Journal of Phonetics 58 (2016) 118–134

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English

Mary Stevens *, Jonathan Harrington

Institute for Phonetics and Speech Processing (IPS), Ludwig Maximilians Universität, Schellingstr. 3, II, 80799 München, Germany

ARTICLE INFO

Article history:
Received 22 December 2015
Received in revised form
3 August 2016
Accepted 7 August 2016

Keywords:
Australian English
Coarticulation
Sibilant acoustics
Sibilant perception
Sound change
/s/-retraction

ABSTRACT

In contemporary spoken English, /s/ can resemble a post-alveolar fricative when it occurs in /str/ clusters e.g. *street*. /s/-retraction in /str/ is known to be widespread in North American English, but the question of how this sound change comes about has attracted only a very small amount of empirical research. This paper investigates the phonetic pre-conditions for /s/-retraction based on the results of two experiments conducted with Australian English. Study 1 shows that the first spectral moment (M1) for /s/ is lower in /spr, str, skr/ and slightly lower in /sp, st, sk/ than in pre-vocalic position. Temporal variation in first spectral moment trajectories suggests that the articulatory movements associated with the rhotic are timed relatively earlier in /str/ vs. /spr, skr/ clusters. Study 2 tested native listener categorization of sibilants produced by multiple talkers. Results show that sibilants originally produced in *stream* and *steam* elicit /ʃ/ responses when spliced into pre-vocalic contexts. There was an interaction with talker gender, with male voices eliciting more /ʃ/ responses. Thus, the conditions for sound change to occur by which /s/ becomes /ʃ/ originate in a synchronic bias for /s/ produced in /sC(r)/ clusters to show increased auditory similarity to /ʃ/. Australian English shows the pre-conditions for /s/-retraction but is not currently undergoing this sound change.

1. Introduction

Synchronic variation for many varieties of English includes the tendency to produce /str/ with a retracted sibilant (e.g. Lawrence, 2000; Shapiro, 1995), which means that the sibilant in *street*, for example, sounds like that in *sheet.* Such auditory similarity to canonical /ʃ/ is typically attributed to the coarticulatory effect of the upcoming rhotic, which in English involves an alveolar or retroflex approximant that can also be strongly rounded. Both tongue tip retraction and lip rounding would lead to a larger oral cavity in front of the articulatory constriction for /s/, increasing the acoustic and auditory similarity post-alveolar /ʃ/. Descriptive evidence suggests that retracted /str/ pronunciations are especially common in North America where they are the norm for many speakers and in several regional varieties (e.g. Philadelphia, Labov, 2001; New York, Kraljic, Brennan, & Samuel, 2008; southwest Louisiana, Rutter 2011), but /s/-retraction in /str/ has also been reported for speakers of New Zealand English (Warren, 1996; Lawrence, 2000) and British English (Cruttenden, 2014). The sound change /str/>/ʃtr/ is not confined to English: /s/ in /str/ has undergone retraction in the historical development of a number of other languages including Standard German and several regional varieties of Italian (e.g. Rohlfs, 1966 for Italian varieties; see Kümmel (2007:232) for a cross-linguistic overview of /s/-retraction).[1] In terms of experimental phonetic evidence of retraction in /str/, there have been two acoustic phonetic studies, both with American English speakers (Baker, Archangeli, & Mielke, 2011; Rutter, 2011). The results of these studies have shown that the sibilant in /str/ has a lower first spectral moment/spectral peak than the same speaker's pre-vocalic /s/. Rutter's (2011) study of Louisiana English documented what happens when /s/-retraction has progressed to completion in a speech community: speakers produce the sibilant in /str/ so that it is acoustically within the range of their own canonical /ʃ/. The earlier stages of this /str/>/ʃtr/ sound change have not been investigated in

* Corresponding author.
E-mail addresses: mes@phonetik.uni-muenchen.de (M. Stevens), jmh@phonetik.uni-muenchen.de (J. Harrington).

[1] Descriptive evidence for historical cases of /s/-retraction shows that it was not confined to /str/ (cf. Kümmel, 2007:232), and the phonetic nature of the rhotic in other languages in which /s/ has undergone retraction often differs from that in English (e.g. Standard German [ʁ]), yet it is the phonetic nature of the rhotic which is understood to be the primary driving force for retraction in /str/ in English. Therefore, the phonetic conditions that brought about other historical cases of /s/-retraction may not necessarily have been the same as those responsible for /str/-retraction in contemporary varieties of English.

0095-4470/$ - see front matter © 2016 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2016.08.003

similar detail, and it is not yet clear how such fully retracted /ʃtr/ pronunciations actually come about. One of the main questions is whether, within the individual, /s/-retraction involves an abrupt categorical change from /str/>/ʃtr/ or whether it progresses along a continuum of phonetically intermediate variants over time. At first glance, the results of Baker et al. (2011) support the idea that /s/-retraction involves a phonetically gradient sound change: all speakers in that study produced sibilants in /str/ such that mean centroid frequencies were intermediate between those of /s/ and /ʃ/ (with the relative proximity to /s/ or /ʃ/ differing between speakers). The existence of such intermediate variants favours the idea that retraction is a gradient process. However, the nineteen American English speaker participants in Baker et al.'s study were from heterogeneous regional backgrounds, having grown up in fifteen different states between them. For this reason, it is not possible to know whether differences between their /str/ pronunciations should be attributed to idiosyncratic or regional variation in American English. More importantly, given /s/-retraction is already very common in many North American English varieties, it is not clear whether evidence of retracted sibilants in Baker et al.'s production data represents the pre-conditions for sound change or the effects of a sound change that is already underway.

The present study aims to shed light on the phonetic pre-conditions which can give rise to the sound change involving /s/-retraction. To do so, we investigate sibilant production and perception in a variety of English that is not undergoing /s/-retraction. In doing so we draw on Ohala's (1981, 2012) model of the origins of sound change, which is founded on two premises: (a) there are biases in speech production that are common to all languages at all times; (b) these biases can affect individual listener perception and can culminate in permanent change to a language's sound system. In other words Ohala makes a direct link between synchronic variability and diachronic change (see also Harrington (2012) and references therein), which is supported by a wealth of experimental evidence (e.g. Beddor, 2009; Solé, 2010). While Ohala did not address the sound change involving /s/-retraction in particular, his model allows two predictions about sibilants in /str/ in varieties that are not undergoing (or have not undergone) /s/-retraction:

(1) Sibilants in /str/ should be slightly retracted in production compared to canonical /s/
(2) Retraction in /str/ will be audible to listeners; e.g. out of context, listeners should categorize the sibilant as /ʃ/

The present study tests both of these predictions based on production and perception experiments with Australian English. It constitutes the first phonetic study on /s/-retraction in Australian English and only the second for a variety of English spoken outside of North America (after Warren (1996), a small-scale acoustic analysis of /str, stj/ in New Zealand English). Australian English was chosen because, to the best of our knowledge, the canonical pronunciation of /str/ involves an alveolar sibilant in this variety. Descriptive evidence suggests that /s/-retraction in /str/ is, however, an idiosyncratic tendency for some Australian English speakers: in her pedagogical text Cox (2012:128) states "/stɹ/ onsets for some speakers are heavily assimilated […] This results in [ʃtɹ]…".[2] At this stage it is not clear whether such idiosyncratic retraction constitutes stable low-level phonetic variation (i.e. the pre-conditions for a sound change) or whether /str/ might already be undergoing a sound change in Australian English. The results of the present empirical study will help to shed some light on this issue by documenting the degree to which /s/-retraction is a shared tendency for Australian English speakers.

Returning to the origins of /s/-retraction, the first prediction listed above raises the question as to *why* there should be a bias towards retraction in the production of sibilants in /str/ contexts in languages and varieties not yet undergoing this sound change. As noted earlier, the auditory similarity between sibilants in /str/ and canonical /ʃ/ is typically assumed to reflect coarticulation with the rhotic. Few sources on /s/-retraction are more precise about the phonetic motivation for this sound change. Kraljic et al. (2008:56) presume that tongue placement for /s/ differs in /str/ contexts such that "/s/ is articulated with a retracted tongue position, anticipating the /r/". Rutter (2011) suggests that there may be articulatory change to the lips as well as the tongue and that because /ʃ/ and /r/ share a similar lip shape in English "a change from /s/ to /ʃ/ […] in the context /ɹ/ could be seen as a harmonizing process". While differing on exactly which articulators are involved in anticipation of the rhotic, both Kraljic and Rutter assume that /s/-retraction involves long-distance assimilation with the upcoming rhotic (also in line with many others e.g. Shapiro, 1995). Cruttenden (2014:202) suggested that /s/-retraction in /str/ is not a long-distance assimilation process but rather that the plosive /t/ – as well as the sibilant - is retracted. Similarly Lawrence (2000) also suggested that the plosive /t/ is retracted in /str/ clusters, but for different reasons: he drew a link with a separate sound change involving /tr/. Post-alveolar affricate pronunciations of /tr/ are, according to Lawrence (2000), common for English speakers, such that the onsets of *train* and *chain* are pronounced with the same sound. There is surprisingly little discussion of the sound change /tr/>[tʃɹ] in the literature, but *Gimson's Pronunciation of English* (Cruttenden, 2014:192) suggests that affrication of /tr/ is typical (leading to foreign learner confusion of words like *trees* and *cheese*).[3] Lawrence (2000) proposed that affricated /tr/ pronunciations would extend to /str/ contexts, such that the plosive in /str/ would also involve a post-alveolar affricate target. In other words, Lawrence suggests that retraction in /str/ is a secondary development following a separate sound change affecting /tr/ in English. Neither of these two hypotheses about the articulatory motivations for /s/-retraction in /str/ (i.e. long-distance assimilation to /r/ vs. assimilation to retracted /tr/) has been tested empirically. Moreover, they make different predictions about varieties not yet undergoing /s/-retraction, i.e. about the production biases that can develop into sound change. If /s/-retraction involves long-distance assimilation to an upcoming rhotic, then a pre-existing bias towards retraction should also apply to sibilants in /spr, skr/. If, on the other hand, /s/-retraction involves assimilation to an already retracted /t/ target, then any pre-existing bias towards retraction should be confined to /str/ clusters. There is some acoustic evidence (Baker et al., 2011) that American English speakers

[2] Audible retraction appears to be confined to /str/ in Australian English because /skr, spr, st, sk, sp/ were all transcribed with an alveolar sibilant in that source.

[3] On the other hand, regarding the affrication of /tr/ in English John Wells states "I don't believe there is any such phonological change in progress" http://phonetic-blog.blogspot.de/2011/03/how-do-we-pronounce-train.html (Wells, 2011).

produce retracted sibilants in /spr, skr, sp, st, sk/, and descriptive evidence suggests that retracted pronunciations in these other cluster contexts can be audible to native listeners e.g. in *school* (Janda & Joseph, 2003). However, for the reason just outlined for /str/ above, we cannot know whether such retracted /spr, skr, sp, st, sk/ pronunciations in American English are due to a pre-existing bias or to a sound change that is already in progress. With this issue in mind, Study 1 seeks to determine whether the pre-conditions for /s/-retraction include a bias towards retraction in the production of /str/ and whether this bias might extend to /spr, skr/ and /sp, st, sk/ as well.

