Perceptual motivations of sibilant harmony

Avery Ozburn
University of British Columbia

Abstract: This paper presents results of a categorization study of sibilants in different contexts, which examines whether perception mirrors typological patterns of sibilant harmony. Results show that listeners have an assimilatory preference in perception of a [s]-[ʃ] continuum when the context contains an [s] or an [ʃ] compared to non-sibilant contexts, but that other characteristics of sibilant harmony are not reflected in perceptual categorization.

Keywords: sibilant harmony, perception, categorization, consonant harmony typology

1 Introduction

Consonant harmony, also known as long-distance consonant assimilation (Rose and Walker 2004) is a phonological phenomenon in which certain consonants within a word are required to agree in a particular phonological feature (Rose and Walker 2004; Hansson 2010). An example is sibilant harmony, in which sibilants within a word may be required to agree in the feature [anterior] (Hansson 2010, Rose and Walker 2004). While many properties of consonant harmony cross-linguistically are known, the reasons that languages develop such patterns are under-investigated. It has been suggested that consonant harmony may potentially arise from phonologized speech errors and coarticulation (e.g. Hansson 2010); however, there has been little investigation into whether consonant harmony could be affected by listener-driven factors (see e.g. Hansson 2008 for an overview of the role of such factors in the development of phonological patterns). While Gallagher (2012) has found that judgements about perceptual similarity parallel certain properties of ejective harmony, such perceptual effects have not been studied more broadly in other types of consonant harmony. The present study examines the perception of sibilant contrasts in different contexts, in order to investigate whether perceptual factors could be driving the typological properties of sibilant harmony. In particular, it investigates the fact that more (featurally) similar segments are more likely to interact harmonically (see e.g. Rose and Walker 2004, Hansson 2010). If harmony patterns are in part determined by misperceptions, then we expect pairs of sounds that contrast in the harmonic feature to be perceived differently in typologically common harmony contexts than in non-harmony contexts. Since sibilant harmony is cross-linguistically the most common type of consonant harmony (Hansson 2010) and is easily tested on English listeners, whose language does not have categorical sibilant harmony to affect judgements, it offers an ideal way to examine these potential effects.

This paper reports on an experiment consisting of a forced choice categorization task using CVCV sibilant continua, where the first consonant was along a continuum between [s] and [ʃ] and the second was one of the following: sibilants [s] and [ʃ] that are highly similar to the consonant being categorized; less similar sibilants [z] and [tʃ]; or non-sibilants [n] and [m]. Results are analyzed to determine the extent to which the 50% crossover boundary in categorization as [s] or [ʃ] of continuum sounds depends on the context consonant. While results generally show a shift towards [s] response with [s] in the context and a shift towards [ʃ] with context [ʃ], they also demonstrate unexpected patterning of the less similar sibilants [tʃ] and [z], as well as an unexpected tendency for [s] to trigger more perceptual shift than [ʃ].

In UBC Qualifying Papers 3 (2013–2015),
University of British Columbia Working Papers in Linguistics 44,
Andrei Anghelescu, Joel Dunham, and Natalie Weber (eds.), 2016.

The paper is organized as follows. Section 2 presents background information on sibilant harmony and previous perceptual studies. Section 3 details the methodology of this experiment, and Section 4 presents the results. Section 5 discusses the results in terms of the typology of consonant harmony, and Section 6 concludes.

2 Background

2.1 Sibilant harmony

In sibilant harmony patterns, two sibilants differing in the feature [anterior], such as [s] and [ʃ], cannot both occur in a single morpheme or word (Hansson 2010). For example, a language with sibilant harmony in [s] and [ʃ] could allow words of form [s…s] and [ʃ…ʃ], but not *[s…ʃ] and *[ʃ…s]. Examples of languages with sibilant harmony include Slovenian, Chumash, and Sarcee[1].

Within the languages with sibilant harmony, as well as those with other types of consonant harmony, several generalizations can be made about which patterns are widely attested and which are rarely attested or not attested at all. The widely attested properties that will be examined in this paper are dominance and similarity-sensitivity. All of these properties are well-documented effects in consonant harmony. Indeed, phonological accounts of harmony like Agreement by Correspondence (ABC; Rose and Walker 2004), a major theory used for consonant harmony, have been based on them. Each of these properties will be discussed below.

First, sibilant harmony tends to be triggered by [-anterior] segments like [ʃ] rather than [+anterior] segments like [s], which is a property known as dominance or trigger asymmetry. For example, in Sarcee, [ʃ] triggers harmony but [s] does not; there are 13 known languages with the Sarcee-type pattern, but only one that is known to have the reverse pattern, where [s] triggers harmony but [ʃ] does not (Kosa 2010). While this property may not seem strong from sibilant harmony alone, generalizations about trigger asymmetries also hold of other types of consonant harmony. Throughout languages with various types of consonant harmony, it is common to have restrictions in which one type of interacting segment, almost always the one considered more phonologically 'marked', triggers harmony, but the other does not (Hansson 2010). For example, in some languages with voicing harmony, such as Ngizim, [+voice] segments trigger harmony, but [-voice] ones do not (Hansson 2010, Schuh 1997).

