Is speech perception influenced by top-down feedback?

Background

The extent to which subjective perception is affected by top-down processes is one of the most fundamental issues in cognitive science. According to *interactive theories*, our low-level perceptual experience is influenced by top-down feedback from higher levels of processing: In other words, what we see and hear is strongly affected by what we expect to see and hear, and by the knowledge we have amassed over the course of our previous experience. *Autonomous theories*, by contrast, claim that our perceptual systems process input in a bottom-up manner that is unaffected by contextual influences. According to one strong statement of the autonomous perspective, “top-down feedback is never necessary” (Norris, McQueen, & Cutler, 2000).

Adjudicating between these theoretical positions has proved extremely challenging (Bowers & Davis, 2004). Some phenomena that were initially claimed to provide strong evidence for top-down feedback have subsequently been shown to be compatible with autonomous accounts. An example of such an effect is the Ganong effect, which refers to the finding that the interpretation of an ambiguous speech sound (segment) is affected by lexical context. For example, listeners who hear a spoken input like “bronchiti?”, where ? denotes a sound that is somewhere between /s/ and /ʃ/ (sounds like “sh”), are biased to interpret the ambiguous segment as the phoneme /s/, consistent with the word “bronchitis”. By contrast, listeners who hear an acoustically identical sound within a spoken input like “demoli?” are biased to interpret the ambiguous segment as the phoneme /ʃ/, consistent with the word “demolish”. This effect (Ganong, 1980) was initially taken as strong support for interactive theories: when perceptual analysis of a segment is ambiguous, the phonemic encoding can be influenced by top-down feedback from lexical knowledge. However, the Ganong effect can also be explained by autonomous theories. According to the Merge model (Norris et al., 2000), the encoding of the ambiguous segment is identical for “bronchiti?” and “demoli?”. Listeners’ bias to interpret the ambiguous /?/ sound in a way that is consistent with a word takes place at a level of processing that occurs after phoneme identification, and the lexical influence of word knowledge never feeds back to the stage of phoneme identification.

The adaptation paradigm. An ingenious means of dissociating the predictions of interactive and autonomous theories was devised by Samuel (2001), using an adaptation paradigm. Selective adaptation to speech stimuli was first shown by Eimas and Corbit (1973), and has since been observed in many experiments. For example, repeated exposure to /s/ shifts listeners’ perceptual boundaries, so that a segment previously categorised as halfway between /s/ and /ʃ/ is more often

To cite this output:
Davis, Colin (2007). Is speech perception influenced by top-down feedback?: Full Research Report.
ESRC End of Award Report, RES-000-22-1656. Swindon: ESRC

reported as /ʃ/. Adaptation effects appear to involve low-level, sublexical changes in phonetic encoding (Samuel & Kat, 1996). It follows that, if listeners actually hear the phoneme /s/ in stimuli like “bronchiti?” and “malpracti?”, repeated presentation of such stimuli should cause adaptation to /s/, whereas repeated presentation of stimuli like “demoli?” and “aboli?” should cause adaptation to /ʃ/. By contrast, if listeners do not actually hear /s/ (or /ʃ/) in these stimuli, but infer it at a later, post-perceptual stage, there should be no adaptation. These contrasting predictions of interactive and autonomous theories were tested by Samuel (2001). Listeners showed adaptation effects in the direction that would be expected if top-down feedback had altered their on-line perception, supporting the prediction of interactive models.

Perceptual learning. Two experiments using a similar methodology were recently reported by Norris et al. (2003). As in Samuel’s (2001) experiment, listeners were exposed to stimuli like “bronchiti?” during a training phase. They then categorised ambiguous segments as either /s/ or /f/. Those participants who had heard the ambiguous segment /?/ in the context of words ending in /s/ were *more* likely to categorise it as an /s/ than participants who had heard the same sound in the context of words ending in /f/. This outcome is in the opposite direction to the selective adaptation effect reported by Samuel (2001), and appears to reflect retuning of listeners’ phonemic boundaries. Norris et al. refer to this phenomenon as “perceptual learning”: in effect, listeners who hear the ambiguous sound in the context of a word that normally contains /s/ learn that the ambiguous sound represents the phoneme /s/ The different outcomes observed by Samuel and Norris et al. presumably reflect procedural differences between these experiments, although it is unclear which of these differences is responsible for the contrary findings. Nevertheless, the results of both Samuel and Norris et al. indicate that there is some lexical influence at work, contrary to a purely bottom-up model. Norris et al. (2003) argued that this reflects a restricted form of top-down feedback that is used only to support learning, and not for on-line perception. They suggested that top-down learning is beneficial in various situations, as in the case of adapting to an individual speaker’s accent. Critically, though, the influence of lexical knowledge is *post-perceptual* – it can retune a listener’s phonemic categories (affecting her future phonemic perception) but it cannot modify or distort the on-line perception of the speech sounds that she has just heard.

