Attentional weighting of Polish and Taiwanese Mandarin sibilant perception

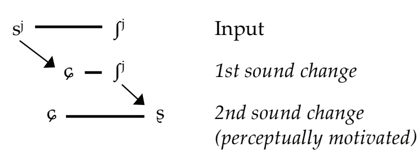
Chenhao Chiu
Department of Linguistics, University of British Columbia

1. Introduction

Polish and Taiwanese Mandarin are two languages well known for the three-way distinction of sibilant fricatives (Ladefoged and Maddieson 1996): apical dental [s], laminal[1] flat post-alveolar (retroflex) [ʂ], and laminal palatalized post-alveolar (alveolo-palatal) [ɕ]. Although these two languages are historically unrelated, their sibilants share a lot of auditory similarities (Ladefoged and Maddieson 1996). These auditory similarities result from similar articulatory gestures as well as different articulatory gestures, depending on sibilants; the alveolo-palatal [ɕ] in both languages requires a tensed tongue, spread lips, and a similar constriction of post-alveolar region of the palate; lip-rounding in PL retroflex [ʂ] and sublingual cavity in TM retroflex [ʂ] have similar acoustic effects. These articulatory gestures contribute to different acoustic properties which are influential in fricative perception. As proposed by earlier research, the acoustic cues, particularly formant transitions and frication noise, have been considered as most influential in frication perception (Nittrouer and Whalen 1989; Nowak 2003; Wagner, Ernestus, and Cutler 2006). Given the auditory similarities of sibilants in Polish and Taiwanese Mandarin, the present study investigates what acoustic cues distinguish one sibilant from another and what kind of role the acoustic cues play in perception.

The three way contrasts of sibilants in these two languages contribute to different historical changes. Polish sibilants as well as Russian sibilant underwent two phases of sound change, as argued by Padgett and Żygis (2007: 292):

> *Due to the change [sʲ] → [ɕ], the perceptual distance between [ʃʲ] and [ɕ] was closer than between the original [sʲ] and [ʃʲ]. In order to create a more optimal contrast, the palatalized palatoalveolar [ʃʲ] changed to [ʂ]. As a result, a perceptually stable contrast between [ɕ] and [ʂ] was introduced into Polish, where it exists today.*



The two perception-driven sound changes in Polish make the two post-alveolars ([ʂ] and [ɕ]) more contrastive. Flemming (2002) also argues that the contrast between the [ʂ] and [ɕ] in Polish is enhanced by the rounding of the retroflex [ʂ] so that [ʂ] and [ɕ] in Polish are auditorily more distinct. On the other hand,

[1] c.f. Ladefoged and Maddieson 1996: 153; Wierzchowska 1980; Rubach 1984; Puppel et al. 1997.

Actes du congrès annuel de l'Association canadienne de linguistique 2010.
Proceedings of the 2010 annual conference of the Canadian Linguistic Association.

not only is Mandarin retroflex lacking of obligatory rounding, but the origin of palatalization in Mandarin also remains controversial. Some research argues that the palatalization was derived from dental sibilants (Hartman 1944; Hockett 1947) while others propose that the palatals have resulted from velar series (Chao 1934). This problem is beyond the present study's scope. Instead of investigating the origin of palatalization in Mandarin, this study rather focuses on the perceptual categorization based on the acoustic differences.

While the three sibilants in PL and TM share a number of acoustic similarities, same places of articulation or same IPA transcriptions do not guarantee identical acoustics. The differences are dependent on contrasts present in each language's inventory (Li, Edwards, and Beckman 2007). In a similar vein, Chiu (2009) found that the acoustic properties of PL and TM sibilants are rather distinct. Considering centre of gravity and formant transitions, the three sibilant in TM are quite different from each other whereas two of the three sibilants in PL are not significantly different. These acoustic cues not only play a role in fricative perceptions but also reflect different attentional weighting from listeners of different ages (Nittrouer and Whalen 1989; Nittrouer 2002). Children weigh formant transition more and frication noise less than adults in perceiving sibilants; as they grow, their cue weighting shifts from the vocal tract movement to the constriction of consonant shapes.

In addition to the attentional weight shifting, the role of these acoustic cues to perception can be also seen in phoneme monitoring tasks (Wagner, Ernestus, and Cutler 2006; Rogers 2009); formant transitions are useful for listeners of languages with spectrally similar fricatives, including Polish. Listeners may be sensitive to mismatched information, but do no necessarily take full advantage of it. The use of formant transitions can be the acoustic make-up of a fricative in relation to all other fricatives in the inventory. Rogers (2009) proposes that English listeners use both transitional and noise cues to monitor the fricatives, and are less affected by noise than Mexican Spanish speakers. In perceiving PL sibilants, frication noise and formant transitions are both claimed to be used (Padgett and Żygis 2006). While the weighting of these two cues was not determined in Padgett and Żygis (2006), Nowak (2003) claims that for Polish listeners, transitional information is more crucial for them to perceive native Polish sibilants. Listeners with different language background may also weight acoustic cues differently in sibilant perception. American English listeners, who lack three-way contrasts in the inventory, showed more difficulties in identifying PL sibilants and depended more on the transitional information from the vowels to identify them (McGuire 2007; McGuire 2008). Mandarin listeners, on the other hand, are suggested to rely more on frication noise in perceiving Polish sibilants. The present study investigates TM listeners' cue weighting in terms of these acoustic properties when perceiving sibilants from languages with similar inventories but different acoustic properties.