If an articulatory bias towards /s/-retraction is to bring about any change to the phonological system, then it must be audible to native listeners. There is some evidence that an adjacent voiceless stop can influence the spectral properties of the sibilant in /sp, sk/ sequences (e.g. Stevens, 1998) and that such spectral differences are audible to listeners (Engstrand & Ericsdotter, 1999). In a small-scale perception study, Engstrand and Ericsdotter spliced sibilants from /sp, st, sk/ and found that listeners ($n=10$) could reliably identify the identity of the upcoming plosive based on the sibilant noise alone. This shows that the place of articulation of an upcoming voiceless stop is encoded in the sibilant noise and that listeners can use this information if necessary. It is not yet clear whether such perceptual sensitivity extends to /s/-retraction: Engstrand and Ericsdotter's study did not include pre-vocalic sibilants or mention whether sibilants were retracted in /sp, st, sk/. Study 2 tests whether the acoustic effects of articulatory retraction are audible to listeners i.e. the second prediction listed above.

In addition to documenting the conditions that potentially give rise to /s/-retraction, the present research aims to make two further contributions to our knowledge of this sound change. First, this study considers /s/-retraction in terms of both static and dynamic parameters. Previous acoustic studies (Baker et al., 2011; Rutter, 2011) have addressed retraction based on static parameters alone: Baker et al. (2011) averaged the first spectral moment (M1) over the temporal middle half of each sibilant token, and Rutter (2011) measured the frequency location of the amplitude peak. While static parameters like mean M1 can reliably distinguish pre-vocalic /s/ and /ʃ/ in English (e.g. Jongman, Wayland, & Wong, 2000; Koenig, Shadle, Preston, & Mooshammer, 2013), research has shown that dynamic acoustic parameters are more effective for capturing contextual differences for sibilants (Haley, Seelinger, Mandulak, & Zajac, 2010; Iskarous, Shadle, & Proctor, 2011; Koenig et al., 2013). Moreover, dynamic attributes of the speech signal can play an important role in perception (e.g. Lindblom, 2004; Divenyi, Greenberg, & Meyer, 2006). These are especially relevant considerations for our study on /s/-retraction, which is a form of coarticulation and should therefore be considered in terms of how it unfolds over time.

Second, this paper addresses the impact of speaker gender on sibilant production and perception in cluster contexts. Sibilants /s/ and /ʃ/ produced by women have higher frequencies than those produced by men (cf. e.g. Koenig et al., 2013:1178 and references therein; Munson, McDonald, DeBoe, & White, 2006). Here we investigate whether there is an interaction between speaker gender and the tendency to retract sibilants in pre-consonantal contexts. The coarticulatory motivation for /s/-retraction should apply equally to male and female talkers, which means that we do not expect to find any differences for gender. However, sociolinguistic research has shown that women lead certain types of sound change, namely sound changes 'from below' (Labov, 2001). Such sound changes involve novel pronunciations that can be attributed to language-internal phonetic pressures rather than external factors such as language contact (cf. e.g. Milroy & Milroy, 1985 on internal vs. external factors in language change). /s/-retraction is a phonetically motivated sound change, which fits the description of a 'sound change from below'. As such, if this sound change is underway in Australian English, we might expect women to produce more /s/-retraction than men. This prediction is consistent with e.g. Clopper and Pisoni (2005:323) who, summarizing the interaction between gender and sound change suggest "speech stimuli from females might be expected to reveal current changes in progress".

2. Production study

Study 1 tests whether the pre-conditions for /s/-retraction include a bias towards retraction in /str/ and whether this bias might extend to /spr, skr/ and /sp, st, sk/ as well.

Tablee 1
Overview of the production data. See Appendix A for the list of target words.

Cluster	Sibilant	Following Segment	Rhotic	No. tokens
/s/	s	/iː, æɪ/	No	397
/st/	s	/t/	No	197
/str/	s	/t/	Yes	200
/sp/	s	/p/	No	199
/spr/	s	/p/	Yes	200
/sk/	s	/k/	No	199
/skr/	s	/k/	Yes	199
/ʃ/	ʃ	/iː, æɪ/	No	398
			Total	**1989**

2.1. Speakers and materials

The production data were collected in Braidwood, New South Wales, a rural heritage town with about 1000 residents which lies about 300 km southwest of Sydney. Braidwood can be taken to represent Australian English; it was chosen as the recording site because personal contacts allowed access to a group of participants with similar linguistic and geographical backgrounds. Braidwood is a close-knit community of mostly monolingual English speakers and affords much less exposure to languages or varieties other than Australian English compared with life in large Australian cities. Such homogeneity between participants' linguistic experiences is important both because we want to document phonetic variation that is typical for a speech community before /s/-retraction takes hold, and because we want to avoid – as far as possible - inter-speaker variation due to external factors such as contact with retracting varieties.

Production data were all collected in the same private home using a headset microphone and a MacBook Pro installed with the SpeechRecorder software (Draxler & Jänsch, 2004). The task for participants involved reading a set of English words aloud in the carrier phrase "*Any*_____". The set of words comprised ten target words with word-initial sibilants and twelve fillers (listed in Appendix A). These words were presented to participants in randomized order on a laptop computer screen with ten repetitions for each word item. This resulted in 20 speakers × 10 repetitions × 10 target words = 2000 sibilant tokens; eleven of these were excluded because of background noise, leaving 1989 for analysis (cf. Table 1). The experiment was self-timed: after participants produced a word item, they (or the experimenter) clicked the mouse button so that the next word item appeared. If a participant produced a different word from the one on the screen, that particular word item was repeated and the older file was overwritten. We report on production data for twenty Braidwood residents who fulfilled the following criteria:

- Born and schooled in Australia and has not lived outside of Australia.
- Resident in Braidwood.
- Monolingual English speaker.
- No speech or hearing problems.

The participants were mostly long-term residents of Braidwood who knew each other, some talking on a daily basis. They comprised seven males (age range 36–48; mean years resident in Braidwood = 11) and thirteen females (age range 29–49; mean years resident in Braidwood = 13). Participants were paid for completing this experiment.

The onset and offset of the sibilant fricative were corrected manually where necessary using Emu Labeller (Winkelmann & Raess, 2014). Sibilants were coded according to segmental context as shown in Table 1.

2.2. Acoustic parameters

The recordings were down-sampled to 32000 Hz and labelled semi-automatically with the Munich Automatic Segmentation System for Australian English (Kisler, Schiel, & Sloetjes, 2012).

As noted in the introduction we analysed sibilants in two ways: in terms of the entire M1 trajectory and in terms of mean M1 (averaged over the temporal middle half). We obtained M1 data using the *emuR* package as follows. We calculated a power spectrum for each sibilant token using DFT with a 40 Hz frequency resolution, a 5ms Blackman window, and a 5ms frame shift. We then calculated M1 on these power spectra throughout the duration of each sibilant token between 500 and 15000 Hz (with the lower limit set at 500 Hz to prevent coarticulatory voicing from influencing M1 values). These M1 trajectories were linearly time-normalized for analysis. For each sibilant, we also calculated mean M1 by averaging M1 over the temporal middle half (averaging between 25%-75% of each sibilant's duration).

The degree of /s/-retraction in /sCr, sC/ clusters (where C = /p, t, k/) was determined by calculating their relative position acoustically between pre-vocalic /s/ and pre-vocalic /ʃ/. For this purpose, we calculated for each fricative the log. distance ratio (see also Bukmaier, Harrington, and Kleber (2014) and Harrington, Kleber, and Reubold (2008)) parameterized as d_{fric} in (1):

$$d_{fric} = log\left[\left(M_{fric.j} - \overline{M_{\int.J}}\right) / \left(M_{fric.j} - \overline{M_{s.J}}\right)\right] \quad (1)$$

where $M_{fric.j}$ is the first spectral moment (averaged over the 25–75% interval) of any given fricative token produced by speaker j and where $\overline{M_{\int.J}}$ and $\overline{M_{s.J}}$ are the mean M1 values across all /ʃ/ and /s/ tokens respectively in prevocalic /ʃV, sV/ contexts produced by the same speaker. When d_{fric} is zero in (1), then a given fricative token is exactly intermediate between these speaker-specific mean M1 values for /ʃV, sV/; when d_{fric} is positive, then the fricative is acoustically closer to /s/ in /sV/ and when d_{fric} is negative, then it is acoustically closer to the same speaker's /ʃ/ in /ʃV/. The hypothesis to be tested was that d_{fric} for /s/ in /sCr, sC/ was lower (i.e. closer to the same speaker's /ʃ/) than the same speaker's /s/ in /sV/: this would be evidence for a greater degree of retraction or at least for a greater degree of similarity in the direction of /ʃ/ for /s/ in /sC(r), sC/ clusters in comparison with the same speaker's pre-vocalic /s/ in /sV/.

The statistical analysis was run by fitting linear mixed models within the R package *lme4* to two separate data subsets: (1) sibilants in rhotic clusters /str, spr, skr/ and pre-vocalic /s/; and (2) sibilants in /st, sp, sk/ and pre-vocalic /s/. We split the data in this way in order to be able to address the effect of rhotic /sCr/ and non-rhotic /sC/ contexts on /s/ separately (in Sections 2.3.1 and 2.3.2). This procedure enabled us to establish the degree of M1-lowering in contexts where it can be explained in purely phonetic terms (the effect of the upcoming /r/) before turning to /sC/ contexts in which a purely phonetic explanation for M1-lowering is less apparent. In

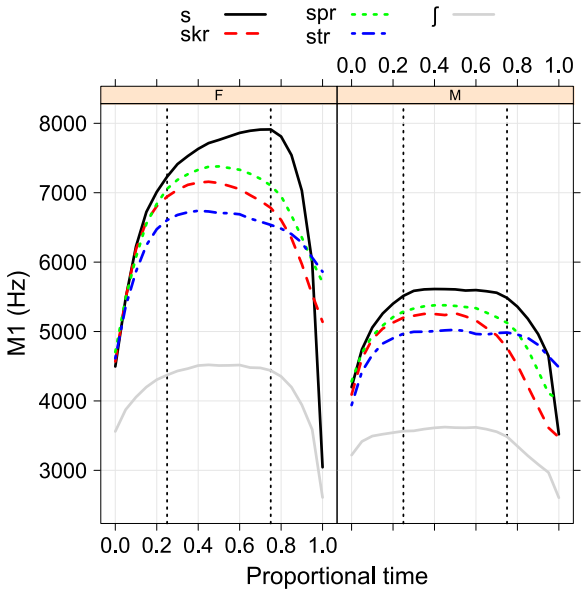

Fig. 1. M1 trajectories for /s/ in pre-vocalic position and in the clusters /spr/, /str/, /skr/, separately for female (F) and male (M) speakers. Data for pre-vocalic /ʃ/ are also plotted (in grey). Tracks were averaged across speakers after averaging within context and speaker. The dotted vertical lines indicate the temporal middle half over which mean M1 was calculated.

both models the log. distance ratio value (d_{fric} in (1)) was the dependent variable, Context (prevocalic vs. pre-consonantal) and Gender were fixed factors and Speaker (20 levels) and Word (5 levels) were random factors. The significance of any term was obtained by testing whether the full model and one without the term being tested differed significantly from each other, with a significance level of 0.05.

2.3. Results

2.3.1. Sibilants in /str, spr, skr/

Fig. 1 shows the M1 trajectory over the entire time course of the fricative noise for /s/ in pre-vocalic position and in /spr, str, skr/ sequences, separately for female and male speakers. Data for pre-vocalic /ʃ/ tokens (cf. Table 1) were also plotted here so that retraction can be seen in terms of proximity to pre-vocalic /s/ (in unbroken black) vs. pre-vocalic /ʃ/ (in grey) over the course of the fricative noise.