Modularity Theory

Another type of autonomous theory is the strong modularity account described by Fodor (1983), according to which perception is subserved by modules that are “informationally encapsulated”, in that they do not receive inputs from external modules, including the higher-level cognitive processes to which they feed their outputs. Evidence of top-down feedback from words to phonemes would not necessarily be inconsistent with Fodorian modularity, because words and phonemes are presumably contained within a single module. However, this theory would be contradicted by evidence that feedback from higher-level conceptual processes alters the internal workings of the phonological system. For example, when hearing a sentence like, "That clown was "?unny", where ? is an ambiguous segment between /f/ and /s/, the context may allow the listener to infer that the ambiguous word was "funny", but it should not cause the listener to actually hear and encode the phoneme /f/, as this would violate the principle of encapsulation of perceptual modules.

Objectives

This project aimed to characterise the influence of top-down feedback on speech perception. The first goal was to investigate whether on-line perception or learning of phonemes (or both) are influenced by top-down lexical feedback. In particular, our experiments aimed to resolve the apparent inconsistency in the recent findings by Samuel (2001) and Norris et al. (2003). We also hoped to learn more about the time-course and persistence of lexically mediated effects. A secondary objective of this project was to investigate whether top-down conceptual feedback influences phonemic perception, in order to test Fodorian modularity theory. Finally, a third goal was to introduce and evaluate an innovative methodological approach for estimating listeners' psychometric functions.

Methods

Participants. All participants were native speakers of English and had no self-reported hearing problems. Experiments 1-3 used a between-groups design, with approximately 20 participants in each group. Experiments 4 and 5 used a repeat measures design, with 20 participants in each experiment.

Recording of Unambiguous Stimuli. We used the same acoustic continua as Samuel (2001) (/ʃ/-/s/) and Norris et al. (2003) (/f/-/s/), embedding the ambiguous segment word-initially (Experiment 1, 4 and 5) and word-finally (Experiments 2 and 3). To construct the test stimuli for Experiment 1, a native British English speaker was recorded saying the words "sunny" and "funny". The initial portion of these recordings was used to provide the /sə/ and /fə/ stimuli that were blended as described below. A similar procedure was used to construct /i:s/ - /i:f/ waveforms for Experiment 2 and /Is/ - /Iʃ/ waveforms for Experiment 3. Adaptor words were recorded by the same native speaker. For Experiment 1, adaptor words were chosen in pairs, so that each initial-/s/ lexical context was paired with an initial-/f/ context that shared the same subsequent vowel (e.g., "senior" -

To cite this output:
Davis, Colin (2007). Is speech perception influenced by top-down feedback?: Full Research Report.
ESRC End of Award Report, RES-000-22-1656. Swindon: ESRC

"fever"). Adaptors were chosen in similar fashion for Experiments 2 and 3, except that the critical phonemes occurred in word-final position (e.g., "police"- "belief"; "bronchitis"- "diminish").

A critical feature of the stimuli in Experiments 4 and 5 was the embedding of ambiguous segments in contexts that are consistent with two words, e.g., "?unny", where /?/ is between /f/ and /s/, such that the stimulus could be perceived as either "funny" or "sunny". Four pairs of f/s words were selected and recorded: "faint"-"saint", "fight"-"sight", "funny"-"sunny", and "feeds"-"seeds". For each critical word pair seven stimulus exemplars were constructed; two unambiguous forms (e.g., "funny" and "sunny") and five ambiguous stimuli composed of various blends of the two unambiguous forms. In Experiment 4, each target stimulus was presented with four /f/ and four /s/ consistent context cue words (e.g., "cloudy" was a contextual cue for "sunny"). In Experiment 5 target stimuli was presented with four /f/ and four /s/ consistent context cue pictures (e.g., a picture of a beach on a sunny day). The effectiveness of the contextual cues was established in pilot studies.

Construction of Ambiguous Stimuli. We were concerned to ensure that the ambiguous adaptors in Experiments 1-3 contained segments that were maximally ambiguous, so as to optimise any lexically mediated shifts. This is not straightforward, as there are large individual differences in phoneme perception. Some individuals will not begin to classify an ambiguous segment as /s/ until very large proportions of /s/ are included in the stimulus, whereas other individuals categorise segments as /s/ as soon as the stimulus includes a relatively small proportion of /s/. This is illustrated in Figure 1, which shows data obtained from two different participants in one of our experiments. As can be seen, there is little overlap in the functions for the two observers. At the level of /s/ input at which Observer 1 consistently makes "s" categorisations, Observer 2 consistently makes "f" categorisations. Furthermore, a stimulus with 40% /s/ input will be treated as unambiguous by both observers; for example, a stimulus "?ilver" that is reported as "silver" by Observer 1 will be reported as "filver" by Observer 2. Neither of these situations is appropriate for producing a Ganong effect.

