THE GRADIENT EFFECT OF TRANSITIONAL MAGNITUDE: A SOURCE OF THE VOWEL CONTEXT EFFECT

Sang-Im Lee-Kim

University of Massachusetts, Amherst
sangimleekim@umass.edu

ABSTRACT

The present study examines the role of transitional magnitude in the identification of consonant place. Transitions in the three vowels /u a ɛ/ following the alveopalatal sibilant /ɕ/ in Polish were systematically manipulated and used as a gradient variable. In an identification task, native Polish speakers were given a choice between /ɕ/ and /ʂ/ for stimuli with varying levels of palatal transitions. The results showed that greater transitions elicit significantly more palatal responses in all vowel contexts. More interestingly, retracted vowels with greater palatal transitions were shown to provide more robust transitional cues than front vowels with weaker transitions.

Keywords: Sibilant, place contrast, transitional magnitude, perceptual cue.

1. INTRODUCTION

The role of vocalic transitions in the perception of consonantal places has been studied extensively in the context of stop place identification [1], English sibilant contrast /s ʃ/ [2, 3], English non-sibilant fricatives /f θ/ [4] and a three-way sibilant place contrast /s ʂ ɕ/ in Shona [5], Polish [6], and Chinese [7]. Using C-V cross-splicing experiments, most studies showed that vocalic transitions play an important role in the identification of consonant place. For example, using stimuli with a continuum of synthetic noise spectra ranging from /ʃ/ to /s/ each combined with transitions of /s/ and /ʃ/, Whalen [1] found that the /ʃ/-to-/s/ categorical boundary shift occurs at much higher frequencies with /ʃ/-transitions than with /s/-transitions.

However, the common methodology, i.e. C-V cross-splicing, is not adequate to isolate the effect of transitions from other factors. For example, while Nowak [6] shows that the identification rate is the highest in the /u/ vowel context, it remains unclear as to whether this particualr vowel effect is based on some properties in the vowel or in the consonant. That is, it is possible that the differences in noise spectra are somehow greater in /u/ (i.e. {ʂ(u)~ɕ(u)} > {ʂ(a)~ɕ(a)} which will be shown to be the case in this study) and thus it is easier to identify consonant place in the /u/ context.

More importantly, C-V splicing is too coarse to precisely probe the role of transitions. Most experiments to date have not systematically controlled the transitional magnitude and moreover random choice of stimuli has given rise to inconsistent results across different studies. Nittrouer and Studdert-Kennedy [8] thus point out that "… the relation between the size of transitional differences found in acoustic measurement and the size of boundary shifts observed in perception warrants more systematic investigation…Vocalic transitions might better be regarded as continuous variables".

Building upon N&S-K [8], the experiments in this study are designed to treat the transition magnitude as a gradient variable through systematic and principled manipulation of the experimental stimuli and examine the effect of transitions in identification of consonantal places in a more rigorous way. In particular, the perception of the contrasting sibilants /ɕ/ and /ʂ/ in Polish was chosen because the acoustic cues in frication noise are relatively weak and listeners tend to heavily rely on the cues in the surrounding vowels [6]. The vowels under examination include /u a ɛ/, since these vowels freely occur with all the sibilants in Polish. The /i/ vowel is not included due to co-occurrence restrictions (i.e. */si/ and */ʂi/). We first establish the acoustic properties of Polish sibilants and surrounding vowels using rigorous acoustic analyses. Based on the acoustic findings, the stimuli are systematically manipulated for the subsequent perception study.

A general prediction is that listeners would give gradient responses in all three vowels, i.e. greater palatal responses for the stimuli with greater transitions. Furthermore, the current study makes a finer prediction based on the acoustic characteristics of the vowels. That is, vowels are different in terms of the magnitude of transitions, and the strength of the cues in transitions to the place of articulation of the consonants is therefore systematically different depending on the vowel type. More specifically, /u/, as a back vowel, is predicted to elicit largest palatal transitions, providing most informative cues for consonant place identification. /ɛ/ is predicted to be least informative due to the small magnitude of palatal transitions. /a/ is predicted to fall in-between.

2. ACOUSTIC STUDY

2.1. Participants

Four female Polish speakers participated in the production study. They were all in their twenties and have lived in the United States less than six months at the time of recording.

