ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 5aSCb: Production and Perception II: The Speech Segment (Poster Session)

5aSCb31. A following sibilant increases the ambiguity of a sibilant continuum
David Fleischer, Michael Wagner and Meghan Clayards*

*Corresponding author’s address: Linguistics & School of Communication Sciences and Disorders, McGill University, Montreal, H3A 1A7, QC, Canada, meghan.clayards@mcgill.ca

We examined the effect of three following contexts: /s/, /ʃ/, and vowel, on the categorization of a /s/-ʃ/ continuum. Unlike previous findings of a shift in category boundary due to context (Mann & Repp, 1980), we found that in the context of a following sibilant, listeners found the target sibilant to be more ambiguous (shallower categorization slopes and responses closer to chance) than when followed by a vowel (p < 0.001). There was also a tendency for the ʃ/ context (which affects pronunciation) to create more ambiguity than the /s/ context (which does not) (p = 0.057). On half of the trials, participants heard the following context as part of the same syntactic phrase as the target (e.g. “Whenever they fra? Shelly gets upset”) and on half heard it was part of a different phrase (e.g. “Whenever they fra? Shelly, John gets upset”). Pronunciation usually is more affected when target and context are in the same phrase (Holst & Nolan, 1995). Listeners tended to perceive target sibilants as more ambiguous when the following sibilant was part of the same phrase (p = 0.01) suggesting a role for top-down knowledge in interpreting segmental information.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Variability in the speech signal remains a primary challenge for spoken word recognition. The task of recognition remains largely unproblematic for listeners despite the signal’s dependence on phonetic, phonological, syntactic, semantic, and pragmatic contexts. One source of variability is the phonetic context following a segment. For example, acoustic analyses demonstrate a dependency of fricative noise spectra on the following vocalic context. Studies with Japanese (Fujisaki and Kunisaki (1978)) and English (Mann and Repp (1980)) have shown that fricatives are produced with lowered noise spectra when followed by phonetic contexts that generate anticipatory lip rounding, such as [u]. The following consonant context has also been found to affect fricative spectra. This paper is concerned with the assimilation pattern observed with sibilants in English (Nolan, 1992). Niebuhr et al. (2011) report that in English, there is assimilation to a following but not to a preceding sibilant, and furthermore, they present evidence for an asymmetry already noted in Nolan (1992). The alveolar fricative [s] assimilates to a following [], but alveopalatal [] does not assimilate to a following [s]. That is, place assimilation of sibilant sequences is asymmetric allowing for sequences such as “glass shoes” from “glash shoes” but not “fiss soup” from “fish soup”. Finally, Niebuhr et al. (2011) found that there was almost always always complete assimilation when [s] was following by [].

The assimilation patterns observed in production raise the question of how listeners deal with this variation when parsing the incoming speech stream. Two very different effects have been found in the previous literature. Based on the coarticulation patterns observed in speech production, Mann and Repp (1980) investigated whether the perception of fricative noise spectra is influenced by the following vocalic context. They found that listeners compensated for the coarticulatory lip rounding that would have resulted from normal production—fricative noises followed by [u] are more likely to be heard to as [] rather than [s] and fricative noises followed by [a] are more likely to be heard as [s] rather than []. The response pattern found in Mann and Repp (1980) shifts the categorization function toward [s] in assimilation contexts. The intuition behind the proposal in Mann and Repp (1980) is that listeners make use of their knowledge that coarticulation makes an [s] more []-like in an assimilation context, and hence the category boundary shifts. In an environment with a high likelihood of complete assimilation one might expect a similar compensatory effect: A [] might be more likely to be interpreted as an [s], if it’s []-ness is attributed to the phonological context rather than its own phonemic status. The pattern predicted under this line of reasoning is illustrated in (1a).

\[ a. \text{“Shift”} \quad b. \text{“Flattening”} \]

\[ \text{FIGURE 1: Two possible patterns for segment categorization in assimilatory environments: An overall shift in the response curve was reported in Mann and Repp (1980), a flattening was reported in Mitterer et al. (2006).} \]

