Article:

Clayards, Meghan, Niebuhr, Oliver and Gaskell, M. Gareth orcid.org/0000-0001-8325-1427 (2015) The time course of auditory and language-specific mechanisms in compensation for sibilant assimilation. Attention, Perception & Psychophysics. pp. 311-328. ISSN 1943-393X

https://doi.org/10.3758/s13414-014-0750-z

The time-course of auditory and language-specific mechanisms in compensation for sibilant assimilation

Meghan Clayards
Department of Linguistics &
School of Communication Sciences and Disorders
McGill University, Canada
1085 Ave Dr. Penfield
Montreal, QC H3A 1A7
(1) 514-398-4235
meghan.clayards@mcgill.ca

Oliver Niebuhr
Department of General and Comparative Linguistics
University of Kiel, Germany

M. Gareth Gaskell
Department of Psychology
University of York, UK

Running Head: Compensation for sibilant assimilation

Authors Note: The authors would like to thank Gerry Altmann, Dirk Kertzel and Uli Fraunfelder for the use of their eye-trackers and lab space and Holger Mitterer for helpful suggestions. Portions of this work were presented at the meeting of the Psychonomics Society in Boston, 2009 and the 10th International Symposium of Psycholinguistics, Donastia San-Sebastian, 2011. This research was supported by the Marie Curie training network Sound to Sense.

The final publication is available at Springer via http://dx.doi.org/10.3758/s13414-014-0750-z

Abstract:

Models of spoken-word recognition differ on whether compensation for assimilation is language-specific or depends on general auditory processing. English and French participants were taught words that began or ended with the sibilants /s/ and /ʃ/. Both languages exhibit some assimilation in sibilant sequences (e.g., /s/ becomes like [ʃ] in *dress shop* and *classe chargée*), but they differ in strength and predominance of anticipatory versus carry-over assimilation. After training, participants were presented with the novel words embedded in sentences, some of which contained an assimilatory context either preceding or following. A continuum of target sounds ranging from [s] to [ʃ] was spliced into the novel words representing a range of possible assimilation strengths. Listeners' perceptions were examined using a visual-world eye-tracking paradigm in which the listener clicked on pictures matching the novel words. We found two distinct language-general context effects: a contrastive effect when the assimilating context preceded the target and flattening of the sibilant categorization function (increased ambiguity) when the assimilating context followed. Furthermore, we found that English but not French listeners were able to resolve the ambiguity created by the following assimilatory context, consistent with their greater experience with assimilation in this context. The combination of these mechanisms allow listeners to deal flexibly with variability in speech forms.

I. INTRODUCTION

Variability in the pronunciation of words is one of the great challenges of word recognition. One type of variability involves changes across word boundaries. The articulation and subsequent acoustic realization of sounds are influenced by phonetic context and languages differ in the extent to which they allow this variation (Steriade, 2001). In each language, some variation can be quite extreme and near-obligatory under certain circumstances. For example, in English, the sibilant /s/ takes on the place of articulation of a following /ʃ/ as in *dress shop* pronounced as *dresh shop* but not a preceding /ʃ/ as in *crash site* (Niebuhr, Clayards, Meunier & Lancia, 2011). Traditionally, more complete and language specific variation (termed assimilation) has been considered part of phonology and thus part of a speaker's language-specific behaviour. Less extreme variation (termed co-articulation) has been viewed as more mechanical in nature and not part of a speaker's language-specific behaviour. A number of papers have called into question this distinction (e.g. Nolan, 1992; Whalen, 1990), so for the present purposes we will refer to all variation along a continuum that increases similarity with phonetic context as assimilation, and distinguish between complete and partial assimilation.

One might expect pronunciation variations like assimilation to be disruptive to word recognition. On the contrary, listeners are quite good at accommodating pronunciation variability (though see Gaskell & Snoeren, 2008 for evidence of persistent ambiguity). For example, in a phoneme monitoring task, Gaskell and Marslen-Wilson (1998) asked English listeners to monitor for coronal sounds such as /t/ in sentences that included words like *freight* or their assimilated forms (e.g. the non-word *frayp*) and found that listeners reported hearing a /t/ in sentences like *frayp bearer* 59% of the time. This is not a simple lexical bias, as reports of /t/ decrease when a viable context (i.e., the /b/ in *bearer*) is missing. On the other hand, more coronal forms were reported when potential assimilations involved words as opposed to pseudowords (e.g. *prayp bearer*). More recently it has been shown that listeners are faster to detect assimilated words in viable than unviable contexts when those words have been recently learned but not when they are unknown (Snoeren, Gaskell, & Di Betta, 2009). In both these cases, context-dependent processing plays a role in allowing the listener to correctly recognize a word (e.g., *freight*) when hearing an assimilated version (e.g., *frayp*).

A. Language-dependent and independent mechanisms

As with all sources of variability in the speech signal, the question of how listeners deal with assimilation arises, and there has been considerable debate over the level of processing responsible. A number of accounts have been given, each proposing slightly different mechanisms or combinations of mechanism. One class of proposed mechanisms contends that compensation occurs at early stages of processing relying on general cognitive processes such as auditory contrast enhancement (e.g., Diehl, Lotto, & Holt, 2004), parsing (grouping) different acoustic cues according to phonological features or gestures (e.g., Gow, 2003; Fowler, 1986), or the auditory integration of the target and context sounds (e.g., Mitterer, Csépe & Blomert, 2006). Feature/gestural parsing accounts capitalize on the observation that many cases of assimilation are not true replacements of one segment (or segmental feature) with another but rather constitute overlapping and/or merging of gestures (e.g. Browman & Goldstein, 1992). These overlapping gestures to varying degrees leave acoustic traces of the original segments/features (Gow, 2002; Niebuhr & Meunier, 2011). The listener then uses grouping mechanisms to separate out the acoustic elements of the first segment from those of the second, recovering the partially obscured segment (Gow, 2003; Gow & Im, 2004). In a similar vein, auditory contrast enhancement posits that auditory processes operate to enhance differences between similar adjacent segments so that target sounds are perceived to be less similar to the context sound (and therefore less assimilated) than they are (Diehl, Lotto, & Holt, 2004). In contrast, Mitterer and colleagues (e.g. Mitterer, Csépe, & Blomert, 2006) have proposed that some contexts may influence perception by partially obscuring target sounds (making assimilated and unassimilated forms harder to distinguish from one another), again through auditory processes they term perceptual integration, but might perhaps be better thought of as masking. Because the target sounds are difficult to perceive accurately, this reduces the salience of the mismatch between an assimilated segment and its unassimilated form, allowing other contextual information to disambiguate. All three accounts predict that context will modulate the perception of assimilated segments and that acoustic details are more predictive of perception than language background. In other words, accommodation of assimilation does not depend on listeners' linguistic knowledge of patterns of variability in their specific language.

Support for language-independent mechanisms comes from studies that have found no effect of language background in compensation. For example, Mitterer and colleagues provided examples of Hungarian liquid assimilation (where /l/ is assimilated to [r] before /r/) to both Dutch and Hungarian listeners in identification and discrimination tasks (Mitterer, Csépe, & Blomert, 2006). They found shallower identification slopes and poorer discrimination performance in the assimilating context, both for Hungarian listeners – for whom the assimilation forms part of their language experience – and for Dutch listeners for whom it does not.. The same stimuli were also used to measure mismatch negativity (MMN) in responses to an oddball stimulus in a passive listening paradigm (Mitterer, Csepe, Honbolygo, & Blomert, 2006). Context-dependent MMN responses were similar for both groups of listeners. Furthermore, Gow and Im tested Hungarian and English listeners on a voicing assimilation pattern found in Hungarian (Gow & Im, 2004).

Both groups of listeners identified a target segment most quickly and accurately when an assimilated segment preceded it. These studies also found that the exact acoustic details of the stimuli are extremely important to the compensation behaviour, and in these cases more important than the language background of the listener, strengthening the claim that low level processes are at work. Further support comes from the findings that non-linguistic contexts can trigger compensation (Lotto & Kluender, 1998) and that non-human species have been shown to exhibit some of the same behaviours (Lotto, Kluender, & Holt, 1997).

A second class of mechanisms has been proposed in which compensation for assimilation is dependent on language experience. One such account is phonological inference (Gaskell & Marslen-Wilson, 1996, 1998) in which listeners learn the acoustic patterns that occur with an assimilating context. When the right acoustic pattern occurs in the assimilating context, listeners use this context to infer the intended utterance. This account was first modeled as a connectionist network (Gaskell, Hare & Marslen-Wilson 1995; Gaskell, 2003) but has recently been couched in models of Bayesian inference (Sonderegger & Yu, 2010; Snoeren, 2011). This mechanism operates pre-lexically making use of knowledge of patterns across phonemes. The account described by Gaskell and Marslen-Wilson (1998) also includes a role for lexical influences to account for the stronger effects in real words than in non-sense words. The crucial prediction of this account is that listeners are better able to compensate for patterns of assimilation that commonly occur in their language than those that are rare (i.e. language-dependent compensation).

Support for the knowledge-based/inference account comes from studies that have found language-dependent compensation for assimilation (e.g. Skorrupa, Mani & Peperkamp, 2013). Darcy and colleagues tested native monolingual English and French listeners as well as beginning and advanced bilingual listeners on obstruent voicing assimilation (found primarily in French) and obstruent place assimilation (found primarily in English) (Darcy, Peperkamp, & Dupoux, 2007; Darcy, Ramus, Christophe, Kinzler, & Dupoux, 2009). When tested on their native language, they found that listeners compensated most for the variation that occurred in their language (Darcy, et al., 2009). Furthermore, English learners of French behaved more like the native French listeners in their perception of assimilation if they were relatively advanced. Note that all of these studies used a word monitoring task and deliberate mispronunciations rather than natural assimilations. Furthermore, the acoustic properties of the stimuli were different for the French and English stimuli, rendering a comparison of the perceptual mechanism of the two groups of participants more difficult.

To summarize, a number of mechanisms have been proposed for listeners' abilities to handle assimilation patterns. They differ along several dimensions while making overlapping predictions. All predict more unassimilated responses in an assimilating context than in a non-assimilating context. One key difference is that the mechanism proposed by phonological inference predicts an important influence of language experience while the other mechanisms do not.

It is important to note that while each of these proposed mechanisms is distinct, most of them are not mutually exclusive. On the contrary, they may each play different roles under different circumstances. Most authors would likely agree that a complete account of compensation for pronunciation variation involves multiple mechanisms (e.g. Coady, Kluender, & Rhode, 2003) or that different mechanisms operate under different circumstances (e.g. see Mitterer, 2011, for an account of different mechanisms operating for different types of variability).

As an example, perceptual integration may reduce the salience of an assimilated pronunciation and phonological inference may further push the listener towards an unassimilated interpretation. Alternatively, feature parsing may be influenced by expectations about likely patterns in a given language (c.f. Gow & Segawa, 2009). Gow and Segawa (2009) investigated the brain structures underlying compensation for assimilation using a combination of neuro-imaging techniques. One of the most striking aspects of their result was the number of different regions involved, implicated in very different aspects of processing. Thus the goal of this paper is not to argue for or against the existence of particular mechanisms, but rather to understand better when and how they may operate. We do this by addressing some methodological limitations of previous studies with a new paradigm (as discussed in Section C below). This paradigm has been designed to better test one key prediction in the proposed mechanisms: the degree to which perception of assimilated speech is driven by specific language experience. It also allows us to test a number of different parameters such as partial and complete assimilation and preceding and following context. Furthermore, we attempt to identify the relative time course of effects by using eye-tracking in a visual world paradigm.