2.2. Stimuli

The stimuli for the production study included 27 target words and 43 filler words. Both target and filler words were in the form of a disyllabic $C_1V_1C_2V_2$ sequence. For the target syllables, C_1 varied between /s ʂ ɕ/ and V_1 varied between /u a e/. The second syllable C_2V_2 consisted of one of the stops /p t k/ and one of the vowels /u a e/. V_2 was always identical to V_1. An example of a triplet would be {sapa ~ ʂapa ~ ɕapa}. The randomized stimuli lists were presented in Polish orthography. Participants read each word in a citation form and were asked to place a stress on the first syllable. The materials were read five times. The recordings were made in a sound-attenuated booth at XXX University. The audio signal was collected as 16 bit audio with a 44 kHz sampling rate.

2.3. Measurements

Formant frequencies were measured using Praat by creating a 20 ms window at five evenly divided time points (T1-T5). T1 (set 5 ms after the onset of the vowel) approximately corresponds to the onset of the vowel, T3 to the midpoint of the vowel and T5 (set 10 ms before the end of the vowel) to the end of the vowel. To ensure good spectral estimates, the multitaper spectral analysis [9, 10] was employed for the analysis of frication noise. Based on the averaged spectra, four spectral moments were computed over the normalized spectra: mean (M1), variance (M2), skewness (L3) and kurtosis (L4).

2.4. Results

Figure 1 demonstrates that the differences in vowel formants following /ʂ/ and /ɕ/ are mostly evident along the F2 dimension, consistent with previous studies [6]. Overall, the magnitude of the F2 transition was significantly greater for /ɕ/ than for /ʂ/. The results also identified a significant role of the vowel context: F2 transition was significantly greater for /u/ (*M* = 680 Hz) than for /a/ (*M* = 346 Hz), followed by the lowest /e/ (*M* = 155 Hz) (all *p* < .001)

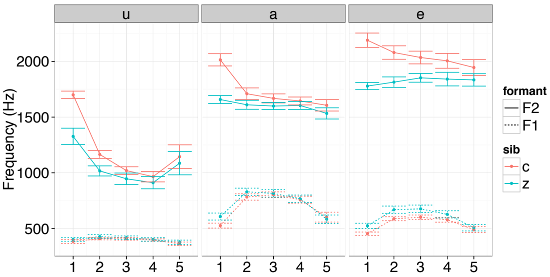


Figure 1: F1 (dotted) and F2 (solid line) trajectories of /ɕ/ ("c") and /ʂ/ ("z").

Figure 2 shows representative multitaper spectra of the three sibilants /s, ʂ, ɕ/ in the three vowel contexts /u, a, e/ taken at mid-phase of frication noise from one speaker. Most relevant to the current study is the effect of the rounded vowel context. In unrounded vowel contexts (i.e. /a/ and /e/), most of the energy distribution of /ʂ/ and /ɕ/ are concentrated at lower frequencies with /ʂ/ being slightly lower than /ɕ/, consistent with the literature [6]. In the rounded vowel context (i.e. /u/), however, the /ʂ/ spectrum has gone through the most considerable changes: the spectral peak becomes a plateau-like peak that covers a wide range of frequencies (i.e. 2-7 kHz). Therefore, the spectral differences between /ʂ/ and /ɕ/ becomes even larger in this context.

Figure 2: Multitaper spectra of one speaker. The three sibilants /s/ (grey), /ʂ/ (black) and /ɕ/ (dark grey) in /a/ (top), /e/ (middle) and /u/ (bottom).

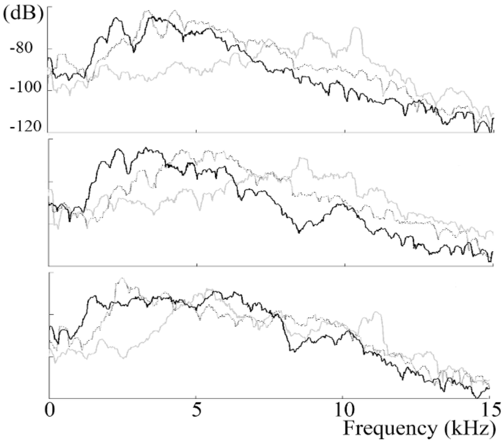


Table 1 summarizes the statistical results showing the relative ranking of the sibilants according to the four spectral moments (">": significant results). Most notably, the spectra of /ʂ/ become more dispersed (greater M2) and left-skewed (lower L3) and less peaked (lower L4) in the rounded context, confirming the patterns observed in the spectra.