More recently, Mitterer et al. (2006) compared the influence of viable and unviable assimilatory contexts on preceding segments and found a very different pattern. Listeners of
Hungarian speech samples were asked whether morphemes ended with /l/ or /r/ in contexts when [r] followed (context viable for assimilation) or followed by an [n] (context unviable for assimilation). They found that unviable contexts generated a steep identification function for the [l]-[r] continuum while viable contexts produced a flattened function, similar to the illustration in (1b). The flattened response curve in the viable contexts were interpreted to be caused by perceptual integration of the target and its following context, an effect which makes liquid identification more difficult. Mitterrer et al. argued that in environments in which an acoustically similar segment follows, contrasts may be harder to perceive for reasons of auditory masking, making them less likely to be maintained, and as a consequence, the emergence of an assimilation pattern is more likely. The connection between the perceptual integration effect and the likelihood of assimilation is akin to the ‘licensing-by-cue’ approach to the distribution of contrast pursued in Steriade (1997). This theory predicts that environments with perceptual integration are likely places for assimilation patterns to emerge, but they do not necessarily have to be assimilation environments.

There is a second possible interpretation of the observed flattening, however, which reverses the explanation. The flattening could be a consequence of a listener’s knowledge that there is assimilation in a certain environment, which results in greater uncertainty about the percept and moves the responses closer to chance. Under this interpretation the flattening effect should only occur in assimilation environments.

Some evidence for a lower level auditory account comes from the fact that Mitterer et al. (2006) found similar degrees of flattening for both word and non-words contexts, indicating that lexical knowledge is not a major mediator for context effects. Furthermore, they found that the effect was comparable for native speakers of Hungarian and for L2 speakers with a native language that do not show assimilation in the relevant environment (Dutch). A more direct test of the different explanations for the flattening effect, however, would be to test the effects of an environment that is acoustically similar and likely to result in perceptual integration but that is not an assimilation context. Clayards, Gaskell and Niebuhr Clayards et al. (2011) tested the effect of following context on sibilant perception and included both [s] which is acoustically similar but does not trigger assimilation, [ʃ] and a control ([p]) context. Their pattern of results was consistent with flattening mostly in the [ʃ] context, therefore not ruling out the possibility that the flattening is due to knowledge of the assimilation pattern.

This paper follows Clayards, Gaskell and Niebuhr (2011) in investigating the effect of following phonological/phonetic context on the perception of sibilants. We examined three phonological contexts: One environment with an acoustically dissimilar segment following, namely a vowel; one environment in which a following segment that is acoustically similar, but does not trigger assimilation, namely [s]; and finally an environment with a following acoustically similar segment that does trigger assimilation, [ʃ].

Another way to examine the nature of the context effect is to look for influences that cannot be auditory, such as influences of syntactic structure. While it is common to find assimilation in cases such as “glash shoes,” sibilant-to-sibilant assimilation appears to be modulated by the strength of the intervening clausal boundary; alveolar–postalveolar transitions at word boundaries are less likely to be assimilated across stronger syntactic junctures at the word boundary than weaker ones (Holst and Nolan (1995); Nolan et al. (1996)). The second goal of this paper is to investigate how the effect of phonological context on perception is modulated by syntactic structure. Scott and Cutler (1984) showed that listeners are able to use segmental information resulting from cross-word phonological processes as a cue for syntax. When presented with syntactically ambiguous sentences, listeners were able to correctly identify the intended meaning of the sentences based on the presence of a [t] (versus [r]) at the disambiguating boundary (e.g. “If you want to eat[t]/[r] early lunch will be served”). We ask the question in the opposite direction: Are listeners able to use syntactic information in interpreting...
This question is interesting from at least two angles. First, a strictly bottom-up model of processing would predict it to be impossible that high-level syntactic information affects phonological processing (let alone auditory processing). A model that the perceptual system will bring all available information to bear on interpreting the input, and not necessarily do so in a bottom-up way, predicts that syntax should be able to affect segmental perception—as long as the syntactic information is available in time for it to exert its effect. Secondly, finding evidence that syntactic structure mediates the phonetic context effect (be it a shift or a flattening of the categorization function) would support a role for top down/knowledge based mechanisms in the effect of phonetic context.