B. The time course of assimilation perception

One way to better understand the contribution of different mechanisms that deal with speech variability involves the time course of the effects. Mechanisms that are hypothesised to operate through auditory processing, such as contrast enhancement and perceptual integration, may be expected to influence the course of speech perception relatively quickly as they rely only on low-level auditory information which generally occur close in time. In contrast, phonological inference may operate over a different time scale, for example requiring the identification of the context sounds or word segmentation before the interpretation of the target sounds can be completed.

Several studies have assessed the time course of perceptual compensation mechanisms using on-line measures. An eye-tracking study using the visual world paradigm presented listeners instructions like “select the cat box” and pictures including both a cat and a cap (Gow & McMurray, 2007). Listeners’ eye-movements were affected by the following context word (e.g. “box”) approximately 560 ms after the onset of the context. This is in contrast to the effects of other manipulations, such as the influence of individual phonetic cues, which have been shown to influence eye-movements within approximately 200 ms of their occurrence in the signal (McMurray, Clayards, Aslin & Tanenhaus, 2008) which corresponds to the expected oculomotor delay. Studies using an oddball paradigm measured the degree of mismatched negativity (MMN) in Dutch listeners hearing assimilation in viable or unviable contexts that they were familiar with (i.e. nasal place assimilation Mitterer & Blomert, 2003) or not (i.e. liquid place assimilation Mitterer et al., 2006). Mitterer and Blomert (2003) did not test the time course directly, but the context effect appeared to be aligned to the onset of the context word before the full context word had been heard. Mitterer and colleagues (2006) found context-dependent modulation of the MMN again peaked before the end of the context word. Gow and Sagawa (2009) used a combination of fMRI, MEG and EEG to focus on brain regions involved in context dependent perception in the first 200ms after the onset of the context word. They found patterns of activity indicating perceptual areas were influence by auditory processing areas as well as areas thought to involve lexical and articulatory representations. Thus these techniques have found effects of context in the earliest moments of processing (though eye-movements were delayed relative to other results using the same paradigm). The present study will build on these results comparing two groups of listeners.

C. Present study

Previous results have found support for both language dependent and independent compensation mechanisms. However these studies have used very different methods (c.f. Gaskell & Snoeren, 2008). Studies that find language-independent results have used the same acoustic materials for all language groups, but with tasks that focus on the acoustic nature of the stimuli. For example MMN and form priming are thought to be sensitive to acoustic similarity (Marslen-Wilson, Moss, & van Halen, 1996; Näätänen, 2001; Utman, Blumstein, & Burton, 2000). Studies that find language-dependent effects have used tasks that focus on higher-level aspects of speech comprehension (e.g. Darcy et al., 2007), but at the cost of needing different materials for different language groups. Snoeren, Segui and Halle (2008a; 2008b) and Mitterer and McQueen (2009) have found dissociations in the kind of context effects found for different tasks (form vs. associative priming and 2AFC vs. the visual world paradigm). Furthermore, studies that find language-independent effects tend to use partial assimilation (e.g. Gow & Im, 2004) – which creates stimuli which are partially consistent with both the assimilated and unassimilated categories – while studies that find language-specific effects tend to use deliberate mispronunciations (e.g. Darcy, et al., 2007). Gow and Segawa (2009) found quite different patterns of neural activity for these two types of stimuli.

The purpose of the study reported here was to address these methodological discrepancies by testing for language specific compensation using acoustically controlled stimuli with a high-level task and a range of assimilation strengths. We chose the sibilants /s/ and /ʃ/ in French and

English as they are not thought to differ along phonologically relevant dimensions across the two languages in neutral contexts. They do however, show different patterns of assimilation. In English, /s/ assimilates to /ʃ/ when it precedes it as in ‘dress shop’ (regressive assimilation) but not when it follows as in ‘crash site’ (progressive assimilation) (Holst & Nolan, 1995; Niebuhr, Clayards, Lancia & Meunier, 2011; Pouplier, Hoole & Scobbie, 2011). A recent production study comparing sibilant assimilation in English and French (Niebuhr, Clayards, Meunier, & Lancia, 2011) found that French does allow assimilation of /s/ to [ʃ] both when /ʃ/ follows /s/ (e.g. ‘classe chargée’) and when /ʃ/ precedes /s/ (e.g. ‘tache super’). Thus French and English differ in the direction of assimilation allowed (French allows both regressive and progressive assimilation while English allows only regressive). Furthermore, while both languages exhibited partial as well as (acoustically) complete assimilation, complete assimilations were much more prevalent in English (Niebuhr, Clayards, Meunier, & Lancia, 2011)[1]. Thus English is dominated by complete assimilations of /s/ in when /ʃ/ follows, while French allows some partial assimilation when /s/ is followed or preceded by /ʃ/. Phonological inference predicts therefore that the most compensation will be observed for the English listeners when /ʃ/ follows the target sibilants. In other words, when the target sibilant is /ʃ/-like, and followed by /ʃ/, we expect English listeners to report hearing /s/ some of the time. The other mechanisms discussed don’t depend on language experience so we expect them to operate in the same way for both French and English listeners. Detailed predictions are discussed at the beginning of the results section.

We also tested the effect of a preceding context on compensation. Progressive assimilation is less common cross-linguistically and compensation for complete progressive assimilation has received little attention in the psycholinguistic literature. In order to investigate the time course of context effects, we used a word recognition paradigm where listeners clicked on objects representing the words they heard and we monitored their eye-movements to the candidate objects as they listened to the instructions.

II. METHODS

A. Participants

Twenty-five monolingual British English speaking participants with no known history of hearing or language problems were recruited from the University of York participant pool in the UK. Twenty-six French speaking participants were recruited at the Université de Genève in Switzerland. All were native speakers of French from either Switzerland or France with limited exposure to English or other languages and no known history of hearing or language problems. All recruiting, testing and other interactions were done in the participants' native language.

B. Artificial Lexicon

We taught both French and English participants the same set of novel words containing sibilants at word boundaries. Previous research has demonstrated that newly learned words can trigger compensation for assimilation (Snoeren, Gaskel & Di Betta, 2009). These words corresponded to novel concepts: six novel shapes with critical consonants at word offset (e.g. "tama***sh***", "pida***s***") and six novel texture 'buttons' with critical consonants at word onset (e.g. "***s***amal", "***sh***innow"). Listeners followed instructions to manipulate the objects, changing the texture of one of the shapes by clicking on it and on the corresponding texture 'button' such as one might do in a computer paint program (see also Revill, Tanenhaus, & Aslin, 2008). These instructions created contexts where assimilation could occur (e.g. /s#ʃ/ in "render the cavee***s*** ***sh***innow please"). The shapes and textures were organized into two sets to test the effect of following context and preceding context. One set of stimuli included a minimal pair of shapes differing in their final sibilant (“caveesh” and “cavees”) and four textures that provided a left context (“shinnow”, “sival”, “pagoon”, “pentuf”). The second set included a minimal pair of textures differing in the onset sibilant (“shamal”,”samal”) and four shapes that provided a right context (“tamash”, “pidas”, “nalip”, “remope”). The novel words were chosen to be roughly equivalently suitable neologisms in each language. This was done by comparing the cumulative bigram frequencies of a large set of candidate words for English and French respectively. All novel words were also restricted to have 0 to 1 real word neighbours. Native speakers in each language then rated a smaller matched set of candidate words “goodness” as a potential word in their language. Six additional French-English bilinguals rated the words on a scale from French to English. Using these ratings, 12 words were chosen (see Table 1).

Table 1: Artificial lexicon of shape and texture names.

	Pronunciation (broad transcription)	English Orthography	French Orthography
Target shapes	/kɑvis/	cavees	cavisse
	/kɑviʃ/	caveesh	caviche
Context textures	/sivɑl/	sival	sivale
	/ʃino/	shinnow	chinno
	/pɑgun/	pagoon	pagune
	/pɛntuf/	pentuf	pentoufe
Target textures	/sɑmɑl/	samal	samale
	/ʃɑmɑl/	shamal	chamale
Context shapes	/pidɑs/	pidas	pidasse
	/tɑmɑʃ/	tamash	tamache
	/rimop/	remope	remope
	/nɑlɪp/	nalip	nalipe

C. Auditory Stimuli

Stimuli were constructed from recordings of one native British English speaker and one native European French speaker, both at the University of York. The native French speaker had limited experience with English and had been in the country for one month. The English speaker had similarly limited experience with French. Recordings were made onto a Marantz CDR 300 digital recorder using a Sennheiser unidirectional table-mounted microphone in a quiet room. The speakers read sentences consisting of instructions containing the novel words. Sentences containing object names were “Render the (shape) purple please” or “Rendez le (shape) pourpre s'il vous plait". Sentences containing texture names were "Render the circle (texture) please" or "Rendez le cercle (texture) s'il vous plait". Both speakers were instructed to place stress on the second syllable of each word as this was expected to be the pattern in French and easier for the English speaker to match. Both speakers produced the sentences in French and English. The purpose of this was to present stimuli from both speakers to both listener groups in their native language. Speakers also produced sentences in which the initial or final sibilants of the minimal pairs were replaced with [f] (eg. "caveef") for use in the stimulus construction as described below. Training and testing stimuli were constructed from these recordings. All splicing and manipulations were done using Praat (Boersma, 2001).

Training stimuli containing context items were naturally produced. The only manipulation was to normalize amplitude – such that all stimuli had the same average SPL – and speaking rate (using PSOLA and equivalent portions of the instructions, e.g., "Render the" and "please", as reference points) – such that all stimuli had the same average speaking rate across items and speakers. Training stimuli containing target words (e.g., "cavees" or "caveesh") were manipulated in the same way. Additionally, the critical words were endpoints of the continuum described below.

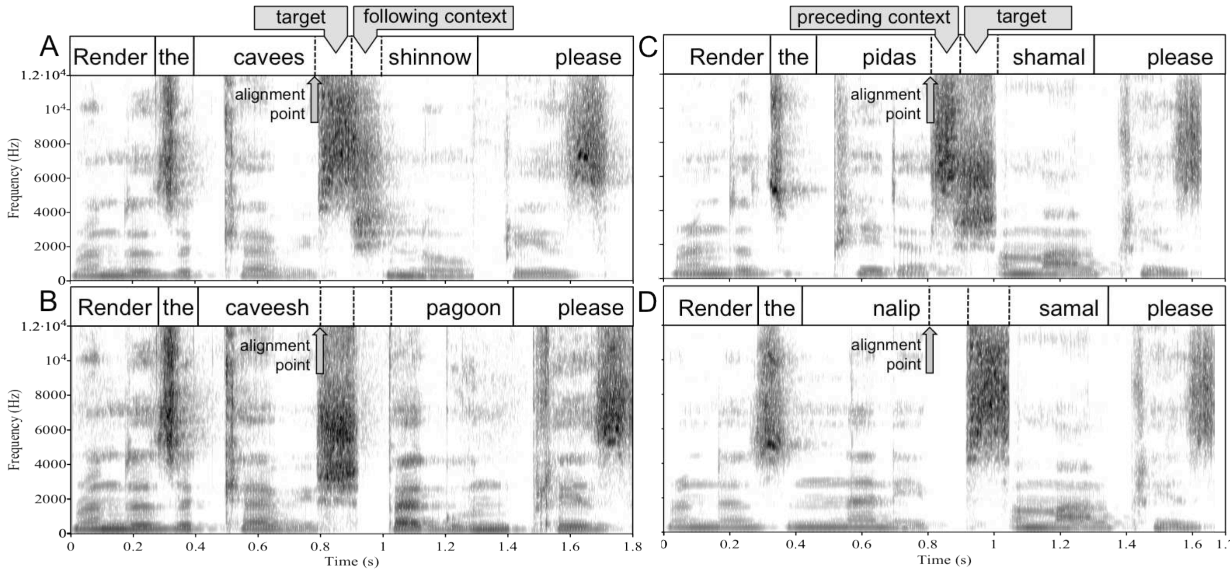

Figure 1: Spectrograms of four sample stimuli spoken by the French speaker in English. A) Right context, target is the [s] end of the continuum, context is natural /ʃ/. B) Right context, target is the [ʃ] end of the continuum, context is /p/ (control). C) Left context, context is natural /s/, target is the [ʃ] end of the continuum. D) Left context, context is /p/ (control), target is the [s] end of the continuum. Alignment points are for eye-movement analysis and refer to 0 ms in the eye-movement plots.