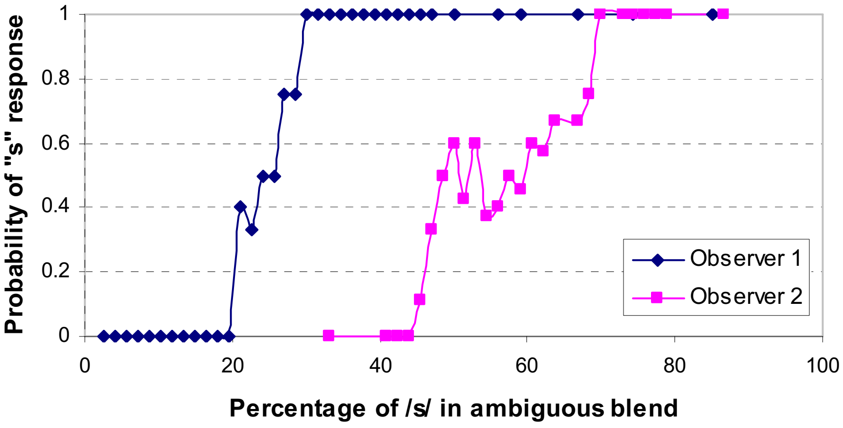

Figure 1: Psychometric functions obtained for two observers.

We therefore wished to determine the maximally ambiguous segment individually for each participant. To achieve this, we used an innovative approach that combined a novel use of equipment, custom-written software, and the application of a psychophysical procedure that has not previously been used in this domain. The first aspect of this novel methodology was an on-line technique that allowed us to vary the proportion of /s/ and /f/ in the signal in a precise fashion. Stereo headphones were customised so that the output of each channel was sent to both ears, and these headphones were used to present wave files in which the two channels contained different stimuli (e.g., the stimulus in one channel was the word “senior” whereas the stimulus in the other channel was “fenior”). Varying the relative volume of the two channels thereby enabled us to create different blend proportions on-line, e.g., shifting the balance entirely to channel 1 resulted in an unambiguous “senior”, whereas shifting the balance entirely to channel 2 resulted in an unambiguous “fenior” (note that, due to the modification to the headphones, the stimulus always sounded identical in both ears).

The second novel component of our methodology was the use of a psychophysical procedure to discover the stimuli that were maximally ambiguous for each participant (adaptive probit estimation; Watt & Andrews, 1981). After each stimulus, the listener made a forced choice, and the algorithm chose a new stimulus (i.e., a new blend proportion) for the next trial by combining a stochastic element with a statistical technique called probit analysis. The probability of “s” responses as a function of objective /s/ input was tabulated, and a sigmoid function fit to the observed points. One parameter of this function corresponds to a measure of bias in classical signal detection theory. This parameter also corresponds to the point at which the two categorisation responses are associated with equal probability, and hence we refer to this measure as the P50.

Procedure. Auditory stimuli were presented over padded headphones at a comfortable listening level. Experiments 1-3 each consisted of three phases. The first phase (Pre-Exposure) was designed to discover the P50 estimate (a categorisation bias expressed as a percentage of /s/ input) for each participant; this required 128 trials. The second phase (Exposure) consisted of an initial presentation of 44 adaptor stimuli, randomly selected from the pool of four adaptors for each condition. This took approximately thirty seconds, during which participants passively listened. The P50 value determined in the first phase was used to blend the two channels. Thus, for the ambiguous stimuli, the critical segment consisted of a blend of /f/ and /s/ (or /ʃ/ and /s/ in Experiment 3); for unambiguous adaptors the speech file contained two identical channels, and hence blending had no effect on the sound of the stimulus. After this initial exposure, participants made a series of 64 categorisations; as in the previous phase, sampling of stimuli was determined by the adaptive probit estimation algorithm, which was also used to re-estimate each participant's categorisation bias. "Top-up" exposures consisting of five adaptors were presented prior to each categorisation during this phase of the experiment. Between the second and third phases there was a five-minute interval during which participants completed a filler task. The presence of this interval enabled us to assess the duration of any lexically mediated shifts in phoneme perception. Finally, the third phase was identical to the first phase, consisting of a further 128 trials over which the P50 was once again estimated.

Experiments 4 and 5 were run using DMDX (Forster & Forster, 2003), with ambiguous stimuli that had previously blended in various proportions. A fully within-subjects design was employed in which all participants were presented with all combinations of critical stimuli and contextual associates. Experimental trials were randomly presented in four blocks of 56 pairs. Lexically ambiguous target words were presented 125 ms after the offset of the context cue.

Dependent measures and analyses. In all experiments, the critical measure was performance on a forced-choice phoneme categorisation task. Participants responded by pressing pre-specified keys (we always used the "f" and "s" keys on a standard keyboard; in Experiment 3, in which the choice was between /s/ and /ʃ/, the "f" key was covered by a label saying "sh").

Results

Experiment 1. This experiment tested perception of ambiguous word-initial sounds that varied along a continuum from /f/ to /s/. There were four adaptor conditions: unambiguous /f/-word adaptors, ambiguous /f?/-word adaptors, unambiguous /s/-word adaptors, and ambiguous /s?/-word adaptors.

