Published in final edited form as:
J Exp Psychol Hum Percept Perform. 2014 February ; 40(1): 200–217. doi:10.1037/a0033182.

Perceptual Learning of Speech under Optimal and Adverse Conditions

Xujin Zhang and
Psychology Department, Stony Brook University, the United States

Arthur G. Samuel
Basque Center on Cognition Brain and Language, Donostia Spain, IKERBASQUE, Basque Foundation for Science, Bilbao Spain, Psychology Department, Stony Brook University, the United States

Abstract

Humans have a remarkable ability to understand spoken language despite the large amount of variability in speech. Previous research has shown that listeners can use lexical information to guide their interpretation of atypical sounds in speech (Norris, McQueen, & Cutler, 2003). This kind of lexically induced perceptual learning enables people to adjust to the variations in utterances due to talker-specific characteristics, such as individual identity and dialect. The current study investigated perceptual learning in two optimal conditions: conversational speech (Experiment 1) vs. clear speech (Experiment 2), and three adverse conditions: noise (Experiment 3a) vs. two cognitive loads (Experiments 4a & 4b). Perceptual learning occurred in the two optimal conditions and in the two cognitive load conditions, but not in the noise condition. Furthermore, perceptual learning occurred only in the first of two sessions for each participant, and only for atypical /s/ sounds and not for atypical /f/ sounds. This pattern of learning and non-learning reflects a balance between flexibility and stability that the speech system must have to deal with speech variability in the diverse conditions that speech is encountered.

Keywords

perceptual learning; conversational speech; clear speech; noise; cognitive load

Listeners have a remarkable ability to understand spoken language despite the highly variable nature of the speech signal. Recent work has shown that listeners can use lexical context to guide their interpretation of atypical sounds in speech (Eisner & McQueen, 2005, 2006; Kraljic & Samuel, 2005, 2006; McQueen, Norris, & Cutler, 2006; McQueen, Cutler, & Norris, 2006; Norris, McQueen, & Cutler, 2003; see Samuel & Kraljic, 2009 for review). For example, not only is an ambiguous sound between /s/ and /f/ reported as /s/ in “Pari_”, and as /f/ in “sheri_”, but this difference also has an effect on later perception of these sounds. The flexibility of the speech system enables listeners to adjust to variations in utterances due to talker-specific characteristics, such as individual identity and dialect. The

Correspondence concerning this article should be addressed to Arthur G. Samuel, Paseo Mikeletegi 69, 2nd Floor, 20009, Donostia-San Sebastián, Gipuzkoa, España. a.samuel@bcbl.eu.

present study investigates how the speech system handles speech variability, as a function of the quality of the speech signal, and concurrent demands on the listener.

Norris, McQueen, and Cutler (2003) demonstrated that listeners will modify their phonetic category boundaries when lexical context provides information about phonetically ambiguous segments. They presented Dutch listeners with Dutch words that were manipulated so that word-final instances of /s/ or /f/ sounded ambiguous (i.e., as midway between /s/ and /f/) in an exposure phase. Subsequently, in a categorization phase, the listeners identified sounds on an /εs/-/εf/ continuum. Listeners who heard the ambiguous sound in /s/-final words categorized more sounds on the continuum as /εs/, while those who heard the ambiguous sound in /f/-final words categorized more sounds as /εf/. They called this effect “perceptual learning” of speech.

Subsequent studies replicated their findings using different types of stimuli, such as ambiguous fricatives between /s/ and /ʃ/ (Kraljic & Samuel, 2005), ambiguous stops between /d/ and /t/ (Kraljic & Samuel, 2006), and ambiguous vowels between /ɒ/ and /aɪ/ (Samuel & Kraljic, 2011). Other studies observed perceptual learning with tasks other than lexical decision in the exposure phase, such as counting the number of words presented (McQueen, Norris, & Cutler, 2006), or listening to a story without any decision task (Eisner & McQueen, 2006). In Eisner & McQueen’s (2006) study, participants passively listened to a 4-minute story in the exposure phase. The story contained 78 /s/ sounds and 78 /f/ sounds. The participants performed a categorization task on an /εs/-/εf/ continuum both before and after listening to the story. Listeners who heard a version of the story in which all the /s/ sounds had been replaced by an ambiguous sound categorized more sounds on the continuum as /εs/ than those who heard a version of the story in which all the /f/ sounds had been replaced by the ambiguous sound. Moreover, perceptual learning remained robust 12 hours after exposure to the ambiguous sounds.

Despite the broad set of conditions that generate perceptual learning, the effect is also restricted in various ways. For example, the listener must have sufficient lexical constraint to support phonetic selection: Exposure to ambiguous sounds at the beginning of words (Jesse & McQueen, 2011) or at the end of nonwords (Norris, et al., 2003) does not bias listeners’ interpretation of the ambiguous sound. Moreover, perceptual learning is talker-specific. Perceptual learning for ambiguous fricatives produced by one speaker is not generalized to the same sound produced by a novel speaker (Eisner & McQueen, 2005; Kraljic & Samuel, 2005).

In addition, perceptual learning for ambiguous sounds is blocked if they are preceded by standard pronunciations of the same sound spoken by the same speaker (Kraljic & Samuel, 2011; Kraljic, Samuel, & Brennan, 2008). For instance, when listeners first hear normally pronounced words, such as “medicine” (with a normally-pronounced /s/), and then hear words with ambiguous sounds, such as “dino?aur” (with an ambiguous sound between /s/ and /f/), by the same speaker, perceptual learning will be prevented. According to Kraljic et al. (2008, 2011), this is because listeners rely heavily on the initial experience with a speaker to build a model of this speaker’s pronunciation. If a model of a speaker has been built for his/her standard pronunciations of a particular sound, later-arriving variation of this sound will have a greatly reduced impact on the sound's representation. Presumably for this reason, several labs (e.g., our own; J. McQueen, personal communication, May 2012) have found that running a baseline (i.e., the categorization task on the continuum) before the exposure phase will also block the learning effect (an exception is Eisner & McQueen, 2006). Because the test continuum contains relatively standard tokens of the critical sound, this exposure during a baseline test can block the later learning that would otherwise be produced by ambiguous pronunciations of the same sound.

There are additional situations that have been shown to limit perceptual learning. Kraljic, Samuel, and Brennan (2008) found that learning did not occur when the speaker had a pen in her mouth while producing ambiguous sounds. The perceptual system also tends to remain stable when ambiguous sounds occur in certain phonemic contexts. For instance, in some dialects of English, place assimilation causes speakers to produce /s/ in a way that it is somewhere between /s/ and /ʃ/ when it is followed by /tr/. When listeners were exposed to ambiguous /s/ sounds only before /tr/, there was no perceptual learning (Kraljic, Brennan, & Samuel. 2008). Samuel and Kraljic (2009) have suggested that the pattern of effects and non-effects in the literature reflects a tension between the need for flexibility in adjusting to changing input and the need for stability in perceptual organization.

The existing literature on lexically induced perceptual learning has provided very useful clues to the balance of flexibility and stability needed for successful spoken word recognition. However, previous studies have tested perceptual learning only under optimal listening conditions, usually using simple stimuli (in almost all studies, single words). It is possible that previous work overestimates listeners' ability to adapt to atypical sounds, because language experience in real life situations is more complicated than what has been tested in previous work. For example, everyday conversation is usually intermixed with various background noises, and other tasks often compete for our attention when we are perceiving speech. Therefore, it is important to see how listeners perceive spoken language under more challenging conditions, using more complicated stimuli.

The purpose of the present study was to investigate perceptual learning effects under various listening conditions, by exposing listeners to ambiguous /s/ and /f/ sounds in sentences instead of single words. In particular, Experiments 1 and 2 tested perceptual learning under relatively optimal conditions, when there was no noise or distraction. Experiment 3a, 4a and 4b tested perceptual learning under relatively adverse conditions, when there was either noise to degrade the speech signal or a secondary cognitive load task to distract attention. Experiments 1, 2, 3a, and 4a each included two experimental sessions for each participant, with each session including two phases: Exposure and Categorization. In the first session, participants were randomly assigned to two experimental groups. During Exposure, one group was exposed to ***?s***-sentences, in which an /s/ sound in one word had been replaced by an ambiguous sound midway between /s/ and /f/. The other group was exposed to ***?f***-sentences, in which an /f/ sound had been replaced by an ambiguous sound midway between /s/ and /f/. During Categorization, both groups were given a continuum of /sΛ/-/fΛ/ sounds and were asked to identify the tokens on the continuum. In the second session, the subjects who had listened to the ***?s***-sentences listened to the ***?f***-sentences during Exposure, and vice versa. They then performed the same categorization task on the same /sΛ/-/fΛ/ continuum during Categorization.

The two sessions were separated by one week. We chose this relatively long delay to see if this separation would eliminate the exposure effect from the first session. The existing literature only tells us that perceptual learning is stable for at least 12 hours (Eisner & McQueen, 2006); there is no information in the literature about the time course of the blocking effect of prior exposure to the test continuum. If any effects dissipate within a week, then we can examine within-subject differences (across sessions) for /s/ versus /f/ exposure. If these effects remain, then we can assess perceptual learning by the difference between the two groups in the Categorization phase within each session.

In addition, to test whether the learning effect is symmetric for the /s/ and /f/ conditions within each session, a Baseline categorization function was established by having a separate group of participants perform the Categorization task on the test continuum without any exposure to ambiguous sentences. The Baseline categorization will be compared to the categorization performance of each experimental group within each session in each experiment. If the learning effect is symmetric, we should see similar sized shifts for the /s/ and /f/ conditions, in opposite directions, relative to the Baseline. As we discussed above, this between-subject comparison is necessary because collecting a within-subject baseline before the exposure phase is likely to block the learning effect.

Preliminary Study: Baseline

Method

Participants—Ninety-five listeners from Stony Brook University participated in the preliminary experiment. All participants were 18 years of age or older. None reported any hearing disorders, and all reported American English as their native language. They received either payment or research credit for participation.

Materials and procedure—A male speaker of standard American English recorded two syllables-/sΛ/ and /fΛ/- in a sound proof booth onto a PC, sampling at 16kHz. Each syllable was edited using Goldwave sound editing software and was saved in its own file. A /sΛ/-/fΛ/ continuum was created by mixing the two syllables using the same procedures used in prior studies (Kraljic, Brennan, & Samuel, 2008; Kraljic & Samuel, 2005; 2006; 2007; 2011; Kraljic, Samuel, & Brennan, 2008). First, the frication of the /s/ and /f/ sounds of each syllable were located and digitally mixed together. Nineteen mixtures with different weightings that varied from 95% /s/ and 5% /f/ to 5% /s/ and 95% /f/ were then constructed, and these were inserted into the two syllables by replacing the /s/ in /sΛ/ and the /f/ in /fΛ/. In this way, there were 19 mixtures with /sΛ/ as the “frame” and 19 mixtures with /fΛ/ as the “frame”. Three experienced researchers in the lab listened to the mixtures in each frame, and independently selected an ambiguous range of eight mixtures from whichever frame provided the cleanest sounding range. Since the choices of the three researchers were all from the /fΛ/ frame, the final set of eight mixtures was taken from that set, with weightings that reflected the central tendency of the three listeners. The selected tokens had weightings that ranged from 75% /s/ and 25% /f/ to 40% /s/ and 60% /f/, in 5% increments.

Up to three participants were tested at the same time in a sound proof booth. The participants listened to the members of the /sΛ/-/fΛ/ continuum and identified each token by pressing the SUH or the FUH button on a button board. For each token, the participants had up to 3s to respond. The next token was played 0.5s after the participants' response. Seventeen randomizations of the eight tokens on the /sΛ/-/fΛ/ continuum were presented. Since listeners often need a few trials to get used to synthetic or processed speech, the first two randomizations were practice and were not analyzed (Pitt & Samuel, 1993; Samuel & Kat, 1996).

Results and Discussion

The percentage of /f/ responses for each token on the continuum was calculated. Data from 15 out of 95 listeners were eliminated because these listeners did not identify the tokens consistently (for these listeners, the percentage difference of /f/ responses between the two endpoints was less than 70%)[1]. The mean percentages of /f/ responses to the eight tokens on the continuum for the remaining 80 participants were calculated. An arcsine transformation was then performed on the proportion data by multiplying the arcsine of the square root of the original proportion by two (Keppel & Wickens, 2004). An ANOVA with token as a within-subject factor was conducted on the transformed data. As expected, a significant main effect of token was found, $F(7, 553)=640.70$, $p<.001$, $\eta^2=.89$, with a lower percentage of /f/ responses on the /sΛ/ endpoint (3%) and a higher percentage of /f/ responses on the /fΛ/ endpoint (95%). Thus, the stimulus construction for the continuum was successful, with the eight tokens spread out nicely on the continuum. More importantly, this preliminary test provides a stable and accurate baseline that we will use in the following experiments to see if any given exposure condition produced labeling of these tokens that is significantly shifted away from the baseline.