Table 1: A summary of the spectral moments

Moments	Unrounded	Rounded
M1	s > ɕ > ʂ	s > ɕ ≈ ʂ
M2	s > ɕ ≈ ʂ	ʂ ≈ s > ɕ
L3	ʂ > ɕ > s	ɕ > s > ʂ
L4	ʂ > ɕ ≈ s	ɕ ≈ s ≈ ʂ

3. PERCEPTION STUDY

Having established the acoustics of Polish sibilants and the following vowels, the present perception study examines whether the relative magnitude of the palatal transitions depending on the vowels is the source of the "vowel effect". To that end, the acoustics of frication noise was strictly controlled for, while vowel transitions were systematically manipulated.

3.1. Stimuli

3.1.1. Manipulation of formant transitions

In order to examine vowel transitions as a gradient variable, nine tokens with /ɕ/ as an initial (3 vowels (u/a/e)* 3 consonants (p/t/k); e.g. /ɕapa/) from one speaker were selected based on clarity of the speech signal. They were used as original tokens for manipulation of transitions. Specifically, the transitional period after /ɕ/ was cut out by two pulses from the vowel onset until most transitions were removed: e.g. a$_0$-a$_2$-a$_4$-a$_6$-a$_8$, where the numbers represent how many pulses were removed from the original signal. The cutouts were made at zero-crossings of every two pulses. A five-step continuum was chosen because it covers most of the transitional period of the vowel.

While this method allows for very natural sounds, the tokens with greater cutouts are necessarily shorter than those with no cutouts in length. To compensate for the pulses cut out (e.g. a$_4$, a$_6$, a$_8$), pulses at around 75% into the vowel were copied and added back in to a similar location at zero-crossings. To maintain the natural flow of the waveform, the addition of pulses was carried out by adding two pulses in an incremental manner. However, no more than six pulses were removed or added to the original token in order to avoid any unnatural signal.

3.1.2. Manipulation of frication noise

The acoustic analysis of Polish sibilants showed that there was a strong coarticulatory effect on frication noise from the following vowel, particularly in the rounded vowel context. Hence, to exclude any influence from the noise period, manipulation of the noise signal is also necessary.

To make the stimuli relatively compatible with distinct vowel contexts, frication noise in all vowel contexts were mixed to create one grand noise signal for each sibilant: grand average of the retroflex sibilant ʂ$_G$ and the alveopalatal sibilant ɕ$_G$. After normalization of duration and intensity, the noise in /a/ and /e/ contexts (e.g. /ʂ$_a$/ and /ʂ$_e$/) was combined first and this mixed signal was combined with the noise in /u/ (e.g. /ʂ$_u$/) in Praat. It should be noted that the grand average of ʂ$_G$ turned out to be closer to the spectra of /ʂ$_u$/ especially at the low frequency range between 1 and 2 kHz in spite of the same weighting of the signals from the rounded and non-rounded contexts. This is likely to make the resulting signal more compatible with /u/, but further manipulation was not implemented in order to be consistent with the method used for the creation of the grand alveopalatal sibilant ɕ$_G$. In addition, a mixed noise S$_M$ from ʂ$_G$ and ɕ$_G$ (amplified by 1.5) was created in order to add an additional condition where the noise property is more ambiguous. The three averaged sibilants (ʂ$_G$ and ɕ$_G$ and S$_M$) were combined with the sequence $V_1C_2V_2$ with different transitions in V_1 created in the previous step (e.g. ɕ$_G$a$_0$ta, ɕ$_G$a$_2$ta, ɕ$_G$a$_4$ta, ɕ$_G$a$_6$ta, ɕ$_G$a$_8$ta).