Second, many consonant harmony patterns, including sibilant harmony ones, are sensitive to similarity in other features. For sibilant harmony, the relevant features are manner and voicing. For example, Wanka Quechua demonstrates harmony in fricative-fricative pairs and affricate-affricate pairs, but combinations differing in manner (fricative-affricate) do not harmonize (Hansson 2010). Similarly, in Nkore-Kiga, sibilant harmony is more limited in combinations of fricatives that differ in voicing (e.g. [ʃ]/[z]) than in those that agree in voicing (e.g. [ʃ]/[s]) (Hansson 2010). Moreover, cross-linguistically, many types of consonant harmony are sensitive to similarity in other features; laryngeal harmony is often sensitive to similarity in place, manner and other laryngeal features, for instance. Similarly, in Kalasha retroflex harmony, fricative/fricative, affricate/affricate, and stop/stop combinations undergo harmony, but stop/affricate, stop/fricative, and affricate/fricative combinations do not (Arsenault and Kochetov 2011). This property of consonant harmony is so crucial that it is the foundation of the Agreement by Correspondence (ABC) theory in all of its forms, with constraints defined in a way that allows

[1] Counting the cases of sibilant harmony in the appendix of Hansson (2010) gives an approximate total of 65 known cases of sibilant harmony cross-linguistically.

more similar segments to interact harmonically while less similar segments do not (Rose and Walker 2004).

Overall, there are a number of properties of sibilant harmony that are important to the design of the present study[2]. These characteristics are also true of a variety of other types of consonant harmony, such as ejective harmony, voicing harmony, aspiration harmony, and so on (Hansson 2010), and they are robust enough to have phonological theories developed to account for them. As such, using experimental work to look for potential acoustic and perceptual correlates of such patterns is not only interesting, but also possibly useful to understanding how these patterns arise.

2.2 Motivations of consonant harmony

The theories on why and how consonant harmony develops as a sound pattern are relatively underdeveloped. In vowel harmony, which is much better studied, research has suggested that both articulatory and perceptual motivations play important roles for different properties and systems (see e.g. Benus 2005 on front/back harmony, Kaun 1996 on rounding harmony, Przedziecki 2005 on ATR harmony, etc.). For example, Kaun (1996) argues that vowels with weak perceptual rounding cues tend to trigger rounding harmony, while those with strong cues tend to be targets. She suggests that such a pattern could have developed from a bias towards making rounding cues as salient as possible; extending rounding from vowels with weak cues to those with strong cues makes it more likely for the rounding feature to be correctly identified (Kaun 1996).

In contrast, relatively few studies have looked at these types of effects in consonant harmony. One of the only studies comes from Gallagher (2012) who provides evidence that misperceptions of ejective/non-ejective contrasts by English speakers mimic the typology of ejective harmony systems. Testing pairs of CVCV words with 1 vs. 2 (e.g. p'itu-p'it'u), 1 vs. 0 (e.g. p'itu-pitu), and 2 vs. 0 ejectives (e.g. p'it'u-pitu), participants were best at discriminating the 2 vs. 0 contrast and worst at the 1 vs. 2 contrast, which mirrors the fact that languages with ejective restrictions, assimilatory or dissimilatory, disprefer the 1 vs. 2 contrast. Such results suggest a potential perceptual motivation for at least some types of consonant co-occurrence restrictions.

In terms of sibilant harmony, Hansson (2010) suggests that the patterns could be due to phonologization of coarticulation and speech errors, which, like sibilant harmony, tend to be regressive and triggered by [ʃ]. In other words, the claim is that it is easier to produce agreeing sibilants than disagreeing ones, creating coarticulatory patterns and speech errors. These coarticulations and errors then become part of the phonology, through a process that is not well understood. However, to date, no studies have investigated potential perceptual motivations for the properties of sibilant harmony described above. Given arguments for the idea of various types of harmony as listener hypocorrection (see e.g. Ohala 1994a, Ohala 1994b, Hansson 2008), exploring perceptual correlates of sibilant harmony could prove crucial to understanding its origins. In particular, the hypocorrection hypothesis suggests that imperfect perceptual compensation for coarticulation, through attributing coarticulatory effects to the target rather than the context, could lead to the phonologization of harmony patterns (Ohala 1994a, Ohala 1994b, Hansson 2008). If that is the case, then knowing how listeners perceive sibilant contrasts in contexts that do and do not trigger harmony typologically could provide a deeper understanding of how sibilant coarticulatory effects, whether they are real or inferred, may become part of the

[2] It is also relevant to note that the choice of CVCV words for this study, and not a variety of word types, is important due to the fact that in some languages, consonant harmony occurs only between consonants in adjacent syllables, and not ones further apart (see McMullin and Hansson 2013 for discussion and references). This fact will not be discussed further here.

phonology of a language. However, no previous research has looked for any perceptual correlates of sibilant harmony patterns.

While no one has investigated potential perceptual motivations for sibilant harmony, evidence for perceptual foundations of consonant interactions has been found for similar patterns. In particular, perceptual studies exist for strictly local sibilant assimilations across word boundaries (Fleischer et al. 2013) and for long-distance dissimilation of liquids (Abrego-Collier 2013). Both studies used forced-choice categorization tasks and were run on English speakers, and thus prompted similar methodological decisions for the current study.

For local sibilant assimilations, Fleischer et al. (2013) found significant flattening of a sibilant perception curve (the curve made by plotting percent [s] response per step along a continuum between [s] and [ʃ]) preceding other sibilants, compared to preceding vowels; participants were less likely to respond [s] at the [s] end and less likely to respond [ʃ] at the [ʃ] end in the sibilant contexts than in the neutral contexts. This effect was similar for both contexts [s] and [ʃ], even though [s] is not a trigger of local assimilation (Niebuhr et al. 2011). Thus, a strictly adjacent sibilant caused more ambiguity in the perception of a preceding sibilant in that experiment, suggesting that local sibilant assimilations could be due to perceptual factors.

In terms of liquid dissimilation, Abrego-Collier (2013) found that the categorization boundary of an [r]-[l] continuum was shifted towards more [r] responses when the context contained another (not strictly adjacent) [r] compared to contexts containing a neutral [d], and shifted towards more [l] response when the context contained another [l]. As such, in non-local perception of ambiguous liquids, there is an assimilatory preference, despite the typological dissimilation tendencies. Abrego-Collier (2013) interprets this result as indicative of hypocorrection, with listeners not correcting for the perceptual influence of context [r] and [l] on the continuum consonant, and therefore interpreting more of the continuum in an assimilatory way. As such, there is evidence that non-local patterns may have perceptual motivations.