Optimal Conditions

We noted in the Introduction that a central goal of the current project is to examine how perceptual learning varies as a function of the listening conditions. We begin by measuring these adjustments when people are hearing words spoken in sentences, under generally good listening conditions. As we pointed out, the vast majority of prior work on perceptual learning has been done with isolated words. And, in laboratory studies, isolated words are almost always recorded under unusually good conditions – noise is carefully controlled, and the speaker is instructed to produce the words clearly. In our first two experiments, we potentially make the listener's task a bit more difficult (and realistic) by presenting the critical sounds within sentences. Across the first two experiments, we contrast two types of speech: "conversational", and "clear" speech. The latter is akin to the type of speech that is typically recorded for speech experiments, whereas the former is more like the speech that listeners hear outside of the lab.

[1] This criterion was chosen so that the remaining participants were clearly those who could discriminate the tokens on the continuum, while not excluding too many participants [see Reinisch, E., Weber, A., & Mitterer, H. (2012) for discussion about the substantial individual differences in this type of task].

Multiple studies have compared speech recognition for these two speaking styles. Conversational speech is usually produced at a speaker's normal speaking rate, as if he/she is talking to native speakers or someone familiar with his/her voice and speech pattern. Clear speech is usually produced more slowly and precisely, to ease listeners' perceptual difficulty due to a noisy/reverberant environment, different language background, or a hearing impairment (Bradlow & Bent, 2002; Bradlow & Alexander, 2007; Smiljanic & Bradlow, 2009). This kind of clear speech, in contrast to conversational speech, involves a wide range of modifications, including the decreased speaking rate, more salient stop consonant releases, an expanded vowel space, a wider dynamic pitch range, the insertion of longer pauses, etc. (Bradlow & Bent, 2002; Smiljanic & Bradlow, 2005).

One of the seminal studies of clear speech was done by Picheny, Durlach, & Braida (1985), who found a significant clear speech benefit (a 17% intelligibility improvement) for hearing-impaired subjects listening to nonsense sentences. Clear speech has been shown to benefit children with and without learning disabilities (Bradlow, Kraus, & Hayes, 2003), elderly listeners with normal hearing and moderate hearing loss (Helfer, 1997), and non-native listeners (Bradlow & Bent, 2002; Bradlow & Alexander, 2007). Young normal-hearing listeners were also better at recognizing words in clear speech than in normal speech when noise and reverberation were present (Bradlow & Bent, 2002; Gagné et al., 1995; Krause & Braida, 2002).

The purpose of Experiments 1 and 2 was to investigate perceptual learning effects with conversational speech and clear speech, respectively. Based on Eisner and McQueen's (2006) results with story contexts, we expected similar perceptual learning to occur for conversational speech. Due to the multiple benefits of clear speech on sentence recognition discussed before, it seems likely that clear speech would produce at least the same, if not more, perceptual learning as conversational speech. On the other hand, it is possible that ambiguous sounds would be more noticeable in clear speech than in conversational speech, and therefore potentially block perceptual learning.

Experiment 1: Conversational Speech

Method

Participants—Fifty-seven listeners from Stony Brook University participated in Experiment 1. All participants were 18 years of age or older. None reported any hearing disorders, and all reported American English as their native language. They received either payment or research credit for participation.

Materials and procedure

Stimulus selection: Seventy-two English critical words were selected, ranging from two to four syllables. Thirty-six of them contained an /s/ sound but no /f/ sound (e.g., "compen**s**ation"), and the other 36 contained an /f/ sound but no /s/ sound (e.g., "counter**f**eit"). Most of the critical segments occurred at the beginning of the third syllable or later, with a smaller number occurring at the beginning of the second syllable. None of the words contained the sound /z/ or /v/, because they share features with the /s/-/f/ contrast. Critical /s/ words had 3.39 syllables on average and critical /f/ words had 3.14 syllables on average. The median SUBLTEX frequency of critical /s/ words was 4.5 and that of critical /f/ words was 6.9 (per million words; retrieved from http://subtlexus.lexique.org/moteur2/). This small difference was not significant, *t*(70)= −1.14, *p*=.258.

A high-predictability English sentence was created for each critical word (see Appendix A for a list of the sentences). In order to provide strong contextual information, each word was presented at or near the last position of the sentence. There were no /s/, /f/, /z/ or /v/ sounds in the sentences, except that the critical word had either an /s/ or /f/. The critical sentences were on average 12.0 words in length. Thirty graduate students who did not participate in the main experiment completed a pilot cloze test for the critical sentences. The mean cloze probability was 55% for critical /s/ words and was 62% for critical /f/ words. As with word frequency, this difference was not significant, *t*(70)= −1.21, *p*=.231. An additional 156 sentences were created, with no /s/, /f/, /z/ or /v/ sounds. Twelve of these were used for practice, and the remaining 144 served as filler items. The average number of words in filler sentences was 11.3.

Stimulus construction: Each of the 72 critical sentences and 156 filler sentences was recorded by the same male speaker who recorded the items used in the preliminary test. The speaker was asked to read the sentences in a conversational way, as if he was talking to someone familiar with his voice and speech pattern (Bradlow & Bent, 2002; Bradlow & Alexander, 2007; Smiljanic & Bradlow, 2009). For the critical sentences, the speaker also read a second version with the critical segment (/s/ or /f/) being replaced by its counterpart. For example, the speaker recorded both "*Due to an injury at work, the employee got compen**s**ation*" and "*Due to an injury at work, the employee got compen**f**ation*". The average speech rate for the conversational speech was 3.19 words per second. Each critical and filler sentence was edited using Goldwave sound editing software and was saved in its own file.

A unique ambiguous mixture was selected for each critical sentence following the same procedure used to create the /sΛ/-/fΛ/ continuum in the preliminary experiment. For each item (e.g., "compen**s**ation" and its mate "compen**f**ation"), the frication portions of the waveforms for the /s/ and /f/ sounds were located and then digitally mixed together. Nineteen mixtures with different weightings that varied from 95% /s/ and 5% /f/ to 5% /s/ and 95% /f/ were constructed for each such pair, and these were then inserted into two "frames" (in this example, the word "compen**s**ation", replacing the /s/, and the nonword "compen**f**ation", replacing the /f/). The same three researchers in the lab listened to the mixtures in each frame, and independently selected the most ambiguous mixture for each. The midpoint of the choices made by the three listeners was used as the most ambiguous mixture, using whichever frame sounded more natural. In the relatively small number of cases when there was not a clear consensus among the three independent judges, two or three additional listeners provided judgments to find the best mixture. The average mixture chosen was 56.5% /f/ for the critical /s/ words and was 53.8% for the critical /f/ words (see Appendix B for detailed information about the mixtures and frames used). The selected mixture was then put back into the sentence, replacing the original /s/ or /f/ sound.

Finally, two lists of sentences were created for use in the Exposure phase. One list had 36 ***?s***-sentences (sentences in which the /s/ sound in each critical word was replaced by the ambiguous sound) and 72 filler sentences. The other list had 36 ***?f***-sentences (sentences in which the /f/ sound in each critical word was replaced by the ambiguous sound) and the other 72 filler sentences. The filler sentences here are comparable to the filler words and nonwords in the lexical decision exposure task used in most perceptual learning studies, and are designed to minimize the likelihood that listeners will notice the critical mispronunciations. In those studies, the critical words usually comprise about 10% of the tokens in the exposure phase. In the current study, with about 12 words per sentence, critical words comprised about 3% of the words heard during the exposure phase. We asked our subjects after they finished the experiment about mispronounced words in the sentences, and virtually none of them reported hearing mispronunciations – the filler contexts were effective.

Procedure: Up to three participants were tested at the same time in a sound proof booth. In the Exposure phase, the participants wore headphones and listened to sets of three sentences read in a conversational way. In each set, a critical sentence was always preceded and followed by a filler sentence. The interval between the offset of one sentence and the onset of the next sentence was 1s. The participants were told to listen to the sentences, and were warned that they would be tested on the meaning of one of the three sentences. After each set, a probe sentence was presented on the screen; as we had warned the subjects, these probes were designed to ensure that they were listening to the sentences. Half of the probes were rewordings of an exposed sentence, with the same meaning as one of the three sentences; half of the probes clearly differed in meaning from the exposure sentences. The probe sentence could relate to any of the sentences in the set (equal probability of referring to the first, second, or third sentence). To avoid any phonological exposure issues, none of the probe sentences included words with /s/, /f/, /z/, or /v/. The participants were asked to decide whether the meaning of the probe sentence matched one of the three sentences they had just heard by pressing “YES” or “NO” buttons on a button board. The probe sentence stayed on the screen until the participant gave a response. If the participant failed to respond within 10s, the next set of sentences began. Otherwise, the next set began 1s after the response. The main experiment was preceded by a practice with four sets of sentences.

In the Categorization phase, the participants performed the same task as the Baseline group had done in the preliminary test. They listened to the members of the same /sΛ/-/fΛ/ continuum used in the preliminary test and were asked to identified each token by pressing a SUH or FUH button on the button board.

The experiment consisted of two sessions, with each session including both the Exposure and Categorization phases. In the first session, participants were randomly assigned to two experimental groups. In the Exposure phase, one group was exposed to ***?s***-sentences and the other group was exposed to ***?f***-sentences. In the second session, subjects who had listened to the ***?s***-sentences listened to the ***?f***-sentences, and vice versa. In the Categorization phase, they performed the same categorization task on the same /sΛ/-/fΛ/ continuum in each session. The two sessions were separated by one week.

Results and Discussion

The accuracies for the sentence probe task and the percentages of /f/ responses for the categorization task were calculated. Seven participants were excluded because they could not identify the items on the continuum consistently (the percentage difference of /f/ responses between the two endpoints was below 70% in either session). For the remaining 50 participants, half listened to the ***?s***-sentences in the first session, and half listened to the ***?f***-sentences in the first session.

Overall, the participants did well in the Exposure phase. Mean accuracies for the sentence probe task were .91 for the ***?f***-sentences and .90 for the ***?s***-sentences. For the Categorization task, the mean percentages of /f/ responses to the eight tokens on the continuum were calculated for each participant. We looked at perceptual learning effects separately for each session. The difference between the average percentages of /f/ responses for the two experimental groups indexed perceptual learning. They were also compared with the Baseline group to examine whether the learning effects were symmetric (Figure 1a &1b). The same arcsine transformation described in the preliminary test was performed on the proportion scores prior to the ANOVAs.

/s/ vs. /f/ Analyses—For the first session, an ANOVA with token (the eight /sΛ/-/fΛ/ sounds) as a within-subject factor and exposure (***?s*** vs. ***?f***) as a between-subject factor was conducted. As on the Baseline, there was a significant main effect of token, F (7, 336)=558.51, p<.001, η^2=.92, with a lower percentage of /f/ responses on the /sΛ/ endpoint (1%) and a higher percentage of /f/ responses on the /fΛ/ endpoint (96%). More crucially, there was a significant main effect of exposure, F (1, 48)=7.83, p=.007, η^2=.14: Listeners who were exposed to the ***?s***- sentences labeled fewer tokens on the continuum as /fΛ/ (44%) than those who were exposed to the ***?f***-sentences (50%). This demonstrates that exposure to ambiguous sounds in conversational sentences induced clear perceptual learning in the first session, which is consistent with the previous study that used sentence stimuli (Eisner & McQueen, 2006). There was also a significant interaction between token and exposure, F(7, 336)=2.89, p=.006, η^2=.06, reflecting larger effects in the ambiguous middle range of the test series.

For the second session, a comparable ANOVA was conducted on the /f/ responses for each group. Recall that the participants who listened to the ***?s***-sentences in the first session listened to the ***?f***-sentences in the second session, and vice versa. The main effect of token was significant, F(7, 336)=474.11, p<.001, η^2=.91. However, there was no main effect of exposure, F<1, and the interaction was also not significant, F<1. Thus, in the second session, the two groups of listeners did not differ in their perception of the tokens on the continuum (52% for those who listened to the ***?s***-sentences vs. 51% for those who listened to the ***?f***-sentences). The null effect in Session 2 suggests that the previous session impaired perceptual learning, even after a week. Thus, in all of the experiments we will focus on the within-session results, and on a comparison of effects to the Baseline.

Baseline vs. Exposure Condition Analyses—The previous analyses show that differential responding occurred for listeners in the ***?s*** and ***?f*** exposure conditions in Session

1, but not in Session 2. We can use the Baseline group to test whether the shifts in Session 1 were symmetric for the ***?s*** and ***?f*** conditions.

To investigate whether participants who listened to different sentences (***?s*** vs. ***?f***) shifted their phonetic category boundaries, we compared the percentage of /f/ responses of the experimental groups with those of the Baseline group. ANOVAs with token as a within-subject factor and exposure (***?s*** vs. Baseline, or ***?f*** vs. Baseline) as a between-subject factor were conducted on the transformed arcsine data, for each experimental group and for each session. As shown in Figure 1a, for the first session, participants who were exposed to the ***?s***-sentences labeled fewer tokens on the /sΛ/-/fΛ/ continuum as /fΛ/, compared to the Baseline group (44% vs. 50%). The difference was significant, *F*(1, 103)=9.63, *p*=.002, η^2=.09. In contrast, participants who were exposed to the ***?f***-sentences produced results quite like those for the Baseline group (50% vs. 50%); the difference was not significant, *F*<1. This result indicates that exposure in the first session to the ***?s***-sentences, but not the ***?f***-sentences, resulted in perceptual learning.