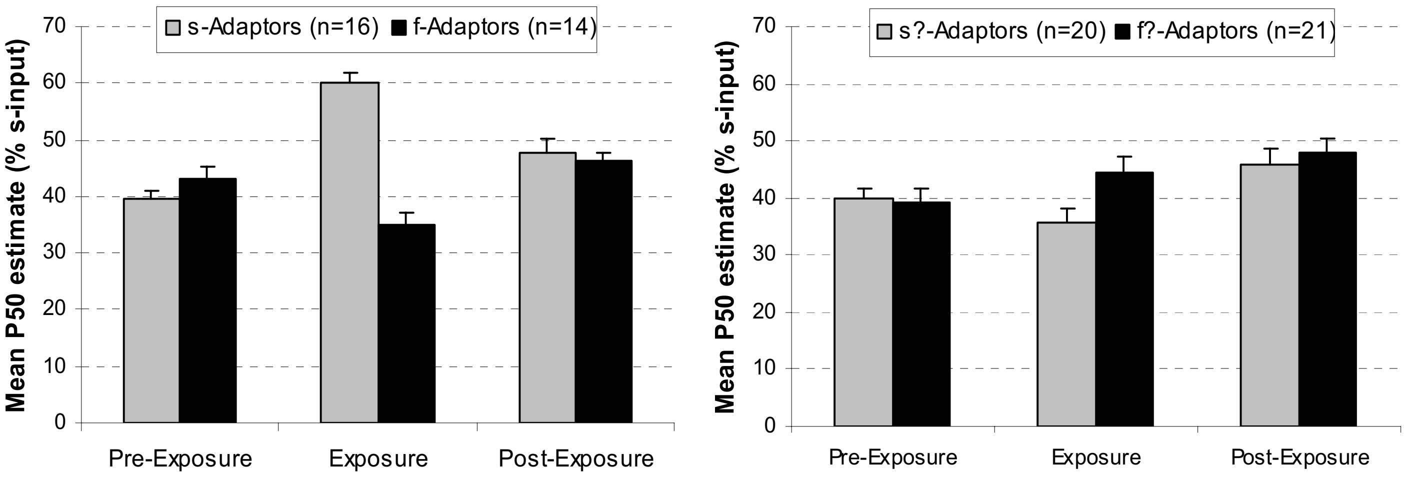

Figure 2: Average P50 estimates prior, during and following exposure to unambiguous adaptors (left panel) and ambiguous adaptors (right panel).

The lefthand side of Figure 2 shows the results for the unambiguous adaptor conditions. As can be seen, the groups showed equivalent biases prior to the exposure phase, but diverged following exposure to the adaptors. The /s/-adaptor condition showed a large increase in their P50 values, indicating that biases shifted away from /s/ (i.e., participants needed more /s/ in the test stimulus before they would make an "s" categorisation), whereas the /f/-adaptor condition showed a decrease in their P50 values, indicating that biases shifted towards /s/. This pattern is consistent with the adaptation effect expected for unambiguous adaptors. The results from the post-exposure phase show that this difference between adaptor conditions faded relatively quickly. The righthand side of Figure 2 shows the results for the ambiguous adaptor conditions. Note that the effect of exposure to these adaptors goes in opposite directions to the effect for unambiguous adaptors. This pattern is consistent with the perceptual learning effect observed by Norris et al. (2003), e.g., listeners who had heard the ambiguous phoneme in the context of /s/-words showed a greater bias to categorise ambiguous test phonemes as /s/, relative to listeners who had heard the ambiguous phoneme in the context of /f/-words. This effect also seems to be relatively short-lived, as it is not evident in the data from the post-exposure phase.

These data were analysed in an 2 x 2 x (3) ANOVA with the between-subject factors Stimulus Type (Unambiguous or Ambiguous) and Adaptor Type (/f/- or /s/-initial words) and the within-subjects factor Phase (Pre-Exposure, Exposure, Post-Exposure). The main effect of Phase was significant, $F(2,67) = 13.84$, $p<.001$, reflecting a trend for listeners' biases to shift away from /s/ (and towards /f/) over the course of the experiment. There was also a significant interaction of Phase x Adaptor, $F(2,67) = 9.35$, $p<.005$, which reflects the finding that, following exposure to the adaptor stimuli, P50 estimates tended to be higher for those participants who had heard /s/ adaptor words than for those who had heard /f/ adaptor words. This overall adaptation effect was driven by the unambiguous adaptor words. The critical interaction was the significant three-way interaction of Phase x Adaptor Type x Stimulus Type, $F(2,67) = 33.83$, $p< .01$. A post-hoc *t*-test on the difference between the /f?/ and /s?/ adaptor conditions during the exposure phase showed a significant difference in a direction opposite to adaptation, $t(39)=2.42$, $p<.05$. This difference had completely disappeared by the time of post-exposure, $t(39)=0.52$, $p>.6$.