Looking first at pre-vocalic /s/ and /ʃ/, these were more strongly differentiated for female than for male speakers, both in terms of frequency and in terms of contour shape. This was primarily due to pre-vocalic /s/ for females, which showed a higher frequency, a much more defined peak and more dynamic change over the course of the fricative noise than the same fricative produced by males. Compared with pre-vocalic /s/, sibilants in /str, spr, skr/ showed lower M1 over almost the entire duration of the fricative noise, with the greatest amount of lowering evident in /str/ clusters. The absolute differences (in Hz) were greater within the female data but the location of the /str, spr, skr/ trajectories relative to those for /s/ and /ʃ/ was similar in the female and male data sets. Thus it appears that gender impacts the frequency of the spectral energy for /s/ but not the tendency towards retraction in cluster contexts. Both male and female speakers showed lower M1 values for /spr, str, skr/ than for pre-vocalic /s/, but not to the extent there was any overlap with pre-vocalic /ʃ/.

Sibilants in /str/ (in blue) differed from those in all other contexts in terms of the way that the M1 maximum was maintained towards the temporal end of the sibilant. Thus sibilants in /str/ did not become more /ʃ/-like towards their offset, which conflicts with the idea that retraction, as a form of assimilation, should *increase* with time.[4] The temporal location of the M1 maximum for /str/ is difficult to identify in Fig. 1 because of the overall flatter contour shape; in this respect the sibilant in /str/ shows a strong resemblance to /ʃ/. Sibilants in /spr, skr/ (but not /str/) showed a relatively early M1 maximum with a subsequent drop during the second temporal half. Articulatory movements can only be interpreted indirectly from M1, but this drop in M1 likely reflects anticipatory lip rounding and/or tongue tip retraction before rhotics. These gestures would increase the size of the cavity in front of the tongue tip constriction, lowering M1 during the fricative noise. Koenig et al. (2013) report a drop in M1 towards the temporal end of the sibilant before labials.

Fig. 2 shows the relative distance between /s/ and /ʃ/ of /s/ in pre-vocalic position and in /sCr/ clusters, calculated using (1). Greater positive values denote a greater proximity to /s/ and greater negative values indicate a greater proximity to /ʃ/.

Fig. 2 shows that the log. distance ratio values for /s/ in /sCr/ clusters were lower than for pre-vocalic /s/, which is evidence of greater acoustic similarity to /ʃ/. We tested this pattern with the mixed model with d_{fric} as the dependent variable, Context (two levels: /sV/ vs. /sCr/ where C was collapsed across /p, t, k/) and Gender as fixed factors and Speaker (20 levels) and Word (5 levels) as

[4] This pattern also conflicts with Warren's (1996) observation (based on visual inspection of spectrograms for /str/ in New Zealand English) that the sibilant became more /ʃ/-like over the course of the fricative noise. Similarly, assimilation in /sj/ sequences (e.g. *confess your*) has also been shown to involve an acoustic change from /s/-like to /ʃ/-like during the sibilant noise (Zsiga, 1995).

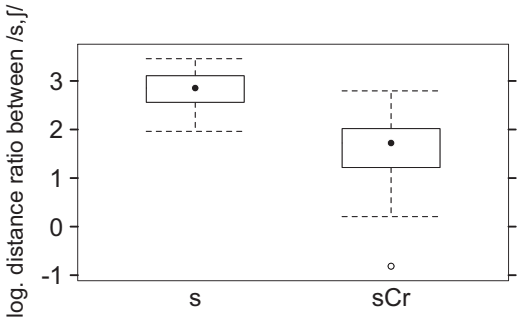

Fig. 2. The log. distance ratio value for /s/ in pre-vocalic position and in /sCr/ clusters calculated using (1). Each box consists of one aggregated point per speaker.

random factors.[5] The fixed factor Gender did not have a significant effect and was removed from the model. The factor Context had a significant effect on the relative proximity of sibilants to /s/ vs. /ʃ/ (χ2[1] = 7.3, p < 0.05). This result shows that /s/ in /sCr/ contexts was acoustically significantly closer to /ʃ/ compared with /s/ in a pre-vocalic /sV/ context. The lack of an interaction with gender shows that the degree of M1-lowering for /s/ in /sCr/ contexts was similar for male and female talkers, relative to the gender-specific /s/ vs. /ʃ/ endpoints.

So far we have seen that /s/ shows the closest acoustic proximity to /ʃ/ when it occurs in /sCr/ clusters and Fig. 1 shows that this centre of gravity lowering is primarily due to /str/. This observation regarding place-specific effects was supported by additional statistical tests with Context as a four-level fixed factor (i.e. without collapsing for stop place): a mixed model[6] with mean M1 as the dependent variable, Context (four levels: t, p, k, Vowel) and Gender (male, female) as fixed factors and Speaker (20 levels) as random factor showed that both Context (χ^2[3]=25.9, p<0.001) and Gender (χ^2[1]=27.4, p<0.001) had a highly significant effect on M1; there was no interaction between these two fixed factors. Post-hoc Tukey tests on the factor Context showed highly significant differences (p<0.001) between all six possible pairs; that is, mean M1 was significantly lower in each cluster context than in pre-vocalic position and the degree of M1-lowering was significantly greater in /str/ than in /spr/ and /skr/ contexts.

The question arises as to whether some speakers might show categorical retraction in /str/ clusters such that the sibilant falls within the range of their own /ʃ/ productions. Assimilation processes in English can involve categorical differences for some speakers (e.g. Ellis & Hardcastle, 2002; Nolan, Holst, & Kühnert, 1996 on consonant place assimilation) and the question of whether /s/-retraction involves a categorical or a gradient change remains open (cf. in particular Rutter, 2011). To address whether some speakers might show categorical behaviour in /s/-retraction, we compared each speaker's /str/ with their own pre-vocalic /s, ʃ/ tokens in terms of mean M1. Each data point in Fig. 3 shows the mean M1 for one sibilant token produced by one speaker.

Looking first at pre-vocalic /s/ (in grey), we can see that there is considerable speaker-specific variation: compare the data for pre-vocalic /s/ produced by speakers M08 and M04 (both male) or by speakers F05 and F01 (both female). This pattern is consistent with the results of many other studies on the sibilant /s/ in English which have also reported individual variation (e.g. Baker et al., 2011; Haley et al., 2010; Niebuhr, Clayards, Meunier, & Lancia, 2011; Zsiga, 1995). As far as /str/ is concerned, M1 values for individual tokens were typically lower than those for the same speaker's pre-vocalic /s/. That is, these acoustic data suggest that the tendency to produce /s/ in /str/ with a retracted articulation was common to all speakers. The magnitude of the acoustic effect of the /str/ context differed across individuals; for example, mean M1 in /str/ was 2914 Hz lower than pre-vocalic /s/ for speaker F05 but only 131 Hz lower for speaker M08. Nonetheless, all but three speakers (F05, M04 and M19) produced /str/ such that the sibilant was within the lower range of their own canonical /s/ or intermediate between their own /s/ and /ʃ/. Thus for all but three speakers there was a complete lack of overlap between /str/ and /ʃ/. This result supports the idea that /s/-retraction involves continuous change to an individual's production target for sibilants in /str/ rather than replacement with /ʃ/. Data for the remaining three speakers (F05, M04, M19) who *did* show some acoustic overlap between /str/ and /ʃ/ tokens may provide the conditions for retraction to be phonologised as /ʃ/. However, of these three speakers only speaker F05 produced /str/ such that the sibilant was consistently (a) within the range of /ʃ/ and (b) distinct from /s/. We cannot know whether this speaker's exceptionally categorical differences between sibilants in pre-vocalic vs. /str/ contexts are the result of an abrupt change or a gradual shift that has already taken place.[7] Acoustic data for all other nineteen speakers suggest that /s/-retraction is a shared, gradient tendency in the production of /str/ clusters.

2.3.2. *Sibilants in /st, sp, sk/*

Fig. 4 shows the M1 trajectories for /s/ in /sp, st, sk/ clusters and in pre-vocalic position. The M1 trajectory for pre-vocalic /ʃ/ was again plotted in grey for comparison. The differences in Hz between /s/ in prevocalic /sV/ and /st, sp, sk/ contexts were much smaller than those for the rhotic clusters, which involved more than 1000 Hz (cf. Fig. 1). Nevertheless, Fig. 4 shows that M1 was slightly lower for sibilants in /sp, st, sk/ than in pre-vocalic position throughout most of the duration of the sibilant noise and especially during the middle temporal half (i.e. between the vertical dotted lines).

[5] The model in R that was used was: dfric ~ Context + (1 | Word) + (1 + Context | Speaker).

[6] The model in R that was used was: M1 ~ Context + Gender + (1+Context | Speaker).

[7] This speaker's data are exceptional not only in terms of the similarity between /str/ and /ʃ/ but also because of the extremely high mean M1 for pre-vocalic /s/. Thus contextual differences for this speaker's /s/ might be due to hyper-articulation in pre-vocalic contexts as much as a tendency to retract in /str/.

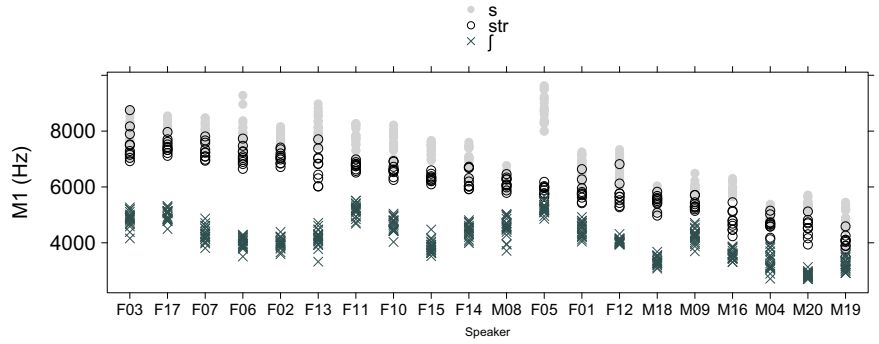

Fig. 3. Mean M1 for pre-vocalic sibilants /s, ʃ/ and for the sibilant in /str/ produced by 20 speakers. One data point per sibilant token: light grey filled circles for /s/, dark grey crosses for /ʃ/ and black open circles for /str/. Speakers are ordered along the *x*-axis in order of decreasing (mean) M1 in /str/.

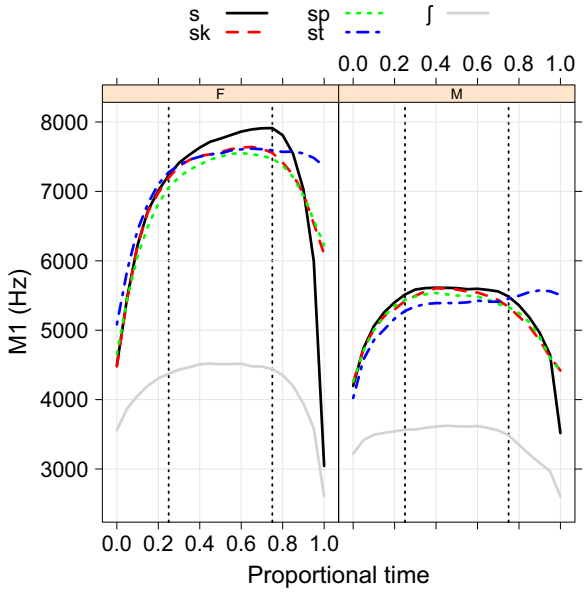

Fig. 4. First spectral moment trajectories during the sibilant in /sV/, /spV/, /stV/, /skV/ and /ʃV/ (where V = vowel). Tracks were averaged across speakers after averaging within context and speaker. The dotted vertical lines indicate the temporal middle half, over which mean M1 was calculated.