The question in the present experiment is whether sibilants across a vowel show similar perceptual effects as shown for strictly adjacent sibilants and for liquids, in a way that suggests a potential perceptual motivation for the typology of sibilant harmony. As such, we might expect to see a flattening or an assimilatory or dissimilatory shift in the perception of ambiguous sibilants in harmony contexts compared to neutral contexts. The current study looks only for shifts, primarily because previous work on long-distance phenomena (the liquid study) has shown shifts.

Following the ideas from these previous studies and arguments, the core hypothesis tested here is that perception of ambiguous sibilants shifts assimilatorily in a way that reflects the typology of sibilant harmony. This general hypothesis can be divided into three separate predictions. First, the sibilants [s] and [ʃ] will show an assimilatory shift compared to the neutral conditions, with more [s] response for [s] contexts and more [ʃ] response for [ʃ] contexts. This prediction follows from Ohala's harmony as hypocorrection hypothesis, as well as analogy with the [r]/[l] results. The idea is that listeners do not correct for the assimilatory coarticulation effects that sibilants have on each other, and therefore perceive more of the ambiguous [s]/[ʃ] tokens as [s] when there is a nearby [s] in the context to affect them. Second, [ʃ] contexts will show more shift than [s] contexts, in that the crossover point will be further from the neutral contexts or significantly different from neutral contexts more often. This prediction follows from the assumption that the perceptual effects will mirror the typology; [ʃ] is a more robust harmony trigger cross-linguistically due to it being a [-anterior] segment. If this asymmetry is due to perceptual effects, and in particular to hypocorrection of the sensory analysis, then it is expected that the shift resulting from hypocorrection should happen more for [ʃ]. Moreover, analogy with the perceptual motivations for trigger asymmetry in rounding harmony (see Kaun 1996) further motivates the idea that perception may be relevant to this property of sibilant harmony. Third, [z]

and [tʃ] will pattern like [s] and [ʃ] respectively, though potentially with less shift. This prediction is strictly typological; cross-linguistically, they either pattern like sibilants of the same value of [anterior] or do not trigger harmony. Again, if this pattern is due to perception, then the expectation is that the harmonic shift triggered by these sibilants that are less similar to the consonant being categorized will be similar to that triggered by most similar [s] and [ʃ].

3 Methodology

To test the above hypotheses, an experiment was conducted using a forced choice categorization task with [s]-[ʃ] continua. The study examines to what extent the point at which participants begin categorizing the continuum as [ʃ] more than [s] depends on the identity of surrounding consonants.

3.1 Participants

There were 74 participants in the study, but 36 participated in a different condition (C2 condition) for which the results are not included in the present paper, and 18 were excluded for being non-native speakers. Included participants were 20 native English speakers (ages 18 to 28; 18 female, 2 male) who reported no speech or hearing disorders. They were recruited from the University of British Columbia linguistics participant pool, and were compensated with course credit.

3.2 Stimuli

A phonetically trained male native speaker of Canadian English from the Vancouver area produced 36 CVCV nonce forms, one with each of [s] and [ʃ] in place of the S in each of the 18 (= 3 vowel contexts x 6 consonant contexts) forms in Table 1. The speaker was presented with the [s] and [ʃ] forms of the same cell from Table 1 (i.e. the continuum endpoints) in pairs (e.g. [sama]/[ʃama], [sisi]/[ʃisi], etc.) and was instructed to produce the members of each pair as similarly as possible, in terms of characteristics such as vowel length, duration, and intonation. This was done in order to make later synthesis of continua easier, because pairs differing in these characteristics gave rise to continua that sounded more unnatural. Each pair was repeated as many times as necessary, sometimes over multiple sessions, to obtain recordings that were of good quality and for which the pairs were as similar as possible to each other. Sound was recorded through a C520 headset microphone into a USB Pre2 pre-amp, with the microphone placed approximately 8cm from the speaker's mouth. The sound was recorded into Praat (Boersma and Weenink 2015) at a 44,100Hz sampling rate.

Table 1 Stimuli list (where S indicates the consonant being categorized)

Context Consonant	**C1 condition**		
	[a]	**[i]**	**[u]**
[s]	Sasa	Sisi	Susu
[z]	Saza	Sizi	Suzu
[n]	Sana	Sini	Sunu
[m]	Sama	Simi	Sumu
[tʃ]	Satʃa	Sitʃi	Sutʃu
[ʃ]	Saʃa	Siʃi	Suʃu

The natural stimuli were then synthesized into continua using the program STRAIGHT in Matlab (Kawahara et al. 2008). For each pair, the entire word of the [s] recording and the entire

word of the [ʃ] recording were morphed together in an 11-step continuum (0% to 100% of the [ʃ] recording, in steps of 10%). Prior to morphing, the [s] and [ʃ] stimuli were time-aligned based on acoustic landmarks, including onset of formants and frication. With 18 pairs of stimuli each used to create an 11-step continuum from [s] to [ʃ], the result was 198 distinct stimuli. Silences at the beginnings and ends of the stimuli were then trimmed in Praat (Boersma and Weenink 2015). Figure 1 shows spectrograms of the initial consonant of the [saʃa]-[ʃaʃa] continuum at Steps 1, 6, and 11, showing how the fricative changes at each step.