In summary, Experiment 1 showed very large and robust adaptation effects for unambiguous adaptor words. The psychophysical estimation procedure allows us to quantify this as a shift in categorisation biases that corresponds to 25% of the range of the continuum. By contrast, however, ambiguous adaptors showed a shift of 11% in the opposite direction. On the surface, this result seems difficult to reconcile with Samuel's (2001) finding of significant adaptation for ambiguous adaptors. The obtained pattern is more consistent with the perceptual learning effect reported by

Norris et al. (2003) and Kraljic and Samuel (2005). However, in contrast to Kraljic and Samuel (2005), who found effects that persisted over the course of 25 minutes, in our experiment the perceptual learning associated with ambiguous adaptors had disappeared following a five-minute break.

Experiment 2. A possible explanation for our failure to replicate Samuel's (2001) adaptation finding in Experiment 1 was that our critical ambiguous segments were in word-initial position rather than word-final position, as in Samuel's experiment. In Experiment 2 we switched to adaptors in which the critical sound was word-final, and the test stimulus was /i:?/ (i.e., a sound varying between "eef" and "eece"). After an initial pilot experiment using unambiguous adaptors with six participants per condition showed a large adaptation effect of 30%, $t(10)=6.98$, $p<.001$, we focused our attention on ambiguous adaptors.

Figure 3 shows the results for Experiment 2. There was a main effect of Phase, $F(2,56)=10.06$, $p<.001$, indicating a general trend for listeners' biases to shift away from /s/ over the course of the experiment (as in Experiment 1). The effects of Adaptor Type and the interaction of Adaptor Type x Phase were both nonsignificant, both *F*s < 1.

In summary, Experiment 2, like Experiment 1, failed to find evidence of an adaptation effect for ambiguous adaptors. Unlike Experiment 1, there was no evidence for a perceptual retuning effect. Although the estimates of bias showed no differences between the adaptor conditions, a post-

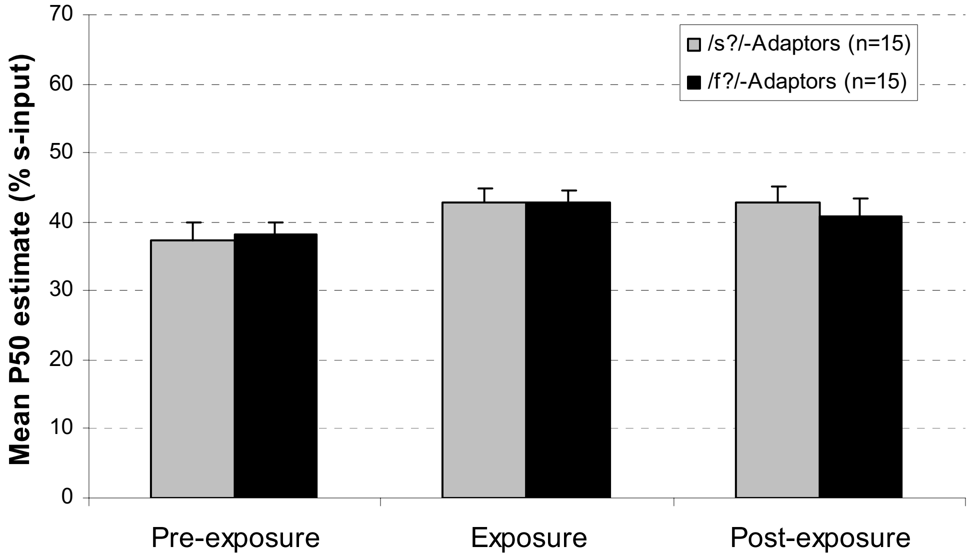

Figure 3: Average P50 estimates prior, during and following exposure to the five adaptor conditions.

hoc analysis did reveal an interesting difference. This analysis examined categorisations of the first segment following exposure to the adaptors (i.e., the first categorisation listeners made during the exposure phase), which, for all listeners, was a 50-50 blend of the /s/ and /f/ channels. 13 out of 15 participants in the /s?/-adaptor condition made “s” categorisations of this stimulus, compared to 7 out of 15 participants in the /f?/-adaptor condition. This bias toward the adaptor was significant, $\chi^2(1) = 5.4$, $p<.05$. This suggests that an initial perceptual learning effect may have been obscured by a later shift in the opposite direction. The significance of this possibility is discussed below.

Experiment 3. In this experiment, we made one final attempt to replicate the adaptation effect of Samuel (2001). In order to do this, we used the same stimuli as Samuel: the same phonemic continuum /Is/-/Iʃ/ and the same word adaptors. The results are shown in Figure 4. Nonsignificant results were observed for the main effects of Phase, $F(2,76)=1.24$, $p>.05$, and Adaptor Type, $F(1,38)=1.86$, $p>.05$. The critical effect was the interaction of Phase x Adaptor Type, which was significant, $F(2,76) = 4.17$, $p<.05$. This reflects the fact that the /s?/-adaptor condition showed a shift towards /s/ over the course of the experiment, whereas the /ʃ?/-adaptor condition showed a slight shift towards /f/. That is, this experiment showed the same perceptual learning effect observed in Experiment 1. One difference between the results of the experiments was the time-course of the effect: in this experiment the learning effect took slightly longer to manifest and persisted longer. The reason for this difference is not entirely clear. Nevertheless, the critical aspect of the present findings is that, once again, we failed to obtain evidence of adaptation for ambiguous adaptors.