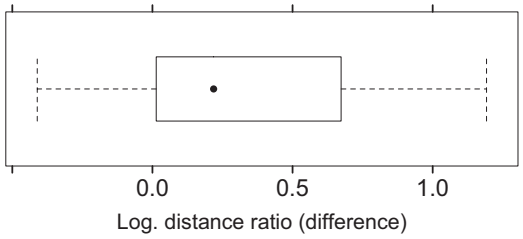

Fig. 5. The plot shows the difference (one value per speaker) between pre-vocalic and pre-consonantal /s/ on the log. distance ratio. Positive values indicate that the log. distance was greater for /s/ in pre-vocalic than in pre-consonantal position.

While the acoustic differences (compared to /sV/) were much smaller in this data set, the trajectory shapes for /sp, st, sk/ were broadly similar to those for their rhotic cluster counterparts, e.g. M1 stayed high before alveolar /t/ in contrast to other contexts. These trajectories, averaged across all speakers, suggest that female speakers showed a relatively similar degree of M1-lowering in /sp, st, sk/, whereas men showed more M1-lowering in /st/ (in blue) than in /sp, sk/.

For the data in Fig. 4, the relative distances of pre-vocalic /s/ (*sV*) and pre-consonantal (*sC*, *C* = /p, t, k/) tokens between /s/ and /ʃ/ were calculated using (1): recall that lower values on (1) mean that a given fricative is acoustically closer to /ʃ/ than to /s/. These relative distances were aggregated by speaker and context (thus giving two aggregated values per speaker, one for *sV*, one for *sC*). The aggregated *sC* was then subtracted from aggregated *sV* in the same speaker. The null hypothesis is that there is no difference on this measure between *sV* and *sC* in which case the resulting distribution in Fig. 5 should be centered at zero. However, the interquartile range (extent of the box) is positive: thus, *sV* was higher than *sC* on (1) which in turn means that pre-consonantal /s/ was closer to /ʃ/ than was pre-vocalic /s/.

We tested whether pre-consonantal /s/ was closer to /ʃ/ with the mixed model as described previously i.e. with the log distance ratio as the dependent variable with Gender and Context (two levels: (/sV/ vs. /sC/ as fixed factors, and with Word and Speaker as random factors.[8] Since neither Gender nor its interaction with Context were significant, they were removed from the model. The log. distance ratio was shown to be significantly influenced by Context ($\chi^2[1] = 6.6$, $p<0.05$). Compatibly with the data in Figs. 4 and 5, this result shows that sibilants in a pre-consonantal /sC/ context (C = /p, t, k/) were acoustically closer to /ʃ/ compared with /s/ in pre-vocalic /sV/ contexts. Consistently with the /sCr/ data analysed earlier, gender did not have a significant effect on the degree of retraction in /sC/.

2.4. Summary and discussion of production results

The first spectral moment for /s/ in /spr, str, skr, st, sp, sk/ was lower than for the same sibilant in pre-vocalic position, which was interpreted as evidence of articulatory retraction in such cluster contexts. The degree of M1-lowering was greater before rhotics (i.e. /sCr/) than in /sC/ and was greatest for /str/ in particular. All participants showed lower M1 for /s/ in /str/ than in pre-vocalic position (cf. Fig. 3), which suggests that /s/-retraction is a shared tendency in the production of /str/. Nonetheless, there was rarely any overlap between /strV/ and the same speaker's /ʃV/; thus /s/-retraction appears to involve a gradual change to an individual speaker's production target for /s/. While gender influenced the spectral energy for /s, ʃ/, with males showing lower M1 as expected (e.g. Koenig et al., 2013:1178), both males and females showed lower M1 values in /sC, sCr/ contexts. Moreover, gender did not have a significant effect on the degree of retraction in /sC/ or /sCr/ clusters.

Specific to the rhotic clusters, we observed different temporal dynamics for M1 trajectories in /str/ vs. /spr, skr/ (cf. Fig. 1 and accompanying text). This difference suggests that two distinct articulatory strategies might bring about the lower M1 values that we observe for /s/ in these clusters compared with pre-vocalic position. More specifically, any articulatory change should show up acoustically as a change in M1 (e.g. Iskarous et al., 2011). Thus the fall in M1 over the second temporal half for /s/ in /spr, skr/ suggests that there was some change in the position of the articulators *during* the sibilant noise. This would be the case if speakers were to begin to form the articulatory position for the rhotic sometime after the onset of oral closure for the plosive in /spr, skr/. In contrast, the M1 trajectory for the sibilant in /str/ formed a plateau, which suggests that the articulators remained relatively stable during the sibilant noise. Moreover, the relatively low M1 for /s/ in /str/ contexts suggests that the articulators were already in a retracted position at the onset of the /s/. Such acoustic and articulatory stability for /s/ in /str/ would be possible if it were to precede a fully developed rhotic, as shown in Fig. 6.

Fig. 6 suggests that the rhotic is formed relatively earlier in /str/ than in /spr, skr/. This representation is consistent with Cruttenden (2014:202) who commented that the plosive /t/ is retracted in /str/, although /spr, skr/ were not addressed in that source. The reportedly common pronunciation of /tr/ with a post-alveolar articulation in English (cf. Cruttenden, 2014; Lawrence, 2000) also supports the idea that /t/ is especially prone to assimilate to an upcoming rhotic. In contrast, the degree of tongue-tip retraction of the rhotic in /skr/ may be attenuated by the need to produce a velar closure. An electromagnetometry study by Kühnert, Hoole, and Mooshammer (2006), for example, reported that tongue tip movements for the lateral occurred relatively later in /kl/ than in /pl/ contexts. These authors attributed this pattern to the idea that the whole tongue is constrained during production of a velar closure for /k/, such that the tongue tip cannot articulate /l/ as early as in the bilabial /pl/. At this stage it is unclear why /s/ is not retracted to the same degree in /spr/ as in /str/ because there is nothing to prevent tongue tip retraction during the bilabial closure. In summary, then, lower frequencies for sibilants in /spr, skr/ appear to be caused by dynamic articulatory movements (lip rounding, tongue tip retraction or both) during the sibilant, whereas the acoustic data suggest that the sibilant in /str/ precedes a fully formed rhotic tongue position.

The production data also showed lower M1 for /s/ in non-rhotic /sC/ contexts. We did not expect to find acoustic evidence for retraction in /sC/ because it is not immediately clear why speakers would do so. Nonetheless, our results are consistent with acoustic evidence of a lower spectral centre of gravity for /spV, stV, skV/ than for /sV/ in American English (Baker et al., 2011). This evidence from two independent acoustic studies favours the idea that there might be a physiological explanation for the lower frequencies that we observed for sibilants in /sCV/ contexts.[9]

Turning first to /spV/, lip approximation during /s/ before /p/ causes a lowering of spectral central of gravity (Fant, 1970). Such an explanation for retraction in /sp/ is consistent with the relevant literature (e.g. Koenig et al., 2013; Stevens, 1998:559) in which the spectral peak for /s/ is reported to drop rapidly before labial closure. Our data showed that M1 did indeed drop towards the temporal offset of the sibilant in /sp/ (cf. Fig. 4). However, the influence of the upcoming labial on the noise spectrum appears to extend further leftwards into the sibilant in our data than has been reported in other sources: closer inspection shows that retraction is actually already present for /sp/ at the 10% time point.

Turning to /skV/, the tongue position for a velar closure is antagonistic to the position for an alveolar fricative (e.g. Recasens & Pallarès, 2001) and might cause tongue tip retraction for the sibilant in /skV/. Alternatively, rather than dragging the tongue tip backwards, M1-lowering in /skV/ could also be explained by perturbations to the noise spectrum as the velar closure is approximated. Stevens (1998:559) explains how aerodynamic turbulence at the velum can introduce a second spectral prominence at about 1600 Hz during the sibilant in /sk/, which would of course have the effect of lowering M1. The M1 trajectory for the sibilant in /skV/ drops at the temporal offset which would correspond to the formation of the velar closure (Fig. 4). However, M1 is actually slightly lower for /skV/ than for pre-vocalic /s/ from the temporal midpoint onwards. This suggests that retraction in /skV/ might be caused by both (a) tongue tip retraction during the sibilant and (b) turbulence at the velar constriction towards the temporal end.

It is more difficult to account for retraction in /stV/ than in /spV, skV/ because there is no immediately apparent reason why alveolar /s/ should be retracted before a constriction at the same place of articulation. Moreover, there is very little phonetic evidence available

[8] The model in R that was used was: dfric ~ Context + (1 | Word) + (1 + Context | Speaker) where Context was the fixed factor and Word and Speaker the random factors.

[9] Baker et al. (2011:357) noted that "[t]he presence of a consonant has a depressing effect on the centroid frequency" but did not offer a phonetic explanation for this effect.

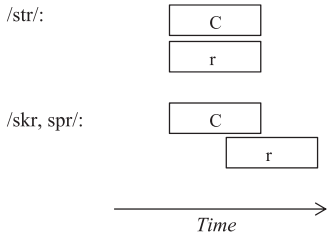

Fig. 6. Schematic representation of rhoticity relative to production of the preceding consonant in /sCr/ sequences where C = /t/(top) and /p, k/(bottom).

for /stV/ sequences aside from an early study on alveolars in British English (Bladon & Nolan, 1977), and the results of that study are not consistent with ours. Bladon and Nolan reported that speakers produced /t/ with an apical articulation but changed to a more laminal articulation after the (laminal) fricative in /st/. In other words the fricative exerted coarticulatory influence on the *stop* in /stV/ and not the other way around (cf. also Recasens and Pallarès (2001) for similar coarticulatory patterns in other CC sequences involving the tongue tip). Nonetheless, the evidence that sibilants in /st/ showed lower spectral frequencies in our data as well as in another study on English (Baker et al., 2011) favours the idea that there might be a bias towards retraction in the production of /stV/ that is common to many speakers. Given that lip rounding is unlikely in *steam,* the lower M1 values for sibilants in /stV/ are more likely to be caused by a slightly retracted tongue tip which would increase the length of the cavity in front of the noise source. It is possible that the combination of a laminal /s/ + apical /t/ would favour dynamic retraction during the sibilant – although this is *not* what Bladon and Nolan found for British English. An articulatory study is necessary to clarify whether lower M1 values reflect tongue tip retraction in /stV/ contexts and why speakers would differentiate sibilant production in this way.

In the following study we address whether listeners are sensitive to the acoustic differences between sibilants in pre-vocalic vs. /sC(r)V/ contexts that we have observed and whether lower frequencies in the latter might cause listeners to categorize the sibilant as /ʃ/. We focus on sibilants in /str/ contexts because the acoustic effects of retraction were strongest and most frequent across speakers in this context; we also included /stV/ tokens to test whether smaller acoustic differences are also audible to native listeners.

3. Perception study

Having shown that speakers produce sibilants in /sC(r)V/ such that they are acoustically more similar to /ʃ/, this second study tests how such sibilants are categorized by native listeners. We address this issue based on production data collected during Study 1 rather than a synthesized /s…ʃ/ continuum, so that any contextual information during the sibilant noise can remain intact (especially any dynamic differences between sibilants in pre-vocalic vs. pre-consonantal positions). In doing so, we follow the experimental paradigm described in detail by Li, Munson, Edwards, Yoneyama, and Hall (2011) who tested listener categorization of pre-vocalic /s, ʃ/ based on multiple talkers' natural productions (child and adult talkers in their case). We limit the scope of this experiment to sibilants originally produced in *stream, steam, seam* and *sheep,* with the latter two serving as the unambiguous endpoints.