For the second session (Figure 1b), there was no significant difference between the participants who listened to the ***?s***-sentences (52%) and the Baseline group (50%), *F*(1, 103)=1.08, *p*=.301, η^2=.01. There was also no significant difference between those who listened to the ***?f***-sentences (51%) and the Baseline group (50%), *F*<1.

One important question that arises here is whether the participants actually shifted their phonetic category boundaries in the second session, with the shift being masked by other factors. To assess this, we begin with the fact that in the first session, only exposure to the ***?s***-sentences produced perceptual learning. From this starting point, there are three possible scenarios for what might have happened in the second session. First, if the first session with /f/ exposure had no consequences, we would have observed perceptual learning for the same group of participants who were exposed to the ***?s***-sentences in the second session. This did not occur. Second, if the learning effect for ***?s*** from the first session was stable after one week and no learning occurred for the same group of participants in the second session (with ***?f***), we would have observed the same results for this group in the second session as the first one – a difference between ***?s*** and Baseline. This also did not occur. Third, if the learning effect for ***?s*** did not last for a week and learning did not happen in the second session, we should observe no perceptual learning for either group in the second session. Our results are consistent with this third possibility. In this scenario, only ***?s*** is effective, and its impact may dissipate within a week. In addition, the testing procedures of the first session seem to block learning in the second session, even for ***?s***. We will return to this effect in the General Discussion.

Experiment 2: Clear Speech

Method

Participants—Fifty-seven listeners from Stony Brook University participated in Experiment 2. All participants were 18 years of age or older. None reported any hearing disorders, and all reported American English as their native language. They received either payment or research credits for participation.

Materials and Procedure—The second experiment used the same sentences in the Exposure phase as in the first experiment, except that the sentences were read in an unusually clear way. The same speaker was asked to read the critical and filler sentences as precisely as possible, as if he were talking to non-native listeners or someone with hearing problems (Bradlow & Alexander, 2007; Bradlow & Bent, 2002; Smiljanic & Bradlow, 2005; Smiljanic & Bradlow, 2009). The average speaking rate for the clear speech was 1.86 words per second, substantially slower than the rate for conversational speech. A unique ambiguous mixture of /f/ and /s/ was made for each clear critical sentence, in the same way as in the first experiment. The /sΛ/-/fΛ/ continuum used for the Categorization task was the same as in the preliminary test and the first experiment. The participants performed the same sentence probe task in the Exposure phase and the same categorization task in the Categorization phase.

The second experiment also consisted of two sessions, one week apart, and each session consisted of an Exposure phase and a Categorization phase. Half of the participants listened to the clear ***?s*** list in the first session, and half of them listened to the clear ***?f*** list. They all listened to the other sentence list in the second session. The Categorization task was the same across sessions.

Results and Discussion

The accuracies for the sentence probe task and the percentages of /f/ responses for the categorization task were calculated. Seven participants were excluded because they could not identify the items on the continuum consistently (the percentage difference of /f/ responses between the two endpoints was below 70% in either session). For the remaining 50 participants, half listened to the ***?s***-sentences in the first session, and the other half listened to the ***?f***-sentences in the first session.

As in Experiment 1, the participants did well in the Exposure phase. Mean accuracies for the sentence probe task were .91 for the ***?f*** sentences and .88 for the ***?s*** sentences. For the Categorization task, the mean percentage of /f/responses to the eight tokens on the continuum was calculated for each participant. The two experimental groups were first compared with each other within each session, and then were each compared with the Baseline group (Figure 2a &2b). The data were arcsine transformed prior to the ANOVAs, as before.

/s/ vs. /f/ Analyses—For the first session, an ANOVA with token (the eight /sΛ/-/fΛ/ sounds) as a within-subject factor and exposure (***?s*** vs. ***?f***) as a between-subject factor was conducted. There was the expected significant main effect of token, *F* (7, 336)=415.44, p<. 001, η2=.90, with a lower percentage of /f/ responses on the /sΛ/ endpoint (4%) and a higher one on the /fΛ/ endpoint (96%). Critically, there was a significant main effect of exposure, *F* (1, 48)=8.11, *p*=.006, η2=.15. Listeners who were exposed to the ***?s***-sentences labeled fewer tokens on the continuum as /fΛ/ (45%) than those who were exposed to the ***?f***-sentences (52%). This indicates that exposure to ambiguous sounds in clear sentences also induced perceptual learning in the first session, and the effect was numerically slightly larger than that of the conversational speech. As with Experiment 1, there was a significant interaction between token and exposure, *F*(7, 336)=2.23, *p*=.032, η2=.04, again reflecting the larger effect in the middle part of the test series than near the endpoints.

For the second session, a comparable ANOVA was conducted on the /f/ responses for each group. As usual, the main effect of token was significant, *F*(7, 336)=621.39.17, p<.001, η2=.93. There was no main effect of exposure or interaction, both *F*s<1. As we found in the first experiment, in the second session, the two groups of listeners did not differ in their perception of the tokens on the continuum (54% for those who listened to the ***?s***-sentences vs. 53% for those who listened to the ***?f***-sentences).

Baseline vs. Exposure Condition Analyses—As in Experiment 1, the /f/ responses of the two experimental groups were compared to those of the Baseline group. As shown in Figure 2a, for the first session, participants who were exposed to the ***?s***- sentences labeled fewer tokens on the /sΛ/-/fΛ/ continuum as /fΛ/, compared to the Baseline group (45% vs. 50%), *F*(1, 103)=4.13, *p*=.045, η2=.04. In contrast, participants who were exposed to the ***?f***-sentences behaved similarly to the Baseline group (52%vs. 50%), *F*(1, 103)=1.63, *p*=.204, η2=.02. As in Experiment 1, this result indicates that only the ***?s***-sentences resulted in perceptual learning; the ***?f***-sentences did not.

For the second session (Figure 2b), there was a marginally significant shift for the participants who listened to the ***?s***-sentences (54%) compared to the Baseline group (50%), *F*(1, 103)=3.76, *p*=.055, η2=.04. Note, however, that this marginal shift is actually in the opposite direction than the standard perceptual learning effect. There was no significant difference between those who listened to the ***?f***-sentences (53%) and the Baseline group (50%), *F*(1, 103)=3.34, *p*=.071, η2=.03.

Collectively, the pattern of results was almost identical to what was found in the first experiment. Although perceptual learning occurred in the first session, it was not observed in the second session (except for the marginal results for ***?s*** versus Baseline, an effect in the opposite direction of other perceptual learning shifts). Within the first session, only the listeners who were exposed to the ***?s***-sentences shifted their phonetic category boundaries to adjust to the atypical sounds. This asymmetry will be discussed in the General Discussion.

Adverse Conditions

The first two experiments showed that perceptual learning effects occurred under relatively optimal conditions (conversational and clear speech), replicating previous findings that showed the flexibility of the speech system to adjust to atypical sounds in utterances when there is no perceptual or cognitive load. However, in real life situations, speech is usually perceived under a wide range of suboptimal conditions.

There are many studies that investigate how speech recognition is affected by the presence of noise or reverberation. For normal-hearing native-speaking listeners, background noise impairs the ability to detect phonemes in words (Broersma & Scharenborg, 2010; Cutler, Weber, Smits, & Cooper, 2004; van Dommelen & Hazan, 2010) and to identify words in sentences (Bradlow & Bent, 2002; Cooke, Lecumberri, & Barker, 2008; Shi, 2010; see Lecumberri, Cooke, & Cutler, 2010, for review). Due to the compensation provided by lexical-semantic information, high-predictability sentences are more intelligible than low-predictability sentences in a noisy background (e.g., Mayo, et al., 1997). If the background noise consists of other languages, such as multi-talker babble and competing speech, a language known to the listener is more disruptive to the target speech than an unfamiliar language. For example, native English listeners were more adversely affected by English babble than by Chinese (Van Engen, & Bradlow, 2007) and Spanish babble (Lecumberri & Cooke, 2006), while Spanish (L1)-English (L2) bilinguals were equally affected by Spanish and English babble (Lecumberri & Cooke, 2006).

Not only is the speech signal usually experienced with perceptual challenges, such as background noises, it may be also experienced with some type of cognitive load. For example, taking notes while listening to a lecture imposes a cognitive load that may affect comprehension. Cognitive load is usually caused by situations that produce divided attention and/or that impose a memory load. In these cases, listeners must divide their attention between the primary task and a secondary task. If processing resources are limited (Kahneman, 1973), the secondary task will impair performance on the main task. A typical divided attention task would be presenting subjects with pictures or sounds and asking them to detect a specific visual target or to monitor for a certain sound, while performing the main task (Fernandes, Kolinsky, & Ventura, 2010; Mattys et al., 2010; Mattys & Wiget, 2011). A typical memory load task would be presenting subjects with a series of words before each trial of the main task and asking them to recall the words after the trial (Matttys, Brooks, Cooke, 2009).

Mattys and his colleagues (Mattys et al., 2009; Mattys et al., 2010) have conducted a series of interesting studies to investigate the effect of different adverse conditions on speech segmentation. These studies show that perceptual load (in their terms, “energetic masking”), such as speech-shaped noise, and cognitive load (“informational masking”) affect the relative weighting of speech segmentation cues differently. In Mattys et al. (2009), listeners heard two-word phrases (e.g., “mild option”), the pronunciation of which varied from being compatible with the lexically acceptable parse (“mild option”) to being in conflict with it (“mile doption”). Listeners used a rating scale to index the degree to which one of the two words was heard (e.g., “mild” or “mile”). In this example, responses favoring “mild” would be taken as an indication that the listener relied predominantly on the lexical-semantic cues of the phrase, while responses favoring “mile” would be taken as an indication that the listener relied predominantly on the acoustic-phonetic cues (because “doption” is non-lexical). Mattys et al. found that when the phrases were presented in strong speech-shaped noise (−8dB SNR), listeners tended to rely more on acoustic-phonetic cues for speech segmentation. However, when a memory load task was used, listeners tended to rely more on lexical-semantic cues. The authors concluded that severe perceptual load (e.g., speech-shaped noise) increases relative reliance on salient acoustic cues and curtails reliance on lexical-semantic information, whereas cognitive load results in the opposite pattern.

Experiments 3 and 4 examined perceptual learning effects for clear speech under different adverse listening conditions to see if the speech system is able to adapt to atypical sounds when listening conditions are more challenging. We chose clear speech, rather than conversational speech because the /s/ vs. /f/ analyses in the previous two experiments suggested that the perceptual learning effects were slightly larger for clear speech than for conversational speech. Our experiments follow Mattys et al. (2009) in terms of contrasting perceptual challenges and cognitive load.

In particular, Experiment 3a tested perceptual learning to clear speech, when signal-correlated noise was added to the speech. According to Mattys et al. (2009, 2010), listeners rely more on acoustic-phonetic information and less on the lexicon under noisy conditions. If listeners do reduce their reliance on lexical processing under these conditions, then perceptual learning should be reduced or eliminated in Experiment 3a.

Experiment 4 tested perceptual learning to clear speech under cognitive load. In Experiment 4a, participants were required to keep five consonants in memory while listening to the exposure sentences. In Experiment 4b the cognitive load was imposed by requiring participants to do a letter searching task while listening to the sentences. If perceptual learning is relatively automatic, we should observe results in the two cognitive load conditions that are similar to what we found in the optimal conditions tested in Experiments 1 and 2. However, if phonetic retuning requires cognitive resources, then we should find a reduction in such learning in the cognitive load conditions.

Experiment 3a: Clear Speech with Noise

Method

Participants—Fifty-five listeners from Stony Brook University participated in Experiment 3a. All participants were 18 years of age or older. None reported any hearing disorders, and all reported American English as their native language. They received either payment or research credit for participation.

Materials and Procedure—In the Exposure phase, the participants listened to the same clear sentences used in Experiment 2, except that signal-correlated noise was added to each of the critical and filler sentences. The noise was created first by generating white noise that had the same amplitude envelope as the original speech and then by adding the noise to the speech. It started and ended at the same time as the sentence.

After the noise-addition process, the original ambiguous mixtures of /s/ and /f/ (taken from the stimuli in Experiment 2) were spliced into their noise-added sentences. Thus, although the rest of the sentence was noise masked, the critical segment was not, so that it could provide the information needed for perceptual learning to occur. The average amplitude of the vowel that followed the critical segment was 67dB before adding noise and was 70dB after adding noise. The average amplitude of the unmasked ambiguous sound was 55dB. Since the critical segments themselves were aperiodic, like the white noise, they did not stand out from the rest of the sentence. The participants performed the sentence probe task in the Exposure phase, as in the first two experiments.

The same /sΛ/-/fΛ/ continuum was used for the Categorization task as in the previous experiments. No noise was added to the tokens on the continuum. The participants performed the same categorization task in the Categorization phase.