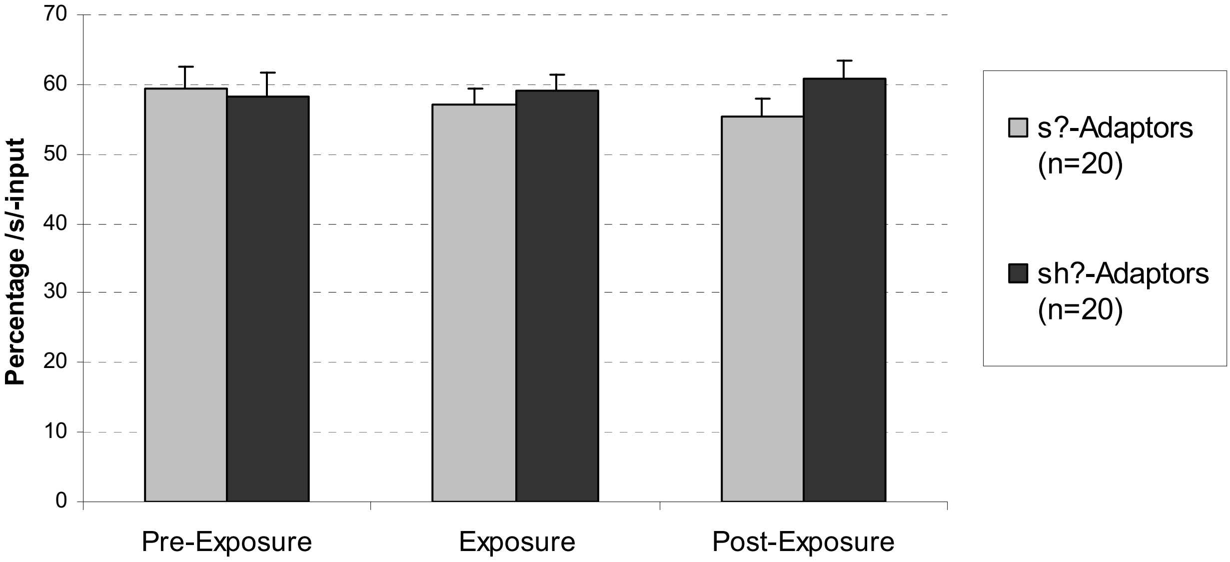

Figure 4: Average P50 estimates prior, during and following exposure to the /s?/ and /ʃ?/ adaptor conditions.

Summary of Experiments 1-3. In three separate experiments we failed to find evidence of lexically-mediated adaptation for adaptors containing ambiguous segments, under the same conditions that produced extremely large adaptation effects for unambiguous segments. This was true whether the ambiguous segment was word-initial (Experiment 1) or word-final (Experiments 2 and 3), and whether it was positioned near the midpoint of an /s/-/f/ continuum (Experiments 1 and 2) or an /s/-/ʃ/ continuum (Experiment 3). Consequently, our findings challenge the claim that evidence for on-line top-down feedback in speech perception can be obtained in an adaptation paradigm.

Although this pattern of results would appear to contradict the findings of Samuel (2001), it is possible that the two sets of findings can be reconciled. Vroomen et al. (2007) recently re-analysed Samuel's (2001) data, which consisted of 24 blocks, each block involving exposure to adaptors followed by test stimuli. They found that although adaptation effects were present in blocks two onwards, the results following the first block of adaptors actually showed a pattern in the opposite direction, that is, a pattern consistent with perceptual retuning. This raises the possibility that the adaptation effects obtained by Samuel (2001) depend on prior perceptual learning. For example, if exposure to the first block of /s?/ words leads a listener to retune their phonetic representations so that the ambiguous /?/ sound corresponds to /s/, exposure to subsequent blocks of the same words may then lead to adaptation in much the same way as for unambiguous /s/.

The post-hoc analysis we conducted on the data of Experiment 2 supports this possibility. Although there was a very strong bias for participants to categorise the first ambiguous stimulus following exposure to the adaptor stimuli in accord with these stimuli (i.e., a perceptual learning effect, as in Experiments 1 and 3), the P50 estimates (which were based on all 64 trials from this phase) showed no difference between the adaptor conditions. This may have been because the top-up procedure that we used following the initial exposure phase caused some shift in the direction of adaptation. We have begun to develop a computational account of these findings in which a Hebbian learning mechanism leads to rapid retuning of phonemic categories, and repeated activation of phoneme detectors results in adaptation effects.