A seven-step continuum of sibilants ranging from [s] to [ʃ] was created to represent a range of assimilation strengths[2]. One token of [s] and [ʃ] was excised for each speaker and intermediate steps were created by weighted averaging of waveforms such that the two sibilants were mixed together in different proportions. These sibilants were then spliced[3] onto versions of the target words produced with a control sound [f]. This was done so that any other cues which may be contained in the word (such as vowel cues identified by Niebuhr & Meunier 2011), would not bias towards either sibilant. The fricative [f] was chosen because it was thought to be equally unrelated to [s] or [ʃ] in terms of place of articulation and because it matched the target sounds in voicing. Six native English and six native French listeners were asked to categorize these stimuli (two alternative forced choice) in a pilot test to determine which steps would be used in the main experiment. Table 2 includes the CoG values (calculated on the entire unfiltered sibilant) of the 7 steps used for the French and English talkers. Context stimuli were always naturally produced tokens. Figure 1 shows spectrograms of a sample of the stimuli including the two endpoints of the sibilant continuum, the context sibilants and the control contexts for the French speaker speaking English.

Table 2: Centre of gravity (Hz) for each step of the fricative continua for each talker.

Talker	step 1	step 2	step 3	step 4	step 5	step 6	step 7
English	6606	8040	8160	8280	8400	8520	8640
French	5503	6900	7020	7140	7260	7380	7500

Test stimuli were constructed by splicing together the first half of a sentence containing a novel object (eg. “Render the nalip...”) and the second half of a sentence containing a novel texture (eg. “...shinnow please”). Filler sentences contained all possible combinations of the context words except those that would create sibilant sequences. Each of the contexts was also combined with the target continua.

D. Visual Stimuli

Shapes and textures were chosen to be easily distinguishable but difficult to name with existing words. The novel shapes were free-form abstract shapes, and buttons included an area shaded with a novel texture. All textures were greyscale in order to prevent participants from relabeling them with a colour adjective. Assignment of labels to shapes and textures was separately randomized for each participant to further reduce the impact of any differences between the visual stimuli.

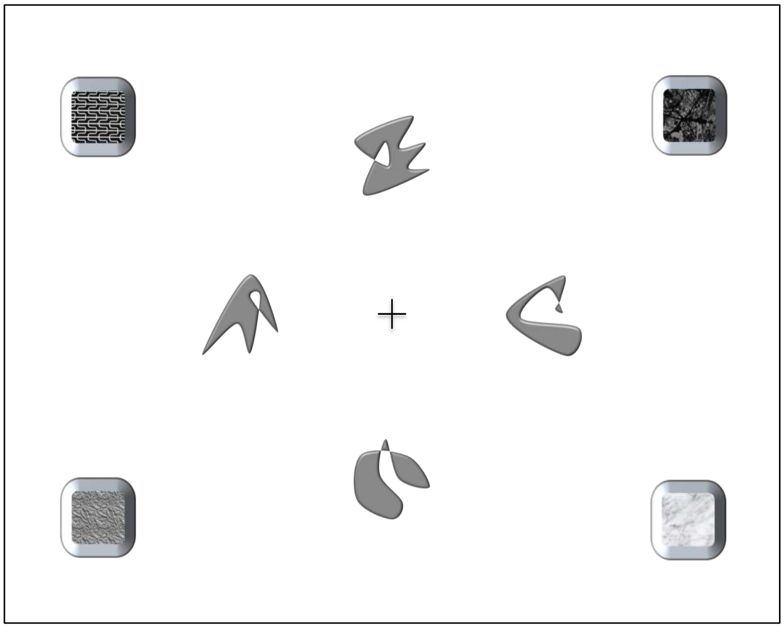

Figure 2: An example display screen from the test phase containing 8 of the 12 visual stimuli.

During training and testing, visual stimuli were displayed in one of eight possible locations on screen. Screen dimensions were 16x12 inches and the resolution was set to 1024x768 pixels. Four of the locations were in the centre of the screen, 150 pixels above, below and to either side of a central fixation cross. The other four were displayed in the four corners of the screen, 80 pixels away from the edge of the screen. All stimuli were 101 by 101 pixels. Figure 2 is an example test display screen. Location was randomized across trials. Either shapes or buttons were displayed in the center four locations and the others were in the four corners. The type of visual stimulus in the center was varied by block as described below.

E. Procedure

1. Training

Participants were first trained on the artificial lexicon. Training was done on a PC computer and auditory stimuli were presented via audio speakers placed on either side of the monitor. Training occurred in a single session lasting approximately an hour. The session began with simplified versions of the ultimate task (2AFC with feedback textures, 2AFC with feedback shapes) and progressed through more complex versions (4AFC) once the participants reached a performance criterion on a test block or once they had completed four training blocks. After finishing the 4AFC training phase, participants were given a sheet of paper with the pictures of the shapes and buttons on one half of the paper and their labels (in the appropriate orthography) on the other half. They were instructed to match the pictures with their labels. The final version of the training task was the same as that used in testing – four novel shapes and four novel texture buttons – but crucially did not contain any sibilant sequences (e.g. "pidas samal"). On half of these trials the four textures were in the centre of the screen and the shapes in the outer positions (see Figure 2), on half the shapes were in the centre and the textures on the outside. The task was always to click on the correct centre object (either shape or texture) first and then the correct peripheral object (shape or texture). This was done for testing reasons explained below. Shape-first and texture-first trials were blocked and half the participants began training with shape-first trials and half with texture-first trials.

2. Testing

Testing was done in two sessions, normally on different days[4], each lasting approximately 45 minutes to an hour. Testing was never done on the same day as training. During each session the participants' eye movements were monitored using an eye-tracker. The English language participants were tested in York with an Eyelink 2000 desktop mounted tracker. The French language participants were tested in Geneva with an Eyelink II head mounted tracker. For both groups the right eye only was tracked unless there were technical difficulties in which case the left eye was tracked. Both trackers sampled at 250 Hz and all other parameters were the same. The first author operated the trackers during all data collection. It was not expected that the difference in trackers would cause any systematic differences in the data collection. Testing was done in a quiet room. All participants were seated in a comfortable position approximately 18 inches from the screen. Auditory stimuli were presented using the same audio speakers as during training at a comfortable listening level. The same audio speakers were used for both groups of participants.

The first session began with a short refresher phase, identical to the last phase of training in which four shapes and four textures appeared together on screen. Participants that made more than 10% errors in this phase were shown the correct answers for the items they got wrong using the matching task paper and the training was repeated maximally one more time.

After the refresher phase, testing began. In the second session testing began immediately after calibration of the eye-tracker. Test trials were similar to the last phase of training but did not include feedback. Each trial began with a drift correction in which participants fixated a central dot and pressed the space bar. The visual stimuli then appeared for 500 ms. After the delay, the auditory instructions were presented such as “Render the pidas pagoon please”. Participants responded by clicking on a shape and a texture just as in training. After they made their response the display continued for 500 ms then was replaced with a blank screen and the next trial started.

Test trials were blocked by whether the participant was instructed to click on the shapes first or the textures first, just as in training. This manipulation was included because it was expected that eye-movements would only be reliably time-locked to the auditory stimulus for the first set of items that was clicked on. Figure 2 shows a configuration for a trial in which the shapes are clicked on first. In general, participants were able to do both versions of the task and made few errors in the order of clicking.

Each test session comprised 200 trials, alternating between blocks of 25 shape-first trials and blocks of 25 texture-first trials for a total of 400 trials across the two sessions. The order of blocks was counterbalanced across subjects just as in training. Participants were given the opportunity to take breaks between blocks. The 200 shape-first trials consisted of 168 critical trials and 32 filler trials. The critical trails were made up of each of the 7 steps of the 'caveesh'-'cavees' continuum paired with each of the three contexts four times for each talker (following context trials). The filler trials contained only the non-minimal pair items (context words) paired with each other. Analogously, the 200 shape-first trials consisted of 168 critical and 32 filler trials. The critical trials consisted of the 7 steps of the 'shamal'-'samal' continuum paired with each of the three contexts four times for each talker (preceding context trials).

III. RESULTS

In pilot testing it was found that some participants failed to categorize the endpoints of the [ʃ] to [s] continua correctly despite having passed all the training criteria. To ensure that participants' responses reflected their interpretation of the target words and not any lack of learning about the form of the words themselves, accuracy on the endpoints of the continua in the control context was assessed. Each participant's accuracy was calculated for each continuum ('shamal'-'samal' and ‘caveesh’-‘cavees’) for each of the talkers separately and an inclusion criterion of 75% correct was set. For the ‘shamal’-‘samal’ continuum, 7 of the English and 8 of the French participants failed to meet the criterion for the English talker and 6 of the English and 4 of the French participants failed to meet the criterion for the French talker. For the ‘caveesh’-‘cavees’ continuum, 12 of the English and 12 of the French participants failed to meet the criterion for the English talker and 5 of the English and 4 of the French participants failed to meet the criterion for the French talker. As the exclusion rate was quite high for the English talker, the data in the control conditions was examined for each talker separately. The biggest differences in responses to the two talkers related to how /s/-like the two sibilant continua sounded. The continuum for the English talker had higher CoGs and consequently both listener groups responded /s/ on a higher proportion of trials for the English talker (M = 67%, SD = 20%) than the French talker (M = 60%, SD = 23%) in the control condition. This higher proportion of /s/ responses, raises the possibility of ceiling effects for many of the steps, making it difficult to observe effects of sibilant context. This was especially true for the English listeners in the ‘shamal’-‘samal’ trials for which the mean /s/ responses to the English talker in the control condition (across all continuum steps) was 77%. Because of these baseline differences in /s/ responses and the difficulty that many listeners had with the English talker in general, the analyses here focus on the results for the French talker alone. Importantly both French and English listeners heard this talker. Additional analyses were performed for data from both talkers and all cases are noted in which the significance of an effect differed between the analyses with the French talker and the analyses with both talkers.

Exclusion of participant data according to the above criteria left data from 22 participants in the French group and 20 participants in the English group. Trials were also excluded from analysis if the participant clicked on any of the objects other than one of the targets. Less than 2% of the trials were excluded and 40% of these were errors of “sival” for “samal” or “shamal”, likely reflecting the partial overlap of these forms. Analyses of the mouse click responses and eye-movements were then performed on the remaining trials.

A. Mouse click responses

Before considering the results it may be helpful to consider the predictions of the different mechanisms for our stimuli. The most basic distinction is that language dependent mechanisms (such as phonological inference) predict differences between language groups while language independent mechanisms (any of the auditory accounts) do not. The phonological inference account predicts the most compensation (/s/ responses to the target continuum) for listeners and contexts where the most assimilation occurs (English listeners where the assimilating context follows the target continuum, i.e. ‘caveesh-cavees’ trials). The other mechanisms predict the same patterns for both language groups.