The design of Experiment 3a was the same as the previous two experiments. In each session, the participants first listened to sentences masked by signal-correlated noise (except for the critical ambiguous sounds) in the Exposure phase, and then identified the tokens on the /sΛ/-/fΛ/ continuum with no noise. The Categorization task was the same across sessions.

Results and Discussion

The accuracies for the sentence probe task and the percentages of /f/ responses for the categorization task were calculated. Five participants were excluded because they could not identify the items on the continuum consistently (the percentage difference of /f/ responses between the two endpoints was below 70% in either session). For the remaining 50 participants, half listened to the ***?s***-sentences in the first session, and half listened to the ***?f***-sentences in the first session.

For the sentence probe task in the Exposure phase, mean accuracies were .89 for the ***?f***-sentences and .88 for the ***?s***-sentences. The high accuracies demonstrate that the level of noise did not prevent listeners from understanding the sentences (Experiment 3b will provide more specific information relevant to this point – see below). For the Categorization task, the mean percentage of /f/responses to the eight tokens on the continuum was calculated for each participant. Again, the two experimental groups were first compared with each other within each session, and then each was compared to the Baseline group (Figure 3a and 3b). All of the mean percentages were arcsine transformed prior to the ANOVAs, as before.

/s/ vs. /f/ Analyses—For the first session, an ANOVA with token (the eight /sΛ/-/fΛ/ sounds) as a within-subject factor and exposure (***?s*** vs. ***?f***) as a between-subject factor was conducted. The main effect of token was significant, *F* (7, 336)=455.20, p<.001, η2=.91, with a lower percentage of /f/responses on the /sΛ/ endpoint (2%) and a higher percentage of /f/responses on the /fΛ/ endpoint (96%). The main effect of exposure was not significant, *F*<1, nor was its interaction with token, *F* (7, 336)=1.28, *p*=.258, η2=.03. Listeners who were exposed to the ***?s***-sentences did not significantly differ from those who were exposed to the ***?f***-sentences in labeling the tokens on the continuum (48% vs. 50%). This indicates that the signal-correlated noise prevented perceptual learning; the noise level was sufficient to have blocked any change in the phonetic category boundary.

For the second session, a comparable ANOVA was conducted on the /f/ responses for each group. The main effect of token was significant, *F*(7, 336)=479.23, *p*<.001, η2=.91. The main effect of exposure was not significant, *F*<1. The two groups of listeners did not differ in their perception of the tokens on the continuum (52% for those who listened to the ***?s***-sentences vs. 52% for those who listened to the ***?f***-sentences). The interaction between token and exposure was significant, *F*(7, 336)=3.39, *p*=.002, η2=.07, due to a slightly sharper labeling function for the /f/ exposure condition than for the /s/ condition.

Baseline vs. Exposure Condition Analyses—Figure 3a shows the results for the first session, along with the Baseline. There was no difference between the listeners who were exposed to the ***?s***-sentences and the Baseline (48% vs. 50%), *F*<1. There was also no difference between those who were exposed to the ***?f***-sentences and the Baseline (50% vs. 50%), $F<1$. Similar results were found for the second session, shown in Figure 3b. No difference was found between ***?s*** exposure and the Baseline (52% vs. 50%), $F<1$, or between ***?f*** exposure and the Baseline (52% vs. 50%), $F(1,103)=1.25$, $p=.265$, $\eta 2=.01$.

In contrast to the results of the previous two experiments, no perceptual learning occurred when the speech was presented with signal-correlated noise. This loss of the lexically-driven shift suggests that under noisy conditions, the perceptual system does not treat the ambiguous pronunciation of the critical segment as providing reliable enough information to retune the categorization boundary. Of course, if the signal was so noisy that listeners could not reliably understand the critical words, it would not be surprising that the lexical context was ineffective. Within Experiment 3a, there is reason to doubt that this is what happened: Performance on the sentence probe task was essentially as good as it had been in the first two experiments, indicating that the words were being heard well, despite the noise. To make sure that the critical words were in fact recognized in noise, we conducted a simple transcription experiment in which listeners were asked to write down the sentences, presented either in the clear or under the noisy conditions used in Experiment 3a.

Experiment 3b: Transcription

Method

Participants—Nineteen listeners from Stony Brook University participated in Experiment 3b. All participants were 18 years of age or older. None reported any hearing disorders, and all reported American English as their native language. They received either payment or research credit for participation.

Materials and Procedure—The transcription task included only the critical sentences: 36 ***?s***-sentences and 36 ***?f***-sentences. The participants listened to half of the ***?s***-sentences and half of the ***?f***-sentences in clear speech in one block, and listened to the other half of the ***?s***-sentences and ***?f***-sentences in signal-correlated noise in the other block. The materials were the same ones used in Experiment 3a. The order of the sentences was randomized within each block, and the order of the two blocks was counter balanced across participants. The participants were instructed to write down every word they heard from the sentence on a piece of paper, and to press a button to hear the next sentence when they were done. They never heard the same sentence in both clear and noisy speech.

Results and Discussion

Although the participants were required to write down the whole sentence, our interest is only in the critical ***?s*** and ***?f*** words: Were these well recognized? The accuracy of the critical words was calculated for clear speech and noisy speech, respectively. Both accuracies were very high and comparable (97% for clear speech vs. 95% for noisy speech), although the difference reached significance, $t(19)=2.33$, $p=.031$, $\eta^2=.23$. These results, together with the unimpaired performance on the sentence probe task in Experiment 3a, make it clear that the participants did have access to the critical words in the sentences, despite the noise. Therefore, the lack of perceptual learning in Experiment 3a was not due to an inability to hear the stimuli. We now examine whether cognitive load has a similar detrimental effect on the ability to adjust to segmental variation.

Experiment 4a: Clear Speech with Cognitive Load (Letter Recognition)

As we noted above, Mattys and his colleagues (Mattys et al., 2009; Mattys et al., 2010) have reported an interesting dissociation between perceptual challenges (e.g., noise) and cognitive challenges (e.g., a concurrent task) on speech segmentation: Noisy conditions led listeners to reduce their reliance on lexical/semantic information, while a concurrent task increased the impact of such information. We have seen in Experiment 3a that lexically-driven perceptual learning is reduced under noisy conditions. In Experiment 4a, we test whether imposing a concurrent memory load task also interferes with perceptual learning.

Method

Participants—Fifty-eight listeners from Stony Brook University participated in Experiment 4a. All participants were 18 years of age or older. None reported any hearing disorders, and all reported American English as their native language. They received either payment or research credit for participation.

Materials and Procedure—The participants listened to the clear sentences from Experiment 2 in the Exposure phase. No noise was mixed with the speech, but an additional cognitive load task was added to the procedure. Before each sentence, five consonant letters were presented visually for 3s. The letters were presented as a horizontal string in the center of a computer screen in front of the participant, in lower case; all consonant letters except f, s, v, x, and z were used[2]. After the sentence was played, four letters were displayed on the screen, spread out horizontally. Three of the four were ones that had been present in the set of five shown before the sentence, and one had not been. The horizontal positions of the letters were randomly selected separately for the initial and test phases, so that participants could not use position as a cue. Participants were asked to report which was the new letter by pressing one of four buttons on a button board, using the left-to-right location of the new letter to match the left-to-right arrangement of the four buttons. The four letters stayed on the screen until the participant gave a response. If the participant failed to respond within 5s, a new trial would start. As in the previous experiments, after each set of three sentences, the participants made a “Yes” or “No” response to a visual probe sentence that assessed comprehension. The /sʌ/-/fʌ/ continuum used for the categorization task was the same as the previous experiments; no memory load task was added to it.

Experiment 4a consisted of two sessions, as in the previous experiments. Thus, the procedure was that same as in Experiment 2, other than the addition of the memory load task during the Exposure phase.

[2] The visual letter “c” was inadvertently included in the memory load task, potentially having some phonological influence because it is sometimes pronounced as /s/, but it only occurred on 7% of the trials (and two thirds of its occurrences would have been during filler sentences).

Results and Discussion

The accuracies for the sentence probe task and the letter recognition task, and the percentages of /f/responses for the categorization task, were calculated. Eight participants were excluded because they could not identify the items on the continuum consistently (the percentage difference of /f/ responses between the two endpoints was below 70% in either session). For the remaining 50 participants, half listened to the ***?s***-sentences in the first session, and the other half listened to the ***?f***-sentences in the first session.

For the sentence probe task in the Exposure phase, mean accuracies were .85 for the ***?f***-sentences and .84 for the ***?s***-sentences, about 5% lower than performance on this judgment in the three experiments without a cognitive load. A one-way ANOVA with Experiment as a between subject factor showed that there was a significant difference among the experiments, *F* (3, 396)=13.28, *p*<.001, η2=.09. Pairwise comparisons showed that performance in the cognitive load condition was significantly lower than that of the other three conditions, *p*s<.001. This demonstrates that the additional letter recognition task successfully provided an adverse listening condition. Moreover, the average accuracy on the letter recognition task was .85, indicating that the listeners also paid attention to the secondary task.

For the categorization task, the mean percentage of /f/responses to the eight tokens on the continuum was calculated for each participant and was arcsine transformed prior to the ANOVAs. Figure 5a & b present the identification functions.

/s/ vs. /f/ Analyses—The same analyses were conducted as before. For the first session, an ANOVA with token (the eight /sΛ/-/fΛ/ sounds) as a within-subject factor and exposure (***?s*** vs. ***?f***) as a between-subject factor was conducted. The main effect of token was significant, *F* (7, 336)=395.34, *p*<.001, η2=.89, with a lower percentage of /f/ responses on the /sΛ/ endpoint (4%) and a higher percentage of /f/responses on the /fΛ/ endpoint (95%). Critically, there was a significant main effect of exposure, *F* (1, 48)=4.59, *p*=.037, η2=.09. Listeners who were exposed to the ***?s***-sentences labeled fewer tokens on the continuum as /fΛ/ (46%) than those who were exposed to the ***?f***-sentences (50%). Thus, despite the substantial cognitive load, exposure to ambiguous sounds produced perceptual learning. The interaction between token and exposure was not significant, *F*(7, 336)=1.15, *p*=.333, η2=.02.

For the second session, a similar ANOVA was conducted on the /f/responses for each group. The main effect of token was significant, *F*(7, 336)=529.98, *p*<.001, η2=.92. The main effect of exposure was not significant, *F*<1, and neither was the interaction between token and exposure, *F*(7,336)=1.04, *p*=.403, η2=.02. As we found in the previous experiments, in the second session, the two groups of listeners did not differ in their perception of the tokens on the continuum (49% for those who listened to the ***?s***-sentences vs. 50% for those who listened to the ***?f***-sentences).

Baseline vs. Exposure Condition Analyses—Again, the /f/responses of the two experimental groups were each compared with the Baseline. As shown in Figure 5a, for the first session, participants who were exposed to the ***?s***-sentences labeled fewer tokens on the /sΛ/-/fΛ/ continuum as /fΛ/, compared to the Baseline (46% vs. 50%), *F*(1, 103)=4.02, *p*=.047, η2=.04. In contrast, participants who were exposed to the ***?f***-sentences behaved similarly to the Baseline group (50%vs. 50%), *F*<1. As in Experiments 1 and 2, the ***?s***-sentences resulted in perceptual learning, and the ***?f***-sentences did not. For the second session (Figure 5b), there were no differences from Baseline (50%) for either the ***?s*** group (49%) or the ***?f*** group (50%), both *F*s<1.

Thus, the results were essentially identical to what was found under optimal listening conditions: 1) perceptual learning occurred in the first session, but not in the second session; and 2) even within the first session, only the listeners who were exposed to the ***?s***-sentences shifted their phonetic category boundaries to adjust to the atypical sounds. The close similarity of the results under cognitive load suggests that the adjustment process is relatively automatic.

Although the results of Experiment 4a suggest such automaticity, there are some issues that should be considered. For example, due to the nature of the memory load task, trials in Experiment 4a were longer than the trials in Experiments 1, 2 and 3a. It is possible that the lower accuracy on the sentence probe task in Experiment 4a was due to the longer time between an exposure sentence and the probe sentence. A more general concern is that perhaps the letter recognition task was not difficult enough – the accuracies for both the primary and secondary tasks were still relatively high (around .85).

To address these concerns, we ran another load experiment, one in which the load task does not affect the duration of a trial. In Experiment 4b, we replaced the letter recognition task with a letter search task which required participants to search for a repeated letter among a large number of consonant letters while listening to the sentences. Responses to the letter search task were only required after each set of sentences. This task was designed to provide a more demanding load task while keeping the duration of each set the same as Experiment 1, 2, and 3a. Because the previous experiments consistently found no perceptual learning effect for the ***?f***-sentences or for the second session, in Experiment 4b we only tested the ***?s***-sentences, in a single session.

Experiment 4b: Clear Speech with Cognitive Load (Letter Search)

Method

Participants—Thirty listeners from Stony Brook University participated in Experiment 4b. All participants were 18 years of age or older. None reported any hearing disorders, and all reported American English as their native language. They received either payment or research credit for participation.