2. Experiment

2.1 Method

Three research questions are addressed: (1) What kind of role do these acoustic properties play in sibilant perception? (2) Do TM listeners favour one acoustic cue over the other? Is there any cue weighting change when perceiving two different languages? (3) Can TM listeners weigh the acoustics cues differently when they acknowledge different language information? An experiment was designed to investigate the weighting of the acoustic cues for TM listeners in sibilant identification task. Cross-spliced sound files were created and used as auditory stimuli. The cross-spliced sounds from both languages were created but only played to TM listeners. The participants were divided into three groups and each group received different instructions. The first two groups were instructed that they would hear either all TM sounds or all PL sounds. The third group, to serve as a control, did not receive any language-specific information.

Following previous studies (Wagner, Ernestus, and Cutler 2006; Nowak 2006; McGuire 2006; McGuire 2007), it is predicted that TM listeners rely more on the frication noise than on vocalic information in perceiving sibilants from two different languages. This prediction can be tested by examining the participants' sensitivity to the sibilants with congruent and with incongruent formant transitions, their responses times, as well as their acknowledgement of language information.

2.2 Materials

The auditory stimuli were composed of a sibilant followed by a low back vowel [a], a phonotactically legitimate context in both languages. Ten tokens of the six syllables ([sa], [ʂa], and [ɕa] from both PL and TM) were recorded[2]. The PL stimuli were recorded by a thirty-two-year-old male native speaker of Polish and the TM stimuli were recorded by a thirty-five-year-old male native speaker of Taiwanese Mandarin. Intensities and durations of all the sounds were normalized (60dB; 180 ms for sibilants and 320 ms for vowels). All the recordings were accomplished in the soundproof room at ISRL, UBC.

When splicing, sibilants and vowels were first segmented, yielding six primitive sounds: [s], [ʂ], [ɕ], [a] from [sa], [a] from [ʂa], and [a] from [ɕa]. Each sound file was taken from different tokens. The transition information from each preceding sibilant was included in the vowel files. Each sibilant was then concatenated with the vowels of three different transitions. The permutation created nine types of stimuli in each language. Based on the splicing condition, the stimuli were classified as congruent (sibilant + vowel with matched transition information, such as [sa$_\text{s}$]) and incongruent (sibilant + vowel with mis-matched transition information, such as [sa$_\text{ɕ}$]). Table 1 gives the full list of materials.

[2] To avoid the judgement bias from the intonation pattern of different languages, all the syllables were produced in a high level tone.

Table 1: Polish and Taiwanese Mandarin experiment stimuli

Context	[a] from [sa]	[a] from [ʂa]	[a] from [ɕa]
[s]	[saₛ]	[saʂ]	[saɕ]
[ʂ]	[ʂaₛ]	[ʂaʂ]	[ʂaɕ]
[ɕ]	[ɕaₛ]	[ɕaʂ]	[ɕaɕ]

2.3 Design and Procedure

The experiment employed a between-subject design along within-subject variables. Each participant was assigned in one and only one of the three groups. The participants received 360 (= 9 stimulus types x 10 tokens x 2 PL blocks x 2 TM blocks) trials in total. The orders of the stimuli and the blocks were randomized. To make sure that the participants are familiar with the experiment procedure, a practices session was provided. The practice trials did not contain any cross-spliced sounds. All the participants received the instruction in Taiwanese Mandarin given by the experimenter. The instruction was also written in traditional Mandarin shown on the monitor in front of them. A signal response box sat between the monitor and the participant. Three buttons on the box were activated as response inputs. From left to right, the buttons were labeled with the phonetic symbols used in Taiwan (Zhuyin) corresponding to [s], [ʂ], and [ɕ] respectively. The participants were asked to identify the sounds they hear as quickly as possible by pressing the buttons.

The experiment trials started with a fixation first appearing at the centre of the monitor for 500 ms, followed by a pre-stimulus interval of 500 ms. The auditory stimulus was played immediately after the interval. The participants had 2000 ms from the onset of the sound file to respond. After another 500 ms post-stimulus interval, their response times were shown on the monitor for 1500 ms. The response time was measured from the onset of the auditory stimulus to the detection of the button pressing. The experimenter observed the practice session; for the experiment blocks, all the participants were tested alone. The entire experiment session took approximately 30 minutes to complete. The experiments were carried out on a PC running E-prime (Schneider, Eschman, and Zuccolotto 2009). The response times were mediated by E-prime operation.