3.2. Procedure

14 native speakers of Polish (12 F and 2 M) participated in the perception study, ages ranging mostly from 25 to 35 with one female speaker at 55. Most participants have lived in the United States for less than a year except for two speakers.

The perception experiment was an identification task implemented in E-Prime. Given an auditory stimulus (e.g. ɕ$_G$a$_2$ta), two forced-choice items, the /ʂ/-onset (/ʂata/) and /ɕ/-onset (/ɕata/), were provided in a written form. The two items appeared on the screen randomly as encoded in the E-prime script.

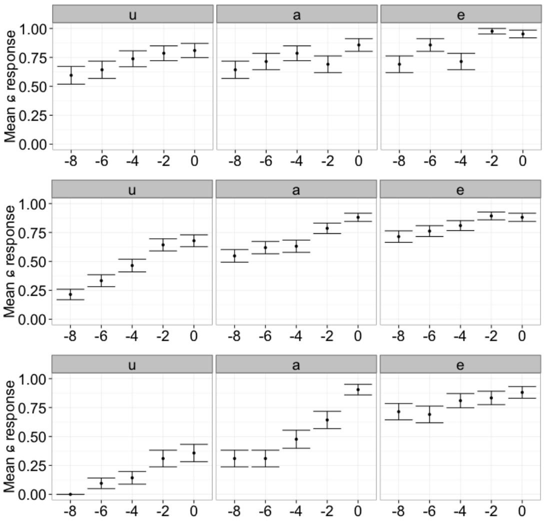
The stimuli were grouped into six blocks. For the first three blocks (30 trials per block), ʂ$_G$-onset and ɕ$_G$-onset stimuli were presented in a random order. For the last three blocks, S$_M$-onset trials were duplicated (90 trials) and they were randomly divided into three blocks again. There were no fillers, and all the stimuli were sibilant-initial words.

Participants were asked to respond as quickly as they could and to choose either choice even when they were not certain. Participants were seated in a quiet laboratory room equipped with PC computers and Sennheiser headphones. A button box marked with 1 or 2 was placed in front of the computer.

3.3. Results

Figure 3 plots mean /ɕ/ responses against transitions in each vowel condition. Overall, the results show that /ɕ/ responses gradually increase as F2 transitions become greater from Step 8 (nearly zero transitions) to Step 0 (full transitions). The results overall reveal that listeners are sensitive to the fine-grained phonetic details and use the information in vowel transitions in a gradient manner.

Figure 3: Mean /ɕ/ responses in ɕ$_G$ (top), S$_M$ (middle), and ʂ$_G$ (bottom) by vowels & transitions



To examine the statistical significance of the role of transitions, a mixed effect binomial regression model was fit to the data separately by onset sibilant (Model: */ɕ/ response ~ Vowel*Transition+(1+Vowel +Transition|Subject)+(1|Item), family="binomial")*. The results showed that vowel transitions are a significant predictor of the identification of consonant place in all vowel contexts (all $p < .05$). Below, we continue an in-depth discussion about the relative strength of vowel transitions.

4. GENERAL DISCUSSION

4.1. Onset bias and the acoustic details in the stimuli

In addition to the general pattern, the results of the perception study also indicate the interaction between sibilant and vowel type. In the /u/ vowel context, mean /ɕ/ responses were significantly greater for ɕ$_G$ (*M*= 71%) than for S$_M$ (*M*= 47%), followed by ʂ$_G$ (*M*= 18%). The mean responses were not significantly different by onset type for other vowels.

The reason that the vowel /u/ patterns differently from other vowels can be inferred from the acoustic details of the stimuli. In the creation of the grand mean of ʂ$_G$, the contribution of /u/ was much larger (see §3.1.2), suggesting that the noise signal ʂ$_G$ is likely to be more compatible with the /u/ vowel context than with others. With the onset pointing toward /ʂ/, participants' response seems to have been strongly biased toward /ʂ/ in the /u/ vowel context. As the contribution of the /ɕ/ onset becomes greater from ʂ$_G$ to S$_M$ and from S$_M$ to ɕ$_G$, significantly greater /ɕ/ responses were elicited as a result.