Most studies that have investigated context effects on target continua (e.g. Mann & Repp 1980; Lotto & Kleunder, 1998) have reported a shift in overall responses away from the context. For example, Lotto and Kleunder (1998) found that a target da-ga continuum varying in F3 was perceived as if it had a higher F3 (more /da/ responses) when preceded by a segment with a low F3 (/r/) or low tone. We refer to this pattern as contrastive – in our case the target continuum would be perceived as more /s/-like in the context of /ʃ/ and more /ʃ/-like in the context of /s/. Such a pattern is predicted by contrast enhancement. It is also possibly to have the opposite pattern – the target is perceived as more like the context so more /s/ responses in the context of /s/. We will refer to this pattern as anti-contrastive. A different pattern was observed by Mitterer and colleagues. They found responses in an assimilating context to be closer to chance across the continuum (flatter categorization function) compared to a non-assimilating context. They argued that the assimilating contexts make the target contrasts harder to discriminate due to interactions of target and competitor. In our case this would predict that in the assimilating context /ʃ/, listeners would respond with /s/ more often at the /ʃ/ end of the continuum and less often at the /s/ end of the continuum. We refer to this pattern as flattening as the categorization function is flatter in the assimilating condition. Thus far, flattening has only been observed when a context that triggers assimilation follows. Here we also test whether it occurs when the context precedes and we also include a context that is acoustically similar (/s/) but does not trigger assimilation. Feature parsing predicts stronger context effects for ambiguous stimuli as they contain traces of the ambiguous segment, though the strongest prediction is for naturally occurring assimilations which we do not test (Gow & Sagawa, 2009).

The proportions of mouse click responses for each target word and for both groups of listeners were submitted to a mixed effects logistic regression using the *lmer()* function of the *lme4* package (Bates, Maechler, & Bolker, 2011) of the R analysis program (R core development team). Two regression models were built, one for the following-context trials and one for the preceding-context trials. Proportion of /s/ responses was the dependent variable. Predictor variables were: participant language background (English vs. French – coded as 0.5 and -0.5), context (/s/ vs. /ʃ/ – coded as 0.5 and -0.5 – and control vs. sibilants – control coded 2/3 and /s/ -1/3 and /ʃ/ -1/3), CoG step (was a continuous variable and was centered, i.e. varied from -3 to 3, to remove co-linearity between it and the interactions) and the two and three way interactions of these variables. Participant was modeled as a random factor and random slopes were included for the within subjects variables (CoG step and context) and their interaction. The /s/ vs. /ʃ/ variable was designed to test whether potentially assimilating contexts (e.g. ‘shamal’) would result in more /s/ responses, indicating that listeners were compensating for the effect of context. Furthermore, interactions between step and context will reveal whether the effect of context is a simple shift in bias towards more /s/ responses, or whether the slope of the categorization function differs by context.

1. Following Context

This is the context where English listeners have experience with the most complete and frequent assimilations and the French listeners have much less experience, thus this is also where we should find differences between language groups, if they exist. Figure 3 shows the results for the following-context trials (‘caveesh’-‘cavees’). We first discuss the effects for both language groups together. As CoG increased, there were more /s/ responses, reflecting the fact that /s/ is produced with higher CoG (CoG step: $\beta = 0.98$, SE = 0.05, $p < 0.0001$). There were fewer /s/ responses in the (average) sibilant contexts than in the control context (sibilant vs. control: $\beta = -0.55$, SE = 0.23, $p = 0.015$) and fewer /s/ responses in the /s/ context than the /ʃ/ context (/s/ vs. /ʃ/: $\beta = -0.58$, SE = 0.26, $p = 0.028$). There was a bigger difference between the sibilants and the control contexts at more extreme CoG values (sibilant vs. control by CoG step: $\beta = -0.60$, SE = 0.08, $p < 0.0001$) and a bigger difference between /s/ and /ʃ/ contexts at more extreme CoG values (/s/ vs. /ʃ/ by CoG step: $\beta = 0.53$, SE = 0.09, $p < 0.0001$). In other words the response curves were closer to chance (flatter) in the ‘shinnow’ context than in the ‘sival’ context and both sibilants contexts produced flatter response curves than the control context.

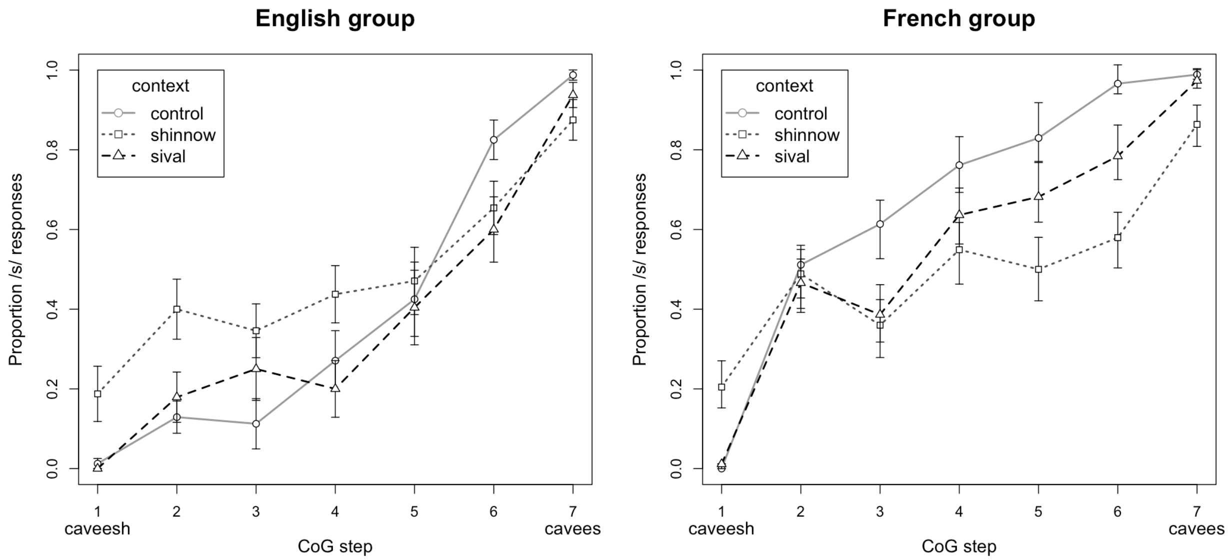

Figure 3: Mouse click responses for the two groups of listeners categorizing the 'caveesh-'cavees' continuum before one of three different contexts, control (solid green), 'sival' (dashed blue) and 'shinnow' (dotted red). Error bars are standard error of the mean by subjects and do not reflect the standard error of the model.

When we considered the effect of language group on the data, we found that there were more /s/ responses overall for the French than the English group (Listener group: $\beta = 0.88$, SE = 0.15, $p < 0.0001$)[5] indicating that the French group heard the continuum as more [s]-like in general but the average slope of CoG step was the same across listener groups (no interaction between Listener group and CoG step). There was also an interaction between language group and sibilant versus control context (Listener group by sibilant vs. control: $\beta = -0.96$, SE = 0.30, $p = 0.001$) and between language group and the two sibilant contexts (Listener group by /s/ vs. /ʃ/: $\beta = 0.93$, SE = 0.30, $p = 0.002$). None of the three-way interactions were significant. The interactions between language group and contexts are shown in Figure 4. The interactions were further broken down in two models, one for the data from each language group, including CoG step, context and their interactions. These models found a main effect of sibilant versus control for the French (β = -3.52, SE = 0.84, $p < 0.0001$) but not the English listeners (β = 0.35, SE = 0.24, p = 0.15). There was also a main effect of /s/ vs. /ʃ/ context for the English (β = -2.21, SE = 0.65, p = 0.0007) but not the French listeners (β = 0.45, SE = 0.34, p = 0.19). Thus the English listeners alone showed a pattern consistent with compensation for assimilation: more /s/ responses in the assimilating context. Consistent with the main model, the models for both listener groups found interactions between CoG step and sibilant context (English listeners: β = 1.05, SE = 0.19, $p < 0.0001$; French listeners: β = 0.37, SE = 0.11, p = 0.0006) indicating that for both groups, categorization of the target continuum was closer to chance in the context of 'shinnow' in line with the flattening observed by Mitterer and colleagues.

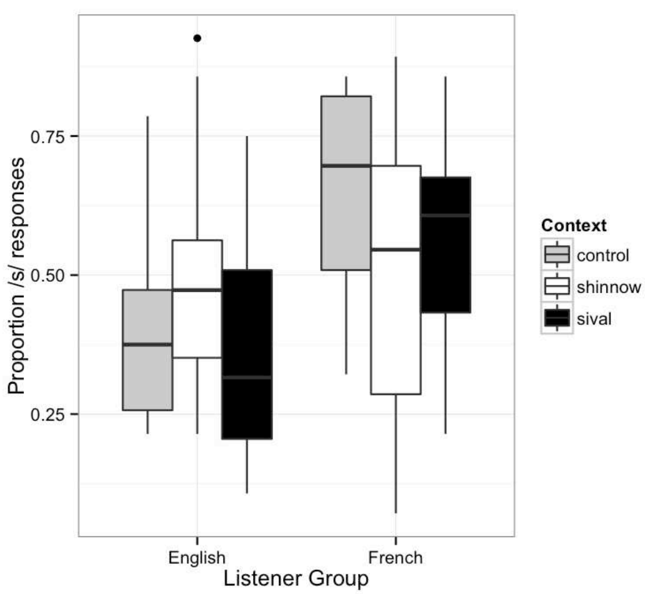

Figure 4: Proportion /s/ responses for each of the contexts and language groups. Data represent distributions of participant mean response proportions. The difference between 'shinnow' and

'sival' contexts is significant for the English listener group. The difference between control and sibilant contexts is significant for the French listener group.

2. Preceding Context

Figure 5 shows the results for preceding context trials ('shamal'-'samal'). In this context neither listener group has extensive experience with assimilation in production, so language-dependent accounts do not predict differences between language groups. As CoG increased, the proportion of /s/ responses increased, again, as expected (CoG step: $\beta = 1.24$, SE = 0.10, $p < 0.0001$). There was no difference in the proportion of /s/ responses between the average sibilant contexts and the control context (sibilant vs. control: $\beta = -0.46$, SE = 0.27, $p = 0.09$) but there were more /s/ responses in the 'tamash' context than the 'pidas' context (/s/ vs. /ʃ/: $\beta = -1.65$, SE = 0.35, $p < 0.0001$).