Not surprisingly, the present series of experiments has not definitively resolved the debate between interactive and autonomous theories of speech perception. Nevertheless, we have provided evidence that questions the interpretation of a finding that is currently perhaps the most important source of support for interactive theories. It seems that different forms of evidence will be required to settle this debate. Our experiments have also contributed to resolving the apparent inconsistency in the recent findings by Samuel (2001) and Norris et al. (2003) by supporting the possibility that adaptation shifts, as observed by Samuel, depend on prior perceptual learning, as observed by Norris et al. We have also established that retuning of phonemic boundaries can occur extremely rapidly, and does not require listeners to actively categorise ambiguous stimuli as words. Contrary to previous findings (Kraljic & Samuel, 2005), we have obtained mixed evidence concerning the persistence of this retuning effect. Further investigation of this issue will help to better understand the flexibility of phonemic representations.

Experiments Investigating Modularity

Thus far, we have conducted two preliminary experiments directed toward testing whether conceptual feedback influences phoneme perception.[1] In Experiment 4, word associates were used to attempt to bias perception of ambiguous segments in lexically ambiguous stimuli like “?unny”, in which the stimulus is potentially consistent with either an “f”-initial or an “s”-initial word (i.e., “funny” or “sunny”). Thus, listeners heard pairs like “humour”-“?unny” and “weather”- “?unny” (to bias them towards either a “funny” or “sunny” interpretation). Experiment 5 followed a similar design, but used pictures as conceptual cues. In our proposal, we suggested using these stimuli in an adaptation paradigm. However, our inability to obtain adaptation effects for ambiguous stimuli in Experiments 1-3 necessitated adopting a different approach. For this reason, we simply asked participants to identify the first phoneme of the critical word by making a forced choice (/f/ or /s/).

Experiment 4. The percentage of “s” responses for each phoneme blend in the /f/ and /s/ biased conditions are displayed in Figure 5.

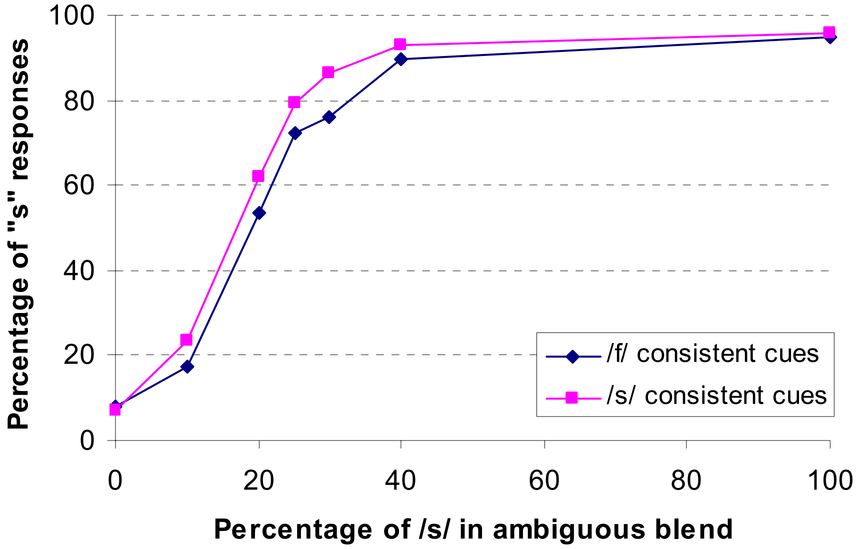

Figure 5: Percentage of “s” responses in Experiment 4 as a function of contextual cues (/f/ or /s/ consistent) and percentage of /s/ in the ambiguous blend.

The data were analysed using a repeated measures (7) x (2) ANOVA with Blend Proportion and Context as factors. There was a significant effect of Context, $F(1,19) = 12.2$, $p<.005$, indicating that the word associates were successful in biasing the classification of the lexically ambiguous stimuli. There was also a significant interaction of Blend Proportion and Context, $F(6,114) = 3.40$, $p<.05$, indicating that the biasing effect was not equally evident across the whole continuum of stimulus ambiguity. In particular, as can be seen in the figure, context exerted no biasing effect at the unambiguous endpoints.

The results of Experiment 4 suggest that phonemic categorisation of ambiguous segments can be biased not just by lexical context, as in the standard Ganong effect, but also by conceptual context. As in the case of the Ganong effect, though, the locus of this effect is unclear. The biasing effect could reflect feedback from semantics influencing phoneme perception, which would falsify the strong version of Fodorian modularity. However, it could alternatively reflect feedback from associations that are stored within the speech perception module. Finally, it could reflect a simple demand effect, whereby listeners respond to ambiguity by categorizing on the basis of the verbal cue. It was hoped that Experiment 5 would provide evidence that might rule out one of these alternatives.

Experiment 5. In most respects the procedure and stimuli used in this experiment were identical to Experiment 4; the only difference was that the contextual cue was supplied by pictures rather than words. Results are shown in Figure 6.