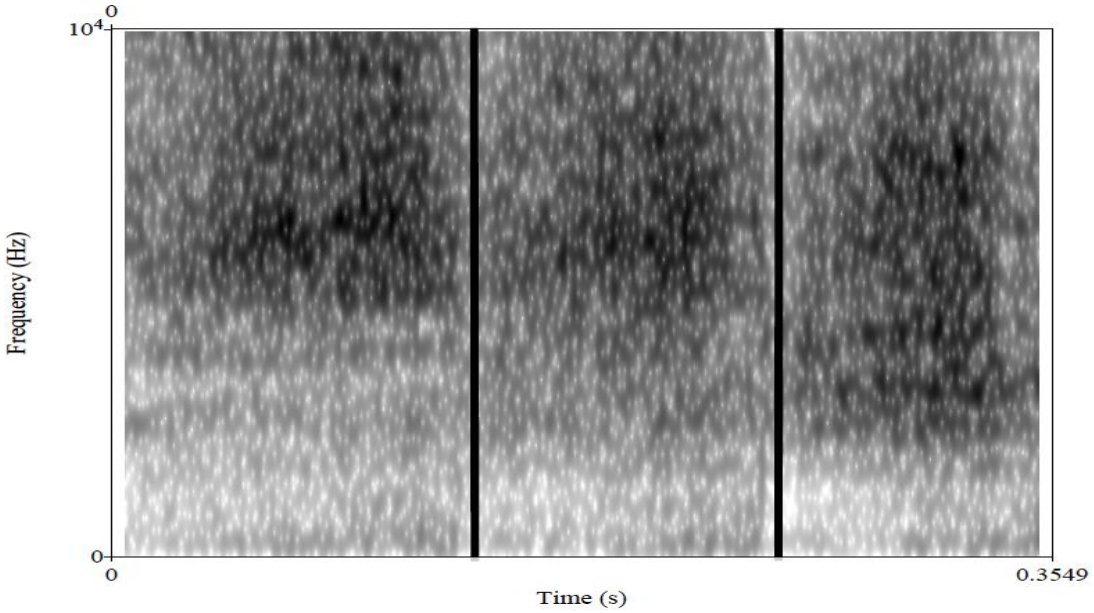

Figure 1 Spectrogram of the continuum consonant at Steps 1, 6, and 11 of the [saʃa]-[ʃaʃa] continuum

3.3 Procedures

The study was presented on a computer using E-Prime (Psychology Software Tools, 2012). Participants were seated in front of a computer and wore AKG K240 headphones. For each trial, participants heard a single stimulus and then a screen displayed to remind them to press 1 on the button box for a "s" response and 5 for a "sh" response. No other buttons were registered as responses. The only feedback given to participants was that their response was registered.

The experiment consisted of a short practice block followed by an experimental block. The practice block consisted of six stimuli, drawn from the same set as the actual stimuli, three at Step 1 of the continuum (unambiguous [s]) and three at Step 11 of the continuum (unambiguous [ʃ]). They were presented in an alternating Step 1, Step 11 pattern. The context consonant in the practice stimuli was either [s] or [ʃ], to give participants the opportunity to practice which of the two consonants they were categorizing. Of the six practice stimuli, there were two with each vowel, and these were the endpoints of the same continuum.

The trial block consisted of all stimuli described above. As such, each participant heard a total of 198 (= 18 x 11) distinct stimuli, which were repeated twice in random order, for a total of 396 fully randomized trials per participant. Participants were instructed to listen to the entire word, to help ensure they listened to the entire stimulus, and then to respond as quickly as possible about whether they heard the first consonant (C1) of the word as [s] or [ʃ]. The experiment automatically moved on if participants did not press 1 or 5 within three seconds from

the trial. Participants were given the opportunity to take a short break nine times during the experiment; the experiment did not proceed to the next stimulus after the break until participants pressed a button to continue. The entire experiment, including the consent form and language background questionnaire, took approximately 35 to 40 minutes to complete.

4 Results

Looking at the results of the practice tokens suggests that participants understood the task, because in almost all cases, the initial consonant at Step 1 was correctly identified as [s] and the initial consonant at Step 11 was correctly identified as [ʃ]. Since the interest of this paper is in consonant harmony, which is about interactions among consonants, the effect of context consonant on sibilant perception is the primary interest. As such, vowel contexts are collapsed for these purposes[3].

Figure 2 below shows proportion [s] response by step for each context consonant. Error bars and confidence intervals are not included and the curves are smoothed using the loess method. This figure demonstrates that there are in fact differences in the curve for the different context consonants; however, it is difficult to determine the nature of these differences by looking at the results in this way.

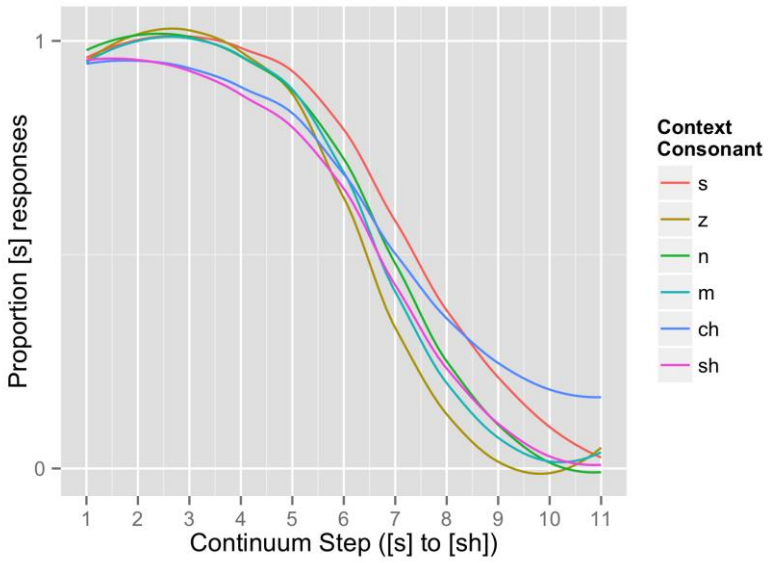

Figure 2 Proportion [s] response by continuum step for each consonant context

Due to low variability in responses for the endpoint steps of the continua, which resulted in models not converging, only responses from steps 4 to 9 were included in the analysis. This choice was made in order to help the convergence of the statistical models, and because these steps were where the crossover points (as described below) occurred based on visual inspection.