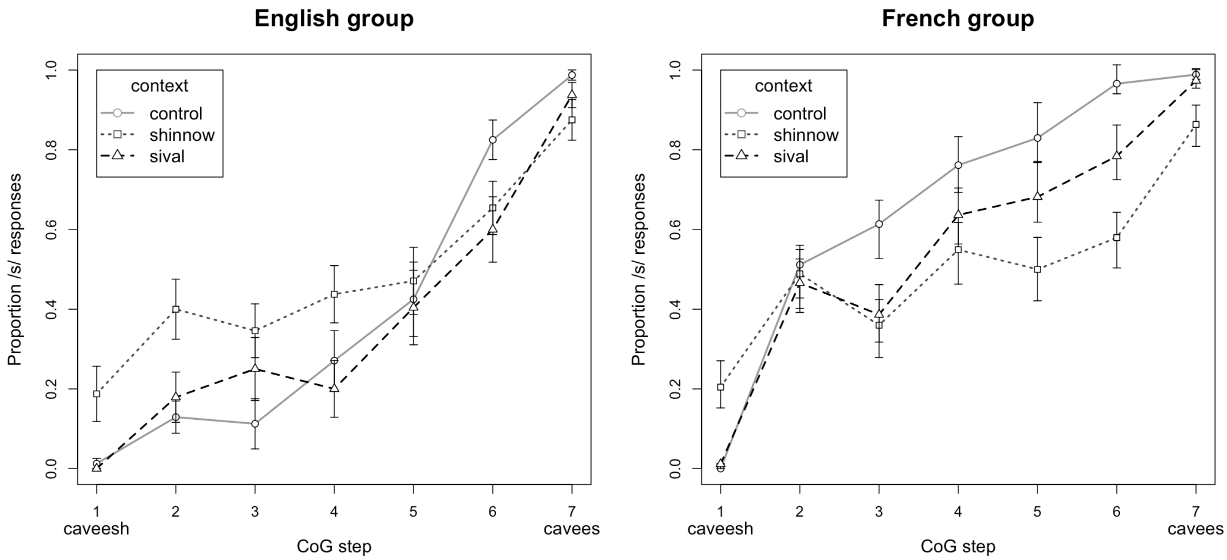

Figure 5: Mouse click responses for the two groups of listeners categorizing the ‘shamal’-‘samal’ continuum after one of three different contexts, control (solid green), ‘pidas’ (dashed blue) and ‘tamash’ (dotted red). Error bars are standard error.

Interestingly, unlike in the ‘caveesh-cavees’ trials, the effect of preceding context did not depend on CoG step (/s/ vs. /ʃ/ by CoG step: $\beta = -0.09$, $SE = 0.13$, $p = 0.47$) indicating that categorization of CoG was no closer to chance for any of the contexts. This time the English listener group had more /s/ responses overall (Listener group: $\beta = 1.78$, $SE = 0.17$, $p < 0.0001$) indicating they heard the stimuli as more [s]-like. There was no significant interaction between the context and listener group – indicating that both groups of listeners compensated for the context in the same way. None of the three way interactions were significant[6].

3. Summary of mouse click data

The pattern of context-dependent mouse click data was strikingly different depending on whether the context preceded or followed the target. Results in the preceding context condition (‘shamal’-

'samal') are best described as more /s/ responses in the context of 'tamash' and fewer in the context of 'sival'. In other words, the target tended to be perceived as less like the context (i.e. contrastively). Both French and English listeners showed the same effect of context despite some differences in the amount of progressive assimilation observed in production between the two languages (slightly more progressive assimilation is observed for French). Such a symmetric effect, restricted to ambiguous sibilants is consistent with an interpretation of a language general process similar to auditory contrast enhancement or feature parsing. Of course it is always possible that a more sensitive test may reveal differences between the groups that reflect their differing experience.

Behaviour in the following context ('caveesh'-'cavees') condition showed a different pattern. First, categorization of the target continuum was closer to chance for the sibilant contexts than the control context and even more so for the 'shinnow' context than the 'sival' context. This is consistent with listeners experiencing more ambiguity in the context of a following sibilant and especially for 'shinnow'. Increased ambiguity is predicted by the perceptual integration account in the context of sounds which trigger assimilation. The fact that an acoustically similar context sound (/s/) also triggers some flattening suggests that the effect may be due to the acoustic properties of the context sound. Note that while there are alternative explanations for this flattening effect (e.g. non-auditory ones), Mitterer, Csepe and Blomert (2006) found corresponding reductions in discrimination accuracy which supports the interpretation of increased ambiguity. Secondly, and crucially, language group also mattered in how listeners were affected by the following context. The English listeners alone responded more often with /s/ in the context of 'shinnow', consistent with the prediction of phonological inference that experience with assimilation patters will trigger compensation.

We will return to the difference between the compensation effects of right and left context in the Discussion. First, to investigate further the role of the different classes of compensation mechanism we examined the time course of the context effects by analysing the eye-movement data. Eye-movement data is often analyzed by comparing looks to target and competitor for different trial types. Our experimental design was quite complex making this kind of analysis cumbersome, and our experimental questions do not lend themselves well to simple target or competitor analysis. Furthermore, we were primarily interested in *when* the effects observed in the mouse click data could be observed in the eye-movement data. For these reasons we adopted an analysis of the eye-movement data which replicated our analysis of mouse-click data at multiple time points, as described below (see Mitterer & Reinisch, 2013, for a similar analysis).

B. Eye-movement results

The same participants and trials were excluded as for the mouse click data. The data were first coded according to which object the participant was looking at for each time sample (every 4 ms) on each trial. Trials were time-aligned at the onset of the sibilant sequence (in the control condition containing only one sibilant, trials were aligned at the beginning of the sibilant in the right context and 100 ms before the sibilant for the left context. This was to control for the fact that the sibilant sequences were approx 100 ms longer than the single sibilants). To obtain a measure similar to proportion of /s/ responses as in the mouse click data, we first computed the proportion of looks to the /s/ object (‘cavees’ or ‘samal’) and the /ʃ/ object (‘caveesh’ or ‘shamal’) across trials for each participant, context and CoG step. The /s/ bias was then computed by subtracting the proportion of looks to the /ʃ/ object from the proportion of looks to the /s/ object. Mixed model regression analyses were then performed. These analyses were identical to the mouse click regression analyses with the exception that they were computed over the proportion of fixations during successive 100ms time windows. Since we expect participants to look at what they click on, we expect the eye-movements to mirror the mouse click data overall (and this is what we found) but because of the temporal nature of this measure, it allowed us to track the development of the effects we found in the mouse click data.

1. Statistical analysis

The bias measure was rescaled to vary from 0 to 1 instead of from -1 to 1 and then transformed to log odds. Linear regression was performed on the log odds data using the *lmer()* function of the *lme4* package (Bates, et al., 2011) of the R analysis program (R core development team). This is equivalent to a logistic regression on a binary outcome but allows for the aggregation of data into time bins. Predictor variables were identical to the mouse click analyses. Regression analyses were performed for 100 ms bins starting from the onset of the sibilant sequence, offset by the expected 200 ms oculomotor delay until 2 seconds after the sibilant onset (i.e. 200-2000 ms after the onset of the sibilant sequence). Median reaction time for the following-context condition was 1812 ms after sibilant onset for English listeners and 2007 ms after sibilant onset for French listeners. Median reaction time for the preceding-context condition was 2016 ms after sibilant onset for English listeners and 2388 ms after sibilant onset for French listeners. Figure 6 panel A shows the estimated p-values calculated from the t-statistic for several of the relevant variables and interactions when the context follows (‘caveesh’-‘cavees’ trials) and Figure 7 panel A shows the estimated p-values for the same variables when the context preceded (‘shamal’-‘samal’ trials). In these figures, a single set of p-values is given for both the French and English listeners as a single model was fit to the data from both language groups.

2. Effect of target sibilant

The CoG step of the target sibilant affected eye-movements within 100 ms of that information becoming available, assuming a 200 ms oculomotor delay. Figure 6 panels C and D show that listeners were more biased to look at 'cavees' and less biased to look at 'caveesh' as CoG step increased. Figure 7 panels C and D show the same pattern for 'shamal' and 'samal'. The coefficient for CoG step became statistically reliable at 300-400 ms after sibilant onset and stayed reliable for the rest of the time windows for the 'caveesh'-'cavees' trials ($\beta = 0.06$, SE = 0.02, $t = 2.60$, $p = 0.009$) and the 'shamal'-'samal' trials ($\beta = 0.04$, SE = 0.02, $t = 2.38$, $p = 0.017$). The timing of this effect is in line with previous eye-tracking results showing that phonetic variables distinguishing contrastive sounds influence looks to the target as soon as they are available (e.g. McMurray, Clayards, Tanenhaus & Aslin, 2008). We can therefore compare the timing of other effects to this benchmark.

As in the mouse-click data, the two listener groups differed in how [s]-like they perceived the target sibilants and this effect became significant at 700-800 ms for the following context (Figure 6 panel B, solid versus dotted lines) and 400-500ms for the preceding context (Figure 7 panel B, solid versus dotted lines).

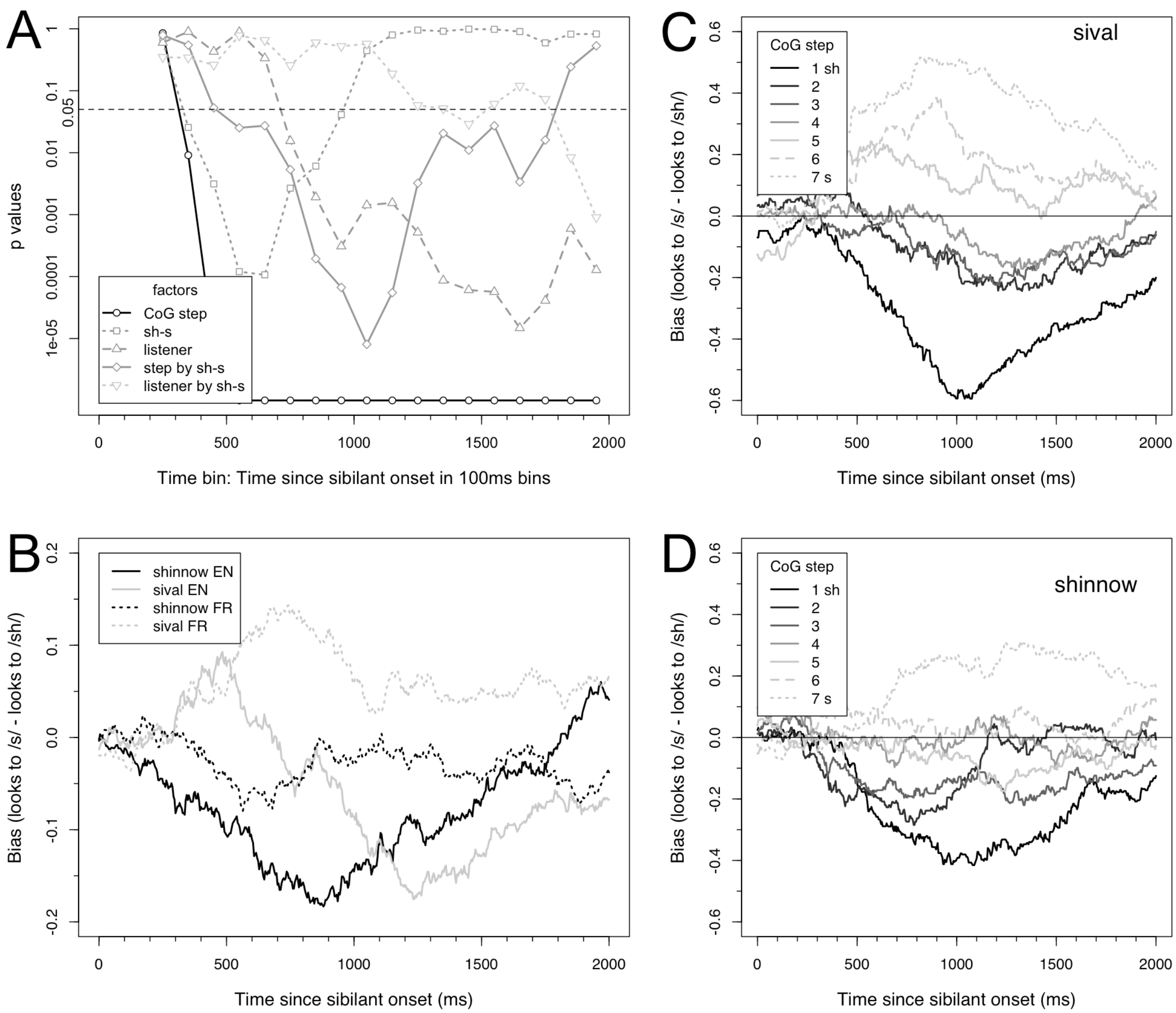

Figure 6: Eye-movements from ‘caveesh’-‘cavees’ trials. A Estimated p-values of key factors of the regression models for each time bin (first bin is 200-300 ms, etc). B Interaction of /ʃ/-/s/ context and listener group on bias measure over time. C Effect of CoG step in the context of ‘sival’. D Effect of CoG step in the context of ‘shinnow’.

