LANGUAGE AND LINGUISTICS 12.3:627-667, 2011
2011-0-012-003-000296-1

Focus on Form as a Pedagogical Framework for Fostering a Native-like Mandarin Tonal Identification System*

Yeu-Ting Liu
National Taiwan Normal University

Second-language acquisition (SLA) research has shown that given short-term intensive form-focused training, non-native Mandarin speakers are able to establish a categorical tonal representation, which allows them to identify systematically features of tones borne by isolated words. However, developing a categorical tonal inventory is a necessary but insufficient condition for a native-like tonal identification system. Native Mandarin speakers' tonal inventory is well connected to the mental lexicon, which allows them to efficiently resolve any tonal ambiguities resulting from phonemic sandhi or phonetic transformation in connected speech. Existing L2 form-focused training programs, as seen in Sun (1997), have not shown positive effects on entrenching the link between the tonal inventory and the mental lexicon. Drawing upon the insights from studies conducted within the framework of *focus on form* (FonF)—a pedagogical intervention used to direct learners' attention to the formal aspect of a linguistic construction in meaning-oriented activities, this paper contends that implicit FonF, when taken into consideration of L2 learners' internal learning agenda and universal processing strategies, can provide an optimal encoding environment for internalizing intricate tonal behaviors. To this end, this paper first elucidates "what do native and non-native tonal speakers actually do?" when identifying tone in connected speech. Based upon the reviewed literature, the paper then discusses issues that need to be considered in L2 tonal instruction, and proposes how FonF pedagogical guidelines may be used to remediate problematic tonal tokens in context and hence foster an efficient tonal identification system for non-native tonal speakers—a domain rarely discussed under the FonF framework and in the SLA literature.

Key words: second language acquisition, spoken-word recognition, tone identification, tone sandhi, focus on form (FonF)

Mandarin Chinese is one of the most widely spoken languages in the world. The profound cultural resources and economic growth and potential in the Chinese-speaking

* I would like to thank ZhaoHong Han, Adam Valenstein, Sue Babcock and Shao-Ju Chen for helpful comments on earlier drafts of this paper. Thanks are also due to two anonymous reviewers for their extremely helpful comments. Any errors herein are my own.

countries have attracted more than 40 million second/foreign language learners of Chinese around the world. This vigorous learning interest, however, is severely underserved by a lack of systematic research on the nature and process of learning, and by insufficient understanding of the kind of pedagogical interventions that may tackle the idiosyncratic learning problems encountered by non-native Chinese learners. Delving into second-language (L2) processing literature and research on the tonal identification process of native and non-native tonal speakers, this paper intends to connect theory and practice, and proposes pedagogical intervention guidelines for the teaching of tone—a domain/issue rarely discussed in the literature exploring L2 Chinese teaching/learning. VanPatten (2005) posits that any instructional efforts that first seek to understand "*what* learners actually do" before developing pedagogical interventions are better than others because they allow us to take stock of insights into "*How should we intervene.*" It is the understanding of the "*what*" that shapes the way we teach and intervene, and it is the aggregate insight of the "*what*" and the "*how*" that allows us to make a difference in learners' language behaviors. In this vein, this paper first reviews relevant theoretical and empirical studies on tone identification, in hopes of shedding light on the tonal identification mechanism at work in connected speech. In view of the above, the paper then discusses the explicitness of pedagogical tonal intervention and constraining factors that may be at work in fostering a native-like tonal knowledge system, based upon insights derived from studies of *Focus on Form* (FonF) and studies on language processing/parsing.

1. Introduction

The world's tonal languages can be classified into two categories based on the shape of their pitch components: (1) *register tonal languages* and (2) *contour tonal languages*. The former is characterized by having *level tones*, and the latter by having *gliding tones*. The distinction between register and contour tonal languages is not absolute. Most languages display some of the qualities of each of these two types. In an ideal register tonal language, the pitch of a syllable hardly ever goes up or down during the production of a particular tone. This is a system in which the majority of syllables maintain the same level or register. As a result, the contour tones marking a string of lexical items can be viewed as transitions from one level to another and the actual contour is irrelevant (Yip 1980). In an ideal gliding contour language system, there is a perceptible fluctuating pitch during the pronunciation of the syllable, such as a rise, a fall, or some combination of a rise and fall (e.g. rising-falling or falling-rising).

Mandarin Chinese is a contour tonal language in which four tonal contrasts are used to distinguish meanings of words. These four distinctive contour tones are usually

named Tones 1, 2, 3, and 4 (or 1st, 2nd, 3rd, and 4th tones). According to Yip (2002), for a tone to be successfully used in spoken-word recognition, three cognitive processes are required: (1) detection of fundamental frequency (F0) fluctuation, (2) perception of pitch differences, and (3) identification of tone (see also Bao 1999). These cognitive processes move from a purely phonetic processing to a linguistic one. Specifically, detecting F0 fluctuation, measured by the number of cycles per second or Hertz (Hz), is the first perceptual process. Next, being the main cue for pitch perception, F0, along with duration, amplitude, and voice quality, results in pitch difference. While perceiving pitch difference is a purely *perceptual* process (Yip 2002), tone identification, on the other hand, is a *linguistic* process during which listeners further interpret the Register and Tone (contour) features from the previously perceived pitch information and then utilize the interpreted information for semantic access. Taking Yip's (2002) views into account, the three cognitive processes involved in tone identification can be illustrated as follows (see Figure 1):

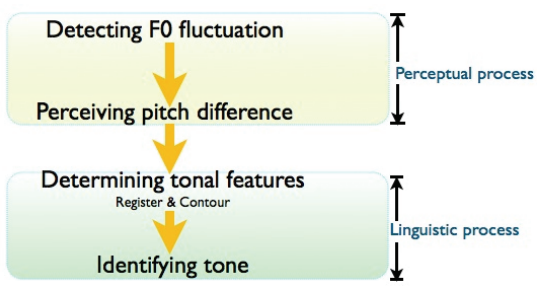


Figure 1: Yip's (2002) conception of the cognitive processes involved in a tonal speaker's identification of tone

According to Yip (2002), in a language where suprasegmental information is not used for linguistic purposes, suprasegmental information is only processed on a strictly perceptual level. At the perceptual level, listeners are only able to distinguish the relative difference between two pitches (i.e. Is the current pitch heard as high or low, the same pitch as the previous portion of the signal or different?). In a tonal language where perceived pitch difference plays a linguistic role, listeners are able to further identify the tonal features of each tonal form (e.g. register and contour features) and/or the contrasts of tonal features between two tonal forms (e.g. the contour shape is low-dipping for the first tone and mid-rising for the second).

For tonal features to be successfully perceived and identified, Mandarin Chinese speakers need to form a stable categorical boundary (for each tone) within which tonal features of a given tone (e.g. Register and Tone features) can be further analyzed in a

relative manner (Gandour 1994). These tonal categories allow speakers to define each tone not only for what it is but also for what it is not, in relation to other tones.

It is important to note that possessing a categorical representation of tone is a necessary but insufficient condition for successful tone identification in connected speech. In fast or connected speech, tones, like any other phonemic units, are not normally realized as a series of static or fixed sequences of underlying features. Tones interact and overlap with each other and with other phonological units of an utterance, such as segments, stress, and intonation. The phonetic realization of the tonal features that occur online may result in either *phonetic* or *phonemic* tonal change, both of which would lead to a partial or even complete mismatch between the input tonal features (either tone register or contour) and the stored tonal features.

1.1 3rd tone sandhi

In syllables whose duration is close to minimal (as in connected speech), contour features of a tone tend to be neutralized (or reduced), thus rendering phonetic sandhi. In cases of phonetic sandhi, tones are recognized mainly on the basis of register features, particularly the register of their starting point (Laver 1994, Gandour 1994). For instance, tone 3, a falling-rising tone, loses the rising part of its contour (and hence resulting in half-tone 3, denoted as 3') before any tone but another tone 3; in this regard, a **T3**-T1-T4 lexical string, e.g. 請收信 (please check your mail), is pronounced as "qing[**3'**]shou[1]xin[4]" instead of "qing[**3**]shou[1]xin[4]" in connected speech. Nonetheless, in such cases, this type of *phonetic change* of tonal form (i.e. phonetic tone 3 sandhi) can still be correctly resolved by a two-way mapping between the input tone and the underlying tonal inventory (without necessarily resorting to lexical cues from the mental lexicon), due to the fact that some major tonal features are still available for relative comparison.

However, in cases of phonemic sandhi, both the register and contour features of a tone undergo dissimilation transformation, thus rendering all the tonal features of the derived tone perceptually different from the underlying form of the original tone. Take phonemic third-tone sandhi (T3S) for instance. T3S occurs when all but the last syllable in a sequence of tone 3 syllables undergo a sandhi process which changes a third tone to a high-rising tone when followed by another tone 3 syllable. For example:

	兩	碗	米	
	liang	wan	mi	('two bowls of rice')
	T3	T3	T3	
(phonemic third-tone sandhi) →	T2	T2	T3	

The derived sandhi tone is perceptually indistinguishable from an underlying tone 2 at the phonetic level. Peng (2000) observed that although Chinese tonal speakers cannot reliably distinguish between tone 2 and third-tone sandhi in a perception task, they uniformly categorize the output of third-tone sandhi as tone 3 in a categorization task (see also Peng 1997). This finding suggests that native Chinese speakers can retrieve and recover the underlying forms of the (sandhi) tones even from radical phonological transformation. It is hard to explain this phenomenon if native Chinese speakers' categorization is simply based upon a phonological mental representation separate from the mental lexicon. It is possible that sandhi transformations of the original tonal forms, along with the default underlying tonal forms, are stored as part of the lexical entry in the mental lexicon (Peng 2000). Thus, in identifying tones undergoing phonemic sandhi, native tonal listeners have immediate access to the lexical information and cross-check the tonal and lexical information (in real time) to disambiguate any confusion caused by the sandhi transformation during the tone identification process. This immediate access to the lexicon allows tonal speakers to quickly categorize an input "tone3'-tone3" (or "tone2-tone3") lexical string as some particular disyllabic morphemes bearing an underlying "tone3-tone3" tonal sequence.

This tonal identification mechanism, however, is not an innate auditory perception mechanism for atonal speakers. Previous research has shown that suprasegmental units (e.g. tone and stress) are acquired prior to segments in first language acquisition (FLA), regardless of the speaker's language background. If a given suprasegmental structure (e.g. tone) is not part of a speaker's first language (L1) phonological inventory for spoken-word recognition, it will constitute one of the most difficult phonological features for the L2 learners to learn and, most likely, the last phonological feature they will acquire (see Harrison 1998, 1999, 2000). Even if the L2 learner can indeed establish a categorical tonal inventory, the acquired tonal phonology may not be totally neutralized in or attached to the mental lexicon. Consequently, difficulties would arise in identifying allotones (positional variant of a toneme) that undergo phonemic changes (e.g. the sandhi tone) in connected speech.

To enhance our understanding of how lexically contrastive prosody is identified in tonal languages, this paper discusses findings of research on tone identification by native and nonnative speakers of Mandarin Chinese. In particular, the connection between tonal inventory and lexicon in native and non-native tonal identification processes will be discussed in greater detail. Additionally, theoretical account and pedagogical implications for fostering a native-like tonal identification system will be discussed under the framework of "Focus on Form" (FonF) (Long 1991).

1.2 Interaction between the tonal inventory and the mental lexicon

From the perspective of phonology, one of the most interesting, and perhaps most challenging hurdles in mastering the Mandarin tonal system is encoding and decoding the sandhi tones. Phonemic third tone sandhi (T3S) is a case in point. The difficulty in encoding/decoding T3S stems not from the relatively simple rules which describe it, but rather from the *cyclical* application of T3S rules of connected speech. The cyclical application of T3S can be illustrated in the following examples (that are both comprised of three 3rd tone syllables). For instance:

(1) 我很好 (wo[**3**] hen[3] hao[3] → (wo[**3'**]) (hen[2] hao[3])
'I am fine.'