To summarize, the goal of the present study was 1) to determine the role of a following sibilant context on preceding sibilant classification—more specifically, whether we see compensation effects or integration effects, and 2) to uncover any interaction effects between phonetic/phonological context and syntactic contexts.

### Methods

#### Participants

24 native English speakers from the McGill University community took part in the experiment.

#### Stimuli

All stimuli were recorded with Praat (Boersma and Weenik (2011)) using a Logitech H390 USB headset by a female native speaker of American English in a soundproof recording booth. A seven step sibilant continuum was constructed in Praat by digitally mixing [s] and [ʃ] tokens in 1/7th increments. The sibilant tokens were taken from carrier utterances of [ʃərən] and [ʃərə].

Six items were chosen for the experiment with each item representing an optionally transitive monosyllabic novel verb. Thirty-six sentences were recorded for the purposes of this experiment and were of the form "Whenever they frat, (Sharon/Sara/Aaron) gets upset" (intransitive) or "Whenever they frat (Sharon/Sara/Aaron), John gets upset" (transitive). Each sentence was originally recorded with one of the six novel verbs ending with [t] in order to avoid formant transitions providing biased sibilant information. The seven step continuum was spliced into the position of the original [t] to create a total of 252 experimental tokens.

In order to investigate the role of following phonetic context on sibilant classification three phonetic conditions were recorded with each item. Each condition represented a [ʃ], [s], or vowel context following the sibilant continuum. Specifically, the proper noun following the verb alternates between an [ʃ]-initial form (Sharon), [s]-initial form (Sara), or vowel-initial form (Aaron).

In order to investigate the role of syntactic structure on sibilant classification two syntactic conditions, a transitive version and an intransitive version, were recorded with each item. The sibilant continuum was spliced into the endpoint of monosyllabic novel verbs in order to avoid participants favouring either sibilant form due to lexical bias. The overall prosody of each sentence as well as the gap between the nonce verb and proper noun were recorded to be as identical as possible in both intransitive and transitive forms.

In sum, the experiment contained 6 items (optionally transitive novel verbs), 6 experimental conditions (3 phonetic conditions × 2 syntactic conditions), and a 7 step sibilant continuum for a total of 252 tokens.
Procedure

Prior to the experiment, participants were given written instructions on the computer screen. All participants were instructed that they were going to hear English sentences containing nonce words that may end in either /s/ or /ʃ/ (e.g. What did you hear? 1=Frass 2=Frash). Following the auditory presentation of each trial they were requested to select the most felicitous option for the verb in the sentence they had heard. Following each trial participants were asked for the judgments of how natural the previous utterance had sounded on a scale from 1 to 5. Data from the naturalness judgements are not discussed here. The experiment followed a pseudorandom design in which all participants saw all items of each condition.

Results

The dependent measure was the rate of /s/ responses across all participants. Figure 2 shows the effect of following phonetic context on /s/ responses across the sibilant continuum. The data pattern clearly looks like the pattern expected with perceptual integration: Participants seem to have been less likely to report having heard /s/ at the most /s/-like continuum step, as well as being less likely to consistently report having heard /ʃ/ at the most /ʃ/-like continuum step in the context of a following sibilant.

![Figure 2: Mean /s/ responses for each of the three following phonetic contexts in Experiment 1. Error bars are constructed using one standard error of the mean.](image)

We analyzed the data using a logistic mixed model regression with the choice between [s] and [ʃ]. We were unable to get convergence with a full model with all three dependent variables: phonological context, the sibilant continuum step and transitivity and their interactions as fixed factors, and the full random effects structure for participant and item. Instead two models were run: one with phonological context and sibilant continuum step and the interaction and the other with transitivity and continuum step and the interaction. These models included the full random effects structure. We were thus unable to test for three-way interactions.