3. Effect of context sibilant: Following context (‘caveesh’-‘cavees’)

Figure 6 panel B shows that after hearing the sibilant sequence listeners were first biased *towards* the context sibilant (more looks to /s/ in the context of ‘sival’ – grey lines – than in the context of ‘shinnow’ – black lines – i.e. an anti-contrastive not a contrastive effect). The effect of /s/ vs. /ʃ/ context became significant at 300-400 ms ($\beta = 0.31$, SE = 0.14, t = 2.23, p = .025; dotted grey line with squares drops below p = 0.05 in Figure 5 panel A), which is again as soon as the context sibilant information became available given the 100 ms duration of the target sibilant and the 200 ms oculomotor delay. This early effect was the *reverse* of the expected compensation effect, and may be due to the nature of sibilant sequences. Because there are no clear boundaries between sibilants, listeners may temporarily interpret the context sibilant as a continuation of the target sibilant. In fact, Whalen (1991) has shown that listeners do not pay attention to the ordering of different frication portions of a sibilant or sibilant sequence and instead respond to the spectral characteristics of the whole and rely on the acoustic context (e.g. formant transitions) to resolve any ordering ambiguity. If this is true for our listeners, they may interpret the [s] in ‘sival’ as evidence for ‘cavees’ initially. Only once the identity of the second sibilant is unambiguously established – the main effect becomes unreliable between 800 and 900 ms after target onset, around the second syllable in ‘sival’ plus the oculo-motor delay – is the sibilant sequence reinterpreted. Note however, the different pattern for English and French listeners. After about one second the English listeners show a reversal of the context effect – more fixations to /s/ in the context of ‘shinnow’, a contrastive effect (Figure 6 panel B, solid lines). An additional set of models was run just on the data from the English listeners. These models found a significant effect of /s/ vs. /ʃ/ context in the contrastive direction starting at 1800-1900 ms. Consistent with this, the main analysis found that the interaction between the /s/ vs. /ʃ/ context effect and listener group became significant from 1800 ms (time: 1800-1900, β =0.63, SE =0.24, t = 2.64, p = .008; Note that the p-value first goes below 0.05 between 1400 and 1500 however it does not stay significant until the later time point thus a more conservative estimate is the later time point). Thus all listeners begin with similar context effects, but this is re-interpreted by the English group.

An additional effect of context, described in the mouse-click data as increased ambiguity or flattening of the categorization function in the context of /ʃ/ (interaction of CoG step and /s/ vs. /ʃ/ context), also shows up in the eye-movements quite early on – within 300 ms of the end of the sibilant sequence and 200 ms after the onset of the main /s/ vs. /ʃ/ context effect just discussed (time: 500-600, β =0.13, SE =0.06, t = 2.24, p = .025). Figure 6 panels C and D show that this interaction is due to listeners looking more equally to both objects (bias closer to 0) in the context of 'shinnow' while being more strongly biased (bigger effect of CoG, i.e. bigger differences between the lines) in the context of 'sival'. This effect weakens around 1000 ms and becomes non-significant by 1800 ms post-sibilant onset (Figure 6 panel A, solid grey line with diamonds). Interestingly, this time-course is mirrored by the increase in strength of the context by language interaction (Figure 6 panel A, dotted light grey line with inverted triangles). These patterns of eye-movements support our interpretation of the mouse-click data in which all listeners experience early ambiguity created by the assimilating context, but only the English listeners end up with a pattern consistent with compensation for assimilation.

One final analysis was done on the data from this context. Because looks to the two target shapes ('caveesh' and 'cavees') likely drop off as listeners plan their response and look to the relevant texture buttons, analyses near the end of the analysis window likely reflect looks from fewer trials. Furthermore the English listeners tended to respond earlier than the French listeners. To confirm that our late occurring language by context interaction is not an artefact of this we performed the same analysis but this time time-aligning trials with the mouse click. This ensures that at a given time point, all looks are the same time from their respective responses. This analysis again found a language group by context interaction that became significant between 400 and 500 ms before the response.

4. Effect of context sibilant: Preceding Context (‘shamal’-‘samal’)

As discussed above, the contrastive effect emerged relatively late for the ‘caveesh’-‘cavees’ trials and the earliest effect was anti-contrastive which may have been due to a temporary mis-parsing of the sibilant sequence. In light of this pattern, the ‘shamal’-‘samal’ trials are interesting. Since the context word comes first, the listener already knows the identity of the first sibilant in the sequence when the target sibilant is heard – which should reduce ambiguity. Indeed, as shown in Figure 7 panel B, the effect of context was only contrastive (solid and dashed black lines are above solid and dashed grey lines). However, the effect again emerged relatively late, (time: 800-900, β =0.28, SE =0.14, t = 2.00, p = .045, Figure 7 panel A dotted grey line with squares) near the end of the target word. Unlike in the mouse click data there was an interaction between the CoG step of the target sibilant and the effect of the context sibilant between 1100 and 1600 ms (time: 1100-1200, β =0.14, SE =0.07, t = 2.26, p = .024, Figure 7 panel A solid grey line with diamonds). Inspection of panels C and D in Figure 7 reveals that looks in the context of ‘tamash’ were generally more biased towards ‘samal’ (bias greater than zero), especially for the steps in the center of the continuum. This is a different pattern than was observed in the ‘caveesh’-‘cavees’ trials in which looks were less biased to either object (closer to 0 bias) in the context of ‘shinnow’.

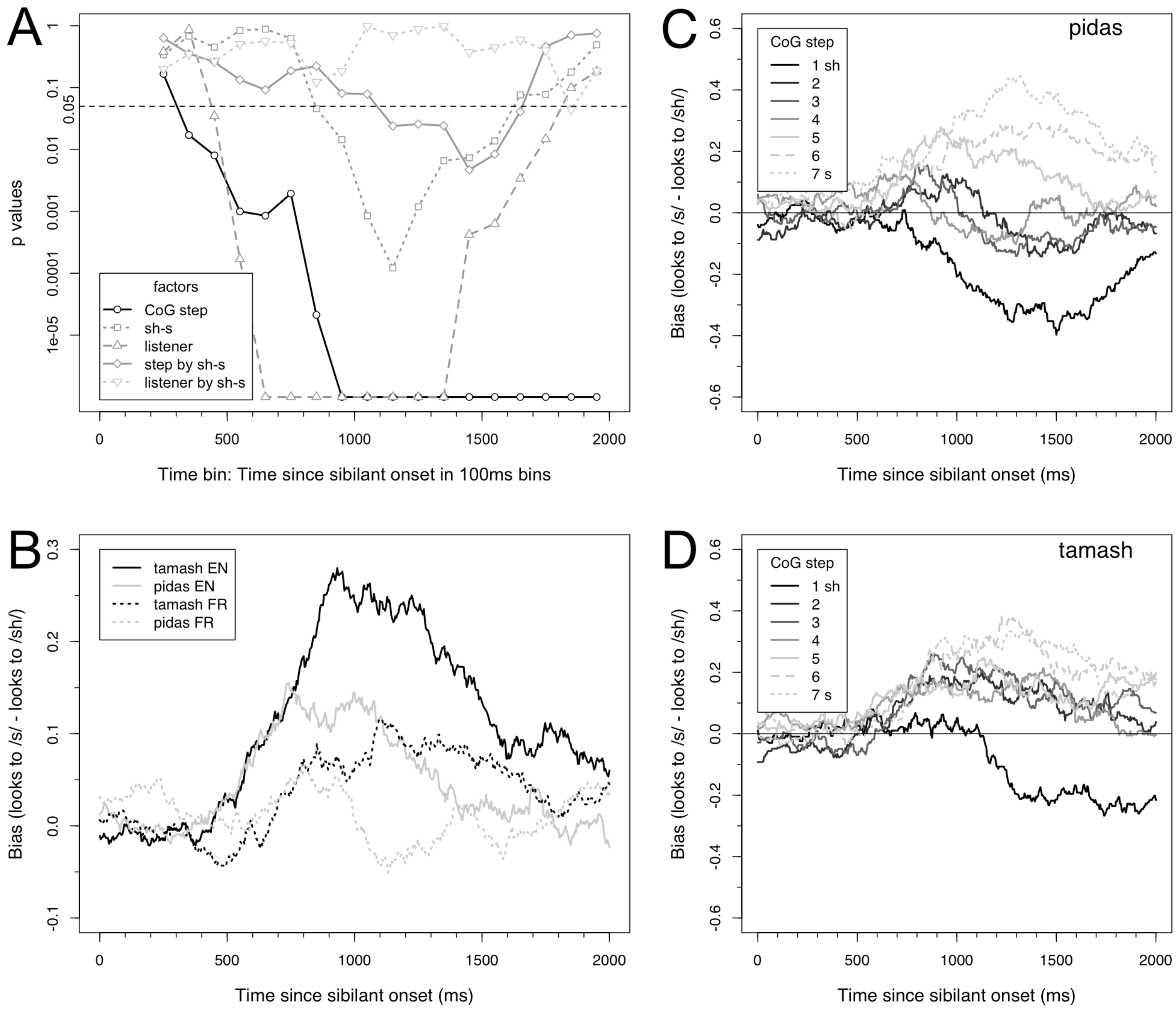

Figure 7: Eye-movements from 'shamal-'samal' trials. A Estimated p-values of key factors of the regression models for each time bin (first bin is 200-300 ms, etc). B Interaction of /ʃ/-/s/ context and listener group. C Effect of CoG step in the context of 'pidas'. D Effect of CoG step in the context of 'tamash'.

5. Summary of eye-movement data

The same pattern of results as seen in the mouse click data, showed up in the eye-movement record. This included contrastive context effects, an increase in ambiguity when the /ʃ/ context followed the target and a language by context interaction. We also found some patterns that were not in the mouse click data.

Figure 8 summarizes the time-course. Listeners' eye-movements showed very rapid effects of the acoustic signal, with the CoG of the target sibilant influencing the probability of looking at the two candidate target objects as early as possible (300-400 ms after sibilant onset). The latency of this effect (in line with previous studies) provides a baseline for evaluating the speed of other effects. In the case where the context followed ('caveesh'-'cavees') a similarly early anti-contrastive (target is perceived *more* like the context) effect (300-400 ms) was observed, possibly due to segmentation ambiguity. This was followed by reduced sensitivity to the CoG continuum in the /ʃ/ context relative to the /s/ context (500-600 ms), the equivalent of flattening in the mouse click data, as predicted by perceptual integration. Finally, a late language-specific context effect was observed (1800 ms). This was due to the English listeners alone resolving the ambiguity caused by the assimilating context in a way consistent with phonological inference. Note that while this effect became significant relatively late, the trend in looks was observable as early as 1000 ms post sibilant onset.

When the context preceded the target ('shamal'-'samal') the segmentation ambiguity should have been reduced, and indeed there is no evidence of an early anti-contrastive effect. However, the contrastive effect (target is perceived *less* like the context) is also late (800-900ms) and there is no effect of context early on. No interactions between context and listener group were observed.