2.4 Participants

Twenty seven student participants (13 males and 14 females) from University of British Columbia and Simon Fraser University were recruited. Nine in each experiment group. They are all native speakers of Taiwanese Mandarin. None of them was reported to have visual or hearing disability. They received ten Canadian dollars for participation.

2.5 Data preparation

Participants who had more than 10 congruent errors (e.g., failed to identify the sibilant in [sa$_s$] as [s]) and more than 20 congruent errors in one languages were removed from analyses. This resulted in the loss of six participants, two in each participant condition. The result analyses were based on the remaining twenty one participants (7 in each group).

All the responses and the reaction times (RT) from every participants were logged by E-prime. By subject rates of hit, miss, false alarm, and correct rejection of each sound type were calculated[3]. The reaction times that were three standard deviations away from each participant's mean were defined as outliers and were removed from the analyses.

2.6 Results

The experiment was a mixed design with one between subject variable (experiment group) and a number of within subject variables (language, sibilant, vowel context, sound type, and congruency). Mixed ANOVAs were performed with d', c, and RT as dependent variables. d' is a measure of the sensitivity to distinguish signals and noise. It is the (z score) difference between the hit rate and the false alarm rate (Macmillan and Creelman 2005). The larger the d', the more sensitive the participant to distinguish signals and noise. The d' values across participant groups are presented in Table 2. No difference across groups was found ($F(2, 18)$ $= 2.45$, $p = 0.11$). Main effects were found in sibilants ($F(2, 36) = 60.67$, $p <$ .01), language ($F(1, 18) = 28.68$, $p < .01$), and their interaction ($F(2, 36) = 41.15$, $p < .01$). Post-hoc analyses summarize that (1) the responses to TM in terms of d' values were statistically higher than those to PL ($p < .01$), (2) in terms of the d' values of PL sibilants, the dental was the highest, followed by the retroflex and then the alveolo-palatal (paired t-tests all $p <.01$), and (3) the d' value for TM retroflex [ʂ] was significantly higher from the dental [s] and the alveolo-palatal [ɕ]

Table 2: d' across participant groups

	No Language Info		**Hearing all PL**		**Hearing all TM**	
Context	PL	TM	PL	TM	PL	TM
dental [s]	2.70	2.32	1.76	1.76	2.15	1.95
retroflex[ʂ]	1.61	3.01	1.1	2.42	1.63	2.26
alveolo-palatal [ɕ]	0.73	1.74	0.55	1.55	0.64	1.51
vowel [a$_s$]	2.26	2.4	1.49	1.82	1.75	1.87
vowel [a$_ʂ$]	1.99	2.49	1.47	1.84	1.81	1.83
vowel [a$_ɕ$]	0.79	2.19	0.45	2.06	0.87	2.01

[3] Calculations of rates, d', and c were based on the sibilant types. Calculations followed the formulae in MacMillan and Creelman (2005).

(both p <.01) whereas these two sibilants were not of any difference from each other (p = 0.16).

When only considering sibilants as a variable, TM listeners are less sensitive to identify PL sibilants than TM ones. Among all three sibilants, TM listeners are least sensitive to identify the alveolo-palatal [ɕ] in PL and TM. Taking the vowel contexts (as opposed to sibilants) as a variable, main effects were also found in vowel contexts ($F(2, 36) = 11.53$, $p < .01$), language ($F(1, 18) = 28.68$, $p < .01$), and their interaction ($F(2, 36) = 13.67$, $p < .01$). As we see in Table 2, the sensitivities to the vowel contexts were not particularly different except for the lowest sensitivity to PL [aɕ]. Post-hoc analyses again found that the d' values for TM were statistically higher than those for PL. In PL, the d' values in the context [aɕ] were significantly lower than those in the contexts [as] and [aʂ] (both p <.01) whereas these two contexts were not of any difference (p = 0.99). No difference was found in TM vowel context with respect to their d' values (all p =.99).

In addition to d', the criterion c was also included as a measure of the participants' bias to say "yes" to the signals. The smaller the value c, the more biased the participants. Table (3) presents the c values across experiment groups. Like d', there was no effect across three experiment groups ($F(2, 18) = 0.04$, p = 0.96). Main effects were found in sibilants ($F(2, 36) = 43.95$, $p < .01$), languages ($F(1, 18) = 45.23$, $p < .01$), and their interaction($F(2, 36) = 17.88$, $p < .01$). In PL, the criterion c for the retroflex [ʂ] was significantly lower than those for the dental [s] and the alveolo-palatal [ɕ] (both p <.01) while the dental and the alveolo-palatal were not any different. On the other hand, the c values for TM alveolo-palatal [ɕ] was significantly higher than those for TM dental [s] and TM retroflex [ʂ] (both p <.01). No difference was reported between TM dental and TM retroflex with respect to their c values (p = 0.99).