(2) 老闆好 (lao[**3**] ban[3] hao[3] → (lao[**2**] ban[2]) (hao[3])
'How are you doing, boss?'

In example (1), phonemic tone 3 sandhi (T3->T2) applies first within the two syllables *hen hao*, changing the tone 3 on *hen* to tone 2. This tone sandhi affects the (phonemic) T3S environment (T3->T2) on the syllable *wo*; *wo*3 precedes *hen*2 (see example (3a)). Thus, phonetic tone 3 sandhi (T3->T3') applies to *wo* instead of phonemic T3S (T3->T2). The resulting lexical string thus carries T3'-T2-T3 tonal pattern (example (3b)).

(3) a. 我很好 wo[3] **hen[3] hao[3]** ➔ wo[3] **hen[2] hao[3]**
b. 我很好 **wo[3] hen[2]** hao[3] ➔ **wo[3'] hen[2]** hao[3]

In example (2), phonemic tone sandhi applies first within the two syllables *lao ban*, changing the tone 3 on *lao* to tone 2 (see example (4a)); the tone on *ban* is still 3 when T3S (3➔2) applies. Thus, unlike example (1), this sandhi process does not affect the phonemic T3S environment for the remaining character (*hao*). The cyclical application of the tone 3 sandhi rule thus yields a T2-T2-T3 tonal pattern (example (4b)).

(4) a. 老闆好 **lao[3] ban[3]** hao[3] → **lao[2] ban[3]** hao[3]
b. 老闆好 lao[2] **ban[3] hao[3]** → lao[2] **ban[2] hao[3]**

If the tone sandhi phenomenon involves simply a static rule that can be universally applied to lexical strings of the same length in connected speech, we would expect uniform tone 3 sandhi (tone2-tone2-tone3) in both of the 3-character lexical strings in examples (1) and (2). However, this is not the case. As seen above, different sandhi processes apply to examples (1) and (2) in connected speech. Early approaches attempt

to account for the above T3S phenomenon purely in syntactic terms (e.g. Cheng 1973), positing that the cyclic application of sandhi rules is determined by syntactic junctures. This explains why the first round of sandhi in example (1) applies to *hen hao* (a predicate) instead of *wo hen* (which is not a legal tone sandhi domain enclosed by two legal syntactic boundaries) and why in example (2) to *lao ban* (subject) instead of *ban hao*. Other researchers, however, contend that the syntactic account does not satisfactorily address why T3S applies first to the predicate (*hen hao*) in example (1), but to the subject (*lao ban*) in example (2) (Brooke, Coppola, Lee & Zhao 2009, M. Chen 2000, Shih 1997). Shih (1997) thus proposed that the cyclic application of T3S is determined by the rhythmic structure of a sentence; T3S applies first to the "lower domain" before the "higher domain" and that the "binary feet" is the lowest minimal rhythmic unit (MRU) within which T3S applies first.

Whether or not researchers draw on a prosodic or syntactic approach to capture the (minimal) T3S domains, they do generally agree that tone sandhi must involve cyclical applications of T3S and that the minimal T3S domain, a tone sandhi domain within which tone sandhi shows no optionality, is a perceived lexical unit bounded by plausible breaks or pauses in connected speech. This suggests that in identifying tones with connected speech (especially in cases of allotones or sandhi tones), native tonal speakers need to actively interact with the lexical/syntactic information (saved in the mental lexicon) and determine the appropriate tone sandhi domains. Without building a tonal inventory inseparable from the mental lexicon, online tonal identification in connected speech is extremely difficult.

The close connection between the tonal inventory and the lexicon in native tonal speakers' tone identification process has been empirically established by Fox & Unkefer's (1985) study. Fox & Unkefer asked native Chinese speakers to determine the lexical status of tokens on sound continua. Sound tokens of a sound continuum were created in the following manner: first, a word (e.g. /hei**1**/; black) and a non-word (e.g. /hei**2**/), which contrasted only in tone, were set as the end-point tokens of each sound continuum. Next, the remaining sound tokens shared the same segments with the two-end point tokens (/hei/), but the tones they carried were synthesized from the tonal continuum between the two end-point tokens (tone 1 and tone 2) and hence were ambiguous in terms of tonal identity (see Figure 2 below). Fox & Unkefer found a "lexical effect" on the Mandarin speakers' tone categorization. Specifically, when presented with synthetic sound tokens that were ambiguous between two tones, the Mandarin speakers' tone categorization was biased toward choosing the tone that would make the ambiguous token a real word (e.g. /hei1/ instead of /hei2/). Their finding, therefore, supports that the acquired tonal inventory interacts with the mental lexicon in resolving tonal ambiguity.

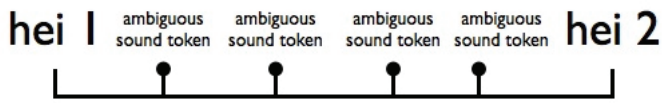


Figure 2: Manipulations of synthetic tones in Fox & Unkefer (1985)

Despite the close connection between the tonal system and the mental lexicon in connected speech, cognitive studies have empirically established that the two mental representations are not stored in the same mental space. Drawing on the speech error data collected from a Chinese radio call-in program over a period of one year, J. Chen (1999) observed that although errors involving segmental constituents were abundant in the corpus (136 slips), errors of tone were rare and were seldom affected by the movement of segments. Moreover, the nature of tonal errors was different from those of segmental units (see also M. Chen 2000). Specifically, errors involving consonants, vowels, and various syllabic constituents included anticipation, preservation, and exchange. However, it was found that, of the twenty-four "suspected" tonal errors, only four could be considered true movement errors (tone anticipation or preservation), the rest could be attributed to other processing errors such as blending, haplology, malapropism, and misapplication of tone sandhi rules. J. Chen (1999) contends that the immunity of tone to speech production errors suggests that lexical tones in Mandarin Chinese are represented and processed differently from other lexical and segmental elements. Accordingly, a native tonal identification system used in connected speech would involve a processing architecture as follows.

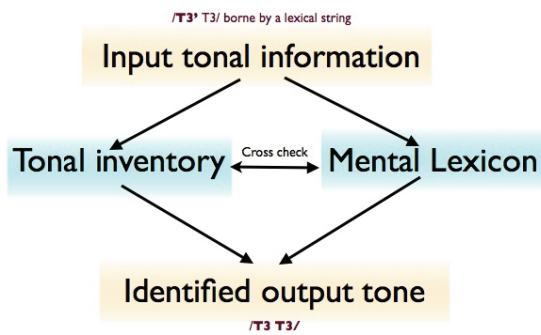


Figure 3: A tonal speaker's identification of tone in connected speech

It is important to note that the strength of the interaction between the tonal inventory and the mental lexicon does not remain the same while native tonal speakers perform listening tasks of a different nature. Some empirical studies have shown that the contextual constraint of the presented stimuli determines the strength of this link.

Findings of some studies have shown that processing spoken words in the contextually constrained context (e.g. words in context or idioms) results in stronger association between the acquired tonal inventory and the lexicon than those in the contextually unconstrained context, e.g. words in isolation (e.g. Ye & Connine 1999). When the association is strong, tone can be more quickly and accurately identified and then more readily processed to constrain lexical competition (early temporal availability of tone). This assumption has been examined in a few empirical studies (e.g. Ye & Connine 1999; see also Lee 2002, 2007; for discussion of this issue regarding written Chinese character recognition, see Liu 2007, 2009). Evidence thereof is meager but suggestive: the interaction between the tonal inventory and the mental lexicon is more crucial in resolving the tonal ambiguity in the contextually constrained condition (e.g. identifying tones in connected speech) than in the contextually unconstrained condition (e.g. identifying tones borne by isolated words).

In Ye & Connine's (1999) study, clear evidence supporting effects of contextual constraints on tone identification was empirically established. In the first experiment (a sound monitoring task), Ye & Connine asked L1 Chinese speakers to monitor whether or not an existent or nonexistent monosyllabic word contained a given tone-plus-vowel combination: "**tone 2 + /a/**" (e.g. /b**a2**/ vs. /r**a2**/). Overall, the reaction times were faster for existent syllable carriers bearing the target "**tone 2 + /a/**" (e.g. /b**a2**/). Non-target (incorrect) stimuli were existent syllables that were mismatched either in terms of the vowel (/a/) or the tone (tone 2), such as /b**o**2/ and /ba**4**/. In rejecting the non-target stimuli, the participants showed faster responses and higher accuracy rates for the vowel mismatch stimuli than the tone mismatch stimuli, suggesting that segmental (vowel) information was available prior to tonal information. The above finding indicates a disadvantage for tone identification in processing isolated lexical string detached from context (late temporal availability of tone).

In the second experiment, separate vowel and tone monitoring tasks were used for target syllables that occurred either in the phrase-final positions of an idiomatic (contextually constrained) phrase (e.g. 過河拆橋 /guo4 he2 chai1 qiao2/) or in a contextually unconstrained phrases. Contextually unconstrained phrases were composed of a string of characters that did not have any lexical entry (i.e. a meaningless character string such as 夏眠師國 /xia4 mian2 shi1 guo2/). In these phrases, vowel monitoring was again much faster and more accurate than tone monitoring. Interestingly, in idiomatic phrases, where an enhanced contextual constraint/support was provided, a reverse observation was made: the tone disadvantage effect disappeared. That is, the presence of a contextually constrained condition renders tonal information more accessible when compared to vowel information and thus permits the privileged role of tone in constraining lexical access to emerge.

Similar findings are also found in Cutler & Chen's (1997) study where native Cantonese-speaking participants were asked to decide whether a given acoustic stimulus (an isolated two-character lexical string detached from a meaningful context) was a word or a non-word. The non-word items could differ from real-word items (e.g. 博士 /bok8-si6/) in various ways, including onset consonant (e.g. /bok8-**j**i6/), nuclear vowel (e.g. /bok8-s**y**6/), tone (/bok8-si**2**/) or some combination of these (e.g. /bok8-**jy2**/ or /bok8-**jy**6/). When only a single difference was present, the participants had the most difficulty in rejecting the non-words if the only difference between them and corresponding real-word items was tonal. The next poor performance came when the only difference was in the nuclear vowel. According to Cutler & Chen, the above finding indicates that tonal information is identified *after* the segmental (vowel) information in processing *isolated* spoken words; this in turn suggests that the link between the tonal system and the mental lexicon was weaker in the contextually unconstrained condition.

Findings from Cutler & Chen (1997) and Ye & Connine (1999) collectively confirm the view that the link between the tonal inventory and the mental lexicon may be more crucial and stronger in the contextualized (e.g. recognizing words in connected speech) than in the decontextualized context (e.g. isolated word recognition) and stress the need to entrench this link and to implement any tonal pedagogical interventions in a context-rich setting.

1.3 Tone identification by non-native speakers

Research using a form-focused training paradigm, in which tonal rules are explicitly explained/taught to learners in de-contextualized settings, has uniformly suggested that tonally naïve adults *can* develop a target-like categorical tonal inventory. Using a tone perception task, Leather (1987) investigated whether Dutch and English speakers who had no previous exposure to a tonal language would be able to establish the same categorical boundary as Mandarin speakers in contrasting words with different Chinese tones. He found that his 10 English and 10 Dutch participants established the same lexical categorical boundary in perceiving the Chinese tones as Mandarin speakers did after receiving short-term intensive training in distinguishing these tones. Similarly, Lu (1992) found that with short-term, explicit tonal instruction, tonally naïve adults could develop a categorical tonal inventory that allows them to distinguish tones in their citation forms—perfect phonetic realizations of tonal features borne by isolated syllables with emphatic stress (see also King 1981, Lin 1985).