[3] A future analysis of these results, however, should take into account vowel differences, since the vowels do appear to be patterning differently.

Step was then centred from the original 1 to 11 steps by subtracting 6 from each step number, to give an analysis range from -2 to 3.

Following the methods of Kleber et al. (2012) using the centred step variable, the 50% [s-ʃ] crossover boundaries were calculated for each subject for each consonant context. The resulting points correspond to the estimate of a mixed effect model for the step along the continuum where, for a particular context, a given subject would respond with [s] 50% of the time and with [ʃ] the other 50% of the time. A higher crossover point indicates a shift towards more of the continuum being heard as [s], while a lower crossover point indicates a shift towards [ʃ], so that hypotheses about a shift towards [s] translate into a higher crossover point, and shift towards [ʃ] translates into lower crossover. In total, 120 (= 20 subjects x 6 context consonants) crossover points were obtained. Figure 3 shows these crossover points (using centred step) graphed by context consonant.

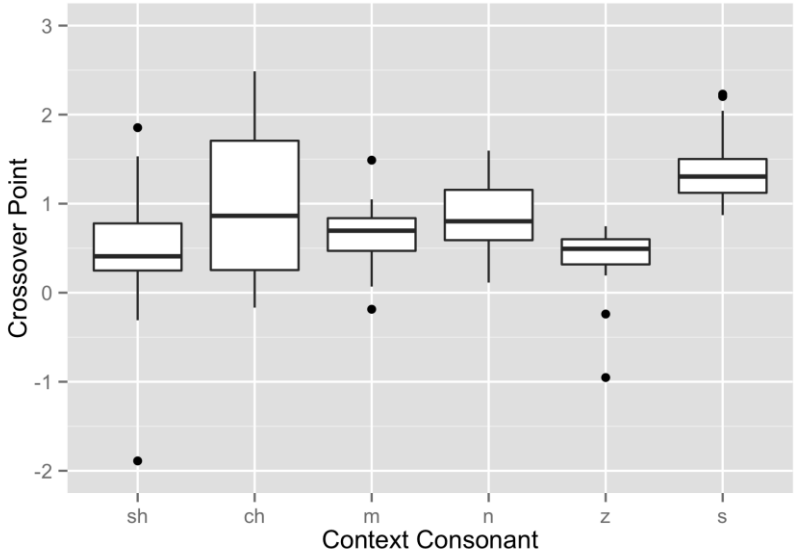

Figure 3 Crossover point by context consonant

In Figure 3, it can be seen that the crossover point medians differ depending on the context consonant. Comparing [s] and [ʃ] contexts to neutral [m] and [n], the distribution of crossover points for [s] is higher than both, while the distribution of crossover points for [ʃ] is at approximately the same level as these neutral contexts. As such, the figure suggests that in the [s] context, the crossover points are generally higher than for the neutral contexts, meaning more of the continuum is heard as [s]; the [ʃ] context is generally more similar to neutral contexts, but in the case where the crossover points are lower, it indicates that more of the continuum is heard as [ʃ]. In terms of the less similar sibilants, the crossover points for the [z] context are relatively low, like for the [ʃ] context, while the [tʃ] context has a large variability in crossover points, and with a median somewhat higher than neutral. As such, the figure suggests that more of the continuum is heard as [ʃ] in [z] contexts and as [s] in [tʃ] contexts, compared to other contexts.

To determine the significance of these differences among the crossover points, a linear mixed effects model with crossover point as the dependent variable, a fixed effect of context consonant, and by-subject random intercepts was fit. The model was fit using the lme4 package in R (Bates

et al. 2014). Furthermore, it was fit twice on each subset, with two different reference levels (of the context consonant predictor variable); in the first fit, all other contexts were compared to [s], while in the second, they were compared to [ʃ]. These two contexts were chosen as reference points because the predictions were based on the behaviour of these two contexts compared to others and of the other contexts compared to these two.

Results of the models are presented in Table 2. As suspected from Figure 3, the crossover points for the [s] context are significantly shifted in the direction of more [s] response (i.e. upwards, higher crossover point) as compared to neutral [m] and [n] (t = -4.14 and -3.10 respectively), as well as compared to [z] (t = -5.53) and [ʃ] (t = -5.17)[4]. However, the [tʃ] context is not significantly different from the [s] context (t = -0.89). In terms of the context [ʃ], again as Figure 3 suggests, there are minimal shifts from neutral, with a borderline significant difference from [n] (t = 2.08) and no significant difference from [m] (t=1.04). Moreover, again as expected, the context [z] is also not significantly different from the context [ʃ] (t = -0.35). Finally, compared to [s] and [tʃ] contexts, the [ʃ] context has significantly more [ʃ] response (t = 5.17 and t = 4.28 respectively).

Table 2 Summary of results (significance shaded)

Reference Context C	Comparison Context C	Estimate	Standard Error	t-value
[s]	**[z]**	-1.01	1.83	-5.53
	[n]	-0.57	1.83	-3.10
	[m]	-0.76	1.83	-4.14
	[tʃ]	-0.16	1.83	-0.89
	[ʃ]	-0.95	1.83	-5.17
[ʃ]	**[s]**	0.94	1.83	5.17
	[z]	-0.06	1.83	-0.35
	[n]	0.38	1.83	2.08
	[m]	0.19	1.83	1.04
	[tʃ]	0.78	1.83	4.28

5 Discussion

5.1 Discussion of hypotheses

The results conformed to some of the predictions made in Section 2, but also showed some effects that were unexpected given the cross-linguistic patterns. The general hypothesis was that the pattern of how crossover points shift (as a measurement of the probability of perceiving an ambiguous sibilant as [s] or [ʃ]) depending on segmental context would reflect typological harmony patterns, giving three distinct predictions: that [s] and [ʃ] contexts would trigger an assimilatory shift in perception of the continuum, the [ʃ] context would trigger more shift than the [s] context, and [tʃ] and [z] contexts would pattern like [ʃ] and [s] contexts respectively. While the first prediction was borne out, the others were not supported by the results.