Materials and Procedure—The participants listened to sets of three sentences in clear speech in the Exposure phase, as in Experiment 4a. The sets contained the 36 ***?s***-sentences and their accompanying 72 filler sentences. A letter search task was added to the primary task. At the onset of each sentence, ten upper case consonant letters, separated by 9 spaces, were presented in a row across the screen. In each set of ten letters, there might or might not be one repeated letter. The participants were instructed to search for a repeated letter while listening to each sentence. All consonant letters except C, V, S, F, X, and Z were included.

At the offset of each sentence, the screen was cleared. Pilot testing was used to determine the number of letters and the spacing needed to keep the participants searching throughout the time that a sentence was playing. The interval between the offset of one sentence and the onset of the next sentence remained the same as Experiments 1, 2, 3a and 4a.

After every three sentences, the participants were asked to select the last repeated letter from four possible choices presented on the screen by pressing one of the four buttons on the button board. The letters stayed on the screen for up to 3s until the participants responded. Then a visual probe sentence was shown on the screen, and the participants made a “Yes” or “No” response to the sentence according to its meaning, as in the previous experiments. For each group of three sentences, there might be one, two or three letter repetitions. Because participants never knew whether there would be a repeated letter presented during the next sentence, they needed to remember each repetition because it might be the last one. Unlike the letter recognition task used in Experiment 4a, which required a response after every sentence, the letter search task only required one response after every three sentences. This ensured that the duration of each set of sentences was essentially the same as it had been in Experiments 1, 2 and 3a.

The /sΛ/-/fΛ/ continuum used for the categorization task was the same as in the previous experiments. Because only the ***?s*** list was tested, there was only one session in Experiment 4b.

Results and Discussion

The accuracies for the sentence probe task and the letter search task, and the percentages of /f/responses for the categorization task, were calculated. Three participants were excluded because they could not identify the items on the continuum consistently (the percentage difference of /f/ responses between the two endpoints was below 70%).

For the sentence probe task in the Exposure phase, the mean accuracy was .76, and for the letter search task, the accuracy was .73. These values are noticeably lower than the corresponding values in Experiment 4a, indicating that the load task in the current experiment met our goal of providing the listeners with a more challenging concurrent task. Figure 4 shows the average accuracy on the sentence probe task for Experiment 1 (conversational speech), Experiment 2 (clear speech), Experiment 3a (noise), Experiment 4a (letter recognition task), and Experiment 4b (letter search task). A one-way ANOVA with Experiment (Experiment 1 vs. 2 vs. 3a vs. 4a vs. 4b) as a between subject factor showed that there was a significant difference among the five experiments, F (4, 421)=24.90, p<.001, η2=.19. Pairwise comparisons showed that performance in Experiment 4b was significantly lower than that in the other four experiments, ps<.001. This confirms that the letter search task did provide a more demanding cognitive load.

For the categorization task, the mean percentage of /f/responses to the eight tokens on the continuum was calculated for each participant and was arcsine transformed prior to the ANOVAs. Since there was only one group of participants, their performance on the categorization task was compared to the categorization by the Baseline group. As shown in Figure 6, participants who were exposed to the ***?s***-sentences labeled fewer tokens on the /

sΛ/-/fΛ/ continuum as /fΛ/, compared to the Baseline (45% vs. 50%), *F*(1, 105)=3.95, *p*=.049, η2=.04. Therefore, even when a more demanding cognitive load task was added to the primary task, perceptual learning remained intact. Thus, across Experiments 3 and 4, we see that signal-correlated noise blocked perceptual recalibration (despite intact sentence comprehension), whereas cognitive load left it intact (while impairing sentence comprehension).

General Discussion

The current study investigated the flexibility of listeners' speech systems to adjust to ambiguous sounds during spoken language processing, as a function of listening conditions. We looked at lexically induced perceptual learning in two relatively optimal conditions, as well as three adverse conditions. We found significant perceptual learning effects for conversational speech (Experiment 1) and for clear speech (Experiment 2). We also found similar learning effects in two cognitive load conditions (Experiments 4a and 4b), but not in a noise condition (Experiment 3a). Across the first three experiments that showed perceptual learning effects (Experiments 1, 2 and 4a), we also found that the results differed across different atypical sounds, and across initial versus later sessions. Only exposure to ***?s***-sentences and only the first session induced perceptual learning.

Signal-Correlated Noise vs. Cognitive Load

The most important finding of the current study is that under the conditions tested here, signal-correlated noise eliminated perceptual learning whereas cognitive load did not. In Experiment 3a, sentences, except for the critical ambiguous sounds, were degraded by signal-correlated noise. Performance on the sentence probe task was essentially unaffected but the perceptual learning effect was blocked: There was no difference between the two experimental groups on the categorization task, or between either experimental group and the Baseline group. The intact performance on the sentence probe task and the very high transcription rates in Experiment 3b demonstrate that the critical words were well recognized in the noise. Thus, with a level of signal-correlated noise that did not impair sentence understanding, perceptual learning was nonetheless abolished. A plausible possibility is that when the signal is quite noisy, the threshold for making a variability-based adjustment is raised. Because all of the speech is very variable (due to the noise), the variability of the ambiguous segment is not interpreted as a reliable cue calling for phonetic retuning.

This breakdown is consistent with the Conservative Adjustment and Restructuring Principle (CARP) that Samuel and Kraljic (2009) have advanced: The speech system must provide both adaptability and stability, and to achieve the latter, recalibration of category boundaries only occurs when there is strong evidence that it should occur. Under optimal conditions like those in Experiments 1 and 2, the possibility of atypical sounds being caused by external factors, such as noise, was very small. The speech system will therefore attribute the atypical sounds to factors specific to the speaker, and adjust its phonetic boundary accordingly. Under the noise condition of Experiment 3a, in contrast, there was considerable signal variability, making it possible that the atypical sounds were a result of whatever external factors were producing the noisy signal. Under these circumstances, a conservative system should not make changes in its categories because the lexical information supporting such changes is of marginal reliability.

Cognitive load produced a very different pattern than what we found with the signal-correlated noise. In Experiment 4a, a letter recognition task was added to the primary task in the Exposure phase. Although lower scores on the probe recognition task showed that sentence comprehension was impaired, the perceptual learning effects were similar to what we had seen under optimal conditions. Even with a more demanding cognitive load task – the letter search task in Experiment 4b – listeners continued to adjust to ambiguous sounds by shifting their phonetic boundaries, despite the evident difficulty posed by both tasks. This indicates that the retuning is a relatively robust process that does not require full attention. This is not to say that perceptual learning can always operate normally regardless of other demands on the system. In fact, in recent work in our laboratory (Samuel, in preparation), we have been able to disrupt such learning by imposing processing demands at exactly the moment when listeners would otherwise be processing the critical phonetic ambiguity.

It is important to note that the different effects of noise (Experiment 3) and cognitive load (Experiments 4a and 4b) on perceptual learning are not a result of the noise condition being “harder” than the load condition. If this were the case, we should have seen better performance on the sentence understanding judgment in Experiments 4a and 4b than in Experiment 3. In fact, we found exactly the reverse – sentence understanding was significantly worse under cognitive load than with noise, yet the perceptual learning was blocked by noise, not cognitive load. Furthermore, the similar effects found for the cognitive load conditions (Experiments 4a and 4b) and the corresponding optimal condition (Experiment 2) are not simply due to the use of the same clear speech stimuli in the two experiments. There are many demonstrations of identical stimuli producing different performance under cognitive load. For instance, Toro, Sinnett, and Soto-Faraco (2005) showed that listeners who performed a concurrent task while listening to speech were worse in extracting words from the speech than those who only passively listened to it. Casini, Burle, and Nguyen (2009) found that vowels were perceived as shorter when attention was divided than when it was not. In these cases, the stimuli remained intact, but cognitive load impaired performance on the primary task. We observed a comparable impact on the average accuracy on the sentence probe task under cognitive load, confirming the effectiveness of the load manipulation, but there was no impact on the perceptual learning effect.

Previous studies have investigated the effect of cognitive load on speech segmentation (Mattys et al., 2009, 2010) and phoneme-identification (Mattys & Wiget, 2011). Mattys et al. (2009) found that listeners tended to segment phrases in a way that was more lexically acceptable when a cognitive load task was added. Mattys and Wiget (2011) asked listeners to decide whether they heard /g/ or /k/ when ambiguous sounds between /g/ and /k/ were presented in the context of “_iss” and “_ift” (Ganong, 1980) and found that the Ganong effect was significantly larger in a cognitive load condition than in a load-free condition. In the current study, perceptual learning was intact, but did not increase, under cognitive load. Thus, cognitive load may have a similar but slightly different effect on perceptual learning than on segmentation or phonetic identification. It neither enhanced lexically induced perceptual learning (as in segmentation or identification) nor disrupted it (as signal-correlated noise had); the /f/ responses in the categorization task were comparable across Experiments 2, 4a and 4b (45% vs. 46% vs. 45%), and all were significantly shifted relative to the Baseline (50%).

First Session vs. Second Session

Across multiple experiments, we consistently found that when perceptual learning occurred, it only occurred in the first session; the identification functions in the second session of all four experiments were indistinguishable from Baseline. We have suggested that the consistently null second-session perceptual learning indicates that any learning from the first session was gone a week later, and that new learning in the second session did not occur after having listened to the stimuli in the first session. Neither of these two proposals has been tested directly in previous studies, but there are some hints in the literature to support at least the second one.

With respect to our first suggestion, there simply is no prior evidence that bears on the question of the long-term durability of lexically-driven perceptual learning shifts. The two most relevant prior studies examined a much shorter time frame: Perceptual learning effects were stable after a 25 minute delay (Kraljic & Samuel, 2005), and after 12 hours (Eisner & McQueen, 2006). Clearly, more research is needed to examine the time course of any change in perceptual retuning over time. It seems reasonable for the speech system to return to normal after a relatively long time, such as a week, and the current results support this conclusion. Future studies should examine intervals between 12 hours and a week.

What has been clearly shown by previous studies is that an atypical sound will not generate retuning if it is preceded by standard pronunciations of the same sound (Kraljic, Samuel, & Brennan, 2008; Kraljic & Samuel, 2011). It is likely that perceptual learning in the second session of the current study was blocked because the listeners had been exposed to relatively standard tokens of the critical sounds in the first session, for the same reason that a within-subject baseline blocks the learning effect. We had included a week's separation between the two testing sessions in the hope that this would be long enough to allow a within-subject test. Our results indicate that the inoculation against retuning caused by initial exposure to good tokens still operates a week later. It is interesting, if somewhat ironic, that the retuning itself does not appear to endure for this time (if it did, we would have seen the /s/-induced shifts from the first session remaining when the same individuals were tested a week later with the ineffective /f/ items; no such shifts were seen).

?s-Sentences vs. *?f*-Sentences

Another consistent finding across the four experiments was that only the listeners who were exposed to the ***?s***-sentences showed perceptual learning, even within the first session. The listeners who were exposed to the ***?f***-sentences, in contrast, did not differ from the Baseline group in the categorization task.

This asymmetry was unexpected. In principle, the null effect for ***?f*** could have occurred if the ambiguous tokens we constructed were somehow not well-positioned. Given that we used the same construction procedures that we have used successfully in a number of other studies (e.g., Kraljic, Brennan, & Samuel, 2008; Kraljic & Samuel, 2005, 2006, 2007, 2011; Kraljic, Samuel, & Brennan, 2008), this seems unlikely. It seems particularly unlikely because the same null effect was found for two different stimulus sets in the current study (conversational vs. clear speech), which were recorded and selected independently.

Having observed this surprising asymmetry, we went back to the literature to check whether a similar asymmetry had been found with stimuli similar to ours. Eisner and McQueen's (2006) study is the only one that used ***?s*** and ***?f*** as the critical sounds that also examined perceptual learning in sentences. The authors compared /ɛs/-/ɛf/categorization for participants who listened to a story containing ambiguous /s/ sounds (***?s*** group) to those who listened to the story with ambiguous /f/ sounds (***?f*** group). Perceptual learning was indexed by the difference in the categorization between these two groups after the exposure phase. They found no difference between the two groups before the exposure phase (54% /f/-responses for the ***?s*** group vs. 55% for the ***?f*** group), but found a significant difference between them after the exposure phase (46% for the ***?s*** group vs. 54% for the ***?f*** group). The post-exposure analyses and results are completely consistent with what we found in our /s/ vs /f/ analyses in Experiments 1 and 2.

Their study did not report a comparison of categorization before and after the exposure phase for each group (***?s*** or ***?f***) separately. We re-plotted the data, based on the figures in the paper, and discovered an asymmetric learning pattern, very similar to what was found in the current study. There was almost no categorization difference before versus after the exposure phase for the participants who listened to the story that contained ambiguous /f/ sounds (based on our replotting, approximately 55% vs. 54%). In contrast, there was a robust difference for those who listened to the story with the ambiguous /s/ sounds (54% vs. 46%). This asymmetric learning pattern for ***?s*** and ***?f*** has not been found in studies using single words (McQueen, Norris, & Cutler, 2006; Norris, McQueen, & Cutler, 2003).