Despite the above possibility, what remains unclear is whether non-native tonal users receiving explicit form-focused tonal instruction can entrench the link between their mental lexicon and tonal phonology and efficiently cross-check the information

from both sources to resolve tonal ambiguity in connected speech. Sun (1997) observed that L2 learners of a tonal language, including highly proficient ones who had received long-term form-focused tonal instruction, often misinterpreted a rising tone (tone 2) as a tone 3 and a tone 3 as a tone 2 in connected speech,[1] while native speakers of Mandarin Chinese only had occasional problems with the latter (tone 2 substitutions for tone 3). According to Sun, the bi-directional error pattern (substitution of tone 2 with tone 3 and substitution of tone 3 with tone 2), as observed in the L2 learners' tone identification, is suggestive of non-optimal working memory capacity and a deficient tonal identification mechanism whose link to the mental lexicon is not firmly entrenched. In other words, while explicit rule-based tonal instruction may help L2 learners develop a categorical tonal inventory, the decontextualized, form-focused tonal instruction setting does not seem to efficiently help the learners entrench the link between the input tone and corresponding lexical-semantic environment—a crucial feature in a native-like tonal identification system.

When such a link is not readily available in online tonal identification process, the tone identification would require more processing cost; L2 learners would have the most difficulty in resolving tonal ambiguity in connected speech where immediate tone identification is required. A possible consequence for this (tonal) processing deficiency is that the semantic candidates that an L2 learner needs to resolve for each Chinese character would involve many "redundant" lexical meanings associated with other phonologically similar characters that differ in tone (Lee 2002, 2007, see also Liu 2009). This in turn would increase L2 learners' processing loads in spoken-word recognition, thereby negatively constraining available mental resources for higher-order thinking. In addition, non-native speakers' tonal processing deficiency may also have a negative impact on their working memory capacity. Psycholinguistic studies have shown that people make use of the phonological loop to rehearse and retain the decoded lexical information in their working memory (Baddeley & Hitch 1974, Xu et al. 1999). When the perceived lexical information (i.e. input tonal features) can be held or rehearsed long

[1] L2 learners' difficulties in identifying Tone 3 may be related to the fact that Tone 3 undergoes the most context-dependent variation of all the four tones (in the non-primary stressed position, the syllable-final rise in the contour is not realized). L2 learners may fail to recognize this so-called "half-third" tone in the non-primary stressed position as a third tone. Even in primary stress positions, where the syllable rise is realized, L2 learners are still prone to perceive third tone as second (rising) tone because the dip in the contour often goes unnoticed. Conversely, second tone is also a difficult tone for L2 learners to perceive because the rising slope in its contour can make it difficult to distinguish from a similar rising slope in the third (dipping) tone. The similarities between these tones explain why beginning L2 learners of Chinese, like young Chinese speakers, have particular difficulty in keeping the second tone distinct from the other third tones.

enough in the working memory, listeners have a better chance to cross-check the underlying tonal inventory with the lexical entries saved in the mental lexicon, thereby resolving the subtle variability tones manifested in running speech (e.g. the allotones resulting from T3S). The efficacy of phonological rehearsal is optimal only if and when the segmental *and* tonal information of a Chinese character can be efficiently accessed online (Xu et al. 1999). Accordingly, helping non-native tonal learners acquire an efficient, native-like tonal identification system is thus a crucial task when developing any pedagogical tonal interventions.

In hindsight, explicit form-focused tonal instruction is quite effective in helping non-native adult Mandarin speakers establish categorical tonal mental representation. In particular, current explicit form-focused tonal training programs have been quite effective in introducing/consolidating knowledge of *novel* tonal features to help tonally naïve speakers. The paper therefore does not intend to propose that explicit lead-in, teaching, and explanation of Mandarin tonal variation are of little merit. Given that explicit form-focused (tonal) instruction is quite effective when inducing learners' attention to *novel* (tonal) formal features, it is therefore more suitable for initial teaching practice, helping tonally naïve learners effectively establish categorical Mandarin tonal inventory in the short term. Despite the initial benefit of explicit tonal instruction, the present paper argues that explicit form-focused tonal instruction alone is a necessary but insufficient condition for fostering a native-like tonal identification system; other (follow-up) complementary/ supplementary pedagogical practice should be considered to help learners notice and internalize the intricate lexical/syntactic environment in which tonal variation occurs and hence entrench the link between tonal inventory and mental lexicon. Explicit form-focused instruction, as seen in Barcroft (2006), is quite limited in helping students to make form-meaning mapping (in the case of developing a native-like tone identification system, mapping the acquired tonal system with relevant lexical/syntactic environment). Barcroft carried out a series of experiments in which participants were asked to learn/remember a set of Spanish (L2) words with corresponding pictures (form-mapping task). One group was asked to study the features of picture-word pairs drawing on their world knowledge and idiosyncratic learning/retention strategies; the other group was asked to write down each word as the word was presented to them. Interestingly, the group that focused primarily on (written) formal features did not outperform the other group at the later retrieval task. Barcroft thus suggested that the learning of form-meaning mapping requires a kind of processing that cannot be easily obtained in exclusive form-focused learning settings.

If explicit instruction is necessary/helpful but insufficient, then, pertinent questions arise: What kinds of 'complementary' pedagogical intervention can be used, along with explicit instruction, to foster a native-like tonal identification mechanism? Why does

decontextualized explicit rule-based tonal instruction only seem to lead to a (non-native) tonal identification system whose link to the mental lexicon is not well-entrenched (and hence does not contribute to automatized/efficient tonal identification)?

2. Focus on (L2 tonal) form

Answers to the above inquiries need to be examined through the information-processing accounts. Researchers working from this perspective have attempted to examine two major issues: (1) whether external manipulation of input (explicit vs. implicit) triggers explicit and implicit modes of internal processing, which are differentiated by the presence (and/or the lack thereof) of awareness on the part of the learner; and (2) whether external manipulation of input fosters explicit and implicit knowledge or mental representation. Explicit (declarative) knowledge is verbalizable and needs to be retrieved with awareness and effort (Paradis 1994); a non-target tonal identification system, whose link to the mental lexicon is not well-entrenched, is based upon such a mental architecture. Implicit (procedural) knowledge, on the other hand, is unverbalizable and can be used without awareness (in an automatic manner); a target-like tonal identification system is assumed to operate on such a mental structure.

Attempting to address the above two issues in a holistic account, Krashen (1985) stipulates that *input*, *processing*, and *knowledge* work in tandem: pure exposure to implicit input (derived from the naturalistic acquisition/immersion environment) was assumed to be processed implicitly (*without* directing learner's attention to the formal aspect of the target structure) and to lead to implicit/procedural knowledge which can only be retrieved and executed in an automatic manner. Explicit input (derived from the form-focused instructional setting) was considered to be processed explicitly (with attention to form) and to result in explicit/declarative knowledge that cannot be accessed in an automatic manner and is only available for offline self-correction. The above views entail that language (sub)systems (in this case, tonal identification system) developed in explicit instructional settings do not lead to an implicit procedural knowledge system and are not readily available for online/real-time processing in connected speech. In this vein of reasoning, explicit/declarative tonal knowledge system cannot be proceduralized, as a result of which, efficient, automatic, and effortless tonal identification in connected speech is not possible for learners only receiving explicit instruction.

The view that explicit input only leads to explicit knowledge and implicit input (pure exposure) only results in implicit knowledge implies that input which aims to direct learners' attention to the target forms cannot provide an optimal environment for fostering a native-like tonal identification system. Research findings, however, have not empirically established the above view; furthermore, there is *no* evidence in the second-

language acquisition (SLA) literature showing that explicit input cannot lead to automatized procedural knowledge (Lightbown 2005). In fact, implicit input which aims to expose learners to the target structure in naturalistic/immersion settings does not always facilitate second language acquisition; and, in many cases, it leads to fossilization, which is more so in adult/late second language acquisition (e.g. Han 2004, Philp 2003). The above findings suggest that pure exposure/immersion may not provide (late) learners of a tonal language with an optimal learning setting, and that there may be some constraining factors in the explicit instructional condition that prevent (late) L2 learners, such as those in Sun (1997), from attaining a native-like language processing (sub)system. In view of the above, the ensuing section will discuss possible constraining factors which may be at work in the explicit instructional setting and then explore how such constraints may be lifted, in hopes of providing pedagogical suggestions for fostering a native-like tonal identification system for connected speech.

The essence of explicit input is to direct learners' attention to features of the L2. While most SLA researchers agree that some kind of attentional process is required for input to become intake, they vary with regard to the kind of practice that can lead to procedural knowledge and "how" attentional resources should be engaged during the pedagogical intervention period (i.e. Focus on form**s** vs. Focus on form).

Focus on form<u>s</u> (FonF<u>s</u>) stems from the traditional way of teaching linguistic elements where decontextualized explicit rule instruction and correction of the target structures are considered to be the best way to induce learners' conscious, and focal attention. Some proponents of FonFs have proposed that noticing at the level of awareness is necessary for converting input into intake and for converting declarative knowledge into procedural knowledge (e.g. Schmidt 1990, 1993). In other words, learners must first demonstrate a conscious apperception of the target form (e.g. ability to perceive Tone 2-Tone 3 contrast and vice versa) before the form can be proceduralized for immediate use/retrieval in subsequent online language processing (automatically tackling tone sandhi in connected speech). If Schmidt's argumentation—that is, directing students' attention to the target structure is a *necessary and sufficient* condition for second language development—is correct, then non-native tonal speakers, such as those in Sun's (1997) study, would not exhibit non-target bi-directional tonal errors (in connected speech) even after receiving a substantial amount of explicit tonal instruction. Perhaps the key issue that is not fully addressed by Schmidt's account revolves around the notion that explicit input from whose perspective is unclear. What is made explicit and salient by the teacher (external salience) may not be perceived as salient by the learners (internal salience). Research has shown that externally created salience (by the teacher) does not always guarantee learners' internally generated salience (focus of attention driven by the learner's internal syllabus or processing preference) (Sharwood-Smith

1993). Thus, attention to the target structure (tone sandhi) is at best a necessary, but an insufficient condition for converting the attended input into intake. There are several constraining factors that need to be considered *a priori* in (re)directing learners' attentional resources. Undoubtedly, 'the context' in which the learners' attention is engaged is a major issue. Taking this issue into consideration, Long and his colleagues (Long 1991, Long & Robinson 1998) proposed **Focus on form** (FonF).

Focus on form (FonF) distinguishes itself from Focus on forms (FonFs) in terms of the context in which the target structure is presented and attended. While practitioners of FonFs often direct learners' attention to the target structure in a discrete manner and in decontextualized context (e.g. teaching tonal contrast in decontextualized context), practitioners of FonF often adopt a task-based or content-based syllabus and only *briefly* draw learners' attention to the (problematic) target structures as they arise in activities whose overriding focus is on meaning or communication (Long 1991). While traditional FonFs approaches tend to lead to limited fluency, and communicative approaches (pure exposure/immersion) tend to result in less accuracy, FonF instruction neatly provides an alternative to the methodologies which treat fluency and accuracy separately. The FonF intervention can be manipulated along what Doughty & Williams (1998) called "*explicitness continuum*" or "*FonF continuum*," ranging from the most explicit metalinguistic feedback, explicit explanation of the rule (**explicit FonF**), to the least explicit FonF feedback (**implicit FonF**). The manipulated FonF technique is more implicit when the well-formedness or salience of the target structure is enhanced by means of reformulation (i.e. recast), increasing its frequency (input flood) or by way of typographic (see Jourdenais et al. 1995) or prosodic input enhancement (see Leeman 2003). Regardless of its orientation (explicit vs. implicit), FonF is premised on the notion that learners must comprehend what they read/hear *before* their (focal or peripheral) attentional resources can be directed to the target form (VanPatten 1996, 2004). Its underlying thrust being to induce occasional attention to the formal aspect of the target linguistic construction in the meaning-oriented (learning) task condition; once the learner's attention is successfully drawn to the target form, what is attended will stand a better chance to translate into acquisition (Alanen 1995, Jourdenais et al. 1995, Leow 2001, Izumi 2002).