3.1. Experiment design and predictions

Because English allows /str/ but not /*ʃtr/, we spliced out the sibilant and prepended it to *–eet,* which enables a forced choice between phonological /s/ and /ʃ/ (with listeners choosing between SEAT and SHEET). Sibilants spliced from *seam* and *sheep* should elicit 100% SEAT and 100% SHEET responses, respectively. Regarding sibilants spliced from *stream* and *steam,* we made two main predictions:

(1) sibilants spliced from *stream* and *steam* will elicit more SHEET responses than those spliced from *seam.*
(2) the number of SHEET responses to sibilants spliced from *stream* and *steam* will increase as M1 decreases (i.e. as sibilants become more /ʃ/-like acoustically).

In addition to these two main predictions, we also investigated the impact on listener response data of (3) individual listener and (4) talker gender. Regarding (4), we saw during Study 1 that sibilants differed predictably between male and female talkers in terms of absolute M1, but that the tendency to show lower M1 in /st, str/ contexts was common to both males and females (Figs. 1 and 4). Given that listeners are known to be relatively good at identifying the gender of the talker and adjusting their perceptual /s/ vs. /ʃ/ category boundaries accordingly (cf. e.g. Munson et al., 2006), we do not expect listener responses to differ according to whether the sibilant was produced by a woman or a man. On the other hand, Fig. 4 shows that the amount of M1 lowering in /st/ was greater in male speech than in female speech (whereas it was relatively similar across genders in /str/). With these issues in mind, we controlled for the gender of the speaker voice in the perception task.

The stimuli for this experiment involved multiple voices and were presented to participants in a completely random order. During such a task it is unlikely that a listener would be able to form speaker-specific perceptual models for sibilants. As such, we tested listeners' categorization of sibilants; we did not test retraction in /st, str/ relative to an individual speaker's typical /s/ pronunciation.

3.2. Methods

3.2.1. Stimuli and task

To make the stimuli, we extracted the word-initial sibilant from each *sheep, steam, seam* and *stream* token produced in Study 1. Each of these sibilants was then prepended to an *–eet* token produced by the same speaker, giving 40 SHEET/SEAT stimuli for each speaker voice. The amplitude was normalized across all 800 SHEET/SEAT stimuli. Importantly, the *–eet* token was kept constant within (but not across) speakers. That is, for each speaker, we spliced all sibilant tokens onto the same *–eet* token produced by the same speaker. The *–eet* tokens were made by extracting the medial /iːt/ sequence out of one repetition of *Peter* produced by each speaker (*Peter* was a filler item in the production experiment). Because of the presence of the preceding bilabial stop in the original production, there was no coarticulatory information that would be likely to bias /s/ or /ʃ/ percepts from *–eet*. Care was also taken to choose an *-eet* token in which the alveolar stop was realised as a stop or an affricate rather than a tap so that it would sound appropriate in word-final position.[10] To keep the duration of the experiment to a reasonable length for participants, the number of SHEET/SEAT stimuli was reduced. For each speaker voice, we included the sibilant from one *sheep,* one *seam*, four *steam* and four *stream* repetitions in the final set. Thus there was a higher number of tokens for the ambiguous *steam* and *stream* tokens compared with the *seam/sheep* endpoints.

1 x /s/ *(seam)*
4 x /s/ *(steam)* + /iːt/ (*Peter)* = 10 SHEET/SEAT × 20 speakers = 200 stimuli
4 x /s/ *(stream)*
1 x /ʃ/ *(sheep)*

None of the 200 items in the final SHEET/SEAT set had any audible silence nor glottalization between the sibilant and the onset of the vowel and no background noise. Differences between stimuli were minimized as far as possible. In particular, all speaker voices were recorded in the same room with the same headset microphone and the amplitude was normalized across stimuli to control for the fact that some speakers spoke louder than others. Thus any change in listener responses to stimuli produced by the same speaker voice can be attributed to differences during the sibilant noise, because the *-eet* token was kept the same within each speaker's (ten) SHEET/SEAT stimuli.

The experiment was conducted using a platform designed for simple online perception experiments by Florian Schiel at the IPS in Munich (adapted from Perlman (1985)). This platform enables audio stimuli to be presented to listeners as a vertical list of play buttons for each stimulus with radio buttons (to the right of each play button) for listener responses. Our 200 SHEET/SEAT stimuli were listed in a randomized order with no repetitions. Listeners were instructed to navigate their browser to the address for the perception experiment. Wearing headphones, listeners clicked on the play button to hear each stimulus and selected a radio button to indicate whether it sounded more like SHEET or SEAT. There was no time limit on the experiment and participants could listen to the stimuli as many times as they wished. Built-in controls ensured that each participant clicked on (i.e. heard) and judged all stimuli. The experiment took about 15 minutes to complete. Results were sent as a spreadsheet to a dedicated email address and imported into R for analysis.

3.2.2. Participants

Twenty two first language Australian English speakers (thirteen female, nine male; age range 20-49 years) completed the perception experiment, for which they were paid AU $10. None of the participants reported any hearing difficulties. Six of the participants were from Braidwood and had also already participated in the production experiment reported in Study 1. These particular participants (4 female, 2 male) may have been able to recognize some or all of the voices as belonging to people from their local community. The remaining participants were recruited through personal contacts and would not have been able to recognise any of the voices. We did not distinguish Braidwood vs. other listener participants in presenting the results, because there was no difference between these two participant groups in terms of the proportion of SHEET responses to each word (i.e. to sibilants spliced from *seat, sheet, steam, stream*; cf. Appendix B).

3.3. Results and discussion

Our first hypothesis is that listener responses will depend on the context in which the sibilant was originally produced and that sibilants spliced from *stream* and *steam* will be categorized as SHEET more often than sibilants spliced from *seam*, owing to the lower M1 that we observed in these contexts. Fig. 7 shows the proportion of listener responses to sibilants spliced from the four word contexts *seam, sheep, steam,* and *stream*.

We begin by examining the unambiguous sibilants spliced from *sheep* and *seam*. Fig. 7 shows that the former were always identified correctly by listeners, whereas sibilants spliced from *seam* elicited a small proportion of SHEET responses (15 of the total 440). These responses to *seam* vs. *sheep* show that there is an asymmetry between /s/ and /ʃ/ in perception: /s/ can resemble /ʃ/, but not the other way around. Examining the *seam* data more closely, we see that only five talker voices elicited SHEET responses from listeners, and that SHEET responses were most frequent for *seam* tokens produced by two talkers in particular: F12 (30%) and M20 (20%).

[10] The vowel in *–eet* was occasionally produced with a prominent onglide, so that it sounded more like [əi]. This realization of /iː/ is typical for Broad i.e. "overtly local" Australian English pronunciation (Cox & Palethorpe, 2007:345). Given they were more common in our data and we wanted to minimize differences between speaker voices, we selected a monophthongal production for each speaker's *–eet* token. For one speaker who consistently produced /iː/ with a prominent onglide, we segmented *–eet* from the onset of the steady state for the target [i], excluding the onglide.

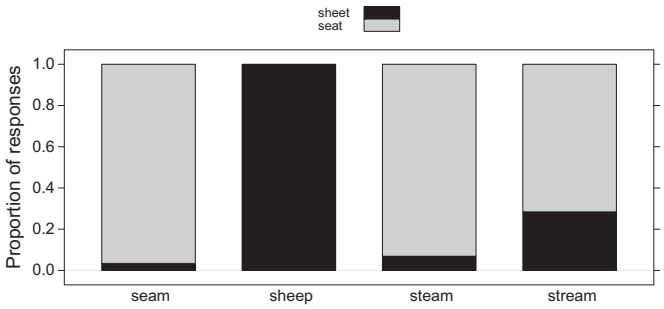

Fig. 7. SHEET responses (in black) to sibilants spliced from *seam, steam, stream* and *sheep*. The total number of stimuli differed across words: there were 440 responses to *seam* and *sheep* vs. 1760 responses to *steam* and *stream.*

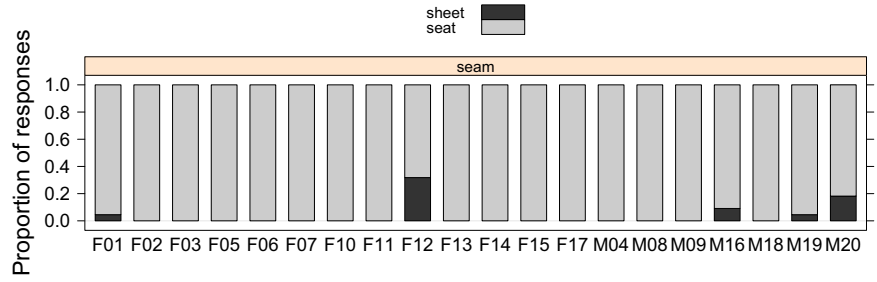

Fig. 8. Proportion of SHEET responses (in black) to sibilants in *seam* produced by twenty talkers (along the *x*-axis).

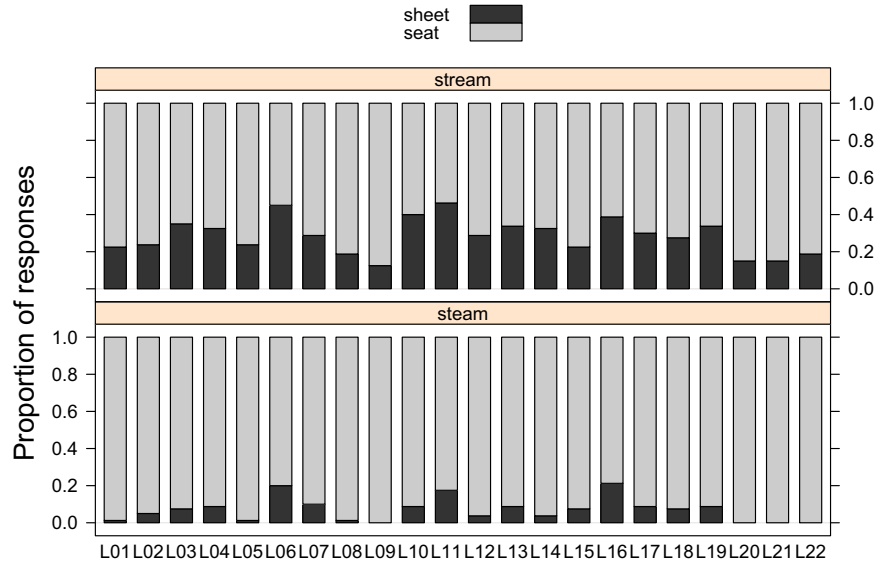

Fig. 9. Proportion of listener SHEET responses (in black) to sibilants in *steam* and *stream,* for each of the 22 listener participants (along the *x*-axis).

Listeners' categorization of F12 and M20's *seam* tokens is consistent with the evidence that they were acoustically more similar to post-alveolar fricatives produced by other speakers of the same gender. More specifically, it can be seen in Fig. 3 that speaker F12 produced pre-vocalic /s/ with a relatively low M1 compared with other female speakers, and the same is true for speaker M20 compared with the other males.

Turning to sibilants spliced from *stream* and *steam*, here listeners responded with SHEET 28% and 7% of the time, respectively (cf. Fig. 7). This result suggests that the acoustic effect of the /str, st/ context can be strong enough to cause native listeners to categorize the sibilant as /ʃ/ – when heard in a pre-vocalic context. Moreover, Fig. 9 shows that all listeners categorized at least some of the *stream* tokens as SHEET, and most listeners categorized at least some of the *steam* tokens as SHEET. As such, the categorization of sibilants spliced from /str, st/ as post-alveolar is a shared tendency amongst listeners rather than an idiosyncratic pattern.