Table 3: c across participant conditions

	No Language Info		**Hearing all PL**		**Hearing all TM**	
Context	PL	TM	PL	TM	PL	TM
dental [s]	0.31	-0.17	0.57	0.08	0.6	0.03
retroflex [ʂ]	0.38	0.27	0.01	-0.17	-0.11	-0.01
alveolo-palatal [ɕ]	0.6	0.78	0.76	0.99	0.79	0.93
vowel [as]	0.66	0.28	0.38	0.34	0.35	0.33
vowel [aʂ]	-0.05	0.23	0.4	0.3	0.42	0.32
vowel [aɕ]	0.68	0.38	0.56	0.26	0.51	0.3

Considering the vowel contexts as a variable, main effects were also found in vowel contexts ($F(2, 36) = 18.42$, $p < .01$), languages ($F(1, 18) = 45.23$, $p < .01$), and their interaction ($F(2, 36) = 15.66$, $p < .01$). Post-hoc analyses did not find any difference across different vowel contexts in PL and in TM.

Table 4 lists the RTs across experiment groups. No difference across experiment groups was found ($F(2, 18) = 0.39$, $p = 0.68$). In comparing sibilant types, main effects were found in sibilants ($F(2, 36) = 44.85$, $p < .01$), languages ($F(1, 18) = 9.76$, $p < .01$), and their interaction ($F(2, 36) = 14.11$, $p < .01$). In PL, the RTs for the retroflex [ʂ] were statistically shorter than those for the dental [s] and the alveolo-palatal [ɕ] (both $p < .01$) whereas no difference was found between those for the dental and the alveolo-palatal ($p = 0.99$). On the other hand, in TM, the RTs for all three sibilants were significantly different from each other (all $p < .01$). The RTs for the alveolo-palatal [ɕ] were the longest, followed by those for the dental [s] and then the retroflex [ʂ].

Table 4: Mean Response Time (RT) across participant conditions (ms)

	No Language Info		**Hearing all PL**		**Hearing all TM**	
Context	PL	TM	PL	TM	PL	TM
dental [s]	668	592	672	599	622	607
retroflex [ʂ]	603	561	619	548	568	546
alveolo-palatal [ɕ]	657	678	684	673	630	651
vowel [a$_{s}$]	622	592	647	599	589	599
vowel [a$_{ʂ}$]	622	591	646	612	589	597
vowel [a$_{ɕ}$]	683	647	683	608	641	606

When taking the vowel contexts as a variable, main effects were found in vowel contexts ($F(2, 36) = 20.2$, $p < .01$), languages ($F(1, 18) = 9.76$, $p < .01$), and their interaction ($F(2, 36) = 4.37$, $p = 0.02$). For both PL and TM, the RTs in the vowel context [a$_{ɕ}$] were the longest while the RTs for the context [a$_{s}$] and [a$_{ʂ}$] were not of any difference (both $p = 0.99$).

3. Discussion

3.1 *d′*, *c*, RTs, and response counts

As the results show, TM listeners were able to perceive the language differences. In perceiving TM, their sensitivities did not vary much across different vowel contexts while their sensitivities to the alveolo-palatal [ɕ] per se were the lowest. These suggests that transitional informations from the vowel contexts were not sufficient to identify the stimuli. The account that TM listeners may have attended to more information, including frication noise in the sibilant as well as the transitional information from the vowel contexts, when perceiving TM [ɕ] is grounded by their lowest sensitivities. Likewise, in perceiving PL, the degree of dependency of more acoustic information was strongest for the alveolo-palatal [ɕ], followed by the retroflex [ʂ] and then the dental [s] (in the reverse order of the *d′* values).

Figure 1 presents the *d′* of each stimulus based on sibilants. Overall, *d′* values for TM are higher than those for PL, except for [sa$_{s}$] and [sa$_{ʂ}$]. The high *d′* values for TM indicates that the TM listeners were more sensitive in identifying TM sibilants than PL sibilants. Since PL [s] has significantly high frication noise than any other sibilants in the study, the frication noise may be responsible for accounting for the high *d′* values in identifying PL dental sibilant. This also suggests that TM listeners may not only pay attention to the frication noise, but are also particularly sensitive to high frequency noise. As shown in Figure 1, one can also notice that there is not much difference between congruent and incongruent stimuli. Overall, the influence from formant transition in PL alveolo-palatal is so strong that it interferes the perception and lowers the sensitivities (*d′*) to PL [ɕa$_ɕ$], [sa$_ɕ$], and [ʂa$_ɕ$].