The asymmetry might be a consequence of the frication cue for /f/ being acoustically much weaker than it is for /s/. Previous research analyzing single words showed that the frication in sibilants (e.g., /s/) is about 10–15 dB greater amplitude than in non-sibilants (e.g., /f/) (Jongman, Wayland, & Wong, 2000). Acoustic analyses on the running speech used in the present study also showed that the frication in /s/-sentences was much stronger than the frication in /f/-sentences, for both conversational speech (+12 dB) and clear speech (+10 dB). Because there is more acoustic variation in sentences than in single words, and the frication cue to /f/ is relatively weak, /f/ may be more susceptible to variation than /s/; listeners might therefore not make adjustments when experimentally induced variation in /f/ occurs because such variation is typical for that sound. There is evidence that non-sibilants (e.g., /f/) are more difficult to identify than sibilants (e.g., /s/) for hearing impaired listeners and for normal hearing listeners under difficult listening conditions (Maniwa, Jongman, & Wade, 2008). These factors all are consistent with our suggestion that the speech system might be less sensitive to missing or atypical /f/ sounds than to missing or atypical /s/ sounds. When ambiguous /s/ and /f/ sounds are tested with less-optimal stimuli, such as in sentences instead of in isolated words, the speech system adjusts only to ambiguous /s/ sounds because non-standard /f/ sounds are the norm, not an exception. This analysis is similar to our explanation for the lack of retuning in the noisy conditions of Experiment 3a: In both cases, the ambiguity of the critical segment is not salient enough, given the context.

Conclusions

Investigating perceptual learning under optimal and adverse conditions provides insight into how the speech system handles speech variability in situations more like those outside the laboratory. The results of the current study confirm (Eisner & McQueen, 2006) that the system does indeed adapt to mispronunciations in relatively good listening conditions, even when the speech is complicated (e.g., in sentences). The system is also flexible enough to adjust for atypical sounds under cognitive load conditions, even when the load condition is relatively demanding. This indicates that the recalibration process is relatively automatic and can operate without needing full access to the listener's cognitive resources, at least within the range that we have tested. In contrast, lexically-driven recalibration broke down under the noise condition.

We have also provided evidence that exposure to standard pronunciations of a sound prevents perceptual learning for the same sound, even after a week's delay. Interestingly, as we have noted, this finding suggests that the blockage from exposure lasts longer than the recalibration effect itself. From a methodological perspective, our results support the use of between-subject designs in future perceptual learning studies, despite the appeal of within-subject pretest-posttest designs. In addition, the asymmetry we have observed demonstrates that although the speech system is flexible enough to adapt to mispronunciations under various conditions, its ability to do so varies with the nature of the sound being tested. An intriguing but as yet unproven notion is that these asymmetries can be leveraged to learn more about the ways that the perceptual system deals with the natural variation among different speech sounds, much as the growing literature on perceptual learning is revealing how the system deals with variation within a speech sound category.

Acknowledgments

This research was supported by NIH grant R01059787. We would like to thank Donna Kat, Elizabeth Cohen, April Pufahl, Annalisa Sciullo and Steven Licardi for help on this project. We thank Sven Mattys and Dennis Norris for very helpful comments on an earlier version of this paper.

Appendix A: Critical /s/- and /f/- Sentences

/s/-sentences
Her commute took more than an hour due to the three car **accident**.
When you need work, you go to the employment **agency**.
Due to the economic downturn, the company declared **bankruptcy**.
In the ring, you need to be able to punch and to take a punch, to be a champion **boxer**.
When the check came in the mail, the appointment with her lawyer could be **cancelled**.
People watched the gladiator combat in the Roman **coliseum**.
Due to an injury at work, the employee got **compensation**.
In order to comprehend her homework, she needed a really quiet room where she could **concentrate**.

/s/-sentences
The arrogant man talked to poor people in a **condescending** manner.
Only the Mint can print American **currency**.
We would be a dictatorship without **democracy**.
By digging in the ground, you could locate a **dinosaur** bone.
Rather than a military approach, the general wanted to try **diplomacy**.
When the patient reacted badly to the medication, the doctor had to lower the **dosage**.
An air conditioner cannot run without **electricity**.
To get an immigration document, you would go to the American **embassy**.
The doctor operated on the patient in the **emergency** room.
Bob liked the new comedy show a lot, and watched each **episode** happily.
The teacher rubbed out the chalk mark with an **eraser**.
Donating money to the church, the lady showed great **generosity**.
Running a mile in under a minute would be **impossible**.
The man in jail claimed to be **innocent**.
You can turn a dimmer down to lower the light **intensity**.
Turn right when you reach the **intersection**.
The camper lit the lamp that ran on **kerosene**.
By the time Michael Jordan retired, he had created an incredible **legacy**.
When you are ill, you take **medicine**.
When the shop completed the repair, the car damage would be barely **noticeable**.
The medical team tried a new drug to kill the roundworm **parasite**.
No one would want a war on the Korean **peninsula**.
The company helped handicapped people by declaring a new **policy**.
The woman bought a crib when she knew about her **pregnancy**.
Due to her not doing her homework, Joan had to talk with the **principal**.
On Broadway, two important people are the director and the **producer**.
To learn your part in a play, you need to work hard at each **rehearsal**.
Meditation can help you when you need total **relaxation**.
/f/-sentences
On her all-natural diet, she could not eat anything with coloring that might be **artificial**.
An army doctor may need to help the wounded while they are on the **battlefield**.
The winner in a beauty pageant usually will be **beautiful**.
When you take a new job, you are no longer able to get any unemployment **benefit**.
To learn about Cleopatra, you could go to the library and read her **biography**.
When playing Pin the Tail on the Donkey, the player should be **blindfolded**.
Many people cannot tell apart a moth and a **butterfly**.
Many people thought that they could get rich during the Gold Rush in **California**.
A chameleon can be **camouflaged** by changing color to match the background.
Out in the wild, you could cook meat in the heat generated by a **campfire**.
When washing a breakable plate, you need to be **careful**.
The merchant marked the hundred dollar bill with a pen to check whether it might be **counterfeit**.

/s/-sentences
The marketing department collected age, income, and other **demographic** data.
With a great memory, and weighing more than a ton, the real king in the jungle might be an **elephant**.
You will need a map or a globe to learn **geography**.
She hadn’t met her grandmother, but she knew her **grandfather** quite well.
It can be important to teach children about delayed **gratification**.
While learning a new language, an electronic dictionary can be quite **helpful**.
With the near-total damage to the body, a dental record would be needed to **identify** it.
A web page can show updated **information** about the weather.
A marine animal with a jelly-like bell-shaped body would be called a **jellyfish**.
Return the radio directly to the **manufacturer**, rather than the retailer.
To be heard by all the people in the hall, you will need a **microphone**.
When the woman died, the doctor checked with her daughter to determine who should be **notified**.
The doctor warned her patient that cleaning the wound would be quite **painful**.
When the actor got hurt, he could not **perform** in the play.
The young lawyer wanted to help the poor and the weak, not the rich and the **powerful**.
Although he had been a great amateur hockey player, he could not hope to compete with a **professional**.
To learn about a potential date, you could read their online **profile**.
In a country with unclean drinking water, you should boil the water to **purify** it.
By doing well in another marathon, the runner **qualified** to compete in the New York Marathon.
During dinner, people do not want to be called on the **telephone**.
When the earth shook, and the building crumbled, the old woman looked **terrified**.
With her hand on the Bible, she took an oath to reply **truthfully**.
On an important occasion, a Marine will wear a white **uniform**.
We’d like to go to Niagara to watch the giant **waterfall**.

Appendix B: Critical Words and Mixtures Selected for Each Word in Conversational and Clear Speech

Critical /s/ words	Conversational speech	Clear speech	Critical /f/ words	Conversational speech	Clear speech
accident	f95	f60	artificial	f65	f55
agency	f75	f65	battlefield	f80	f65
bankruptcy	f75	f65	beautiful	s75	s50
boxer	f5	f40	benefit	f55	f45
cancelled	f5	f45	biography	f60	f70
coliseum	f40	f65	blindfolded	s80	f50
compensation	f60	f55	butterfly	s45	s55
concentrate	f45	f65	California	f40	f15
condescending	f45	f55	camouflage	s55	f65
currency	f55	f85	campfire	s85	s35
democracy	f75	f55	careful	f30	f50
dinosaur	s95	f50	counterfeit	f55	f65

Critical /s/ words	Conversational speech	Clear speech	Critical /f/ words	Conversational speech	Clear speech
diplomacy	f80	f55	demographic	f60	s85
dosage	f50	s50	elephant	f60	f60
electricity	f50	f50	geography	f45	f80
embassy	f55	f55	grandfather	f5	f50
emergency	f80	f55	gratification	f35	f65
episode	f75	f50	helpful	s75	f50
eraser	f5	f40	identify	s75	f50
generosity	f50	f65	information	s60	s36
impossible	f35	f55	jellyfish	s35	f50
innocent	f35	f50	manufacturer	s90	s50
intensity	f15	s65	microphone	f35	f15
intersection	s85	f65	notified	s95	f55
kerosene	s75	f35	painful	f65	f10
legacy	f65	f50	perform	f10	f50
medicine	f45	f10	powerful	f5	s75
noticeable	f40	f60	professional	s60	s15
parasite	f45	f55	profile	s85	s85
peninsula	s80	f50	purify	f5	f50
policy	f40	f65	qualified	f55	f50
pregnancy	f95	f60	telephone	s50	f40
principal	s90	f70	terrified	s90	f45
producer	s90	s70	truthfully	s50	f40
rehearsal	f70	f65	uniform	s50	s55
relaxation	f15	f50	waterfall	f15	f5

Note that the letter before the number refers to the "frame" that the mixtures were inserted into, and the number refers to the percentage of /s/ in the mixture (out of 100).

References

Bradlow A, Alexander J. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. Journal of the Acoustical Society of America. 2007; 121:2339–2349.10.1121/1.2642103 [PubMed: 17471746]

Bradlow A, Bent T. The clear speech effect for non-native listeners. Journal of the Acoustical Society of America. 2002; 112:272–284.10.1121/1.1487837 [PubMed: 12141353]

Bradlow A, Kraus N, Hayes E. Speaking clearly for children with learning disabilities: sentence perception in noise. Journal of Speech, Language, and Hearing Research. 2003; 46:80–97.10.1044/1092-4388(2003/007)

Broersma M, Scharenborg O. Native and non-native listeners' perception of English consonants in different types of noise. Speech Communication. 2010; 52:980–995.10.1016/j.specom.2010.08.010

Casini L, Burle B, Nguyen N. Speech perception engages a general timer : Evidence from a divided attention word identification task. Cognition. 2009; 112(2):318–322.10.1016/j.cognition.2009.04.005 [PubMed: 19457480]

Cooke MP, Lecumberri MLG, Barker J. The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. Journal of the Acoustical Society of America. 2008; 123:414–427.10.1121/1.2804952 [PubMed: 18177170]

Cutler A, Weber A, Smits R, Cooper N. Patterns of English phoneme confusions by native and non-native listeners. Journal of the Acoustical Society of America. 2004; 116:3668–3678.10.1121/1.1810292 [PubMed: 15658717]

Eisner F, McQueen JM. The specificity of perceptual learning in speech processing. Perception & Psychophysics. 2005; 67(2):224–238.10.3758/BF03206487 [PubMed: 15971687]

Eisner F, McQueen JM. Perceptual learning in speech: Stability over time. Journal of the Acoustical Society of America. 2006; 119(4):1950–1953.10.1121/1.2178721 [PubMed: 16642808]

Fernandes T, Kolinsky R, Ventura P. The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. Attention, Perception & Psychophysics. 2010; 72:1522–1532.10.3758/APP.72.6.1522

Gagné JP, Querengesser C, Folkeard P, Munhall KG, Zandipour M. Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: An investigation of conversational and clear speech. The Volta Review. 1995; 97:33–51.

Ganong WF. Phonetic categorization in auditory word perception. Journal of Experimental Psychology: Human Perception and Performance. 1980; 6(1):110–125.10.1037/0096-1523.6.1.110 [PubMed: 6444985]

Helfer KS. Auditory and auditory-visual perception of clear and conversational speech. Journal of Speech, Language, and Hearing Research. 1997; 40:432–443.

Jesse A, McQueen JM. Positional effects in the lexical retuning of speech perception. Psychonomic Bulletin & Review. 2011; 18:943–950.10.3758/s13423-011-0129-2 [PubMed: 21735330]

Jongman A, Wayland R, Wong S. Acoustic characteristics of English fricatives. Journal of the Acoustical Society of America. 2000; 108:1252–1263.10.1121/1.1288413 [PubMed: 11008825]

Kahneman, D. Attention and effort. Englewood Cliffs, N.J.: Prentice-Hall; 1973.

Keppel, G.; Wickens, TD. Design and analysis A researcher's handbook. 4. Upper Saddle River, NJ: Pearson Prentice Hall; 2004.