Having identified the source of the onset bias in the /u/ vowel context, it enables a richer and more nuanced interpretation of the results of the perception study. Specifically, it guides us to predict under what condition the role of transitions would emerge most prominently. In particular, with the ʂ$_G$ onset, i.e. a good /ʂ/ for /u/ but less so for /a/ or /e/, it is possible that transitional cues are more important for the latter due to the non-canonical and thus less reliable noise cues. With the onset S$_M$, a reverse pattern is predicted. This onset is a less canonical /ʂ/ for /u/, and therefore cues in /u/ vowel transitions would become more informative. In contrast, the role of /a/ and /e/ transitions is predicted to decrease, as the cues in the sibilant become more compatible with the vowels. Lastly, given the acoustic findings that ɕ$_G$ was largely compatible with all vowels, vowel transitions are predicted to be less important for all three vowels and listeners may rely mostly on the spectral cues, largely resulting in /ɕ/ responses.

4.2. The relative magnitude of transitions

Table 2 presents the coefficient values in each model and the statistical significance of the interaction between vowel and transitions.

Table 2. Transition coefficient and their relation across different vowel contexts

Onset	Comparison between vowels
ɕ$_G$	e(0.53) ≈ u(0.36) ≈ a(0.25), e > a
S$_M$	u(0.47) > a(0.29) ≈ e(0.29)
ʂ$_G$	u (0.55) ≈ a (0.45) > e (0.22)

Based on the predictions developed in 4.1, two conclusions are drawn from the results summarized in Table 2. First, the results of the ʂ$_G$ onset provide clear and unambiguous evidence that /a/ transitions are stronger than /e/ transitions. In this context, both vowels are predicted to show a relatively strong effect of vowel transitions, but /a/ was significantly stronger than /e/. The results of the S$_M$ onset are in accordance with this interpretation. This context is predicted to show a decreased effect of transitions for /a/ and /e/. Indeed, /a/ and /e/ are no longer different from one another, and /a/ transitions become weaker than /u/ transitions. Second, the overall results further confirm that the /u/ transitions are the greatest among all vowels. Even in the onsets ɕ$_G$ and ʂ$_G$ where the /u/ transitions are not predicted to be the most robust, the magnitude of /u/ transitions are not smaller than other vowels in comparison. In the context of S$_M$, /u/ transitions arise as the strongest as predicted. Taken together, it is concluded that the strength of vowel transitions are the greatest for /u/, intermediate for /a/, and the weakest for /e/. This is in accordance with the relative magnitude of F2 transitions that were in the same order of /u/, /a/ and /e/.

5. REFERENCES

[1] Dorman, M.F., Studdert-Kennedy, M., Raphael, L.J. 1977. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context dependent cues. *Percept. Psychophys.* 22, 109–122.

[2] Whalen, D.H. 1981. Effects of vocalic formant transitions and vowel quality on the English [s]–[š] boundary. *J. Acoust. Soc. Am*. 69, 275–282.

[3] Whalen, D.H. 1991. Perception of the English /s/–/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices. *J. Acoust. Soc. Am*. 90, 1776–1785.

[4] Babel, M., Grant, M. 2013. Listener expectations and gender bias in nonsibilant fricative perception. *Phonetica* 70, 117–151.

[5] Bladon, A., Clark, C., Katrina, M. 1987. Production and perception of sibilant fricatives: Shona data. *J. Int. Phonet. Ass.* 17, 39–65.

[6] Nowak, P.M. 2006. The role of vowel transitions and frication noise in the perception of Polish sibilants. *J. Phonet.* 34, 139–152.

[7] Chiu, C., Babel, M. 2010. Effects of syllable positions on Taiwanese Mandarin sibilant perception. *Proc. 7th ISCSLP,* Taiwan.

[8] Nittrouer, S., Studdert-Kennedy, M. 1987. The role of coarticulatory effects in the perception of fricatives by children and adults. *J. Speech Lang. Hear. Res.* 30, 319–329.

[9] Blacklock, Oliver S., and Shadle, Christine H. 2003. Spectral moments and alternative methods of characterizing fricatives. *J. Acoust. Soc. Am*. 113, 2199.

[10] Blacklock, O.S.B. 2004. Characteristics of variation in production of normal and disordered fricatives using reduced-variance spectral methods. University of Southampton.