First, the [s] and [ʃ] contexts did indeed show some assimilatory shift compared to the neutral [n] and [m] contexts. The [s] context crossover was significantly higher from both neutral contexts, indicating a shift towards more of the continuum being heard as [s]. This corresponds to the idea that the [s] context will influence perception of ambiguous continuum sounds, making

[4] Note that significance is defined as $|t| \geq 2$, since p-values are not given with the lmer() function in R.

them sound more like [s], and that listeners, who might be expected to correct for that influence due to knowledge of coarticulatory patterns, in fact do not correct for it. The [ʃ] context showed fewer differences; it was significantly different from neutral only for the context [n]. However, in that case, the crossover was indeed lower than the neutral context, corresponding to more of the continuum being heard as [ʃ]. Again, this result is what the hypothesis predicted.

On the other hand, the hypothesis about trigger asymmetry did not hold. Recall that the prediction was that the [ʃ] context should trigger greater shift from neutral than the [s] context, and that this idea stemmed from the asymmetrical cross-linguistic preference for [-anterior] triggers and the fact that trigger asymmetry in vowel harmony has been argued to be perceptually motivated. The results showed the opposite tendency to what was expected. Indeed, while the crossover for the [s] context was significantly different from both neutral contexts, the [ʃ] context was significantly different only from one neutral context. This result is opposite to the prediction that the [ʃ] context, which triggers harmony more often cross-linguistically than [s] contexts do, would be more different from neutral than the [s] context.

In terms of the predictions about similarity, one aspect of the prediction was supported, while the other was not. The general prediction, based on the typology, was that [z] and [tʃ] contexts should pattern like [s] and [ʃ] contexts respectively, though to a lesser extent. This hypothesis can be separated out into two sub-hypotheses. First, the crossover points for [z] and [tʃ] contexts will not be shifted from neutral to the same extent as those of [s] and [ʃ] contexts. Second, these contexts should show the same kinds of effects as [s] and [ʃ] contexts respectively; in other words, any effect will be in the same direction. The first hypothesis is supported: while the [s] context shows a shift towards [s] compared to neutral [m] and [n] contexts, there is no corresponding effect for the [z] context. Similarly, while the [ʃ] context shows some shift towards [ʃ] compared to neutral [n], the [tʃ] context does not show this effect. On the other hand, the second prediction is contradicted by the data; [z] and [tʃ] seem to show a dissimilatory perceptual effect, while [s] and [ʃ] show an assimilatory one. In particular, the opposite pattern to the prediction held, with the [z] context patterning more like the [ʃ] context and the [tʃ] context more like the [s] context. Indeed, the crossover point for the [z] context was not significantly different from that of the [ʃ] context, but was from the [s] context, and similarly, the crossover point for the [tʃ] context was not significantly different from that of the [s] context, but was from the [ʃ] context. In other words, [z] and [tʃ] seem to show dissimilatory effects of perception, compared to the assimilatory effects of [s] and [ʃ]. Thus, overall, [tʃ] and [z] seem to pattern similarly to [s] and [ʃ] respectively, which is opposite to the expectation of the direction of their patterning.

In summary, while the general assimilatory hypothesis held for [s] and [ʃ] contexts compared to neutral ones, some of the other results were unexpected based on the harmony typology.

5.2 Possible explanations

The original harmony as hypocorrection hypothesis suggested that harmony patterns are caused when coarticulatory effects of sibilants on each other across an intervening vowel are not corrected for. However, given the results of this experiment, this idea cannot hold for all of the typologically motivated predictions. As such, it is important to consider possible explanations that might account for the effects seen here. It may be that the results are a reflection of sibilant perception in harmony contexts, though it is also important to consider characteristics of the stimuli, influence of English, and ways of analyzing the data.

First, the effects may be a genuine reflection of how sibilant harmony contexts affect perception and/or articulation of other sibilants[5]. If that is the case, then it is necessary to explain why the results are different from the patterns that were expected based on the consonant harmony typology. Two possibilities will be discussed for the effects seen in [tʃ] and [z] contexts.

Recall that the [z] and [tʃ] contexts showed the opposite pattern compared to the [s] and [ʃ] contexts: the former reflected a dissimilatory effect, while the latter showed an assimilatory effect. Reconciling these opposite effects under the assumption that both result from influences of non-adjacent sibilants on each other could go in two possible directions: either there is a general assimilatory effect and [z] and [tʃ] are 'exceptional', or else there is a general dissimilatory effect and [s] and [ʃ] behave differently. Both of these explanations could be thought of as involving a similarity threshold, or a boundary determined by a particular degree of (featural) similarity to the continuum consonant [s]/[ʃ]. With such a threshold, segments on different sides of the boundary behave differently with respect to their perceptual influence on the continuum consonant. In this case, the boundary would occur between [tʃ]/[z], which differ in manner/voicing features from the continuum consonant, and [ʃ]/[s], which do not.