Kraljic T, Samuel AG. Perceptual learning for speech: Is there a return to normal? Cognitive Psychology. 2005; 51(2):141–178.10.1016/j.cogpsych.2005.05.001 [PubMed: 16095588]

Kraljic T, Brennan SE, Samuel AG. Accommodating variation: Dialects, idiolects, and speech processing. Cognition. 2008; 107:54–81.10.1016/j.cognition.2007.07.013 [PubMed: 17803986]

Kraljic T, Samuel AG. Generalization in perceptual learning for speech. Psychonomic Bulletin & Review. 2006; 13:262–268.10.3758/BF03193841 [PubMed: 16892992]

Kraljic T, Samuel AG. Perceptual adjustments to multiple speakers. Journal of Memory & Language. 2007; 56:1–15.10.1016/j.jml.2006.07.010

Kraljic T, Samuel AG. Perceptual learning evidence for contextually-specific representations. Cognition. 2011; 121:459–465.10.1016/j.cognition.2011.08.015 [PubMed: 21939965]

Kraljic T, Samuel AG, Brennan SE. First impressions and last resorts: How listeners adjust to speaker variability. Psychological Science. 2008; 19:332–338.10.1111/j.1467-9280.2008.02090.x [PubMed: 18399885]

Krause JC, Braida LD. Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. Journal of the Acoustical Society of America. 2002; 112:2165–2172.10.1121/1.1509432 [PubMed: 12430828]

Lecumberri MLG, Cooke MP. Effect of masker type on native and non-native consonant perception in noise. Journal of the Acoustical Society of America. 2006; 119:2445–2454.10.1121/1.2180210 [PubMed: 16642857]

Lecumberri MLG, Cooke MP, Cutler A. Non-native speech perception in adverse conditions: A review. Speech Communication. 2010; 52:11–12. 864–886.10.1016/j.specom.2010.08.014

Maniwa K, Jongman A, Wade TW. Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners. The Journal of the Acoustical Society of America. 2008; 123(2):1114–1125.10.1121/1.2821966 [PubMed: 18247912]

Mattys SL, Brooks J, Cooke MP. Recognizing speech under a processing load: Dissociating energetic from informational factors. Cognitive Psychology. 2009; 59(3):203–243.10.1016/j.cogpsych.2009.04.001 [PubMed: 19423089]

Mattys SL, Carroll LM, Li CKW, Chan SLY. Effects of energetic and informational masking on speech segmentation by native and non-native speakers. Speech Communication. 2010; 52:11–12. 887–899.10.1016/j.specom.2010.01.005

Mattys SL, Wiget L. Effect of cognitive load on speech recognition. Journal of memory and Language. 2011; 65:145–160.10.1016/j.jml.2011.04.004

Mayo L, Florentine M, Buus S. Age of second-language acquisition and perception of speech in noise. Journal of speech, language, and hearing research. 1997; 40(3):686–693.

McQueen J, Norris D, Cutler A. The dynamic nature of speech perception. Language and speech. 2006; 49(1):101–112.10.1177/00238309060490010601 [PubMed: 16922064]

McQueen JM, Cutler A, Norris D. Phonological Abstraction in the Mental Lexicon. Cognitive Science. 2006; 30(6):1113–1126.10.1207/s15516709cog0000_79 [PubMed: 21702849]

Norris D, McQueen J, Cutler A. Perceptual learning in speech. Cognitive Psychology. 2003; 47(2)(03): 204–238. 00006–9.10.1016/S0010-0285 [PubMed: 12948518]

Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing. I. Intelligibility differences between clear and conversational speech. Journal of Speech and Hearing Research. 1985; 28:96–103. [PubMed: 3982003]

Pitt MA, Samuel AG. An empirical and meta-analytic evaluation of the phoneme identification task. Journal of Experimental Psychology: Human Perception and Performance. 1993; 19:1–27.10.1037/0096-1523.19.4.699

Reinisch E, Weber A, Mitterer H. Listeners retune phoneme categories across languages. Journal of Experimental Psychology: Human Perception and Performance. 2012; 39(1):75–86. doi: 10/1037/a0027979. [PubMed: 22545600]

Samuel AG. When does interrupting processing prevent lexically-driven recalibration of speech sounds?. (in preparation).

Samuel AG, Kat D. Early levels of analysis of speech. Journal of Experimental Psychology: Human Perception and Performance. 1996; 22:676–694.10.1037/0096-1523.22.3.676

Samuel AG, Kraljic T. Perceptual learning for speech. Attention, Perception, & Psychophysics. 2009; 71(6):1207–1218.10.3758/APP.71.6.1207

Samuel, AG.; Kraljic, T. Accents, Assimilation, and Auditory Adjustments. The Psychonomic Society; Seattle: 2011 Nov.

Shi L. Perception of Acoustically Degraded Sentences in Bilingual Listeners Who Differ in Age of English Acquisition. Journal of speech, language, and hearing research. 2010; 53(4):821–836.10.1044/1092-4388(2010/09-0081)

Smiljani R, Bradlow AR. Production and perception of clear speech in Croatian and English. Journal of the Acoustical Society of America. 2005; 118:1677–1688.10.1121/1.2000788 [PubMed: 16240826]

Smiljani R, Bradlow AR. Speaking and hearing clearly: Talker and listener factors in speaking style changes. Language and linguistics compass. 2009; 3(1):236–264.10.1111/j.1749-818X.2008.00112.x [PubMed: 20046964]

Toro JM, Sinnett S, Soto-Faraco S. Speech segmentation by statistical learning depends on attention. Cognition. 2005; 97:B25–34.10.1016/j.cognition.2005.01.006 [PubMed: 16226557]

van Dommelen WA, Hazan V. Perception of English consonants in noise by native and Norwegian listeners. Speech Communication. 2010; 52:968–979.10.1016/j.specom.2010.05.001

Van Engen K, Bradlow A. Sentence recognition in native-and foreign-language multi-talker background noise. Journal of the Acoustical Society of America. 2007; 121:519–526.10.1121/1.2400666 [PubMed: 17297805]

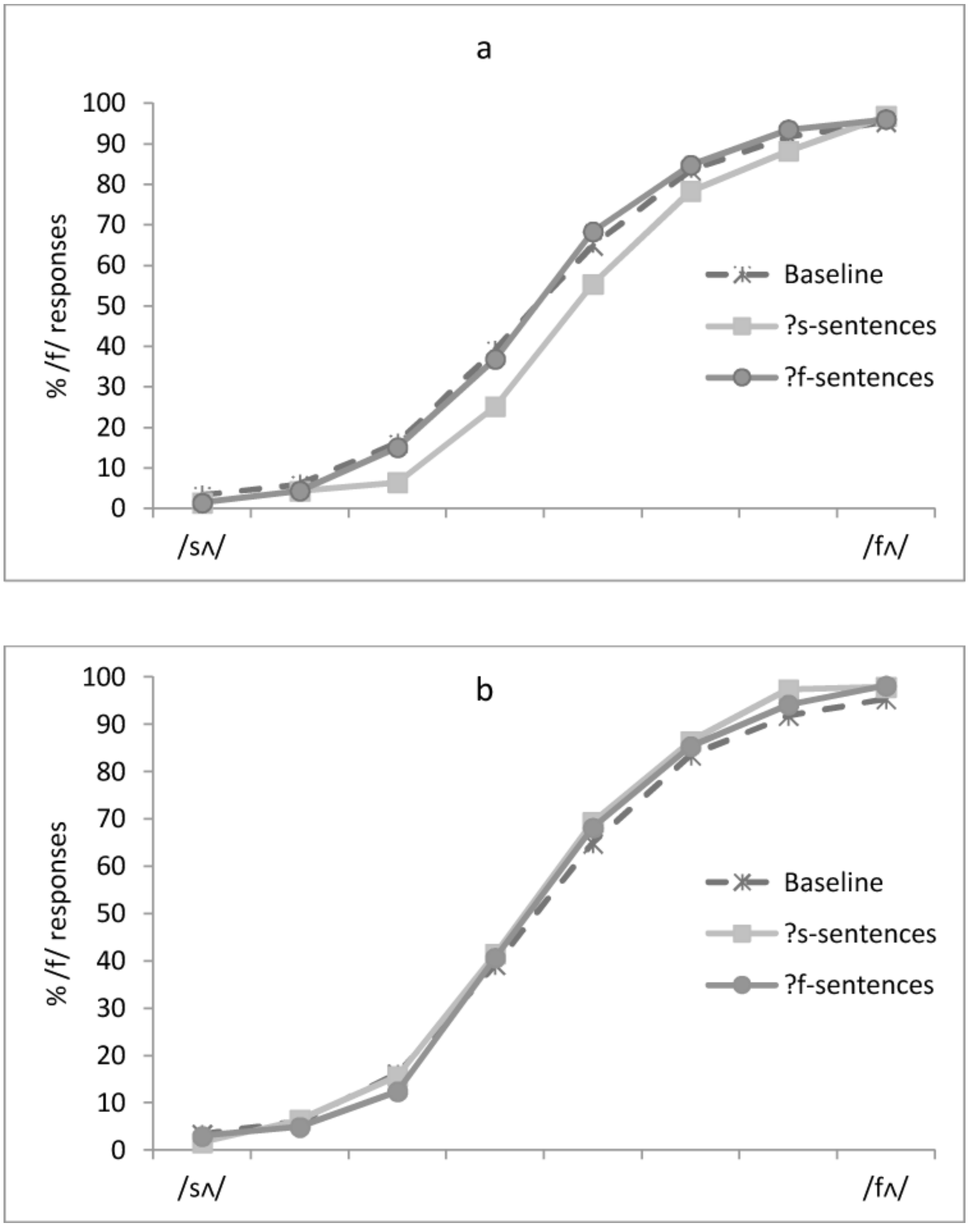

Figure 1.

Percentage of /f/ responses to each of the eight tokens on the /sΛ/-/fΛ/ continuum as a function of the Exposure phase (***?s*** vs. ***?f*** vs. Baseline) in the first session (Fig. 1a) and in the second session (Fig. 1b) in Experiment 1.

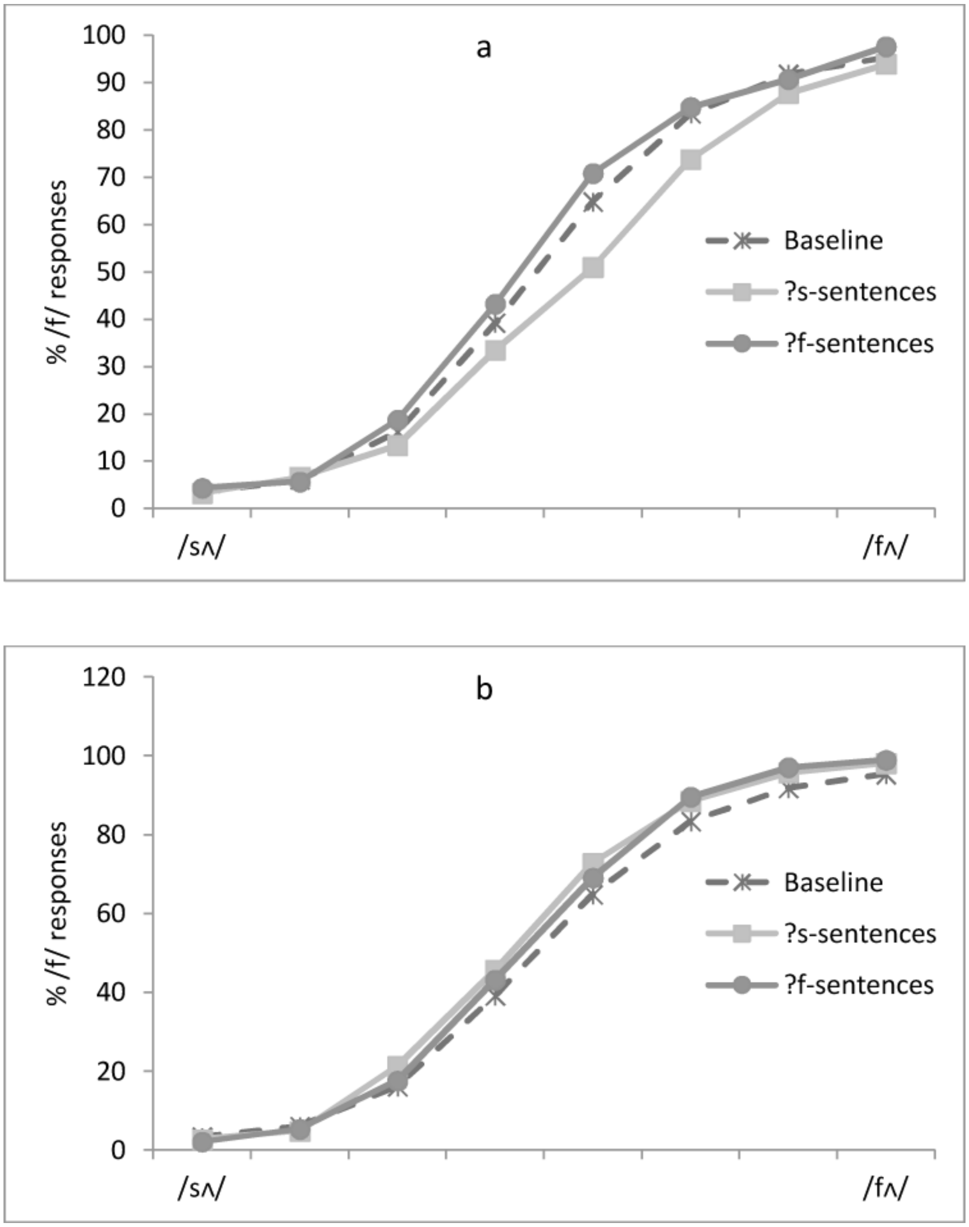

Figure 2.