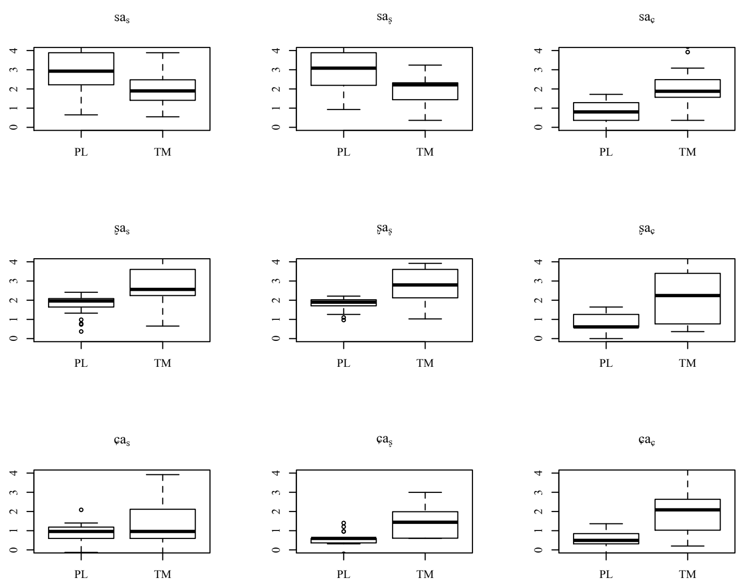


Figure 1: *d′* values across stimuli

With respect to bias, the post-hoc analyses reported that vowel contexts seemed not to be a factor influencing TM listeners' identification of sibilants. Figure 2 shows the bias for each stimulus. As the figure illustrates, in each row, the *c* values for the congruent stimuli were the lowest. Participants showed a stronger tendency (lower *c*) to identify the congruent stimuli as opposed to the sibilants cross-spliced to incongruent vowel contexts. When provided with incongruent stimuli, as long as there is alveolo-palatal formant transition (i.e., [sa$_ɕ$], [ʂa$_ɕ$]), they were less biased. The participants, however, were resistant from the influences of the [a$_s$] and [a$_ʂ$]. Their bias to correctly identify [sa$_ʂ$] and [ʂa$_s$] were still solid. This again suggests that the formant transitions from [s] and [ʂ] do not influence TM listeners perceptual judgements. For both languages, the participants were worse in identifying the alveolo-palatals. Only when the alveolo-palatal is followed by the congruent formant transition are the participants biased to give "yes" responses. The formant transitions from [s] or [ʂ] seemed not contribute to the perception of the alveolo-palatal, no matter which language the TM listeners were listening to. The strong bias towards [$\text{ɕa}_{\text{ɕ}}$] and the weak bias towards [ɕa_{s}] and [$\text{ɕa}_{\text{ʂ}}$] imply that to identify alveolo-palatals, both frication noise and formant transitions are important.

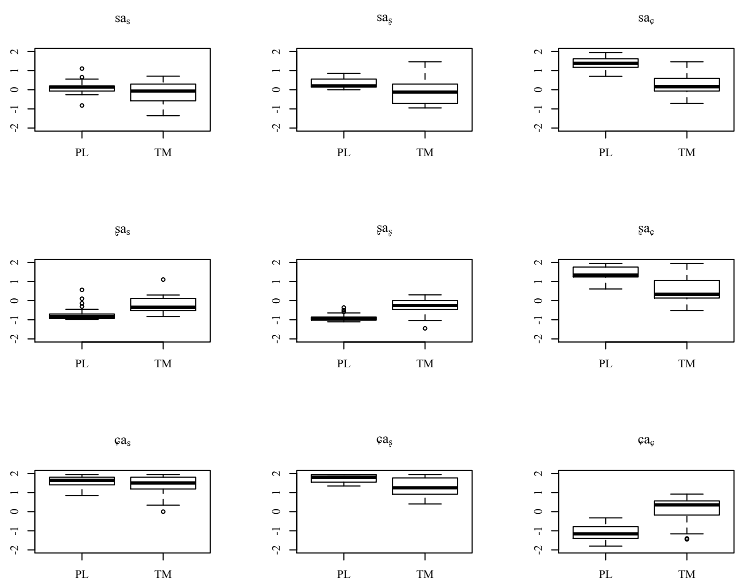


Figure 2: *c* values across stimuli

TM listeners were faster in identifying TM sounds than PL sounds. As shown in Figure 3, no substantial RT difference is observed between congruent and incongruent stimuli between both languages. Vowel contexts seemed not as influential as frication noise to TM listeners' perception. The fact that the RTs for [ɕ] were the longest is correspondent to their sensitivity and bias. It suggests that more information were required for the identification of [ɕ]. Meanwhile, the longest RTs for [$\text{a}_{\text{ɕ}}$]] showed its strong influence on the identification. On the other hand, the RTs for [ʂ] were the shortest. This, however, induces a tradeoff between shorter RTs and higher hit rate as well as higher false alarm rate. This tradeoff is thus considered responsible for the lowest *c* values of [ʂ]. These RT patterns of PL and TM sibilants, as shown in Figure 3, were quite comparable to each other.