We tested the effect of word on listener decisions by carrying out logistic regression within a generalized linear mixed model with Response (SHEET/SEAT) as the dependent variable, Word (3 levels: *seam, steam, stream*) and Speaker Gender as fixed factors and Speaker Voice (20 levels) and Listener Participant (22 levels) as random factors.[11] Note that Word had only three levels in this particular test

[11] The model in R that was used was: RESPONSE ~ WORD + VOICE_sex + (1+WORD|VP) + (1|VOICE), where WORD is the fixed factor and VP and VOICE the random factors (VP = listener participant, VOICE=speaker voice).

Table 2
Wald *z*-statistic and adjusted probability scores of Tukey post-hoc tests for the pairwise comparisons between words.

	z value	Adjusted *p* values
Responses to female speakers		
steam–seam	−0.8	0.727
stream–seam	5.9	0.000
stream–steam	9.9	0.000
Responses to male speakers		
steam–seam	2.8	0.01
stream–seam	7.1	0.00
stream–steam	12.0	0.00

because the *sheep* data were excluded from the statistical analysis. This model showed that listener responses were significantly affected by the word from which the sibilant was spliced ($\chi^2[2]=657.9$, $p<0.001$) and by the gender of the speaker ($\chi^2[1]=6.25$, $p<0.05$). There was a significant interaction between these two fixed factors ($\chi^2[2]=10.9$, $p<0.01$). Post-hoc pairwise comparisons showed significant differences ($p<0.05$) between all word/gender pairs except for *seam-steam* spoken by females (Table 2).

Therefore, our first hypothesis regarding the effect of word context on listener responses was fully supported by the data for male speakers, but only partially supported by the data for female speakers which showed significantly more SHEET responses to sibilants spliced from *stream* than *seam* but no significant difference between *steam* and *seam*. The interaction between word context and speaker gender can be seen in Fig. 10, which shows listener responses to sibilants spliced from *seam, steam, stream* and *sheep* produced by male and female talkers.

It is clear from Fig. 10 that sibilants in *stream* and *steam* spoken by male talkers elicited more SHEET responses than the same items spoken by female talkers. This gender difference in the perception of sibilants was unexpected in the sense that Study 1 showed that gender did not impact the degree of M1-lowering in /sC/ or in /sCr/. To compare the perception and production data more reliably, we need to first examine the degree of M1-lowering in the /st, str/ tokens, in particular, and whether it differed between male and female speech (recall that Study 1 compared /sp(r), st(r), sk(r)/ with /sV/). We tested this issue with two further mixed models, one for the /st/ tokens ($n=197$) and one for the /str/ tokens ($n=200$). In both models, $_d$*fric* was the dependent variable, Gender was the fixed factor and Speaker was a random factor.[12] In neither model did the fixed factor Gender have a significant effect on $_d$*fric*. In other words, there was no difference in the degree of M1-lowering in *steam, stream* contexts between men and women in the production data. Yet sibilants produced by men in these contexts elicited more SHEET responses: one reason for the higher proportion of SHEET responses to male voices may be that listeners did not fully adjust their /s/ vs. /ʃ/ category boundaries for the lowering effects of male voices on the spectral centre of gravity, thus categorizing relatively more of the stimuli produced by male voices as /ʃ/. A different interpretation, suggested by an anonymous reviewer, would be that the smaller absolute acoustic separation between /s/ (in any context) and /ʃ/ in male speech favoured SHEET responses. That is, listeners may have compensated fully for the gender differences but contextual M1-lowering in /st, str/ would have created more ambiguity for male voices because the absolute distance to /ʃ/ was smaller. However, one presumes that the *relative* acoustic distance between any particular sibilant and the appropriate /s/ and /ʃ/ endpoints would be maintained despite gender normalization. Therefore, because the relative location of /st, str/ tokens between these endpoints did not differ significantly between males and females, we favour insufficient gender normalization as the explanation for this result, but it is of course possible that both explanations may have played a role in increasing SHEET responses to male talkers in this experiment. We return to this issue in the general discussion, together with some potential implications for models of sound change.

Our second hypothesis regarding the perception data was that responses to sibilants spliced from *stream* and *steam* will be correlated with M1, with SHEET responses increasing as M1 decreases. We already saw some evidence in favour of such a correlation, because pre-vocalic /s/ tokens with the lowest mean M1 elicited some /ʃ/ responses (Fig. 8). To test our second hypothesis directly, we calculated the mean M1 for each speaker's sibilants in *stream* and *steam* and plotted it against the proportion of SHEET responses to that speaker's *stream* and *steam* tokens. We then fitted logistic regression models to these data, separately for word (*stream* and *steam*) and talker gender. These data are shown in Fig. 11.

It can be seen in Fig. 11 that speakers whose sibilants in *stream* and *steam* are more /ʃ/-like (with lower Hz values) elicited more SHEET responses from listeners than speakers whose sibilants are more /s/-like (with higher Hz values). The results of the logistic regression showed that the proportion of SHEET responses was significantly affected by M1 for all four combinations: *stream* female $\chi^2[1]=276.4$, $p<0.001$; *stream* male $\chi^2[1]=303.2$, $p<0.001$; *steam* female $\chi^2[1]=77.5$, $p<0.001$; *steam* male $\chi^2[1]=20.3$, $p<0.001$. This result confirms our second hypothesis that listener categorization of sibilants was influenced by the talker's M1.

4. General discussion

This investigation sought to document the phonetic pre-conditions that can give rise to /s/-retraction, taking Australian English to represent languages or varieties in which this sound change has not yet taken hold. Study 1 showed that the sibilant /s/ has a lower spectral centre of gravity in /spr, str, skr, sp, st, sk/ than in pre-vocalic position, which indicates that it was produced with a retracted tongue position (and possibly lip rounding). Study 2 showed that when spliced into a pre-vocalic context, listeners occasionally

[12] The model that was used for the /st/ and /str/ data in R was $_d$fric ~ Gender + (1|Speaker).

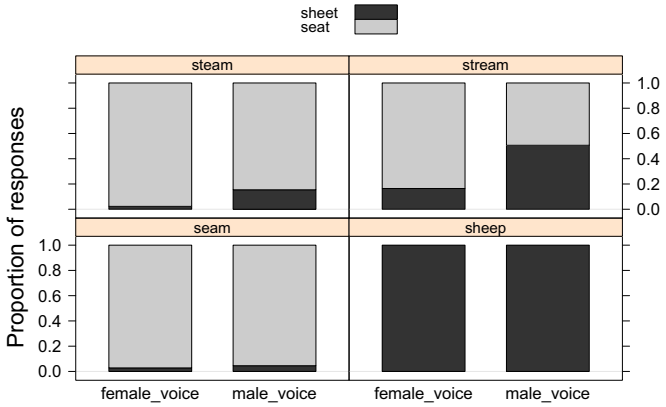

Fig. 10. Listener responses to sibilants spliced from *seam, steam, sheep* and *stream*, produced by male and female speaker voices.

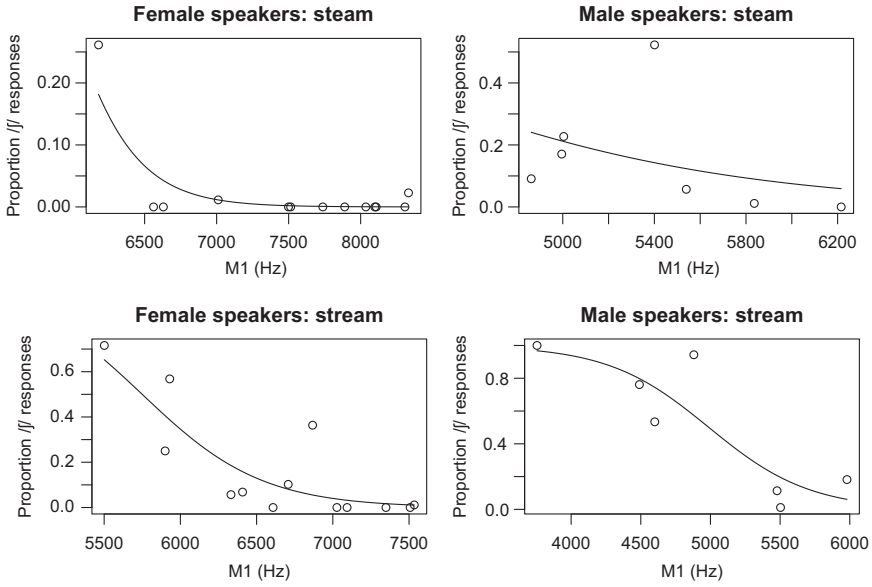

Fig. 11. Proportion of SHEET responses with fitted sigmoid functions and mean M1 for sibilants spliced from *steam* (upper row) and *stream* (lower row). Each data point represents one speaker. To aid legibility the x- and y-scales differ across the four response curves.

perceived sibilants that were originally produced in /str, st/ as post-alveolar. Thus the results of the present investigation with Australian English support both of our predictions about the pre-conditions for /s/-retraction, which were couched in terms of Ohala's listener-driven model of sound change: there is a bias towards retraction in the production of sibilants before voiceless stops and especially before rhotics, and the effects of this bias are audible to native listeners (when heard out of context). Thus although /s/ has been shown to be resistant to coarticulation in production (e.g. Byrd, 1996; Hoole, Nguyen-Trong, & Hardcastle, 1993; Keating, 1990; Recasens, Pallarès, & Fontdevila, 1997) and even though listeners have been shown to be more accurate at identifying the place of articulation of sibilants /s, ʃ/ than in stops and nasals (Hura, Lindblom, & Diehl, 1992), the present study nevertheless clearly shows that /s/ is subject to contextual variation in speech production and perception. Our finding is therefore consistent with other existing studies of fine-grained contextual differences in the production (Baker et al., 2011; Iskarous et al., 2011; Koenig et al., 2013) and perception (Engstrand & Ericsdotter, 1999) of /s/.