When employing overt, rule-based techniques, practitioners of explicit FonF aim to induce learners' conscious, focal attention to the formal aspect of the target structure. Practitioners of implicit FonF, on the other hand, intend to trigger learners' cognitive registration of the stimuli without perforce involving awareness (i.e. a cognitive state coined *detection*). The issue of whether conscious focal attention is required for second language acquisition (SLA) has an important bearing on the overtness of the FonF technique. Many studies have shown that explicit, overt FonF may facilitate the

acquisition of categorical rules, e.g. categorical tonal contrast (e.g. DeKeyser 1995, Jourdenais et al. 1995, Robinson 1996, 1997). However, it is important to note that such beneficial effects are observed only in form-oriented tasks which are *not* highly *communicative* per se, in which case immediate access to the acquired knowledge is often not required, e.g. grammaticality judgment tasks (Alanen 1995), multiple-choice (Bowles 2003), picture-cued written production (Jourdenais et al. 1995).

The above finding provides an important implication for tone instruction: the intake, if any, of the learners receiving explicit FonF instruction (e.g. providing brief metalingustic explanation of tonal rules as learners attempt to make form-meaning connection in a communicative task) would not allow them to *efficiently* retrieve the learned (tonal) knowledge in communicative tasks that require *immediate* lexical access—and this is a result similar to the outcome of rule-based (tonal) FonFs instruction. According to Lightbown (2005), learners encode both aspects of the learning contexts (e.g. rule-based vs. token-based learning)[2] and the learning processes present during learning; retrieval of the acquired (tonal) knowledge is facilitated only if and when the retrieval context and processes resemble those that were present during learning; if learners' attention is directed to the target structure via rule-based meta-linguistic FonF instruction (i.e. explicit FonF), they would encode the target structure in the form of "rules" (or declarative knowledge) and require more processing cost to retrieve and/or compute rule-based knowledge (which have not been proceduralized) for immediate language use.

In a similar vein, Segalowitz & Lightbown (1999) posit that "the learning environment that best promotes *rapid*, accurate retrieval of what has been learned is that in which the psychological demands placed on the learner resemble those which will be encountered later in *natural* [communicative] settings" (p.51; emphasis added). The above view has been empirically established by cognitive processing research. Morris et al.'s (1977) study is a case in point. In this study, participants were presented with two types of questions such as "Does the word rhyme with *train*?" (phonological orienting questions) and "Is the word a type of flower?" (semantic orienting questions) before they were asked to remember words on a list ('word learning task'). While phonological orienting questions aimed to indirectly direct learners' attention to the formal feature of a word, semantic orienting questions intended to induce learners to process the meaning of a word. Then, the participants were asked to identify new words that rhymed with words they had seen during the learning phase. Morris et al. found that those whose

[2] In token-based learning, if a sequence type "tone1-tone2-tone-3-tone4" appears 10 times and "tone1-tone2-tone4" twice in the corpus, "tone1-tone2" is counted as 12 times, "tone2-tone3", "tone3-tone4" 10 times, and "tone2-tone4" twice; in type-based learning, on the other hand, "tone1-tone2" is only counted twice, "tone2-tone3", "tone3-tone4" and "tone2-tone4" once.

attention was directed to the form of a word (rhyme) were signficantly more successful than those whose attention was directed to the meaning. Morris et al. therefore concluded that successful retrieval was achieved when the retrieval condition was similar to the learning condition. Following this line of reasoning, if the connection between the tonal inventory and the mental lexicon is crucial for resolving tonal ambiguity in communicative connected speech, such a connection will be much more likely to be *efficiently* and *successfully* retrieved for resolving tonal identity in connected speech only when the connection is attended and encoded in a similar "noisy" communicative environment—a learning/encoding context that can be made possible via implicit FonF instruction.

Similar results were also obtained in Trofimovich (2005). Trofimovich's participants were presented with a list of L2 (Spanish) words under three learning/encoding conditions. One group was merely asked to listen to the words (Just Listen). A second group was asked to ponder over each word they heard and then assign a score to each word according to its 'pleasantness' (semantic group). The third group was asked to give each word a rating according to the sound quality of the recording (Auditory). Following the learning phase, each participant was asked to repeat/name lists of (Spanish) words, as quickly and accurately as possible, and the word lists included both primed words (those that they heard during the learning phase) and new words (familiar words that were not on the original word list). Trofimovich found that the participants in the Just Listen and Auditory groups, whose attention was directed to the sound quality of the stimuli, exhibited significantly faster response times when naming the primed words. However, the Semantic group, whose attentional resource was directed to the semantic information of the words during the learning phase, did not name the primed words significantly faster than the new words. Trofimovich's findings again suggested that optimal learning occurred when the cognitive processes that were engaged during the learning context resembled those that were called on during the retrieval context; and this promoted automatisation of the learned information (as evidenced in the learners' faster response time)—an important feature for native-like tonal identification system (for similar findings, please also see Barcroft 2002, 2004).

Thus, based on the evidence from cognitive processing literature, we can say that the most primary purpose of implementing implicit FonF intervention, which unobtrusively integrates form focus into communicative context, is to provide atonal learners with a context-rich environment to attend and encode the intricate tonal behaviors along with relevant linguistic environments. In so doing, instructors create a match between the processes and conditions that are present during learning and those that are present at the time of retrieval. Cognitive processing research (e.g. Barcroft 2002, 2004, 2006, Morris et al. 1977, Trofimovich 2005) has empirically established that the match in

learning and retrieval condition is conducive to promoting automaticty in language retrieval (thereby allowing tonal speakers to efficiently retrieve/access relevant lexical and tonal information for online tonal identification in connected speech). More importantly, it can result in "richer, more contextulized (mental) representation of the target form," which will provide ready-to-use retrieval cues for later contextualized language processing tasks (Lightbown 2008). Following this line of reasoning, the intake, if any, resulting from unobtrusive *implicit* FonF may be better able to promote a procedural (tonal) knowledge system which can be accessed with ease and rapidity.

Here, an empirical question arises: how do we ensure that implicit FonF, which is often brief and unobtrusive, can indeed tap into learners' attention to the highlighted tonal patterns and hence promote the correct encoding of the observed tonal behaviors along with relevant lexical/syntactic environments? As Williams & Evans (1998) point out, "not all forms are equal in terms of the effectiveness of FonF activities" (p.151). If this is the case, is the tonal knowledge/processing system amenable to implicit FonF interventions? Although answers to the above inquiry have not been empirically established, findings of available studies have produced evidence that may shed some light on this inquiry.

Studies that explore the effect of implicit FonF (e.g. textual input enhancement) have shown mixed findings, which seem to be modulated by the "semantic value" of the target structure in question. Drawing upon data collected from the picture-based writing task and concurrent think-aloud report, Jourdenais et al. (1995) found that English learners of Spanish assigned to typographic input enhancement group noticed and produced more target forms (Spanish preterite and imperfect verb forms) than those assigned to the control. Similarly, manipulating a number of implicit typographical input enhancement techniques (color-coding, underlining, highlighting), Leeman et al. (1995) examined the effect of implicit FonF on the acquisition of Spanish preterite and imperfect tenses, and observed that students assigned to the content-based FonF class outperformed students assigned to the regular content-based class (control). Despite the above positive evidence, employing similar implicit FonF techniques, White (1998) and Alanen (1995) did not successfully promote L2 learners' acquisition of English third person singular possessive determiners and the locative suffix and consonant alternation rules of a semi-artificial language. A first glance at the findings of the above studies seems to suggest that implicit FonF does not always lead to learners' noticing of the target form. However, a closer look at the target structure explored in these studies may elucidate the constraint(s) that may be at work in achieving successive implicit FonF. While the target structure in Jourdenais et al. (1995) and Leeman et al. (1995) concerns a frequent and semantically important grammar focus, the possessive determiner structure in White's (1998) study is, comparatively speaking, semantically redundant and the

consonant alternation rule in Alanen's (1995) study is "semantically empty" (p.269). This observation underscores the fact that the implicit FonF intervention is more successful and effective in inducing learners' attention to linguistic forms which carry high semantic value and learners are more prone to notice linguistic structures with high semantic value; in particular, in processing meaningful linguistic structures/forms, learners are more likely to process (notice/detect) forms that make semantic distinction (e.g. tone) before they process "semantically *redundant* forms" (e.g. stress) (for further details, see VanPatten 2004, 2005).

Granted, existing implicit Focus on Form (FonF) studies have focused primarily on other formal aspects of language (i.e. morphosyntactic structures); although these studies were not specifically conducted to investigate the effect of implicit FonF instruction on tonal learning, their findings do, however, shed light into the efficacy/feasibility of implicit FonF tonal instruction. As noted above, existing FonF studies have empirically established that implicit FonF intervention is quite effective especially when inducing learners' attention to morphosyntactic forms which carry high semantic value; in fact, some studies even have found that implicit FonF is able to remediate grammatical forms that have low perceptual salience and limited communicative value (e.g. Spanish binary gender markers for inanimate vs. animate nouns in Leeman 2003). It is thus conceivable that implicit FonF (such as prosodic input enhancement) can be quite effective when drawing learners' attention to tone—a decisive, meaningful pitch pattern which makes semantic distinctions and is hence rich in semantic value. Such an argument is based on findings of empirical FonF studies that deal with formal aspects of language that are similar to tone: the target structures at focus are all formal aspects of language that make semantic distinctions and are therefore rich in semantic value (e.g. Leeman et al. 1995, Leeman 2003, Jourdenais et al. 1995); their actual surface realization (in running speech) is highly sensitive to contingent lexical/syntactic environment and does not simply involve a static/regular rule application (e.g. Leeman 2003). It is therefore feasible to expect that non-target tonal features should be amenable to implicit FonF interventions.

In addition to the semantic value inherent in tone, learners' developmental readiness is, *inter alia*, a major factor that crucially determines the success of implicit FonF in drawing learners' attention to the allotone or tone sandhi in connected speech. Williams & Evans (1998), Révész (2007), and Park (2004) all observed that learners who make the greatest gains with implicit FonF tend to be those who already have *partial* mastery of the target form, rather than *ab initio* learners (e.g. in the case of tonal instruction, learners who have partially mastered the target tonal patterns). Accordingly, while explicit form-focused (tonal) instruction is more effective in inducing learners' attention to novel (tonal) formal features and hence is more suitable for the initial phase of

teaching practice (King 1981, Leather 1987, Lin 1985, Lu 1992), implicit FonF is more effective when the target (tonal) form is partially acquired by the learners (and is therefore more suitable for the follow-up remedial intervention). Learners are more ready to notice and internalize target structures that they are ready to learn; after all, noticing is primarily driven by the individual learner's developmental readiness vis-à-vis the target form (internal salience), rather than externally induced salience of the target form. This highlights the importance of matching instructional intervention to developmental readiness. The empirical question that follows is how this match can be achieved when applying implicit FonF. To successfully tap into learners' (sub)conscious saliency (and hence striking a harmony between externally and internally generated salience), teachers need to be *selective* when determining the focus of the implicit FonF tonal intervention. Specifically, teachers should be aware of the learner's built-in syllabus, and should focus primarily (but not exclusively) on the developmentally appropriate tonal features for learners at focus and/or the tonal structures that learners have *partially* mastered.[3]

The above reasoning also suggests that to enhance the congruence between the externally and internally generated salience, there is a need for a *pro-active* or planned FonF intervention. Practitioners of pro-active (implicit) FonF usually have a specific predetermined linguistic focus in mind (in this case, developmentally appropriate tonal features partially acquired by learners) and will design communicative tasks that will ensure that opportunities to attend to and produce the (problematic) target tonal features will indeed arise. Reactive or incidental FonF, in contrast, does not target any specific focus; feedback is usually given to students in an unplanned manner—teachers provide immediate feedback on *any* problematic features as they arise (in a communicative context). Thus, in applying incidental or reactive implicit FonF to tone teaching, teachers may attempt to cover a variety of tonal errors, including both incorrect tonal patterns resulting from "slip of the tongue" (coincidence) and incorrect tonal patterns due to incomplete mastery (persistent error). As a result, many forms rather than a particular form will be made salient to learners within any given communicative activity; in this case, learners' attentional resources may not be consistently directed to the most problematic pattern(s). It is important to note that research on implicit FonF has shown that the effectiveness of implicit FonF depends primarily on *focused* and *consistent* treatment of the same target form; when a single form is targeted, learner's uptake is more likely to occur (Doughty 2004, Han 2002).