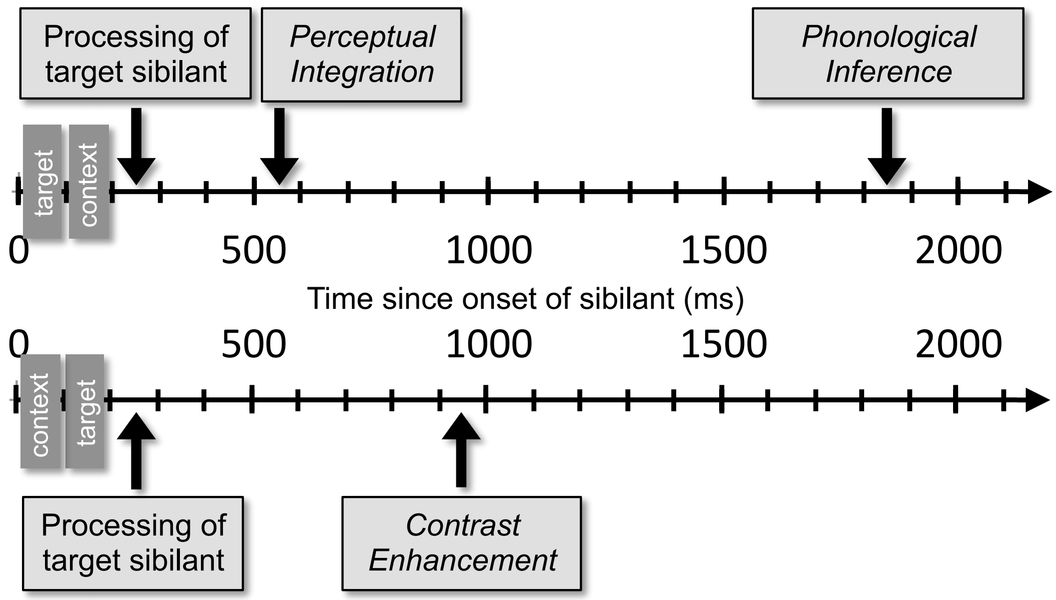

Figure 8: Time course of the three mechanisms according to when they influenced eye-movements. The top timeline is for contexts that follow the target (regressive assimilation) and the bottom timeline is for contexts that precede the target (progressive assimilation).

IV. DISCUSSION

The goal of this paper was to compare context-dependent compensation for assimilation in two groups of listeners with different linguistic backgrounds. For the first time, we were able to equate for the two groups: the task, the phonetic properties of the stimuli and the listeners' lexical knowledge in order to provide an unbiased measure of the strength of compensation effects in the two languages. We looked at effects of both following and preceding context on sibilant place assimilation. We also compared the time course of each of the effects we observed.

Our results suggest that multiple mechanisms have some role to play, but with a strong dependence on the nature of the ambiguity and the contextual circumstances.

First, when the phonetic context preceded the target sibilant (‘shamal’-‘samal’), a contrastive effect was found (fewer /s/ responses in the context of ‘pidas’ than ‘tamash’). This was restricted to ambiguous stimuli. No language-specific effects were found. This pattern is consistent with accounts that posit language general mechanisms such as contrast enhancement and feature parsing. Interestingly, previous studies with the strongest evidence for an auditory locus were restricted to effects of preceding context (see Holt & Kluender 2000 for a review; Kingston, Kawahara, Mash, & Chambless, 2011). One study comparing preceding and following contexts found a contrastive effect only for preceding contexts (Kingston, Kawahara, Chambless, Key, & Watsky, 2008). Our results support the notion that such a low-level mechanism could help listeners deal with small amounts of co-articulation carried over from a previous segment. It should be noted however, that this effect was not observed as early as other effects clearly driven by the acoustics of the signal. It therefore remains unclear whether this was due to an additional early auditory effect in the opposite direction (canceling it out), if the effect was delayed by some aspect of the signal (e.g. ambiguity), or whether a non-auditory locus (e.g. learned co-variation) is a better explanation.

In the following context condition (‘caveesh’-‘cavees’) we found a strikingly different pattern of results. The eye-movement data revealed that the earliest context effect was anti-contrastive (more /s/ responses in the context of ‘sival’ than ‘shinnow’). Its speed, and the fact that it was not affected by language background suggests that is also has an auditory locus. Mitterer (2006) observed a similar effect when sibilants were followed by a steady state sine wave mimicking F3, further supporting the idea that it is auditory in nature.

The anti-contrastive effect was followed by a second language-general effect in which the CoG of the target sibilant had a weaker effect on looks in the context of 'shinnow' than in the context of 'sival'. In the mouse click data, this manifested as a shallower categorization function for the CoG continuum in the 'shinnow' context, similar to what was reported for Hungarian liquid assimilation (Mitterer, Csépe & Bloomert 2006). In the case of Hungarian, both Hungarian and Dutch listeners found it more difficult to categorize a liquid-trill continuum when it was followed by /r/ than when it was followed by a labial. Hungarian listeners also responded in a similar way when the continuum was followed by a non-speech analog of /r/ suggesting that the locus may be auditory and work through some kind of backwards masking. The flattening result of the sibilant continuum has recently been replicated with a different set of participants and stimuli (Fleischer, Wagner & Clayards 2013).

Finally, we found evidence for language-specific processing. As listeners made their final decisions about the target sibilant, eye-movements revealed a late language-specific effect (interaction of language group and /s/ vs. /ʃ/ context) which also showed up in the mouse-click responses. The language-specific effect was the result of English but not French listeners perceiving the target more often as /s/ in the assimilating context.

Taken together these results suggest an intriguing relationship between auditory and knowledge-driven mechanisms when the assimilation context follows. Auditory mechanisms such as backwards masking (called perceptual integration by Mitterer and colleagues) may decrease the salience of a perceptual contrast in the context of the masker, thus increasing ambiguity for the listener. Those listeners who have experience with extensive context-dependent assimilation, are then able to resolve this ambiguity, as proposed by phonological inference. It is interesting to note that the masking was only observed when the context followed[7]. Our production study

(Niebuhr et al., 2011) also found more assimilation when the context followed, especially for English, and this is also a cross-linguistic tendency (Steriade, 2001). This may be due in part to a preference to preserve word onsets over word offsets. However, it also supports the hypothesis that languages allow assimilation (or other kinds of neutralization) in contexts where a phonological contrast is already weak (Mitterer, Csépe, Honbolygo & Bloomert, 2006, Steriade, 2001), in this case the /s/-/ʃ/ contrast is particularly weak when followed by /ʃ/.

Most importantly, however, in situations where strong assimilation creates complete or near-complete changes in the phonemic identity of consonants (namely, English regressive assimilation) none of the auditory accounts is sufficient. In this circumstance we found the clearest language-specific effects, supporting a phonological inference model (Gaskell, 2003). The current results provide a nice delineation of the conditions in which phonological inference is found. Although French shows more evidence of assimilation of /s/ to [ʃ] than English when /ʃ/ precedes /s/ (e.g. 'tache super'; Niebuhr, Clayards, Lancia, & Meunier, 2011), these assimilations are mostly partial and do not lead to language-specific compensation, presumably because auditory mechanisms can resolve the ambiguity in the vast majority of cases. In contrast, English, more than French, exhibits strong regressive assimilations of /s/ to [ʃ]. The ambiguity in this context cannot be completely resolved via contrast enhancement or perceptual integration. In these circumstances it seems that listeners learned to compensate for the increased ambiguity based on their experience of similar ambiguities. This compensation requires access to higher-order knowledge, and consequently operates over a longer timescale.

It should be acknowledged that the description just sketched out is fairly complex, especially for perceiving the target sibilant in the context of a following sibilant, as are the data. Thus it remains possible that alternative explanations exist. For example, we have described the language-specific effect of context as a re-interpretation of the stimuli on the part of the English listeners, once they have identified the potentially assimilating context. However other mechanisms could lead to the same pattern of results. For example, it may be that the English listeners shift their processing of the target in the presence of the ‘sival’ context instead of the ‘shinnow’ context, leading to the same context-specific effect; or that the differences are due to processing of the two contexts with different time-courses rather than a shift in processing. Future studies could help untangle these possibilities, for example by varying aspects of the context stimuli.

We also found baseline differences between the two language groups that may have their origin in compensation for contextual effects. French listeners gave more /s/ responses than English listeners in the control condition of the following-context trials (‘caveesh’-‘cavees’) and more /ʃ/ responses than English listeners in the control condition of the preceding-context trials (‘shamal’-‘samal’). This pattern was also observed in pilot testing with a larger set of non-word continua and depended on the vowel context. In pilot testing, French listeners gave more /s/ responses when the vowel context was /i/ (as in ‘caveesh’-‘cavees’) than when it was /a/ (as in ‘shamal’-‘samal’) and English listeners showed the opposite pattern (more /s/ responses to /a/ than /i/). Re-analysis of the spectral CoG data from the control context of Niebuhr et al. (2011) revealed a small (non-significant) difference in the spectra of sibilants between /a/ and /i/ vowels for both language groups. In French the CoG of /s/ was slightly lower in the context of /i/ (8408 Hz vs. 8664 Hz) and in English it was slightly lower in the context of /a/ (9355 Hz vs. 9447 Hz). The effect of vowel context has been reported for English before (Soli, 1981) but the French pattern has not been confirmed. One possibility is that the baseline differences seen in the present study are the result of listeners compensating for this contextual effect in a language specific way. Interestingly, the main effect of listener group was very early for the ‘shamal’-‘samal’ data (400-500 ms) and also fairly early for the ‘caveesh’-‘cavees’ data (700-800 ms). Although the experiment was not designed to test for effects of adjacent vowels, these data suggest they may be having a relatively early and language-specific effect on the perception of the target sibilants in contrast to the relatively late (1800-1900 ms) language-specific effect of the assimilation (following sibilant) context. This discrepancy may be due to the nature of the two effects. The vowel context occurs within words and thus is likely a relatively robust characteristic of vowel-sibilant sequences, occurring frequently across the lexicon and in normal speech. In contrast, the assimilation context occurs only across a word boundary, thus it is not a feature of any given lexical item and should occur relatively infrequently in normal speech. For example, Dilley and Pitt (2007) estimated that word final stops and nasals occur in an assimilating context in 3% of words in a corpus of English conversational speech and are assimilated 9% of the time on average. Future research will need to determine whether either of these factors (within vs. across lexical items, frequency of occurrence) determine how quickly language-specific context effects influence processing. It is of interest to note that a previous study investigating the time course of language-specific phonetic processing within words also found an early effect (Kharlamov, Cambell & Kazanina, 2011).

V. CONCLUSION

Previous investigations of context-dependent compensation for assimilation have found support for conflicting accounts, both language-general and language-dependent. Our study was designed to look for both classes of effects to better understand their role in compensation for assimilation. We found evidence for both language-general and language-dependent mechanisms with each playing a complementary role. Classic contrast enhancement was restricted to the preceding context and could compensate for small amounts of context-dependent assimilation. An assimilating context following the target decreased the perceptibility of the target sibilant, increasing ambiguity. The English but not the French listeners also consistently reported hearing /s/ more often in this context reflecting their experience with extensive assimilation, presumably via language-specific knowledge.

The results of this study also expand considerably on the previous literature looking at the time course of context-dependent compensation. We were able to observe early effects that were clearly driven by auditory/phonetic processing as well as later effects that were language-specific. We suspect that this interplay between language-general and language-specific mechanisms characterises speech perception when faced with many types of assimilation or reduction. Swift, universal, auditory mechanisms do a good job of resolving ambiguity when changes are partial and leave residual phonetic cues. Slower mechanisms learned from the specific statistics of the language environment then operate when auditory mechanisms are insufficient and ambiguity is more complete.