We interpret the contextual differences for /s/ in Australian English as evidence of the stable phonetic variation that can, potentially, bring about sound change, but will not necessarily do so. In terms of *why* there should be a tendency towards retraction in the production of sibilants in pre-consonantal contexts - before /s/-retraction takes hold in any variety - we outlined in the introduction how both long-distance assimilation with the upcoming rhotic and assimilation to an affricated [tʃr] (the result of a separate sound change) have been put forward as possible explanations. Note that both of these explanations apply specifically to /str/ which is the main context in which /s/-retraction in English has been presumed to occur. One of the aims of this study was to test whether either of these ideas about the origins of /s/-retraction could be supported with empirical evidence for a variety representing the pre-conditions for this sound change. The pervasiveness of retraction in /str/ in the Australian English data, as shown by consistently lower M1 values for sibilants in this context, supports the idea that an immediately adjacent /tr/ provides especially favourable conditions for /s/-retraction. However, comparison of the M1 trajectories in Fig. 1 and Fig. 4 shows that M1-lowering was greater in /spr, skr/ than in clusters without a rhotic. Therefore, lower M1 in /sCr/ clusters appears to be caused by long-distance anticipatory coarticulation due to the rhotic. This effect may be more marked in /str/ than in /skr/ because the rhotic causes a shift in the tongue position for /t/ in a

way that it cannot to /k/. The reason why M1 was lower in /str/ vs. /spr/ remains unclear since the tongue position would be unconstrained by the upcoming bilabial (cf. Fig. 6 and the accompanying text). A physiological study using e.g. electromagnetometry is needed to investigate the precise articulatory movements associated with the production of /sCr/ clusters and their temporal coordination. Nevertheless, the idea that different articulatory strategies might be responsible for the acoustic differences between /str/ vs. /spr, skr/ clusters is interesting in terms of the production biases that can develop into sound change. As just noted above, most sources on /s/-retraction are restricted to /str/. Janda and Joseph (2003) widen the scope of the discussion to include retraction in /spr, skr, sp, st, sk/ clusters but argue that only retraction in /str/ is motivated by a production bias; according to these authors, non-phonetic factors like analogy must be sought to explain the occurrence of retracted pronunciations in other contexts. By contrast, the results of Study 1 suggest that retraction in other /sCr, sC/ clusters may also have a physiological explanation, but that the precise articulatory movements may differ from those for /str/. The idea that /s/-retraction in /sC/ may also have a phonetic motivation is supported by typological evidence that pre-consonantal /s/ has undergone retraction in the historical development of some other languages, e.g. Standard German (see Kümmel (2007:232-7) for cross-linguistic information; Stevens, Bukmaier and Harrington (2015) address pre-consonantal /s/-retraction with perception and acoustic data from German and English). Moreover, our production data were drawn from a variety in which the sound change /str/ > /ʃtr/ has not yet taken hold, further supporting the idea that any pre-existing phonetic bias towards retracted articulations might apply (albeit to different degrees) to all /sCr, sC/ clusters. Therefore it may not be necessary to invoke analogy to account for instances of /s/-retraction in /sp, st, sk/ (and /spr, skr/) in English. Nonetheless, the acoustic differences were often small in our data, especially for /sC/, and further research is necessary to fully understand whether the phonetic effects that we found would be sufficient to bring about the phonologization of /s/-retraction over time. Phonologization of /s/-retraction is most likely in /str/ in English because the acoustic effect is strongest in this context. The acoustic effect was also most common across speakers in /str/ in our production data: all speakers showed lower M1 values in this context compared with their own pre-vocalic /s/ (cf. Fig. 3). Patterns across speakers were slightly less consistent in the other rhotic clusters: most but not all speakers showed lower M1 in /spr, skr/ than in /sV/. M1-lowering appeared to be more common across speakers in /st/ than in /sp, sk/, but eight speakers nonetheless showed *higher* mean M1 in /st/ than in pre-vocalic contexts, i.e. no evidence of articulatory retraction at all. In sum, our acoustic data suggest that the tendency to produce /s/ in /str/ with a retracted articulation (i.e. tongue-tip retraction or lip rounding or both) is common to all speakers whereas it is a speaker-specific tendency in /spr, skr, sp, st, sk/ contexts.

With the exception of one small-scale study by Engstrand and Ericsdotter (1999), previous sources that have reported acoustic variability for /s/ have not tested the impact of this acoustic variability on the perception of /s/. Iskarous et al. (2011:953) for example assume that "the magnitude of [acoustic] variability is still not high enough for perceivers to confuse /s/ and /ʃ/". However, Study 2 showed that this was not always the case: when spliced into pre-vocalic position, sibilants originally produced in /st, str/ can elicit /ʃ/ responses from listeners. This result is consistent with that of Engstrand and Ericsdotter (1999) who also found that listeners are sensitive to contextual information present in the acoustic signal for /s/ (but did not address retraction). The tendency to categorize sibilants as /ʃ/ was stronger when listeners heard male voices and we address this potentially important interaction with gender in more detail below. The fact that sibilants originally produced in /st, str/ can elicit /ʃ/ responses when spliced into pre-vocalic contexts does not mean that a sound change will necessarily take place, of course: listeners are very good at normalising for the effects of context on speech sounds when given the appropriate contextual information (e.g. Ohala, 2012:25 and references therein). American English listeners, for example, can detect nasalization during a vowel but will categorize that same vowel as oral when it precedes a nasal consonant (Beddor, 2009). Along the same lines, we predict that if provided with the appropriate contextual information, listeners should be able to connect the phonetic effect (retraction) with the source (the following /tr /) that gives rise to it and adjust their perception accordingly. This issue could be tested using the same paradigm from Study 2 except with sibilants spliced into clusters. These clusters would need to be formed over a word-boundary, e.g. *mass trial* vs. *mash trial*, because /*ʃtr/ is not permitted word-internally in English. The tendency to categorize sibilants with a relatively lower spectral centre of gravity as /ʃ/ that we saw for pre-vocalic position should disappear once listeners hear the same sibilants in the context that causes retraction (i.e. more *mass trial* responses). This prediction is supported by results reported by Kraljic et al. (2008), who exposed American English listeners to passages containing acoustically ambiguous s/ʃ sibilants. Listeners for whom all instances of /s/ were replaced with an ambiguous sibilant showed perceptual learning: they subsequently categorized more of a six-step [s]…[ʃ] continuum as /s/ than a control group. No such shift in the location of the perceptual boundary was observed for listeners who were exposed to ambiguous sibilants in /_tr/ contexts only (e.g. *distract*). According to Kraljic et al. (2008:62) these listeners treated "ambiguous /s/ tokens as a form of assimilation [which] lead to successful interpretation of the phoneme, without any need to change the underlying representation". In other words, listeners in that study attributed retraction to the effects of context and successfully recovered the identity of the /s/ despite the ambiguous acoustic signal. As Kraljic et al. (2008:61) noted, their participants were all resident in New York State and would have been familiar with extremely retracted /str/ pronunciations which are typical of the New York dialect and this may have facilitated perceptual processing of ambiguous stimuli. Whether individuals who are less familiar with retraction in /str/ would also recover an /s/ remains to be tested (see e.g. Docherty, Langstrof, & Foulkes, 2013 on the cognitive processing of fine-grained phonetic variation). Indeed, it is possible that our Australian English listener participants did not fully compensate for the influence of gender on sibilant production (see below), which suggests that they might not compensate sufficiently for contextual effects on sibilants, either. A mismatch between the production and perception of sibilants could provide the conditions for the phonologization of /s/-retraction in Australian English.

The results of Study 2 showed that talker gender interacted with the perceptual categorization of sibilants spliced from *stream* and *steam,* with male voices eliciting more SHEET responses (cf. Fig. 10). This result cannot be attributed to coarticulatory differences in production because both males and females produced /s/ with lower M1 in pre-consonantal contexts.[13] We pointed to two other

[13] An anonymous reviewer suggested that increased SHEET responses for male voices may be due to an interaction between M1 and F0. However, it is unlikely that pitch differences would have influenced M1 in this study because it was calculated between 500–15000 Hz, thus excluding any low-frequency energy associated with voicing.

explanations for the impact of talker gender on the perceptual categorization of sibilants: (1) incomplete normalization in perception for inherent gender differences, and (2) the acoustic proximity of /s/ (in any context) to /ʃ/, for male speakers. For either (or both) of these reasons, ambiguous sibilants such as those originally produced in *stream* would be more likely to be categorized as /ʃ/ when spoken by male voices. Further perception tests are needed to tease apart the role of relative vs. absolute acoustic differences in listeners' categorization of sibilants produced by men and women. Nonetheless, we favour (1), because we presume that relative differences between /s/ and /ʃ/ would be most important in categorizing ambiguous *steam, stream* sibilants, and these were the same for men and women. We need to examine (1) more closely because this interpretation conflicts with the results of the other empirical studies (e.g. Mann & Repp, 1980; Munson et al., 2006) which show that English listeners are in fact very good at normalizing for gender differences in the categorization of /s/ and /ʃ/. There is also no reason to assume that talker gender would have been ambiguous for listener participants in Study 2, given male and female voices can be separated based on the pitch during *–eet.*[14] Nonetheless the (isolated word) stimuli may have been too short for listeners to identify the gender of the speaker with confidence, and perception tasks involving multiple talker voices such as the one used in Study 2 are known to be associated with a higher processing load (e.g. Cutler, Eisner, McQueen, & Norris, 2010). These factors together may have reduced the opportunity for listeners to adjust sufficiently for talker gender in categorizing the ambiguous *stream* sibilants.

Future experiments are needed to test whether this asymmetry between production and perception extends to sibilants when listeners hear them in /str/ contexts (remembering that sibilants were all spliced into pre-vocalic contexts for Study 2). That is, whether listeners might be more likely to categorize /str/ produced by men as /ʃtr/, even when the reason for lower M1 (the upcoming /tr/) is available in the signal. While this issue remains to be addressed experimentally, at this early stage we would still like to suggest that it is possible that the combined acoustic effect on /s/ of gender (M1-lowering due to a male voice) and context (M1-lowering due to /tr/) may favour "ʃ" responses when listeners hear /str/ produced by males. That is, we would like to suggest that insufficient compensation in perception for the inherent acoustic differences between male and female speech *might* play a role in the initiation of sound change and in /s/-retraction in particular. This is a novel idea because it suggests an overlap between the (phonetic) factors typically thought to be responsible for the initiation of sound change and the (social) factors that help it to spread throughout a speech community (cf. e.g. Janda & Joseph, 2003 on this division). Such an interaction is of particular interest in terms of theories of sound change for three reasons.

First, gender is not normally thought to play a role in the very earliest stages of sound change. Indeed, the phonetic biases responsible for the initiation of a sound change are assumed to apply equally to all members of a speech community (e.g. Bybee, 2012:221 on articulatory biases). Sociolinguistic research on the spread of sound change shows that gender plays an important role, but differences between men and women are not thought to emerge before members of a particular community are aware of the novel variant. Cheshire (2002:429), for example, explains how "in the early stages of a change, sex differentiation is relatively small, but […] as the new forms become more widespread and speakers become consciously aware of them, sex differentiation becomes more marked". In light of this research, it was somewhat surprising to find an interaction of any sort between gender and /s/-retraction in Australian English, a variety in which this sound change is not yet underway. Nonetheless, results showing that listeners are more likely to categorize sibilants spoken by men as /ʃ/ suggest that they might also be likely to link /s/-retraction with male speech in their cognitive representations. This idea is consistent with an exemplar approach in which speech is linked not just to phonetic but also to speaker indexical categories (e.g. Johnson, 2006; Munson, 2010). Indexing retraction for gender would facilitate the spread of the novel variant through the community (with the difference between men and women's pronunciations eventually disappearing once the sound change is completed, e.g. Labov, 2001:283). Thus the results of this study suggest that the social factors that help to spread sound change can have their roots in phonetic biases and can be evident well before a sound change takes hold in a speech community. This idea is consistent with recent discussions in the sound change literature that have argued that the traditional division between the role of phonetic and social factors in sound change might be unnecessary or at least difficult to uphold (e.g. Stevens & Harrington, 2014; Thomas, 2011; Yu, 2013).

Second, based on the sociolinguistic literature on the role of gender in sound change, we did not expect to find an innovative form like /s/-retraction to be associated with male speech in particular. Sound changes that can be attributed to language-internal phonetic forces – like /s/-retraction - are often led by women (e.g. Labov, 2001; Thomas, 2011:291). Labov (2001:292) refers to these types of change as 'change from below' and based on a survey of the available data concludes that "women use higher frequencies of innovative forms than men do". Instead, as we just outlined above, the tendency to perceive sibilants produced by men as /ʃ/ in Study 2 suggests that members of the speech community would be more likely to associate /s/-retraction with male speech i.e. that males would lead this sound change were it to take place in Australian English.

Third, we did not expect the innovation to be located in speech perception: to the best of our knowledge, gender differences reported in the sociolinguistic literature always involve speech production (e.g. Labov, 2001). Instead, this study shows evidence that an effect of gender can emerge without any change in speech production behaviour (as far as can be determined based on one acoustic parameter) because the tendency towards M1-lowering in /str/ was common to men and women in production. This idea is broadly in line with the terms of Ohala's model of sound change – although he did not address gender – because /s/-retraction would be seen to originate in listener hypo-correction and not to involve any initial change to the speaker's production target.