First, and unsurprisingly, there was a main effect for the continuum step (|z|=0.75, p<0.001). Second, there was a significant difference in overall [s] responses in the vowel versus the sibilant (/s/ and /ʃ/) context (|z|=0.43, p= 0.02). Most importantly, response rates for the two sibilant contexts were closer to chance across the continuum than those for vowel contexts (|z|=0.93, p< 0.001), which supports that flattening took place. Note that the flattening in the [s] context is unexpected if flattening only occurs in assimilation environments: Niebuhr et al. (2011) showed that [s] is in fact not a trigger for assimilation, and yet the degree of flattening in [s] contexts is more similar to [ʃ] contexts than vowel contexts. This supports the auditory interpretation in terms of perceptual integration.

However, there was also a small and marginally significant difference between [s] and [ʃ] contexts, such that the response curve was slightly shallower in [ʃ] contexts (|z|=0.02, p=0.058). This could be an effect of being in an assimilatory environment, which adds further uncertainty...
about the identity of the perceived segment—unless of course there was a lower level auditory reason why [ʃ] should lead to a greater masking than [s]. There was also a main effect of less [s] responses in sibilant contexts which is not expected under either theory (|z|=0.43, p=0.02).

The second goal of the experiment was to identify the influence of syntactic structure on sibilant identification (Fig. 3). There was indeed a significant interaction such that flattening was greater in the transitive than in the intransitive condition (z>0.12, p< 0.01).

![Figure 3: Mean /s/ responses between syntactic conditions for all phonetic contexts in Experiment 1. Error bars are constructed using one standard error from the mean.](image)

**CONCLUSION**

Our results for the role of phonetic context on sibilant classification support earlier findings that assimilatory environments show a perceptual integration phenomenon which results in a flattening of the categorization function (Mitterer et al., 2006; Clayards et al., 2011). The results are compatible with the view that environments in which a contrast is hard to hear for auditory reasons are prone to lose the contrast, but also suggests that they are not necessarily lost in such environments, as evidence by the flattening in the [s] context which favours the auditory interpretation of the flattening.

However, there was some evidence that there was more flattening in [ʃ] contexts. This effect could be an top-down effect: The knowledge that /s/ is likely to assimilation in this environment contributes to the uncertainty about the percept and hence leads to performance that is even closer to chance. However, it could also be that [ʃ] simply has a greater auditory masking effect than [s]. There was no sign of a shift of the categorization function similar to those observed in Mann and Repp (1980).

There was also a significant effect of syntax. This suggests first that syntactic top-down information does affect segmental perception, contrary to what would be expected in a strictly bottom-up model of speech perception. However, the effect appears to be observable in both [s] and [ʃ] contexts, which would mean that listeners used their knowledge that environments with weaker prosodic boundaries (the transitive case) usually lead to greater masking. This is unexpected under the original perceptual integration proposal in Mitterer et al. (2006), which argues that the flattening effect is due to a low level auditory mechanism. Under normal circumstances, there indeed should be a greater perceptual integration effect in transitive conditions, since the sibilants will be shorter and separated by a smaller prosodic boundary—so maybe the perceptual system can factor in knowledge about how likely it is that a stimulus will have been affected by integration effects, and even syntactic information is available in making this estimate.

In our stimuli, the syntactic relation to the following word could not yet be known to the
listener at the point in which the sibilant occurred, which probably weakened the effect that the syntactic manipulation had in this study. The time course may have made it difficult for syntactic information to affect the percept of the sibilant. Further experiments are under way that manipulate whether or not listeners were able to anticipate the syntax of the sentence they were listening to and generate expectations about whether or not assimilation should occur.

**References**


