, 329–363.
Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, *15*, 165–179.
Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research*, *10*, 361–392.
Sheen, Y. (2007). The effects of corrective feedback, language aptitude and learner attitudes on the acquisition of English articles. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 301–322). Oxford, UK: Oxford University Press.
Sheldon, A., & Strange, W. (1982). The acquisition of /ɹ/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*, 243–261.
Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language Learning*, *63*, 296–329.
So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*, 273–293.
Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *29*, 73–87.
Spada, N. (2011). Beyond form-focused instruction: Reflections on past, present and future research. *Language Teaching*, *44*, 225–236.
Spada, N., & Lightbown, P. (2008). Form-focused instruction: Isolated or integrated? *TESOL Quarterly*, *42*, 181–207.
Spada, N., Lightbown, P. M., & White, J. L. (2005). The importance of form/meaning mappings in explicit form-focused instruction. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 199–234). Berlin, Germany: Mouton de Gruyter.
Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*, 263–308.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–484). Mahwah, NJ: Erlbaum.

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning*, *62*, 1231–1258.

Trofimovich, P., Collins, L., Cardoso, W., White, J., & Horst, M. (2012). A frequency-based approach to L2 phonological learning: Teacher input and student output in an intensive ESL context. *TESOL Quarterly*, *46*, 176–186.

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in L2 Spanish word processing: Implications for pronunciation instruction. *Modern Language Journal*, *90*, 519–535.

VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, *52*, 755–803.

VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, *15*, 225–243.

Walley, A. C. (2007). Speech learning, lexical reorganization, and the development of word recognition by native and non-native English speakers. In O.-S. Bohn & M. Munro (Eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 315–330). Amsterdam, the Netherlands: Benjamins.

Wang, Y., Jongman, A., & Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, *113*, 1033–1043.

Wang, Y., Spence, M., Jongman, A., & Sereno, J. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*, 3649–3658.

Wong, M. K. S. (1982). *Tone Change in Cantonese* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Wu, X., Munro, M., & Wang, Y. (2011). The effects of linguistic experience on perceptual assimilation of lexical tone. *Proceedings of 17th International Congress of Phonetic Sciences* (pp. 2169–2172). Retrieved from http://www.icphs2011.hk/ICPHS_CongressProceedings.htm

Yu, A. C. L. (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology*, *24*, 187–214.

Zee, E. (1999). Chinese (Hong Kong Cantonese). In International Phonetic Association (Ed.), *Handbook of the International Phonetic Association* (pp. 58–60). Cambridge, UK: Cambridge University Press.

APPENDIX

The 39 target words (i.e., 13 triplets) used in the pre- and posttests and the training tasks:

Word No.	Chinese	Syllable	Tone	Gloss	Session
1	方		1	square	
2	房	/faŋ/	2	house	
3	放		4	to place	
4	貓		1	cat	
5	毛	/mau/	2	animal hair	
6	帽		4	hat	
7	輕		1	light	
8	晴	/tɕʰiŋ/	2	sunny	
9	慶		4	celebrate	Training
10	花		1	flower	
11	滑	/xua/	2	to slide	
12	畫		4	painting	
13	雞		1	chicken	
14	吉	/tɕi/	2	lucky	
15	寄		4	to mail	
16	星		1	star	
17	行	/ɕiŋ/	2	to walk	
18	姓		4	last name	
19	豬		1	pig	
20	竹	/tʂu/	2	bamboo	
21	柱		4	post, pillar	
22	八		1	eight	
23	拔	/pa/	2	to pull	
24	爸		4	father	
25	包		1	bag	Training & pre- and
26	薄	/pau/	2	thin	posttests
27	報		4	newspaper	
28	飛		1	to fly	
29	肥	/fei/	2	fat	
30	費		4	fee	
31	街		1	street	
32	節	/tɕjɛ/	2	festival	
33	借		4	to borrow	Pre- and posttests
34	趴		1	to lie	only
35	爬	/pʰa/	2	to climb	
36	怕		4	fearful	
37	七		1	seven	
38	騎	/tɕʰi/	2	to ride	
39	汽		4	steam	