Table 5 gives the response counts to PL sibilants. For [s] and [ʂ], the congruency was not a factor to the perception. TM listeners appeared to be more attentive to frication noise than to transitional information. The only influential information from the transition is [$\text{a}_{\text{ɕ}}$]). As for the alveolo-palatal [ɕ], high error rate implies that the identification of [ɕ] is heavily dependent on its vocalic infor-

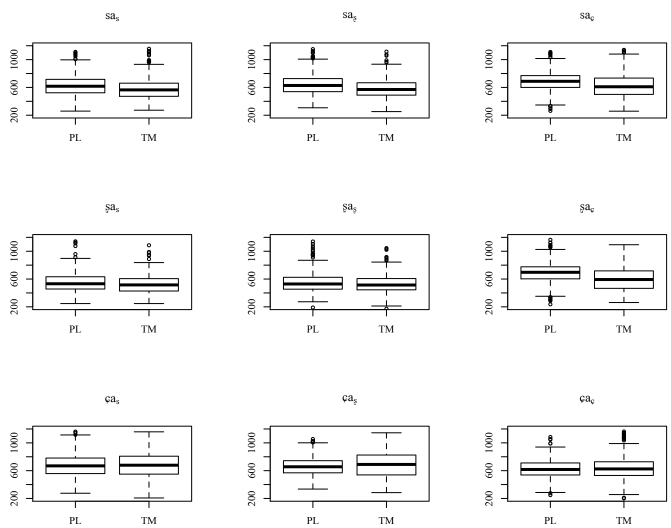


Figure 3: RTs (ms) across stimuli

mation in transition.

Table 5: Polish response counts of nine stimuli. The transitional information is labeled in subscript.

	Polish								
	[sa$_{s}$]	[sa$_{ʂ}$]	[sa$_{ɕ}$]	[ʂa$_{s}$]	[ʂa$_{ʂ}$]	[ʂa$_{ɕ}$]	[ɕa$_{s}$]	[ɕa$_{ʂ}$]	[ɕa$_{ɕ}$]
[s]	364	349	78	32	11	21	45	32	13
[ʂ]	41	51	6	381	405	75	**307**	**358**	31
[ɕ]	12	14	**333**	6	1	**318**	59	26	374

Similar patterns were found in TM sibilant perception (as in Table 6); frication noise seemed to be the dominant cue (over its vowel transitional information) to identify TM [s] and [ʂ]. Like in Polish, the transitional information [a$_{ɕ}$] is influential to perception, and to identify [ɕ] relies heavily on its transitional information in the vowels.

3.2 Perceptual cues and language differences

Figure 1 - 3 show that TM listeners are better at identifying TM sibilants than PL sibilants. They were more sensitive to the frication noise and more biased to say "yes" to the TM sibilants. No significant difference in RTs suggests that the listeners were not biased by the language information given. No familiarity

Table 6: Taiwanese Mandarin response counts of nine stimuli. The transitional information is labeled in subscript.

	Taiwanese Mandarin								
	[sa$_\text{s}$]	[sa$_\text{ʂ}$]	[sa$_\text{ɕ}$]	[ʂa$_\text{s}$]	[ʂa$_\text{ʂ}$]	[ʂa$_\text{ɕ}$]	[ɕa$_\text{s}$]	[ɕa$_\text{ʂ}$]	[ɕa$_\text{ɕ}$]
[s]	350	344	296	29	17	28	**192**	**169**	69
[ʂ]	56	57	31	388	397	269	**97**	**102**	22
[ɕ]	12	14	85	3	4	**118**	125	135	324

effect or novelty effect was observed. Nevertheless, the tendency of longer RTs for alveolo-palatals as well implies that frication noise by itself may not be enough for the listeners to identify the alveolo-palatal sounds. Longer processing times are thus required for the vocalic information following the sibilant.

As suggested by the results, the frication noise from different sibilant types have stronger effects in sibilant perception than the vowel contexts do. TM listeners seemed to be less attentive to the vowel contexts. If TM listeners do pay attention to vocalic info, they should have difficulty in distinguishing PL [s] and PL [ʂ] since the formant transitions for these two sibilants are not significantly different (c.f. Chiu 2009), which is contrary to our findings here. TM listeners were good at distinguishing the dental sibilants and the retroflex sibilants, for both PL and TM. The d' values of the three vowel contexts were consistent across vowel contexts. The particularly low d' value for PL [a$_\text{ɕ}$] suggests a strong effect from the vocalic information in transition. More errors in identifying PL sibilants with [a$_\text{ɕ}$] than identifying the TM sibilants in the same environments suggest that the influence from PL [a$_\text{ɕ}$] is stronger than TM [a$_\text{ɕ}$].