In the case that there is a general assimilatory effect, but [z] and [tʃ] are exceptional, it is possible that their patterning is due to an overcompensation for sibilant coarticulatory effects. Since these segments are sibilants, they are likely to trigger behaviour in same direction as other context sibilants [ʃ] and [s] respectively. However, since they are less similar than [s] and [ʃ] to the ambiguous sibilant being categorized, it is conceivable that listeners do compensate for the effects of sibilants on each other with these segments, and in fact overcompensate, resulting in them behaving like sibilants of the opposite value of [anterior]. In other words, listeners process perceptual and acoustic coarticulatory effects because of the fact that these context consonants are sibilants, but they know that these sibilants differ from the consonant being categorized in more ways than just the harmonic feature. The response is then adjusted to account for the original hearing of harmony (i.e. un-harmonized), but goes too far in the opposite direction, resulting in dissimilation. As such, listeners overcompensate for the harmony and are more likely to judge the continuum consonant as dissimilar from the context consonant. If this is the case, we might expect to find cases of consonant harmony cross-linguistically where, given a similarity threshold, segments above that threshold harmonize, but those below the threshold undergo dissimilation. The only such case in phonological consonant harmony patterns is total-identity exemptions (discussed below), but looking further at possible thresholds in consonant harmony, particularly statistical patterns, might be useful to further examine this possibility.

Under the other possibility, in which the general effect is dissimilatory but [s] and [ʃ] are exceptional, their patterning could be due to a total-identity exemption. Such exemptions, in which there is a general dissimilatory preference, but fully identical segments are permitted, are a known phenomenon in consonant harmony systems. For example, in ejective harmony, there are cases of languages that permit two ejectives within a word if they are at the same place of articulation, but do not allow two ejectives of different places to co-occur within a word (see e.g. Gallagher 2014). If the underlying perceptual pattern in the results here is dissimilatory effects of sibilants on each other, but with a total-identity exemption, then we should expect listeners to favour [s…tʃ] and [ʃ…z] over [ʃ…tʃ] and [s…z] respectively (dissimilation), but these dissimilatory effects should disappear with the contexts [s] and [ʃ], so that listeners should prefer [s…s] and [ʃ…ʃ] (total-identity exemption). This result is exactly what was found in this experiment. Total-identity exemptions are perhaps best known from Gallagher's work on ejective

[5] Articulation is a possible explanation because the stimuli were created from entire CVCV tokens synthesized and combined, not from the same continuum spliced into all contexts.

co-occurrences; however, if this interpretation of these results holds, it suggests that looking for total-identity exemptions in sibilant co-occurrence patterns could be a fruitful avenue for future consonant harmony research.

Thus, overall, there are several possible explanations of these results for [tʃ] and [z], even though they were opposite to the hypothesis originally made for the experiment. At present, it is difficult to distinguish between these hypotheses; more experiments and analysis will be needed to do so. In order to look more carefully at the harmony explanations, studies should be done of statistical sibilant co-occurrences, in order to determine whether some of the patterns predicted by a harmony-based explanation of these results exist at any level in any languages.

While these possible explanations provide future directions to explore in understanding the relationship between perception and sibilant harmony patterns, it is also important to consider whether certain particularities of the present experiment might have caused the results. The remainder of this section considers the stimuli, English speakers, and the analysis methods.

First, it is possible that some property of these particular stimuli aside from the intended manipulations caused the unexpected results. However, listening to the stimuli both before and after the experiment was run revealed no immediately apparent qualities that might explain the results. Future work will examine the spectral properties of the stimuli, including the COG values of the manipulated fricatives and the formant values of the vowel transitions from the fricatives, in order to more closely examine possible stimuli-related explanations of the results. One possibility is that, due to coarticulation in production, the COG values of the initial consonant of the entire continuum for the [ʃ] context are lower than the values for neutral contexts at the corresponding continuum steps, which are in turn lower than the values for the [s] context. In that case, it may be that the COG values at the crossover point for all contexts are in fact equal.

A second possibility is that English lexical restrictions are influencing the judgements of the listeners in this experiment. The possibility of (statistical) sibilant co-occurrence restrictions in the English lexicon was tested using the Phonological CorpusTools software (Hall et al. 2015). Using the IPhOD corpus of English (Vaden et al. 2009) counts were obtained of words in which the sibilants [s] and [ʃ] appear before other sibilants [s], [ʃ], [z], and [tʃ] on a consonant tier[6]. The resulting counts were compared to overall numbers of [s] and [ʃ] in the corpus using a binomial test, in order to determine the significance of the differences in counts in each context. There was overall a greater frequency of [s] than [ʃ] in the corpus, but in some contexts, there was a significant change in the degree of bias. Results of this test showed that [s] and [ʃ] are not significantly over- or under-represented before [s] and [ʃ] compared to the general bias in the corpus, but that [ʃ] is significantly under-represented, and [s] significantly over-represented, before both [z] and [tʃ], compared to the overall counts. This result could help to explain the effect seen for [tʃ]; since [ʃ] before (a vowel then) [tʃ] is quite rare in English compared to [s] before [tʃ], it is possible that listeners are accessing these lexical restrictions in making their decisions about ambiguous [s]/[ʃ] in the context [tʃ], and are therefore more likely to select [s] in this context. However, this idea still leaves the behaviour of the [z] context unexplained, since it predicts that [z], too, should have greater [s] response, which is opposite to the result.

[6] A consonant tier was chosen in order to make the distance between the sibilants in the forms being tested approximately the same as the distance used in the experiment. In the experiment, the sibilants were just across a vowel; on a consonant tier, the sibilants are either just across a vowel or strictly adjacent, which is less common word-internally in English. In particular, a sibilant tier was not used because then the sibilants could be separated by multiple syllables. For instance, the word ‘horseradish’ has adjacent sibilants on a sibilant tier but not on a consonant tier.

Finally, it could be that crossover point is simply not the appropriate way to get at the differences among the curves. As Fleischer et al. (2013) found, sibilant continua in local assimilatory contexts can change through flattening rather than through shifts. Such flattening cannot be observed by strictly calculating crossover point, because it requires knowledge of the entire shape of the curve and in particular of the behaviour of the continua endpoints. While Figure 2 does not show any clear flattening behaviour, except perhaps with the context [tʃ], future work on the results from this experiment should check for such effects.