[3] The evidence obtained from Williams & Evans (1998) and Park (2004), however, should not be stretched to mean that focus on form that does not match learners' developmental stage is necessarily counter-productive *in all cases*.

Therefore, to maximize the efficacy of implicit FonF, the claim of early FonF proponents—drawing learners' attention to (any) problematic forms "as they arise accidently in lessons whose focus is on meaning or communication" (Long 1991)—should be changed. Teachers ought to pro-actively identify problematic form(s) that require(s) further treatment and decide the focus of implicit FonF before designing any teaching or intervention activities (Doughty & Williams 1998).[4] In other words, prosodic input enhancement (and/or other implicit FonF feedback) should *selectively* target particular forms—ideally one at a time—as opposed to any forms *indiscriminately* (Han et al. 2008). Problematic exemplars, in the scenario of tone identification in connected speech, include, but are not limited to, tone2-tone3 distinction (and vice versa), and incorrect application or overuse of phonemic Tone 3 sandhi in the obligatory context (兩碗米 pronounced as */liang3 wan2 mi3/ instead of /liang2 wan2 mi3/). In each treatment session, teachers ensure that opportunities to attend to and produce the target form (e.g. T3S involving two disyllabic units, 老闆很好) will indeed arise in the course,[5] and provide persistent and focused feedback for the target form before a new/different problem is focused on.

It is important to note that providing focused, sustained FonF treatment does not necessarily mean that the teacher should intensively bombard students with the highlighted (target) forms throughout the *whole* treatment session. Careful spacing out of and controlling for the instance distribution of the target tonal token avoids enhancing the target tonal token in an unnatural context and making it excessively salient within the input, thereby eliminating the problem of 'over-enhancing.' When the target tonal token is over-enhanced, it debilitates, rather than facilitates, learners' processing of the target form. Overstreet's (1998) study is a case in point. In this study, the participants in the enhancement group were not only intensively exposed to instances of the target form in a short text (input flood), but each instance was typographically highlighted in three ways (underlying, boldface/shadowing, size enlargement). Results obtained from the measures of production, recognition, and comprehension showed a negative effect on comprehension as well as on the intake of the target form. Similarly, findings from S. K. Lee (2007) and Shook (1999) also show that input enhancement,

[4] This does not mean that teachers should refrain from providing any sorts of implicit FonF feedback for an important problem that arises in the treatment session. While learner's focal attention should be primarily directed to the target feature, learners' peripheral attention can be drawn to "unexpected" ill-formed/problematic tokens that repetitively occur in the treatment session. That is, there is a need for pro-active syllabus, which is also reactive in nature.

[5] To this end, teachers probably have to plan on the lexical items that can bring out the target tonal tokens. However, this does not mean that the conversation in the classroom should be scripted; such a method is not communicative per se and hence is not desirable under the FonF framework.

regardless of how implicit it is, may temporarily disrupt learners' comprehension of the text which contains the enhanced form; when the target form is over-enhanced, processing of meaning—the prerequisite for successful implicit FonF—will be frequently disrupted. Given the aforementioned trade-off between intake and comprehension, finding the right balance between saliency and over-enhancing is therefore a difficult but necessary concern for practitioners in planning and structuring the enhanced input.

To avoid over-enhancing the target form to the detriment of disrupting the meaning processing, Han (2008) suggests teachers switch gears to "non-intensive" FonF treatment once there is a sign of uptake on the part of students' production. Furthermore, Han (2001) urges that the implementation of implicit FonF technique, such as prosodic input enhancement or recast, should be strictly confined to providing information on the well-formedness of their utterances in *obligatory* context (didactic function), and should not be overused to seek or confirm information regarding the content of the utterances (communicative function). When prosodic input enhancement (or other implicit FonF feedback) is used to perform both communicative and didactic functions, implicit FonF feedback often leads to ambiguity and obscurity from the learners' perspective and may have confused, rather than enlightened, the learners (Han et al. 2008, see also Han 2007). In this case, overuse of input enhancement, as seen in Alanen (1995) and Leeman et al. (1995), can easily result in learner's overgeneralization of the target form, inducing them to abuse the noticed feature(s) both in obligatory and nonobligatory contexts.

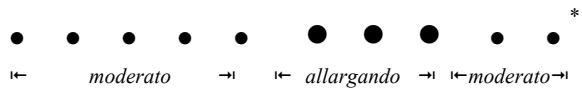
3. Guidelines for implicit FonF tonal intervention

As noted earlier, to help non-native tonal speakers entrench the link between the tonal system and the mental lexicon (and hence acquire a native-like tonal identification system), implicit FonF tonal interventions need to be implemented in a context-rich, communicative setting; teachers should attempt to provide positive evidence of the target tonal token in the form of meaning-based utterances. In the ensuing section, I shall illustrate how prosodic input enhancement may be used, along with other implicit FonF strategies (e.g. recast), to implicitly direct L2 learners' attention to tone sandhi (or other problematic phonetic/phonemic tonal transformation) in the context of communicative interaction. Enhanced segments that are stressed, sonorous, or uttered with high pitch (intonation) are more salient to L2 listeners than unenhanced segments (Leeman 2003). Perceptual salience of the target form in oral input is often enhanced by raising the tone or intonation when uttering the target structure (e.g. Leeman 2003) and/or by repeating/reformulating the target form in question (e.g. Lyster 1998). In the case of tone instruction, teachers can prosodically highlight the target tonal token (e.g.

T3S phenomenon) by temporarily raising the intonation or volume when uttering the words involved in tone sandhi (tone sandhi domain). To make the T3S domain or other problematic tonal tokens (to which prosodic input enhancement is applied) even more salient to learners, a *tempo rubato*[6] reading technique can be employed (cf. Leeman 2003); specifically, the target of the implicit FonF (in this case, T3S domains) should be read at *allargando* (slow but fuller in tone) while the rest of the sentence is read at *moderato* (normally paced; moderate tempo).[7] The combinatory use of the *tempo rubato* reading technique and stress leads to 'compound enhancement' of the saliency of the target form, which is more effective than simple enhancement—e.g. using stress alone—in inducing further processing and noticing and of the target form (Alanen 1995, Williams 1999, cf. Leeman et al. 1995). Simple enhancement, as reported in Leow (2001) and Leeman et al. (1995), may fail to impose a noticeable impact on learners' intake, especially when learners have limited prior knowledge of the form in question. To illustrate:

In a treatment session that targets the T3S on T3-T3-T3 lexical strings, the teacher should be pro-active in planning instances of the target tonal structure (e.g. 展覽館 'exhibition hall') in accordance with the theme of the session (e.g. visiting an art gallery).[8] In uttering the sentence in which a given instance is embedded, the teacher could emphatically, but slowly read the words involving the designated T3S token (展覽館) with a fuller tone, but read the remaining words with a normal pace (see below).

(5)



Teacher: 經 理 昨 天 去 展 覽 館 視 察。

jing1 li3 zuo2 tian1 qu4 zhan2 lan2 guan3' shi4 cha2

'The manager went to the exhibition hall to supervise (the operation of the exhibition stand) yesterday.'

[6] The temporary disregarding of strict tempo to allow an expressive quickening or slackening, usually without altering the overall pace.

[7] In this case, the concept of *allargando* and *moderato* is only a relative, rather than an absolute, concept, which may vary from one person to another.

[8] This paper draws on Merrill Swain's (1985) Comprehensible Output Hypothesis, according to which learning or reconstruction of learner's existing language system (in this case, tonal system) takes place when learners encounter a gap between what they want to say and what they are able to say. By uttering something, learners test their existing tonal knowledge and receive corrective feedback from their interlocutors (teachers). In other words, output produced by the teacher and the student serves as pedagogical input and jointly contributes to the reconstruction of learners' tonal system.

In implementing implicit FonF interventions, teachers need to be extremely cautious with and sensitive to the length of the portion that is being enhanced/highlighted (i.e. the parts of words involved in T3S). The target form (T3S variation and the lexical environment in which it occurs) will appear more salient and therefore more readily noticeable by the learner if it does not involve a substantial portion of the sentence (see Philp 2003); sentences such as (**老闆想**)要參考這本書的設計 (in which only the initial 2-3 characters are involved in T3S) are more likely to tap into learners' internal salience than alliterative sentences such as (老闆想找兩百五)十本書 (in which 7 out of 10 characters are involved in T3S). Implicit FonF will be more effective when the target form can be highlighted in a succinct and minimal manner. Other factors which are also closely tied to and hence may contribute to the salience of the target form are frequency and grammatical complexity (Bardovi-Harlig 1987, Goldschneider & DeKeyser 2001, VanPatten 2005).

Note also that existing research has empirically established that the length of tone sandhi domain is very sensitive to the speech rate and varies depending upon the tempo (e.g. Brooke et al. 2009, Hayes & Lahiri 1991). When the tempo is slow (moderato), a (perceived) tone sandhi boundary/break often appears at strong syntactic junctures, e.g. between the subject and predicate (Speer, Shih & Slowiaczek 1989). As the tempo increases, the sandhi domains in a sentence would be merged, as a result of which sandhi break and length of T3S domain also change (see examples (6) & (7) below). Most importantly, the tone contour shape of the new T3S domain in slower speech will differ from that in faster speech. As can be seen in examples (6) and (7) below, the tones carried by the first four characters are T2-T3'-T2-T3' in moderato speech, but turn to T2-T2-T2-T3' in allegro speech.

(6) 老 闆 獎 賞 員 工 (possible T3s domain in moderato or normal speech)
lao2 ban**3'** jiang2 shang3' yuan2 gong1

(7) 老 闆 獎 賞 員 工 (possible T3s domain in allegro or brisk speech)
lao2 ban**2** jiang2 shang3' yuan2 gong1

The fact that the tone sandhi domain and contour shape may vary due to different speech rates and contexts suggests that certain kinds of postlexical computation is required. However, it is important to note that in order to efficiently sandhi a phrase and/or identify the tones in connected speech, the language processor also needs to perform some kind of advance prelexical linguistic processing, predicting possible lexical/phonological breaks and applying relevant (sandhi) rules several syllables *ahead* (Brooke et al. 2009). Results of such advance processing/computation (different T3S possibilities that may be applied to different potential T3S domains in a sentence) will

be temporarily stored in the Working Memory until vocal production is heard or uttered.[9] Such online prelexical and postlexical processing—which cannot be completed without active appeal to the mental lexicon—accounts for why/how listeners/speakers are able to quickly and efficiently determine the *first* (low-level) T3S domain to which cyclic T3S rule is initially applied in examples (1) and (2) (see below).

(1) 我很好 (wo[**3**] **hen[3] hao[3**] → (wo[**3'**]) (hen[2] hao[3]) 'I am fine.'