In identifying PL sibilants, when the [a$_\text{ɕ}$] is present, the listeners often mistakenly identified the sound as [ɕ]. Nonetheless, this pattern was not as strong as in identifying TM sibilants. TM listeners more accurately identified the sibilants even when the sibilants were incongruently matched with the [a$_\text{ɕ}$]. This is in response to the findings in Nowak (2003). For PL, when the vocalic cues are available, it may override the frication noise. Particularly, he also points out that the effect of PL alveolo-palatals on the neighbouring vowels seems to be rather different from that of both retroflexes and dentals. The results in the present study showed that the vocalic cues from PL alveolo-palatal indeed overpowered the frication noise during perception.

3.3 Discussion about [ʂ] and [ɕ]

Among all three sibilants, the retroflexes in both languages indeed attract a lot of attention. As shown in Figures 1, 2, and 3, TM listeners show highest sensitivities (high d'), strongest bias (low c), and shortest RTs for retroflexes than for the other two sibilants. In Nowak (2006), it is suggested that PL listeners would first decide whether the sound they hear is a dental or non-dental on the basis of frication noise. The next decision is to identify the non-dental sound as a retroflex or an alveolo-palatal based on frication noise along with F2 transition. If TM listeners used the same strategy, namely to distinguish dental from post-alveolar sounds based on their acoustic properties, we should have found supports from the RT patterns. If there were two stages of sibilant identification involved, RTs for dental sibilants would have been significantly shorter than those for retroflexes and alveolo-palatals. In this experiment, RTs for retroflexes were the shortest. Along with RTs, the sensitivity and bias results suggest that TM sibilants are most sensitive to retroflex sibilants.

Velarization is a secondary articulation when the tongue is drawn back towards the velum (Laver 1994; Trask 1996). As pointed out by previous research (Wierzchowska 1980; Rubach 1984), retroflexion involves the velarization as well as the coarticulation effects with the neighbouring vowels. Hamilton (1980: 21) uses velarization and retroflexion interchangeably for postalveolar fricatives in Polish. If the coarticulation effect between the retracted sibilant and velarized vowels does affect perception, one would expect the RTs and accuracy should be disturbed when those coarticulatory information are present. That is, the vocalic information from the retroflex should interfere the sibilant categorization. The experiment results show that the accuracy rates for retroflexes are high. Therefore in identifying retroflexes, we found fewer confusions with the other two sibilants (i.e., higher accuracy) and shorter RTs. The high percentage of accuracy for retroflex in all kinds of vowel contexts suggest that the vocalic information from the velarization (due to the retraction rule, (Rubach 1984)) do not induce any influence in perception. It is the fricative noise that affects the perceptual judgements.

Alveolo-palatals in both languages puzzled the TM listeners the most. They delivered the lowest *d′* and *c*, along with the longest RT. The most common errors were [s] and [ʂ] spliced with [a$_{\text{ɕ}}$]. Other types of error are rare. This also suggests that in order to successfully identify alveolo-palatals, the TM listeners rely on both cues from the frication noise and the formant transitions. When the target [ɕ] is followed by [a$_{\text{s}}$] or [a$_{\text{ʂ}}$], the participants tend to erroneously identify it as a retroflex [ʂ]. Similar F2 transitions between the retroflex and alveolo-palatal may be responsible for this confusion.

Nowak (2006) argue that the removal of the transitional cues in the second vowel in VCV words makes PL alveolo-palatal and PL retroflexes very confusable. Meanwhile, his study finds that the transitional information from alveolo-palatal sounds is not sufficient to induce the subjects to perceive retroflexes as alveolo-palatals. Nowak (2003) conjectures that this may be due to a more dominant effect from the frication noise of retroflexes. Similar patterns are also found in this study. Across three participant conditions, when a PL retroflex was present, we found high response rate of PL alveolo-palatals; when a PL alveolo-palatal was present, the response rate of retroflex was even higher than the response rate of alveolo-palatal. Our results suggest that although the F2 information from alveolo-palatals may have masking effect in identification retroflexes and alveolo-palatals, the frication noise from retroflex still over powers the vocalic influences.