FOOTNOTES

[1] Note however that this was only true when the frication portions of the sibilants were considered. Acoustic measurents of the surrounding vowels found small but reliable traces of the unassimilated segments, (Niebuhr & Meunier, 2011). These vowel differences, which can serve as acoustic cues to place of articulation, were not part of the stimuli used in the present study.

[2] Partially assimilated sibilants with intermediate frication of this type were observed in Niebuhr et al. (2008, 2011) and (Holst & Nolan, 1995). Tokens in which the [ʃ] portion of the sibilant sequence was longer than would be expected were also observed but this type of partial assimilation was not used here.

[3] Modified sibilants were 65 dB and the amplitude of the words they were spliced onto ("-amal" and "cavee-") was 70 dB. The amplitude of the first 4 pitch cycles of "-amal" was ramped slightly to prevent artifacts from a sudden

onset. The sibilants and the words were spliced together with a 5ms overlap to produce a natural sibilant-vowel transition.

[4] A few participants did both testing blocks on the same day but in those cases they were separated by at least a half hour break.

[5] This effect was not significant when the data from the English speaker was included but is not of interest.

[6] Sibilant context by Step interaction and the interaction between these two factors and Listener group were significant for the analysis including the data from both talkers. This was likely due to the fact that for the English listeners, all but the most [ʃ]-like of the English talker's target sibilants were classified as /s/. This produced a ceiling effect for the /ʃ/ context which made the effect of step different for this context in comparison to the others.

[7] Note that an interaction between CoG step and context was found in the eye-movement data for the 'shamal'-'samal' trials which could indicate some masking/perceptual integration for this context as well. However both the pattern and the time-course were somewhat different than in the 'caveesh'-'cavees' trials and we do not believe it is the same mechanism.

REFERENCES

Baayen, H., Piepenbrock, R., & van Run, H. (1993). *The Celex data base [CD-ROM]*. Philadelphia: Linguistics Data Consortium, University of Pennsylvania.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9), 341-345.

Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica, 49*, 155-180.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Speech perception reflects optimal use of probabilistic speech cues. *Cognition, 108*, 804-809.

Clayards, M., Fleischer, D. & Wagner, M. (submitted) Can syntactic structure influence compensation for assimilation? *Speech Communication*.

Coady, J. A., Kluender, K. R., & Rhode, W. S. (2003). Effects of contrast between onsets of speech and other complex spectra. *The Journal of the Acoustical Society of America*, *114*, 2225.

Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules. Perceptual compensation for assimilation in late L2-learners. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 411-442). Berlin: Mouton de Gruyter.

Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & Dupoux, E. (2009). Phonological knowledge in compensation for native and non-native assimilation. In F. Kügler, C. Féry & R. v. d. Vijver (Eds.), *Variation and Gradience in Phonetics and Phonology* (pp. 265-309). Berlin, New York: Mouton de Gruyter.

Diehl, R., Lotto, A., & Holt, L. (2004). Speech perception. *Annual Review of Psychology, 55*, 06.01-06.31.

Dilley, L. C., & Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America, 122*, 2340-2353.

Duyck, W., Desmet, T., Verbeke, L., & Brysbaert, M. (2004). WordGen: A tool for word selection and non-word generation in Ducht, German, English and French. *Behaviour Research Methods, Instruments & Computers, 36*(3), 488-499.

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition.*Journal of cognitive neuroscience*, *16*(4), 577-583.

Fleischer, D., Wagner, M., Clayards, M. (2013) A following sibilant increases the ambiguity of a sibilant continuum. *Proceedings of Meeting on Acoustics (POMA), Vol. 19.* doi: 10.1121/1.4800753

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics, 14, 3–28.

Gaskell, M. G. (2003). Modeling regressive and progressive effecs of assimilation in speech perception. *Journal of Phonetics, 31*, 447-463.

Gaskell, M. G., Hare, M., & Marslen-Wilson, W. D. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science*, 19, 407-439.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Expreimental Psychology: Human Perception and Performance, 22*, 144-158.

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 380-396.

Gaskell, M. G., & Snoeren, N. D. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1632-1647.

Gow, D. (2002). Does English coronal place assimilation create lexical ambiguity? *Perception and Psychophysics, 65*, 575-590.

Gow, D. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics, 65*(4), 575-590.

Gow, D., & Im, A. M. (2004). A cross-linguistic examination of asimilation context effects. *Journal of Memory and Language, 51*, 279-296.

Gow, D., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. In J. S. C. J. Hualdo (Ed.), *Papers in Laboratory Phonology 9* (pp. 173-200). New York: Mouton de Gruyter.

Gow, D. W., & Segawa, J. A. (2009). Articulatory mediation of speech perception: a causal analysis of multi-modal imaging data. *Cognition*, *110*(2), 222-236.

Holt, L. L., & Kluender, K. R. (2000). General auditory processes contribute to perceptual accommodation of coarticulation. *Phonetica, 57*, 170-180.

Holst, T., & Nolan, F. (1995). The influence of syntactic structure on [s] to [ʃ] assimilation. In B. Connell & A. Arvaniti (Eds.), *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence* (pp. 315-333). Cambridge: CUP.

Hura, S. L., Lindblom, B., & Diehl, R. L. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech, 35*(1-2), 59-72.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America, 108*, 1252.

Kharlamov, V., Campbell, K., & Kazanina, N. (2011). Behavioral and electrophysiological evidence for early and eutomatic detection of phonological equivalence in variable speech inputs. *Journal of Cognitive Neuroscience, 23*(11), 3331-3342.

Kingston, J., Kawahara, S., Chambless, D., Key, M., & Watsky, S. (2008). The independence of auditory and categorical effects on speech perception. *Poster presented at LabPhon11, Victoria University of Wellington*.

Kingston, J., Kawahara, S., Mash, D., & Chambless, D. (2011). Auditory Contrast versus Compensation for Coarticulation: Data from Japanese and English Listeners. *Language and Speech, 54*(4), 499-525.

Lotto, A., & Kluender, K. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception and Psychophysics, 60*(4), 602-619.

Lotto, A., Kluender, K., & Holt, L. (1997). Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). *The Journal of the Acoustical Society of America, 102*, 1134.

Marslen-Wilson, W., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human perception and performance, 22*(6), 1376.

McMurray, B., Clayards, M., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin and Review, 15*(6), 1064-1071.

Mitterer, H. (2006). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception and Psychophysics, 68*, 1227-1240.

Mitterer, H. (2011a). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 496.

Mitterer, H. (2011b). Recognizing reduced forms: Different processing mechanisms for similar reductions. *Journal of Phonetics*, *39*(3), 298-303.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimiliation in speech perception: Evidence for early compensation. *Perception and Psychophysics, 56*(6), 956-969.

Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *The Quarterly Journal of Experimental Psychology, 59*(8), 1395-1424.

Mitterer, H., Csepe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimliated words does not depend on specific language experience. *Cognitive Science, 30*(3), 451-471.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*(1), 73-103.

Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 244.

Mitterer, H., Yoneyama, K., & Ernestus, M. (2008). How we hear what is hardly there: Mechanisms underlying compensation for/t/-reduction in speech comprehension. *Journal of Memory and Language, 59*(1), 133-152.

Mitterer, H., & Reinisch, E. (2013). No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language*, *69*(4), 527-545.

Naatanen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*(1), 1-21.

New, B., Pallier, C., Segui, J., Ferrand, L., & Rastle, K. (2004). *Lexique* 2: A new French lexical database. *Behaviour Research Methods, Instruments & Computers, 36*, 516-524.

Niebuhr, O., Clayards, M., Lancia, L., & Meunier, C. (2011). On place assimilation in sibilant sequences - comparing French and English. *Journal of Phonetics, 39*, 429-451.

Niebuhr, O., & Meunier, C. (2011). The phonetic manifestation of French /s#ʃ/ and /ʃ#s/ sequences in different vowel contexts: On the occurrence and the domain of sibilant assimilation. *Phonetica, 68*(1-28).

Niebuhr, O., Meunier, C., & Lancia, L. (2008). On place assimilation in French sibilant sequences. *Proceedings of the 8th International Seminar on Speech Production (ISSP08)* 8-12 December 2008, Strasbourg, France, 2221-2224.

Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. In D. R. L. G. J. Docherty (Ed.), *Papers in Laboratory Phonology 2* (pp. 261–280). Cambridge: CUP.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204-238.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrik & M. F. Miller (Eds.), *Papers from the Parasession on Language and Behavior* (pp. 178-203). Chicago: Chicago Linguistic Society.

Ohala, J. J. (1993). Coarticulation and phonology. *Language and Speech, 36*(2-3), 155.

Ohala, J. J. (Ed.). (1990). *The phonetics and phonology of aspects of assimilation*. Cambridge: Cambridge University Press.

Pitt, M. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language, 61*, 19-36.

Pouplier, M., Hoole, P., & Scobbie, J. M. (2011). Investigating the asymmetry of English sibilant assimilation: Acoustic and EPG data. *Laboratory Phonology*, *2*(1), 1-33.

Revill, K. P., Tanenhaus, M. K., & Aslin, R. N. (2008). Context and spoken word recognition in a novel lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1207.

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Listening to different speakers: On the time-course of perceptual compensation for vocal-tract characteristics. *Neuropsychologia*.

Skoruppa, K., Mani, N., & Peperkamp, S. (2013). Toddlers' processing of phonological alternations: Early compensation for assimilation in English and French. *Child development*, *84*(1), 313-330.

Slis, I. H., & Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction. *Language and Speech, 12*(2), 80-102.

Snoeren, N.D. (2011). Psycholinguistique Cognitive de la Parole Assimilée . *Editions Universitaires Européennes.*

Snoeren, N. D., & Gaskell, M. G. (2008). Computational Modeling of Assimilated Speech: Cross-Linguistic Evidence. In D. S. McNamara & J. G. Trafton (Eds.), *29th Annual Conference of the Cognitive Science Society* (pp. 1509-1514). Nashville TN: Cognitive Science Society.

Snoeren, N. D., Gaskell, M. G., & Di Betta, A. M. (2009). The perception of assimilation in newly learned novel words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 542-549.

Snoeren, N. D., Segui, J., & Halle, P. A. (2008a). On the role of regular phonological variation in lexical access: Evidence from voice assimilation in French. *Cognition, 108*(2), 512-521.

Snoeren, N. D., Segui, J., & Halle, P. A. (2008b). Perceptual processing of partially and fully assimilated words in French. *Journal of Experimental Psychology, 34*(1), 193-204.

Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative vowel coarticulation. *The Journal of the Acoustical Society of America, 70*, 976.

Sonderegger, M., & Yu, A. C. L. (2010). A rational account of perceptual compensation for coarticulation. In S. O. R. Catrambone (Ed.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 375–380). Austin, TX: Cognitive Science Society.

Steriade, D. (2001). Directional asymmetries in place assimilation. In K. J. E. Hume (Ed.), *The Role of Speech Perception in Phonology* (pp. 219-250). London: Academic Press.

Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Attention, Perception, & Psychophysics, 62*(6), 1297-1311.

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics, 18*, 3-35.
Whalen, D. H. (1991). Perception of the English /s/-/sh/ distinction relies on fricative noises and transitions, not on brief spectral slices. *The Journal of the Acoustical Society of America, 90*, 1776-1774.
Yu, A. C. L. (2010). Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PloS one, 5*(8), e11950.