(2) 老闆好 (**lao[3] ban[3]** hao[3] → (lao[**2**] ban[2]) (hao[3]) 'How are you doing, boss?'

Given that tone sandhi is tempo-dependent, as learners' sensitivity toward and control of the tone sandhi improve, the teacher may need to accelerate the speech rate and apply the implicit (prosodic) input enhancement techniques noted above to enhance the saliency of the tone sandhi domain in allegro speech; however, before learners' perception of tone in moderato speech improves, the teacher should refrain from abruptly switching to allegro speech, thereby providing consistent and focused implicit FonF treatment. Considering that speech rate is a subjective and relative concept, rule-based explicit FonF is perhaps not an optimal way to foster learners' procedural knowledge of the tempo-dependent T3S variation. Systematic application of implicit FonF to T3S tokens in speeches delivered at different speech rates would allow learners to be aware that the T3S phenomenon does not simply involve a static suprasegmental rule computation in online speech. Most importantly, it will provide learners with a context-rich environment to internalize the intricate interaction between tonal behaviors and relevant lexical/tempo information.

Up to this point, I have mainly discussed the implementation of implicit FonF in terms of how salience of the target form can be enhanced by external manipulation (the teacher). If formal features of various tonal variations in connected speech must be attended to (or detected) before they can be internalized, as is widely agreed by studies on adult SLA, then it stands to reason that—all things being equal—forms that are visually or prosodically more salient stand a better chance of being detected/attended to and acquired. However, this reasoning does not necessarily entail that learners will automatically attend to all salient forms and that pro-active, implicit FonF will necessarily draw learners' attention to all salient linguistic formal constructions. Although pro-active, implicit FonF activities may externally enhance perceptual salience (or

[9] Brooke et al. (2009) argue that due to working memory capacity constraint, such advance processing cannot occur "without bound"; after all, you cannot always know how the sentence is going to unfold when you begin it.

noticeability) of the target form, learners may continue to "ignore a vast mass of [explicit] evidence and continue, obstinately, [and] to operate with a[n] [*internal processing*] *system* that is in contradiction with the target norms as manifest in the input" (Sharwood-Smith 1993:168, emphasis added). Sharwood Smith's statement underscores the fact that externally generated salience does not always converge with internally generated salience and highlights the need to consider the learner's internal language (processing) system in achieving successful FonF. Whether the enhanced input will ultimately induce learners' attention and successfully trigger relevant mental representation (re)construction crucially depends on whether implicit FonF is implemented in consideration of learners' universal linguistic parsing/processing strategies (VanPatten 2005), without which "forms may be noticed *perceptually*, but not *linguistically*" (Leeman et al. 1995:219). In other words, enhanced input may attract attention, but may fall short of deeper linguistic processing: in the case of fostering a target-like tonal system, associating the encoded tonal features and relevant lexical environment—a crucial encoding condition that later determines if pitch information can be *linguistically* used for semantic access (Yip 2002).

Providing implicit FonF in consideration of learners' universal processing principles[10] will maximize the chance that learners *act upon* the noticed feature and spark a chain of cognitive processes to further analyze the attended input. Results from cognitive studies have empirically established that in performing a *comprehending* task, learners tend to notice items in the sentence-initial position before those in the final and medial position (a phenomenon called the primacy effect; for details, see VanPatten 2005), but that in *retention* tasks, learners tend to notice the items in the sentence-final position, recalling those items best (a phenomenon called the recency effect). Initial and final positions are therefore privileged in terms of processing; learners usually do not attend to the target forms embedded in the sentence medial position—a position where many grammatical features occur. The above cognitive processing bias suggests that learners will allocate (or redirect) their attentional processing resources in response to different task demands and that location of the target form intimately interacts with the availability of attentional resources. Depending on the purpose of the learning or listening task (comprehension vs. retention), the teacher should carefully manipulate the input, placing and highlighting the target form in locations that will make the form more processable and salient to the learner whenever possible, i.e. placing and highlighting the target form in the sentence-initial position in a comprehension task, but in sentence-final position in a retention task.

In cases of mishearing or mispronunciation (e.g. incorrect application of T3S or

[10] For instance, learners may be unconsciously attracted to fix on formal features of high-frequency words or consciously pay attention to words.

the lack thereof in the obligatory context), implicit FonF feedback should be swiftly provided wherever appropriate to allow learners to *immediately* hear the gap between (and hence compare) the target and non-target tonal tokens before transient memory traces of the target vs. non-target structure are dampened in the Working Memory (Han 2002, 2008); note, however, that if the problematic (or ill-formed) tonal tokens appear in the sentence-medial position (see example (8a)), teachers should attempt to recast the sentence by moving the part which requires further (FonF) treatment to a salient position (see the underlined characters in examples (8a) and (8b)). To illustrate:

(8) a. Student: 這 書 的 設 計 是 **老 闆 想** 參 考 的。
Zhe4 shu1 de5 she4 ji4 shi4 lao3* ban3* xiang3* can1 kao3 de5
'The design of the book is what the boss would like to refer to.'

In highlighting the problematic T3S instance produced by the student (老闆想 in example (8a)), the teacher reformulates the student's production by moving the instance to the sentence-initial position (see example (8b) below).

b. Teacher: **老 闆 想** 參 考 這 本 書 的 設 計？
lao2 ban2 xiang3' can1 kao3 zhe4 ben3 shu1 de5 she4 ji4
'The boss would like to refer to the design of the book?'
Student: 沒 錯，他 是 這 麼 說 的。
mei2 cuo4 ta1 shi4 zhe4 me5 shuo1 de5
'That's right. This is what he said.'

* The fonts highlighted in red are the words taken from the student's original production.

It is important to note that for implicit FonF to work at its best, the lexical-semantic and/or syntactic environment in which the (enhanced) target tonal token is embedded needs to be carefully 'structured.' Specifically, in responding to the learners' ill-formed tonal token (such as "**老闆想**" in example (8a)), teachers should immediately recast the problematic sentence (that contains the ill-formed T3S token) using similar/identical structure and wording, minus the (T3S) error (see example (8b) above);[11] in example (8b), the well-formed T3S token can be prosodically highlighted using stress/intonation and/or the *tempo rabato* reading technique mentioned earlier. In so doing, the reformulation

[11] It is important to note that positive feedback (such as "Okay") to and verbatim repetition of the learner's production (and error) in example (8a) would potentially send a confusing message to the student(s), who may inappropriately perceive the repetition or the positive feedback as a confirmation check or approval to the learner's production (see Han 2001) or hear the recast as a continuation of the conversation.

(i.e. recast) still addresses the same lexical-semantic content (thus, will not impose extra processing cost for semantic load) and involves limited processing for forms that do not require further attention. Consequently, focal attention can be reserved for the target form that requires further treatment (i.e. T3S). Following the first modeling (example (8b)), the teacher may want to continue modeling the same (T3S) token 2-3 times in the following conversation turns (of course, with the token being prosodically highlighted and embedded in the salient position). Any "one-shot" input enhancement may not successfully draw learners' attention to the formal features in the input; it is problematic to expect that the reconstruction of the learner's (tonal) mental structure will take place from a single treatment. In the second and third modeling, teachers can *progressively* change/expand the wording and sentence structure in the second (or the third) modeling (see examples (8b) → example (8c)), but should refrain from abruptly referring to too much novel semantic-lexical information.

(8) c. Teacher: 老 闆 想 參 考 的 書 太 多 了。
lao2 ban2 xiang3' can1 kao3 de5 shu1 tai4 duo1 le5
'There are so many books that the boss would like to refer.'

Student: 爲 什 麼 老 闆 特 別 想 要
wei4 she2 me5 lao3* ban3* te4 bie2 xiang3* yao4
參 考 這 本 書？
can1 kao3* zhe4 ben3 shu1
'But, why does our boss decide to refer to the design of this book?'

* The fonts highlighted in red are the words taken from the student's original production.

d. Teacher: 老 闆 想 要 類 似 那 本 書 的 風 格。
lao2 ban2 xiang3' yao4 lei4 si4 na4 ben3 shu1 de5 feng1 ge2
'The boss prefers a style that is similar to the design of that book.'

After highlighting and modeling the same token (老闆想…) 2-3 times following the structured input enhancement guideline noted above, the teacher may want to switch to *other* sandhi instances in the following conversation turns (e.g. 老李想; see example (8e) below).

e. Teacher: 老 李 想 知 道 這 本 書 的 細 節。
lao2 li2 xiang3 zhi1 dao4 zhe4 ben3 shu1 de5 xi4 jie2
'Mr. Li would like to know more details about this book.'

The above structured (consecutive) input (examples (8b), (8c), (8d)) would create a semantically/lexically familiar environment in which attentional resources could be gradually freed up and re-allocated to the reprocessing of the target vs. ill-formed tonal tokens. In this case, the structured input would stimulate an *intentional* focus on form and an incidental/peripheral focus on meaning, allowing learners to make more efficient use of their attention resources for form decoding (Han et al. 2008; see also Gass 1997). If input enhancement is not structured in the above manner—upon hearing the problematic tone (in example (8a)), teachers immediately recast/highlight the target (sandhi) tonal token in a sentence that shares little lexical/semantic information with the learner's original production (e.g. example (8d)), learners will allocate a substantial amount of attentional resources to process the input for meaning (as a result of their default language processing predilection), even in defiance of the teachers' instruction and any input enhancement intervention techniques; in this case, the processing of formal aspect of the enhanced input will only occur *incidentally*, as a byproduct of comprehension. Accordingly, structured input enhancement does not just involve temporal order of meaning-based and form-based processing of input; most importantly, it calls for incidental or intentional processing of the formal aspect of the target form (Han et al. 2008).

The view that processing input for form presupposes that the meaning of the input has been clarified is not only in line with the fundamental tenets of cognitive theory of information processing, but has also been substantiated in empirical research (e.g. Doughty 1991, Izumi 2002). Information processing studies have shown that: (1) information processing is selective: learners process input *first* for *meaning*; (2) learners can simultaneously handle two different types of information (e.g. decoding tonal patterns and meaning at the same time) only if and when one of the information types is clarified/automatized and hence requires little, if any, processing/attentional resources; (3) simultaneous processing of two types of information that are not automatized can result in deficient decoding/encoding of both types of information (Han & Peverly 2007, VanPatten 1996, 2004, 2005). Given that simultaneous processing of form and meaning is often not automatized in (adult) L2 learners, any input enhancement technique which aims to direct the learner's attention to form and meaning at the same time may not be an optimal consciousness-raising technique (Han et al. 2008, VanPatten 1990).[12] Therefore, input enhancement is more likely to tap into the learner's attention to the formal aspect of the target structure and hence promote growth in comprehension and acquisition when meaning is clarified *before* a focus on form, namely, sequential processing of meaning and form (Han et al. 2008).

[12] Patsy Lightbown (1998), however, contends that dual focus on forms and meaning would not disrupt learners when the form in question is crucial to the meaning being conveyed.

Of course, the aforementioned structured implicit FonF tonal treatment would have a limited effect on (re)constructuring learners' tonal system if the provision/timing of implicit FonF treatment is entirely based upon the learner's output; that is, teachers only provide implicit FonF treatment when a problematic tonal token is detected in the learners' production, in this case, FonF is reactive in nature. What learners can produce, in most cases, generates only a limited context; worst of all, there may be cases of avoidance on the learners' part—not producing obligatory contexts for the target tonal token. Even if learners can conjure up their own contexts for the target tonal token in question, the student-supplied context is oftentimes not authentic enough or is quite limited in showcasing sufficient relevant/obligatory lexical environment. (Pro-active) teacher-supplied input therefore is imperative for providing a context-rich environment to internalize the intricate tonal behaviors. However, this does not mean that the learners' production does not play a role in their learning. In order to further bolster the effect of implicit FonF, teachers still need to create sufficient *online* processing opportunities whereby learners are forced to (re)produce the target tonal token during and after the FonF treatment (see Nitta 2008). Izumi (2002) observed that while comprehension-based learning tasks only result in rehearsing of the attended form at a relatively shallow level of processing, which then leads learners to experience only short-term retention, output production tasks trigger deeper and more elaborate processing of the form, which leads learners to establish stronger memory traces within the long-term memory. As Lightbown (1998) noted, "through exposure to [FonF] instruction, learners have acquired 'knowledge' of ...[the target] features but have not yet acquired 'control' of them…opportunities to use… [the features] in discourse-appropriate contexts help learners get this control" (p.183).