In summary, this study has documented the phonetic pre-conditions that can give rise to the sound change involving /s/-retraction, and thus provides further information to complement existing studies on /s/-retraction which are based on varieties for which this sound change is either well established or has already progressed to completion (e.g. Rutter, 2011). Study 1 showed that the sibilant

[14] The mean pitch ranged between 163 – 223 Hz for the thirteen female talkers and between 94 and 125 Hz for all but one male speaker (M09) whose mean pitch during *-eet* was 179 Hz, falling within the range of the female data.

/s/ shows lower M1 in /str, spr, skr, sp, st, sk/ than in pre-vocalic position. This was interpreted as evidence of articulatory /s/-retraction, which involves gradient phonetic changes to an individual speaker's production target. Study 2 showed that sibilants in /st, str/ can be categorized as /ʃ/ by native listeners. The categorization of sibilants as /ʃ/ was more likely for male voices; this interaction between gender and contextual effects in the perception of sibilants suggests an overlap between the factors that give rise to sound change and those that facilitate its spread that is worthy of further investigation.

Acknowledgments

We thank Associate Editor Marija Tabain and three anonymous reviewers for their helpful comments and suggestions. This research was supported by ERC Grant no. 295573 (2012-2017) and Deutscher Akademischer Austausch Dienst/Group of Eight (Australia) travel grant 56266404. We are grateful to all of our participants and especially to Kate Stevens for invaluable assistance with data collection in Braidwood.

Appendix A. Production stimuli

Target words:

- /s/: *seem, sane*
- / ʃ/: *sheep, Shane*
- /st, sp, sk/: *steam, Spain, scheme*
- /str, skr, spr/: *stream, sprain, scream*

Fillers:

- *tree, train, car, chief, fish, flail, flame, goat, Australia, Peter, mouse, plan*

Appendix B. Perception according to familiarity with speaker voices

	Seam	*Sheep*	*Steam*	*Stream*
Braidwood	4%	100%	7%	32%
Other	3%	100%	7%	27%

Proportion of SHEET responses to sibilants spliced from *stream, steam, sheet, seat* for residents of Braidwood (*n* listeners = 6) and for the rest of the listener participants who would not have been familiar with any of the voices (*n* listeners = 16).

References

Baker, A., Archangeli, D., & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, *23*, 347–374.
Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*(4), 407–428.
Bladon, R. A. W., & Nolan, F. J. (1977). A video-flourographic investigation of tip and blade alveolars in English. *Journal of Phonetics*, *5*, 185–193.
Bukmaier, V., Harrington, J., & Kleber, F. (2014). An analysis of post-vocalic /s-ʃ/ neutralization in Augsburg German: evidence for a gradient sound change. *Frontiers in Psychology*, *5*, 1–12.
Bybee, J. (2012). Patterns of lexical diffusion and articulatory motivation for sound change. In M.-J. Solé, & D. Recasens (Eds.), *The initiation of sound change. Perception, production and social factors*, 323 (pp. 211–234). Amsterdam/Philadelphia: John Benjamins.
Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*, 209–244.
Cheshire, J. (2002). Sex and gender in variationist research. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *Handbook of Language Variation and Change* (pp. 423–443). Oxford: Blackwell.
Clopper, C. G. and D. B. Pisoni (2005). Perception of dialect variation. The Handbook of Speech Perception (pp. 313–337). D. B. Pisoni and R. E. Remez.
Cox, F., & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association*, *37*(3), 341–350.
Cox, F. (2012). *Australian english: transcription and pronunciation*. Melbourne: Cambridge University Press.
Cruttenden, A. (2014). *Gimson's pronunciation of English* ((8th edition). Oxford/New York: Routledge.
Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). *How abstract phonemic categories are necessary for coping with speaker-related variation. Variation, detail, and Representation. (LabPhon 10)* (pp. 91–111)Berlin: Mouton de Gruyter91–111.
Divenyi, P., Greenberg, S., & Meyer, G. (Eds.). (2006). *Dynamics of speech production and perception*. Amsterdam: IOS Press.
Docherty, G. J., Langstrof, C., & Foulkes, P. (2013). Listener evaluation of sociophonetic variability: Probing constraints and capabilities. *Linguistics*, *51*(2), 355–380.
Draxler, C. and K. Jänsch (2004). SpeechRecorder - A universal platform independent multichannel audio recording software. *Proceedings of the fourth international conference on language resources and evaluation,* Lisbon, Portugal.
Ellis, L., & Hardcastle, W. J. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data. *Journal of Phonetics*, *30*(3), 373–396.
Engstrand, O. and C. Ericsdotter (1999). Explaining a violation of the sonority hierarchy: Stop place perception in adjacent [s]. *Proceedings of the XIIth Swedish phonetics conference (FONETIK 99)*(pp. 49–52). Göteborg.
Fant, Gunnar (1970). *Acoustic theory of speech production*. The Hague/Paris: Mouton De Gruyter.
Haley, K. L., Seelinger, E., Mandulak, K. C., & Zajac, D. J. (2010). Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach. *Journal of Phonetics*, *38*, 548–554.

Harrington, J. (2012). The relationship between synchronic variation and diachronic change. In A. Cohn, C. Fougeron, & M. Huffman (Eds.), *Handbook of laboratory phonology* (pp. 321–332). Oxford: Oxford University Press.
Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: an acoustic and perceptual study. *Journal of the Acoustical Society of America*, *123*, 2825–2835.
Hoole, P., Nguyen-Trong, N., & Hardcastle, W. (1993). A comparative investigation of coarticulation in fricatives: electropalatographic, electromagnetic, and acoustic data. *Language and Speech*, *36*(2-3), 235–260.
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.
Iskarous, K., Shadle, C. H., & Proctor, M. I. (2011). Articulatory–acoustic kinematics: the production of American English /s/. *Journal of the Acoustical Society of America*, *129*(2), 944–954.
Janda, R. D., & Joseph, B. D. (2003). On language, change and language change. In R. D. Janda, & B. D. Joseph (Eds.), *The handbook of historical linguistics* (pp. 3–180). Malden/Oxford/Melbourne /Berlin: Wiley-Blackwell.
Johnson, K. (2006). Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. *Journal of Phonetics*, *34*, 485–499.
Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, *108*(3), 1252–1263.
Keating, P. (1990). The window model of coarticulation: articulatory evidence. In J. Kingston, & M. Beckman (Eds.), *Papers in Laboratory Phonology I* (pp. 451–470). Cambridge University Press.
Kisler, T., F. Schiel and H. Sloetjes (2012). Signal processing via web services: the use case WebMAUS. *Digital Humanities 2012*, Hamburg, Germany.
Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R. (2013). Toward improved spectral measures of /s/: results from adolescents. *Journal of Speech, Language, and Hearing Research*, *56*, 1175–1189.
Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: dialects, idiolects, and speech processing. *Cognition*, *107*, 54–81.
Kühnert, B., P. Hoole and C. Mooshammer. (2006). Gestural overlap and C-center in selected French consonant clusters. *Proceedings 7th International Seminar on Speech Production* (pp. 327–334). Belo Horizonte, UFMG.
Kümmel, M. J. (2007). Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen. Wiesbaden, Reichert.
Labov, W. (2001). *Principles of linguistic change: social factors*. Oxford UK: Blackwell.
Lawrence, W. P. (2000). /str/->/ʃtr/: Assimilation at a distance?. *American Speech*, *75*(1), 82–87.
Li, F., Munson, B., Edwards, J., Yoneyama, K., & Hall, K. (2011). Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: implications for cross-language differences in speech-sound development. *Journal of the Acoustical Society of America*, *129*(2), 999–1011.
Lindblom, B. (2004). *The organization of speech movements: specification of units and modes of control. Proceedings of 'From sound to sense'* (pp. 86–97)MIT86–97.
Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]–[s] distinction. *Perception & Psychophysics*, *28*, 213–228.
Milroy, J., & Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics*, *21*, 339–384.
Munson, B. (2010). Levels of Phonological Abstraction and Knowledge of Socially Motivated Speech- Sound Variation: A Review, a Proposal, and a Commentary on the Papers by Clopper, Pierrehumbert, and Tamati; Drager; Foulkes; Mack; and Smith, Hall, and Munson. *Journal of Laboratory Phonology*, *1*(1), 157–177.
Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). Acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics*, *34*, 202–240.
Niebuhr, O., Clayards, M., Meunier, C., & Lancia, L. (2011). On place assimilation in sibilant sequences—comparing French and English. *Journal of Phonetics*, *39*, 429–451.
Nolan, F., Holst, T., & Kühnert, B. (1996). Modelling [s] to [ʃ] accommodation in English. *Journal of Phonetics*, *24*, 113–137.
Ohala, J. (2012). The listener as a source of sound change. An update. In M.-J. Solé, & D. Recasens (Eds.), *The initiation of sound change. Perception, production and social factors* (pp. 21–36). Amsterdam/Philadelphia: John Benjamins.
Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the Parasession on language and behavior* (pp. 178–203). Chicago: Chicago Linguistics Society.
Perlman, G. (1985). Electronic surveys. *Behavior Research Methods, Instruments, and Computers*, *17*(2), 203–205.
Recasens, D., & Pallarès, M. D. (2001). Coarticulation, assimilation and blending in Catalan consonant clusters. *Journal of Phonetics*, *29*, 273–301.
Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America*, *102*(1), 544–561.
Rohlfs, G. (1966). *Grammatica storica della lingua italiana e dei suoi dialetti: fonetica*. Pisa: Einaudi.
Rutter, B. (2011). Acoustic analysis of a sound change in progress: the consonant cluster /st/ in English. *Journal of the International Phonetic Association*, *41*, 27–40.
Shapiro, M. (1995). A case of distant assimilation: /str/→/ʃtr/. *American Speech*, *70*(1), 101–107.
Solé, M. J. (2010). Effects of syllable position on sound change: an aerodynamic study of final fricative weakening. *Journal of Phonetics*, *38*, 289–305.
Stevens, K. (1998). *Acoustic phonetics*. Cambridge, Massachusetts: The MIT Press.
Stevens, M., V. Bukmaier, and J. Harrington (2015). Pre-consonantal /s/ retraction. *Proceedings of the 18th international congress of phonetic sciences* (pp. 1–5). Glasgow, August 10-14.
Stevens, M., & Harrington, J. (2014). The individual and the actuation of sound change. *Loquens*, *1*(1).
Thomas, E. R. (2011). *Sociophonetics: an introduction*. London: Palgrave Macmillan.
Warren, P. (1996). /s/-retraction, /t/-deletion and regional variation in New Zealand English /str/ and /stj/ clusters. *11th Australian International Conference on Speech Science & Technology* (pp. 466–471). P. Warren and C. Watson. Auckland.
Wells, J. (2011). How do we pronounce train? Retrieved 03 November, 2015, from ⟨http://phonetic-blog.blogspot.de/2011/03/how-do-we-pronounce-train.html⟩.
Winkelmann, R. and G. Raess (2014). Introducing a web application for labeling, visualizing speech and correcting derived speech signals. *Proceedings LREC 2014*, Reykjavik.
Yu, A. C. L. (2013). Preface. Origins of Sound Change: Approaches to Phonologization. A. C. L. Yu. Oxford, Oxford University Press.
Zsiga, E. C. (1995). An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. Papers in laboratory phonology IV (pp. 282–302). Cambridge: Cambridge University Press.