4. Conclusions

Aside from examining the role of acoustic cues playing in perception, the present study also asks whether or not Taiwanese Mandarin listeners favour one acoustic cue over the other and if there is any cue weighting change when perceiving sibilants from two different languages. In a phoneme monitoring task, listeners who attended to formant transitions are more affected by mismatched transitions. This induces longer RTs and more misses (Wagner, Ernestus, and Cutler 2006; Rogers 2009). The present study finds that even though the listeners were sensitive to the congruency, longer RTs were only found in the context with [aɕ]. This suggests a dominancy of frication noise in sibilant perception, at least for [s] and [ʂ]. To correctly identify [ɕ], in contrast, requires more information from transitions as its higher criterion *c* values and longer RTs showed. Weighting changes of the acoustic cues are seen in perceiving different sibilants. For TM listeners, no matter which language they are listening to, different attentional cues were drawn to perceive [s] and [ʂ] from [ɕ]. Vowel contexts (except for [aɕ]), as opposed to frication noise, do not contribute much to the perceptual difference. Vocalic information is applied only when frication noise is not sufficient. The formant transition from [aɕ] is influential in sibilant perception and always bias listeners' identification of sibilants towards alveolo-palatals [ɕ]. With respect to language acknowledgement, the results lead us to conclude that TM listeners acknowledgement of language information does not affect the low level acoustic perception. The cue weighing of sibilant perception is determined by the native language inventory.

References

Chao, Y-R. 1934. The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of Institute of History and Philosophy* 4:363–397.

Chiu, C. 2009. Acoustic and auditory comparison of Polish and Taiwanese Mandarin sibilants. In *The Proceedings of the Acoustics Week in Canada 2009*, vol. 37, 142 – 143.

Flemming, E. 2002. *Auditory Representations in Phonology*. London; New York: Routledge.

Hamilton, W. S. 1980. *Introduction to Russian Phonology and Word Structure*. Columbus, Ohio: Slavica Publishers.

Hartman, L. M. 1944. The segmental phonemes of the Peiping dialect. *Language* 20:28–42.

Hockett, C. F. 1947. Peiping phonology. *Journal of the American Oriental Society* 67:253–267.

Ladefoged, P., and I. Maddieson. 1996. *The Sounds of the World's Languages*. Blackwell.

Laver, J. 1994. *Principles of Phonetics*. Cambridge; New York, NY: Cambridge University Press.

Li, F., J. Edwards, and M. Beckman. 2007. Spectral measures for sibilant fricatives of English, Japanese and Mandarin Chinese. In *Proceedings of the XVIth International Congress of Phonetic Sciences*, vol. 4, 917 – 920.

Macmillan, N. A., and C. D. Creelman. 2005. *Detection Theory: A User's Guide*. New Jersey: Lawrence Erlbaum Associates, Inc.

McGuire, G. 2006. Selective attention and English listeners' perceptual learning of the Polish post-alveolar sibilant contrast. Under revision.

McGuire, G. 2007. Phonetic cue and integrality. *Journal of Acoustical Society of America* 122:2971.

McGuire, G. 2008. Phonetic category learning. Doctoral dissertation, The Ohio State University.

Nittrouer, S. 2002. Learning to perceive speech - how fricative perception changes and how it stays the same. *Journal of Acoustical Society of America* 112:711–719.

Nittrouer, S., and D. H. Whalen. 1989. The perception effects of child-adult differences in fricative-vowel coarticulation. *Journal of Acoustical Society of America* 86:1266–1276.

Nowak, P. M. 2003. The role of vocalic context in the perception of Polish sibilants. In *Proceedings of the 15th International Congress of the Phonetic Sciences*, vol. 3, 2309–2312.

Nowak, P. M. 2006. The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics* 34:139–152.

Padgett, J., and M. Żygis. 2006. A perceptual study of Polish fricatives, and its relations to historical sound change. In *Poster presented at 10th Conference on Laboratory Phonology.* Paris.

Padgett, J., and M. Żygis. 2007. The Evolution of Sibilants in Polish and Russian. *Journal of Slavic Linguistics* 15:291–324.

Puppel, S., J. Nawrocka-Fisiak, and H. Krassowska. 1977. *A Handbook of Polish Pronunciation for English Learners*. Państvove Wydavnictwo Naukowe, Warsaw.

Rogers, B. 2009. Phoneme monitoring in noise: Evidence for language-specific fricative perception. Unpublished qualifying paper, UBC Linguistics.

Rubach, J. 1984. *Cyclic and Lexical Phonology*. Dordrecht: Foris.

Schneider, W., A. Eschman, and A. Zuccolotto. 2009. E-prime: User's Guide (version 2.0) Psychology Software Tools. [Computer program].

Trask, R. L. 1996. *A Dictionary of Phonetics and Phonology*. London; New York: Routledge.

Wagner, A., M. Ernestus, and A. Cutler. 2006. Formant transition in fricative identification: The role of native fricative inventory. *Journal of Acoustical Society of America* 120:2267–2277.

Wierzchowska, B. 1980. *Fonetyka I Fonologia Języka Polskiego*. Wrocław: Zakład Narodowy imienia Ossolińskich.