Taken together, remedial implicit FonF will be more effective when the target tonal token is partially acquired by the learners, embedded in a meaningful communicative task, prosodically highlighted for didactic purposes, placed in the end or beginning of a grammatically simple sentence (composed of high-frequency words), and involves only a minimal portion of the sentence. To achieve successful FonF, teachers do not just determine a potentially problematic area (pro-active syllabus), but also determine *why* it is a problem vis-à-vis learners' universal processing strategies; teachers should provide implicit FonF feedback/treatment in accordance with the learners' processing strategies and most importantly, "push" learners away from non-optimal processing strategy when implicit FonF feedback for an ill-formed structure is required. In other words, the spirit of the FonF pedagogical framework is "to work *with*, *and* more importantly, *against* second language learners' natural, meaning-exclusive tendency for input processing" (Han et al. 2008:597).

4. Conclusion

This paper has set out to explore how tone is identified and processed in spoken word recognition by native and non-native tonal speakers of Mandarin Chinese. It is argued that speakers of a tonal language do not simply rely on their built-in tonal inventory to interpret tones extracted from the acoustic input. In order to better identify and process tone in connected speech, in addition to acquiring a categorical tonal inventory, another challenge that learners of a tonal language face is to establish the link between the acquired tonal inventory and the mental lexicon. The connection of the tonal inventory to the lexicon allows learners of a tonal language to quickly resolve any tonal ambiguities resulting from phonemic sandhi or phonetic transformation in spoken-word recognition. Explicit input (i.e. metalinguistic tonal rule instruction), as seen in many intensive training programs, may help non-native tonal speakers successfully acquire Mandarin tonal contrasts in the short term; however, declarative knowledge of such categorical rules may not help non-native tonal speakers efficiently resolve tonal ambiguity in connected speech (as seen in Sun 1997). Furthermore, the power of explicit feedback/instruction is often constrained by the nature of settings in which tonal features are encoded/attended (detached from context), the nature of the pedagogical treatment (pro-active vs. reactive), and by the lack of concern with learners' internal syllabus (focusing on developmentally inappropriate features) and processing strategies. All of the above factors collectively constrain the efficacy of any consciousness-raising activities. On the other hand, implicit input (exposure-only), as seen in many adult SLA studies, also does not seem to lead to any beneficial effects: there are certain elements in the language that will never be noticed by learners through purely communicative interaction. The traditional implicit-explicit distinction is therefore too limited a dichotomy in the discussion of the pedagogical intervention for fostering a native-like tonal identifying system.

This paper has argued that explicit form-focused tonal instruction and implicit FonF intervention are both required for successful acquisition of the Chinese tonal system. Successful acquisition of a Chinese tonal identification system involves two major hurdles: (1) establishment of a categorical tonal inventory; and (2) internalizing intricate tonal behaviors along with correspondent lexical/syntactic environment. While explicit tonal instruction (or explicit FonF) may faciliate the initial acquisition of novel tonal features and the establishment of categorial tonal inventory (the first hurdle), implicit FonF tonal intervention seems to be more effective in remediating non-target tonal features that are partially acquired by non-native Mandarin speakers and in providing a context-rich environment to encode the lexical/syntactic environment in which tonal variation occurs (the second hurdle). Implicit FonF tonal intervention would

be most effective when the target tonal structure is meticulously chosen (problematic tokens partially acquired by the learners), and echoes the learners' developmental readiness. To the above end, teachers need to have a sound understanding of the learners' built-in learning agenda, gauge the areas of problems that require further pedagogical interventions, and then design implicit FonF activities that unobtrusively enhance the perceptual salience of the target form in a contextualized, communicative setting (e.g. task-essential practice). To successfully apply implicit FonF to tone teaching, teachers first need to identify potential problematic (perceptual and productive) tonal tokens and then design pro-active communicative activities in which focused, implicit FonF feedback can be consistently and unobtrusively provided to enhance the salience of a given problematic tonal token in consideration of learners' processing strategies. Only then can we truly achieve "consciousness-raising" in the true sense of the word, bridging the gap between external salience of pedagogical intervention and the learners' internal salience, and most importantly, enhancing the target form without producing aberrant noticing to the detriment of acquisition. In this paper, I have argued that to achieve successful implicit FonF for fostering procedural tonal knowledge system, the construction of activities should follow a set of guidelines:

1. Keep meaning in focus (attention to form should occur in lessons where the primary focus is on meaning or communication); the (implicit) FonF treatment should be 'masked' under meaning-oriented task conditions. Despite their primary focus-on-meaning, teachers should create opportunities to switch learners' attention from meaning to form under some 'structured' conditions.
2. Adopt a pro-active syllabus: (pre)determine the problematic tonal tokens that require further treatment or tokens that learners have partially mastered; and ensure the opportunities to attend/produce the target tonal token in a communicative task will indeed arise.
3. Employ compound, rather than simple, implicit FonF strategy, when appropriate (e.g. prosodical input enhancement in combination with *tempo rubato* reading technique).
4. Focus on one thing at a time and provide consistent, frequent, and focused implicit FonF feedback for the target tonal token only in the *obligatory* context (solely for didactic purposes).
5. Ensure that implicit FonF feedback is provided in a brief, unobtrusive and minimal manner.
6. Keep learners' processing/parsing strategies in mind when providing implicit FonF feedback.

Taking into consideration *what* learners actually do in online language processing (i.e. learning's parsing strategies) is the prerequisite for achieving successful focus on form. It is the understanding of the "*what*" that allows us to formulate instruction guidelines that can successfully bring about the procedural knowledge of the target form. As Han (2008) aptly noted:

> "The ultimate goal of teaching is to equip our students with implicit procedural knowledge that they can actually retrieve/use automatically. We don't want our students to walk away from the classroom with explicit rule-based knowledge which constantly monitors their production/perception, because it hurts their fluency."

It is hoped that the discussion of this paper will not only shed light on ways to foster a procedural knowledge system of Mandarin tones, but can also be extended to the teaching of other linguistic features—features that are amenable to communicative FonF interventions.

References

Alanen, Riikka. 1995. Input enhancement and rule presentation in second language acquisition. *Attention and Awareness in Foreign Language Learning*, ed. by Richard Schmidt, 259-302. Honolulu: University of Hawai'i Press.

Baddeley, Alan D., and Graham J. Hitch. 1974. Working memory. *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 8, ed. by Gordon H. Bower, 47-89. New York: Academic Press.

Bao, Zhiming. 1999. *The Structure of Tone*. New York & Oxford: Oxford University Press.

Barcroft, Joe. 2002. Semantic and structural elaboration in L2 lexical acquisition. *Language Learning* 52.2:323-363.

Barcroft, Joe. 2004. Theoretical and methodological issues in research on semantic and structural elaboration in lexical acquisition. *Form-Meaning Connections in Second Language Acquisition*, ed. by Bill VanPatten, Jessica Williams, Susanne Rott & Mark Overstreet, 219-234. Mahwah: Lawrence Erlbaum Associates.

Barcroft, Joe. 2006. Can writing a new word detract from learning it? More negative effects of forced output during vocabulary learning. *Second Language Research* 22.4:487-497.

Bardovi-Harlig, Kathleen. 1987. Markedness and salience in second-language acquisition. *Language Learning* 37.3:385-407.

Bowles, Melissa A. 2003. The effects of textual input enhancement on language learning: an online/offline study of fourth-semester Spanish students. *Theory, Practice, and Acquisition: Papers from the 6th Hispanic Linguistics Symposium and the 5th Conference on the Acquisition of Spanish and Portuguese*, ed. by Paula Kempchinski & Carlos-Eduardo Piñeros, 395-411. Somerville: Cascadilla Press.

Brooke, Julian, Greg Coppola, T.-Y. Vivian Lee, and Yang Zhao. 2009. Phrasing, tempo, and focus: a prosodic-hierarchy approach to Mandarin tone sandhi domains. *SFU Working Papers in Linguistics*, Vol. 2: *Proceedings of ICEAL 2*, ed. by David Potter et al. Retrieved from http://www.sfu.ca/gradlings/wp_2.html.

Chen, Jenn-Yeu. 1999. The representation and processing of tone in Mandarin Chinese: evidence from slips of the tongue. *Applied Psycholinguistics* 20.2:289-301.

Chen, Matthew Y. 2000. *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge & New York: Cambridge University Press.

Cheng, Chin-Chuan. 1973. *A Synchronic Phonology of Mandarin Chinese.* The Hague: Mouton.

Cutler, Anne, and Hsuan-Chih Chen. 1997. Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics* 59.2:165-179.

DeKeyser, Robert M. 1995. Learning second language grammar rules: an experiment with a miniature linguistic system. *Studies in Second Language Acquisition* 17.3:379-410.

Doughty, Catherine. 1991. Second language instruction does make a difference: evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition* 13.4:431-469.

Doughty, Catherine. 1998. Acquiring competence in a second language: form and function. *Learning Foreign and Second Languages: Perspectives in Research and Scholarship*, ed. by Heidi Byrnes, 128-156. New York: Modern Language Association of America.

Doughty, Catherine. 2004. Effects of instruction on learning a second language: a critique of instructed SLA research. *Form-Meaning Connections in Second Language Acquisition*, ed. by Bill VanPatten, Jessica Williams, Susanne Rott & Mark Overstreet, 181-202. Mahwah: Lawrence Erlbaum Associates.

Doughty, Catherine, and Jessica Williams. 1998. Pedagogical choices in focus on form. *Focus on Form in Classroom Second Language Acquisition*, ed. by Catherine Doughty & Jessica Williams, 197-261. Cambridge & New York: Cambridge University Press.

Fox, Robert Allen, and James Unkefer. 1985. The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics* 13.1:69-90.

Gandour, Jackson T. 1994. Phonetics of tone. *The Encyclopedia of Language and Linguistics*, ed. by R. E. Asher & J. M. Y. Simpson, 3116-3123. Oxford & New York: Pergamon Press.

Gass, Susan M. 1997. *Input*, *Interaction*, *and the Second Language Learner*. Mahwah: Lawrence Erlbaum Associates.

Goldschneider, Jennifer M., and Robert M. DeKeyser. 2001. Explaining the "natural order of L2 morpheme acquisition" in English: a meta-analysis of multiple determinants. *Language Learning* 51.1:1-50.

Han, ZhaoHong. 2001. Fine-tuning corrective feedback. *Foreign Language Annals* 34.6: 582-599.

Han, ZhaoHong. 2002. A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly* 36.4:543-572.

Han, ZhaoHong. 2004. *Fossilization in Adult Second Language Acquisition*. Clevedon: Multilingual Matters.

Han, ZhaoHong. 2007. On the role of meaning in focus on form. *Understanding Second Language Process*, ed. by ZhaoHong Han, 45-79. Clevedon: Multilingual Matters.

Han, ZhaoHong. 2008. Error correction in second language teaching: towards a differential approach. Symposium conducted at Queensborough Community College, November 14, 2008. Bayside: Queensborough Community College, CUNY.

Han, ZhaoHong, Eun-Sun Park, and Charles Combs. 2008. Textual enhancement of input: issues and possibilities. *Applied Linguistics* 29.4:597-618.