Percentage of /f/ responses to each of the eight tokens on the /sΛ/-/fΛ/ continuum as a function of the Exposure phase (***?s*** vs. ***?f*** vs. Baseline) in the first session (Fig. 2a) and in the second session (Fig. 2b) in Experiment 2.

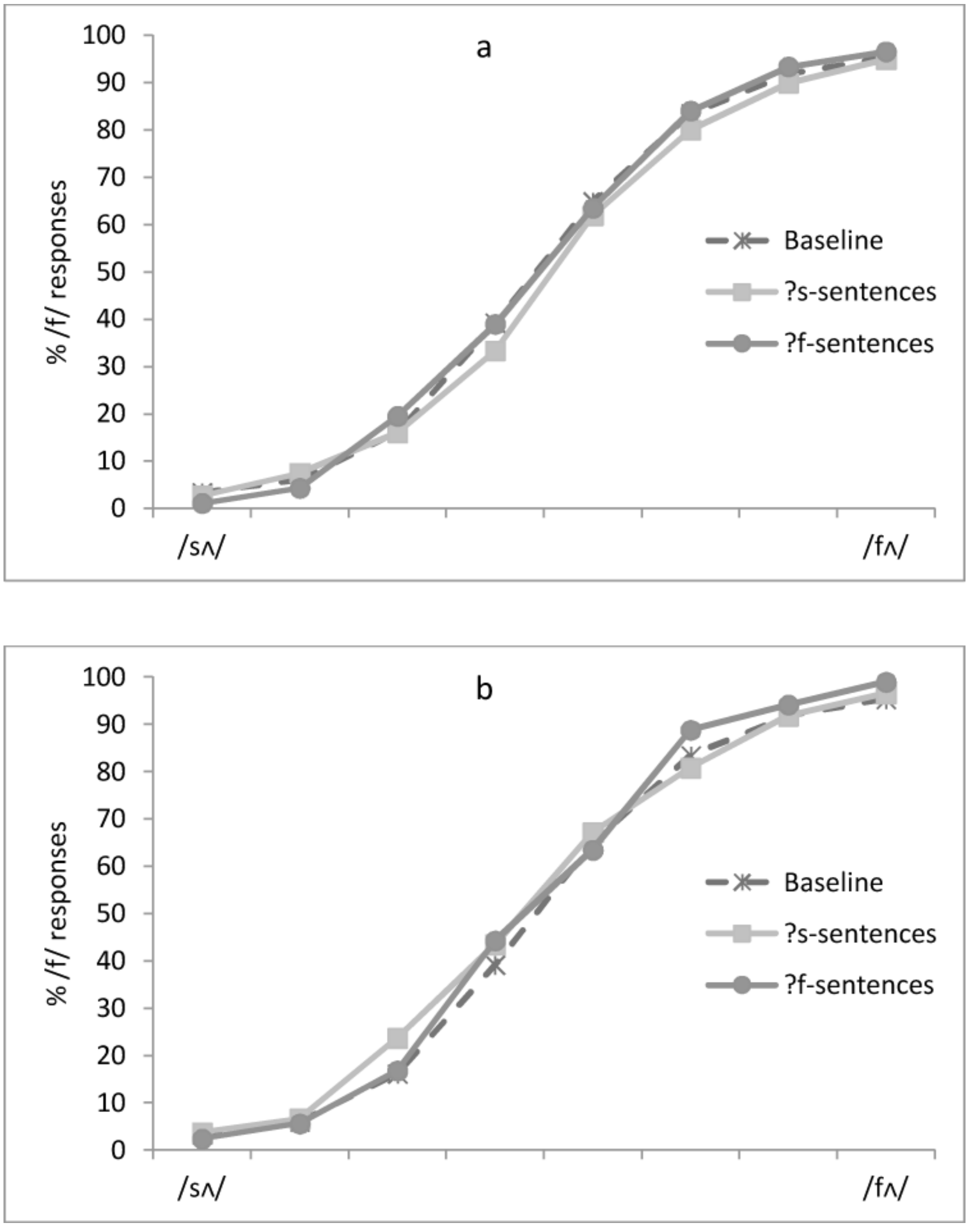

Figure 3.

Percentage of /f/ responses to each of the eight tokens on the /sΛ/-/fΛ/ continuum as a function of the Exposure phase (***?s*** vs. ***?f*** vs. Baseline) in the first session (Fig. 3a) and in the second session (Fig. 3b) in Experiment 3a.

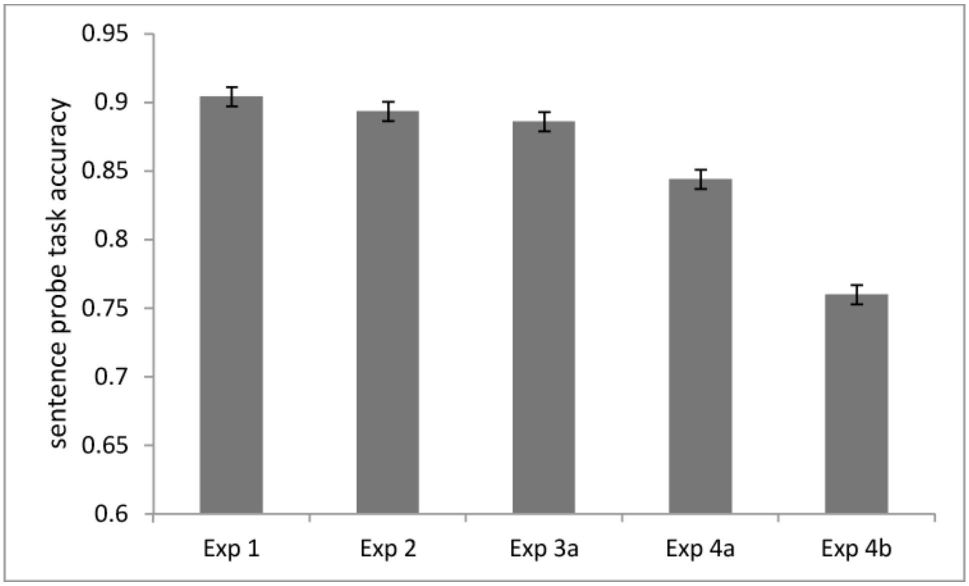

Figure 4.

Average accuracies of the sentence probe task in the Exposure phase across five experiments (Exp1: conversational speech; Exp2: clear speech; Exp3a: signal-correlated noise; Exp4a: cognitive load_letter recognition; Exp4b: cognitive load_letter searching), with error bars representing standard error of the mean.

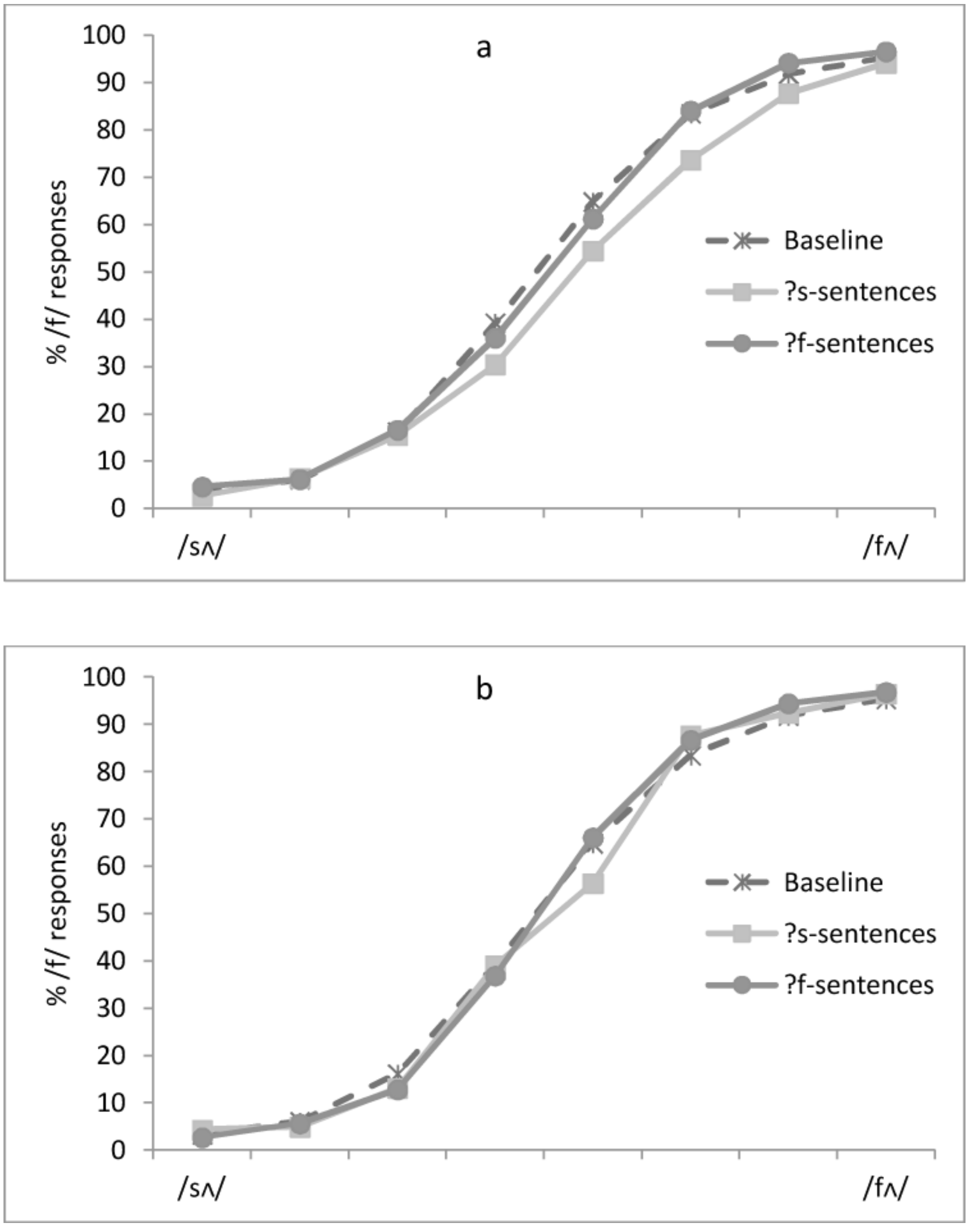

Figure 5.

Percentage of /f/ responses to each of the eight tokens on the /sΛ/-/fΛ/ continuum as a function of the Exposure phase (***?s*** vs. ***?f*** vs. Baseline) in the first session (Fig. 5a) and in the second session (Fig. 5b) in Experiment 4a.

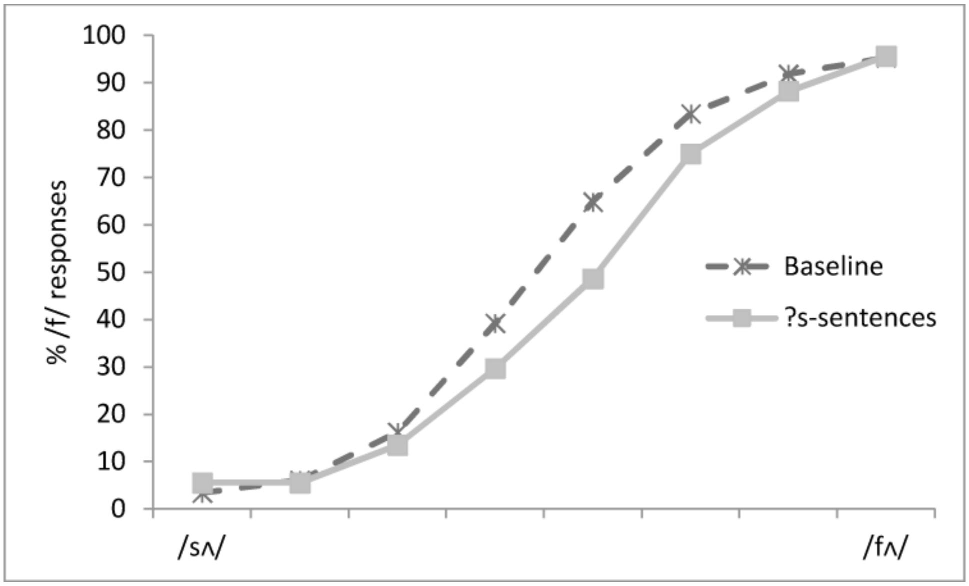

Figure 6.

Percentage of /f/ responses to each of the eight tokens on the /sΛ/-/fΛ/ continuum as a function of the Exposure phase (***?s*** vs. Baseline) in Experiment 4b.