Overall, in order to determine whether these results are truly due to facts about how sibilant contrasts are perceived in harmony contexts, it is necessary to examine the stimuli, native language of the participants, and methods of analysis more closely. In order to rule out these explanations for the results, the experiment should be repeated with different stimuli, such as from different speakers or different repetitions from the same speaker, and with speakers of languages other than English, and results should be analyzed using methods other than crossover point. Furthermore, in order to disambiguate among all of the possibilities presented here, the acoustic properties of the stimuli should be examined, and a follow-up experiment should potentially be run in which continua are created using CV syllables and then spliced into the same contexts as in the present experiment, to determine how coarticulation in the original stimuli affects or confounds the results. All of these future studies will help to deepen our understanding not only of these results, but of the motivations of sibilant harmony patterns more broadly.

6 Conclusions

In conclusion, this study used a [s]/[ʃ] categorization task to examine possible perceptual correlates of the typology of sibilant harmony. The general pattern of assimilatory perception in contexts with [s] and [ʃ] did hold as expected in the results. However, there were also some patterns that do not reflect the typology, with more changes in crossover point triggered by [s] than by [ʃ], and with [tʃ] and [z] contexts patterning like [s] and [ʃ] respectively rather than the opposite. These findings suggest a need to further explore the perception, as well as the acoustics, of sibilant contrasts in the contexts relevant to sibilant harmony, in order to understand how such clear typological patterns might occur despite these perceptual results.

Acknowledgements

Thank you to my committee members Molly Babel, Kathleen Hall, and Gunnar Hansson, to the LING 530A class, to Gunnar's lab group, to Michael McAuliffe for his statistics help, and to Michael Schwan for recording stimuli.

References

Abrego-Collier, C. (2013). Liquid dissimilation as listener hypocorrection. *Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society*, 3–17.

Applegate, R. B. (1972). *Ineseño Chumash Grammar*. Doctoral dissertation, University of California, Berkeley.

Arsenault, P. (2012). *Retroflex Consonant Harmony in South Asia*. PhD dissertation, University of Toronto.

Arsenault, P. and Kochetov, A. (2011). Retroflex harmony in Kalasha: agreement or spreading? In S. Lima, K. Mullin, and B. Smith (Ed.), *Proceedings of the North East Linguistic Society* 39 (pp. 55–66), Amherst: GLSA.

Bates D., Maechler M., Bolker B., and Walker S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7, http://CRAN.R-project.org/package=lme4.

Benus, S. (2005). *Dynamics and transparency in vowel harmony*. Doctoral Dissertation, New York University.

Boersma, P. and Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.08, http://www.praat.org/

Fleischer, D., Wagner, M., and Clayards, M. (2013). A following sibilant increases the ambiguity of a sibilant continuum. *Proceedings of Meetings on Acoustics*. Vol. 19. No. 1. Acoustical Society of America, 2013.

Gallagher, G. (2012). Perceptual similarity in non-local laryngeal restrictions. *Lingua* 122(2): 112–124.

Gallagher, G. (2014). Evidence for an identity bias in phonotactics. *Laboratory Phonology* 5(3): 337–378.

Hall, K. C., Allen, B., Fry, M., Mackie, S., and McAuliffe M. (2015). Phonological CorpusTools, Version 1.0.1. [Computer program]. Available from https://sourceforge.net/projects/phonologicalcorpustools/.

Hansson, G. Ó. (2008). Diachronic explanations of sound patterns. *Language and Linguistics Compass* 2(5): 859–893.

Hansson, G. Ó. (2010). *Consonant harmony: long-distance interaction in phonology*. Berkeley, CA: University of California Press.

Kaun, A. (1995). *The Typology of Rounding Harmony: An Optimality Theoretic Approach.* Doctoral Dissertation, UCLA. Published as UCLA Dissertations in Linguistics, No. 8.

Kleber, F., Harrington, J., and Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*, 55(3), 383–405.

Kosa, L. A. (2010). *Sibilant harmony: investigating the connection between typology and learnability*. Proceedings of the 2010 annual conference of the Canadian Linguistic Association.

Kawahara, H., Morise, M., Takhashi, T., Nisimura, R., Irino, T., and Banno, H. (2008). TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings*, 3933–3936.

Mann, V. A., and Repp, B. H. (1980). Influence of vocalic context on perception of the [sh]-[s] distinction. *Perception and Psychophysics* 28: 213–228.

McMullin, K. and Hansson, G. Ó. (2013). Locality in long-distance phonotactics: evidence for modular learning. Talk presented at the 44th Meeting of the North East Linguistic Society (NELS 44), University of Connecticut, October 18-20, 2013.

Niebuhr, O., Clayards, M., Meunier, C., and Lancia, L. (2011). On Place Assimilation in Sibilant Sequences–Comparing French and English, *Journal of Phonetics* 39, 429–451.

Ohala, J. J. (1994a). Towards a universal, phonetically-based, theory of vowel harmony. *ICSLP* 3, Yokohama, pp. 491–494.

Ohala, J. J. (1994b). Hierarchies of environments for sound variation; plus implications for 'neutral' vowels in vowel harmony. *Acta Linguistica Hafniensia* 27: 371–382.

Przezdziecki, M. (2005). *Vowel harmony and coarticulation in three dialects of Yoruba: Phonetics determining phonology*. Doctoral dissertation, Cornell University.

Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from http://www.pstnet.com.

Rose, S. and Walker, R. (2004). A typology of consonant agreement as correspondence. *Language*, 475–531.

Schuh, R. G. (1997). Changes in obstruent voicing in Bade/Ngizim. Ms. University of California, Los Angeles.

Vaden, K. I., Halpin, H. R., and Hickok, G. S. (2009). Irvine Phonotactic Online Dictionary, Version 2.0. [Data file]. Available from http://www.iphod.com/.