Han, ZhaoHong, and Stephen T. Peverly. 2007. Input processing: a study of *ab initio* learners with multilingual backgrounds. *International Journal of Multilingualism* 4.1:17-37.

Harrison, Phil. 1998. Yorùbá babies and unchained melody. *UCL Working Papers in Phonetics* 10:33-52.

Harrison, Phil. 1999. *The Acquisition of Phonology in the First Year of Life*. London: University of London dissertation.

Harrison, Phil. 2000. Acquiring the phonology of lexical tone in infancy. *Lingua* 110.8: 581-616.

Hayes, Bruce, and Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language and Linguistic Theory* 9.1:47-96.

Izumi, Shinichi. 2002. Output, input enhancement, and the noticing hypothesis: an experimental study on ESL relativization. *Studies in Second Language Acquisition* 24.4:541-577.

Jourdenais, Renée, Mitsuhiko Ota, Stephanie Stauffer, Beverly Boyson, and Catherine Doughty. 1995. Does textual enhancement promote noticing? A think-aloud protocol analysis. *Attention and Awareness in Foreign Language Learning*, ed. by Richard Schmidt, 183-216. Honolulu: University of Hawai'i Press.

King, Brian. 1981. Tone acquisition: some general observations. *Proceedings of the Second International Congress for the Study of Child Language*, ed. by Carolyn Echols Johnson & Carol L. Thew, 78-87. Washington, D.C.: University Press of America.

Krashen, Stephen D. 1985. *The Input Hypothesis: Issues and Implications*. London: Longman.

Laver, John. 1994. *Principles of Phonetics*. Cambridge & New York: Cambridge University Press.

Leather, Jonathan. 1987. F0 pattern inference in the perceptual acquisition of second language tone. *Sound Patterns in Second Language Acquisition*, ed. by Allan R. James & Jonathan Leather, 59-81. Dordrecht: Foris.

Lee, Chao-Yang. 2002. *Lexical Tone in Spoken Word Recognition: A View from Mandarin Chinese*. Providence: Brown University dissertation.

Lee, Chao-Yang. 2007. Does horse activate mother? Processing lexical tone in form priming. *Language and Speech* 50.1:101-123.

Lee, Sang-Ki. 2007. Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning* 57.1:87-118.

Leeman, Jennifer. 2003. Recasts and second language development: beyond negative evidence. *Studies in Second Language Acquisition* 25.1:37-63.

Leeman, Jennifer, Igone Arteagoitia, Boris Fridman, and Catherine Doughty. 1995. Integrating attention to form with meaning: focus on form in content-based Spanish instruction. *Attention and Awareness in Foreign Language Learning*, ed. by Richard Schmidt, 217-258. Honolulu: University of Hawai'i Press.

Leow, Ronald P. 2001. Do learners notice enhanced forms while interacting with the L2? An online and offline study of the role of written input enhancement in L2 reading. *Hispania* 84.3:496-509.

Lightbown, Patsy M. 1998. The importance of timing in focus on form. *Focus on Form in Classroom Second Language Acquisition*, ed. by Catherine Doughty& Jessica Williams, 177-196. Cambridge & New York: Cambridge University Press.

Lightbown, Patsy M. 2005. Transfer-appropriate processing as a framework for classroom SLA. Paper presented at the Second Language Research Forum (SLRF 2005), October 7-9, 2005. New York: Teachers College, Columbia University.

Lightbown, Patsy M. 2008. Transfer appropriate processing as a model for classroom second language acquisition. *Understanding Second Language Process*, ed. by ZhaoHong Han, 27-44. Clevedon: Multilingual Matters.

Lin, William C. J. 1985. Teaching Mandarin tones to adult English speakers: analysis of difficulties with suggested remedies. *Regional English Language Center Journal* 16.2:31-47.

Liu, Yeu-Ting. 2007. *Phonological Recoding in Sentence-Level Chinese Character Recognition by Advanced Adult L2 Chinese Learners.* New York: Columbia University dissertation.

Liu, Yeu-Ting. 2009. Attainability of a native-like lexical processing system in adult second language acquisition: a study of advanced L2 Chinese learners. *Language and Linguistics* 10.3:489-520.

Long, Michael H. 1991. Focus on form: a design feature in language teaching methodology. *Foreign Language Research in Cross-cultural Perspective*, ed. by Kees de Bot, Ralph B. Ginsberg & Claire Kramsch, 39-52. Amsterdam & Philadelphia: John Benjamins.

Long, Michael H., and Peter Robinson. 1998. Focus on form: theory, research, and practice. *Focus on Form in Classroom Second Language Acquisition*, ed. by Catherine Doughty& Jessica Williams, 15-41. Cambridge & New York: Cambridge University Press.

Lu, Jya-Mei. 1992. *On the Perception and Production of Mandarin Tones by Adult English-speaking Learners*. East-Lansing: Michigan State University dissertation.

Lyster, Roy. 1998. Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition* 20.1:51-81.

Morris, C. Donald, John D. Bransford, and Jeffery J. Franks. 1977. Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior* 16.5:519-533.

Nitta, Ryo. 2008. Incidental focus-on-form in task-based language teaching: a rationale for strategic and on-line planning processes. *NUCB Journal of Language*, *Culture and Communication* 9.2:31-42.

Overstreet, Mark H. 1998. Text enhancement and content familiarity: the focus of learner attention. *Spanish Applied Linguistics* 2:229-258.

Paradis, Michel. 1994. Neurolinguistic aspects of implicit and explicit memory: implications for bilingualism and SLA. *Implicit and Explicit Learning of Languages*, ed. by Nick C. Ellis, 393-419. London: Academic Press.

Park, Eun-Sung. 2004. Constraints of implicit focus on form: insights from a study of input enhancement. *Teachers College*, *Columbia University Working Papers in TESOL & Applied Linguistics* 4.2. Retrieved January 10, 2010, from http://journals.tc-library.org/index.php/tesol/article/viewPDFInterstitial/59/65

Peng, Shu-Hui. 1997. Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25.3:371-400.

Peng, Shu-Hui. 2000. Lexical versus 'phonological' representations of Mandarin sandhi tones. *Acquisition and the Lexicon*, ed. by Michael B. Broe & Janet B. Pierrehumbert, 152-167. Cambridge & New York: Cambridge University Press.

Philp, Jenefer. 2003. Constraints on "noticing the gap": nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Learning* 25.1:99-126.

Révész, Andrea. 2007. *Focus on Form in Task-based Language Teaching: Recasts*, *Task Complexity*, *and L2 Learning*. New York: Columbia University dissertation.

Robinson, Peter. 1996. Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition* 18.1:27-67.

Robinson, Peter. 1997. Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition* 19.2:223-247.

Segalowitz, Norman, and Patsy M. Lightbown. 1999. Psycholinguistic approaches to SLA. *The Annual Review of Applied Linguistics* 19:43-63.

Schmidt, Richard. 1990. The role of consciousness in second language learning. *Applied Linguistics* 11.2:129-158.

Schmidt, Richard. 1993. Awareness and second language acquisition. *Annual Review of Applied Linguistics* 13:206-226.

Sharwood-Smith, Michael. 1993. Input enhancement in instructed SLA. *Studies in Second Language Acquisition* 15.2:165-179.

Shih, Chilin. 1997. Mandarin third tone sandhi and prosodic structure. *Studies in Chinese Phonology*, ed. by Jialing Wang & Norval Smith, 81-123. Berlin & New York: Mouton de Gruyter.

Shook, David J. 1999. What foreign language reading recalls reveal about the input-to-intake phenomenon. *Applied Language Learning* 10.1-2:39-76.

Speer, Shari R., Chilin Shih, and Maria L. Slowiaczek. 1989. Prosodic structure in language understanding: evidence from tone sandhi in Mandarin. *Language and Speech* 32.4:337-354.

Sun, Sylvia Henel. 1997. *The Development of a Lexical Tone Phonology in American Adult Learners of Standard Mandarin Chinese*. Honolulu: University of Hawai‘i dissertation.

Swain, Merrill. 1985. Communicative competence: some roles of comprehensible input and comprehensible output in its development. *Input in Second Language Acquisition*, ed. by Susan M. Gass & Carolyn G. Madden, 235-253. Boston: Heinle & Heinle Publishers.

Trofimovich, Pavel. 2005. Spoken-word processing in native and second languages: an investigation of auditory word priming. *Applied Psycholinguistics* 26.4:479-504.

VanPatten, Bill. 1990. Attending to form and content in the input: an experiment in consciousness. *Studies in Second Language Acquisition* 12.3:287-301.

VanPatten, Bill. 1996. *Input Processing and Grammar Instruction: Theory and Research*. Norwood: Ablex.

VanPatten, Bill. 2004. *Processing Instruction: Theory*, *Research*, *and Commentary*. Mahwah: Lawrence Erlbaum Associates.

VanPatten, Bill. 2005. Processing instruction. *Mind and Context in Adult Second Language Acquisition: Methods*, *Theory*, *and Practice*, ed. by Cristina Sanz, 267-281. Washington, D.C.: Georgetown University Press.

White, Joanna. 1998. Getting the learners’ attention: a typographical input enhancement study. *Focus on Form in Classroom Second Language Acquisition*, ed. by Catherine Doughty & Jessica Williams, 85-113. Cambridge & New York: Cambridge University Press.

Williams, John N. 1999. Memory, attention, and inductive learning. *Studies in Second Language Acquisition* 21.1:1-48.

Williams, Jessica, and Jacqueline Evans. 1998. What kind of focus and on which forms? *Focus on Form in Classroom Second Language Acquisition*, ed. by Catherine Doughty & Jessica Williams, 139-155. Cambridge & New York: Cambridge University Press.

Xu, Yaoda, Alexander Pollatsek, and Mary C. Potter. 1999. The activation of phonology during silent Chinese word reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25.4:838-857.
Ye, Yun, and Cynthia M. Connine. 1999. Processing spoken Chinese: the role of tone information. *Language and Cognitive Processes* 14.5-6:609-630.
Yip, Moira. 1980. *The Tonal Phonology of Chinese*. Cambridge: MIT dissertation.
Yip, Moira. 2002. *Tone*. Cambridge & New York: Cambridge University Press.

[Received 21 September 2010; revised 4 January 2011; accepted 11 January 2011]

Department of English
National Taiwan Normal University
162, Sec. 1, Heping East Road
Taipei 106, Taiwan
yeutingliu@ntnu.edu.tw

Focus on Form 教學觀點 在第二語中文聲調辨識教學的應用

劉宇挺

國立台灣師範大學

藉由中文聲調的理論與研究，本報告認為母語和非母語人士為了能夠迅速正確地辨識中文聲調，除了得建立一套清楚、有系統的聲調心理表徵外 (categorical tonal representation)，還得將這套聲調表徵和心理詞彙 (mental lexicon) 作連結；藉由這連結，聽者才能迅捷、有效地辨識平常口語中的“變調現象”(tone sandhi or allotone in connected speech)。但是，現今的聲調訓練課程往往著重在無前後文語境的聲調辨識練習 (decontextualized drills)，因而無法有效地幫助非中文母語人士建立聲調表徵以及心理詞彙之間的連結；即使在情境化的聲調辨識教學環境中，教師往往在沒有考慮學習者「線上語料處理策略」(online processing strategies) 的情況下，使得教師的外在操縱 (external input enhancement/manipulation) 不見得能夠有效地切入學習者的內在聚焦 (internal salience driven by the learner's attention)。本篇報告參酌了 Focus on Form (FonF) 的教學原則與認知心理學對學習者線上處理模式的見解，提出一套適用於中文聲調教學的建議，希望能幫助語言教師針對非中文母語人士的線上語料處理模式，提供更有效率的聲調（或是變調）教學（補救）策略。

關鍵詞：第二語口語教學，華語聲調教學，聲調辨識，聲調變調現象，Focus on Form