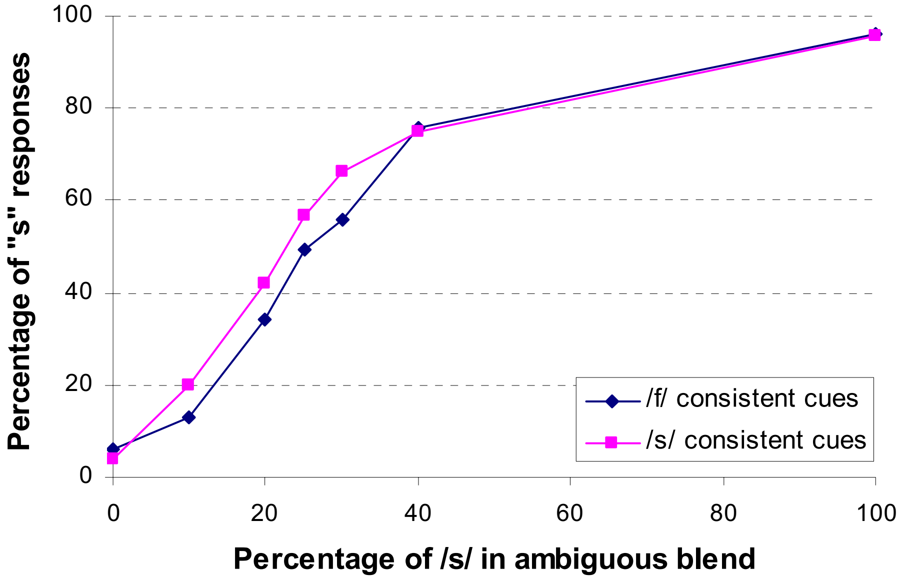

Figure 6: Percentage of "s" responses in Experiment 5 as a function of contextual cues (/f/ or /s/ consistent) and percentage of /s/ in the ambiguous blend.

To cite this output:
Davis, Colin (2007). Is speech perception influenced by top-down feedback?: Full Research Report.
ESRC End of Award Report, RES-000-22-1656. Swindon: ESRC

As can be seen, there was a clear biasing effect of picture context. This effect was statistically significant, $F(1,19) = 5.14$, $p<.05$, and also interacted with Blend Proportion, $F(6,114) = 3.19$, $p<.05$. The latter interaction reflects the fact that the picture context biased phoneme categorization for intermediate proportions of /s/, but not for extreme points along the continuum.

The results of this experiment suggest that it is not necessary for contextual cues to be lexical in order to obtain a contextual-bias effect in phonemic categorisation of ambiguous segments. On the grounds of parsimony, this challenges the conclusion that the biasing effect observed in Experiment 4 was the product of lexical associations stored within the speech perception module. It may instead be that phoneme perception was influenced by conceptual context, which would contradict the strong form of Fodorian modularity. However, we are also unable to eliminate the possibility that the results of both Experiments 4 and 5 reflect demand effects. These putative effects are not so strong as to determine listeners' responses to all stimuli; when the stimulus is somewhat ambiguous, though, it is possible that listeners rely on contextual cues to interpret the stimuli. This possibility is of course analogous to the interpretative difficulties posed by the Ganong effect. This aspect of our research program is ongoing, and below we outline our plans for further experiments that may shed more light on this issue.

Activities

Parts of this work have been presented in internal seminars at the University of Bristol and at Royal Holloway University of London. It has also been discussed with leading experts at other institutions, including Norris in Cambridge. We have also discussed the experiments and the potential technological implications of rapid perceptual retuning with computer scientists at Royal Holloway University of London.

Other Outputs

An article describing Experiments 1-3 is in preparation, and will be submitted to the Journal of Memory and Language. A separate article describing Experiments 4 and 5 awaits further experiments, as described below. Datasets for the five experiments described above are in the process of being transferred to the UK Data Archives. The software developed to estimate perceptual boundaries will be made available to other researchers; it is expected that, with appropriate customisation, this software will prove to be a valuable tool for researchers in a variety of psycholinguistic domains.

Impacts

Once published, the results of this research will be of great interest to other researchers in the field of speech perception, as well as cognitive scientists more generally. The demonstration that listeners unconsciously retune their phonemic boundaries in response to relatively few exposures to an ambiguous segment, and the quantification of this effect are findings that have both theoretical significance, as well as practical significance for language learners and for the development of speech recognition software.

Future Research Priorities

One of our objectives in this project was to introduce a novel methodological approach for measuring listeners' speech perception biases. This objective has been met: we have used the new approach to quantify the shift produced by adaptation to unambiguous adaptors, as well as the shift in the opposite direction produced by exposure to ambiguous adaptors. In future work, we hope to use this methodology to investigate other factors that may produce perceptual shifts in speech perception, and to use a variation of this approach to investigate whether learned perceptual categories show a similar fluidity in other domains, including recognition of visual patterns such as letters.

Another objective of this project was to assess the evidence for on-line top-down feedback in speech perception using the selective adaptation paradigm. Our experiments have established that this paradigm does not provide strong evidence for top-down feedback. They have also partially satisfied another objective, by suggesting a possible resolution of the apparent empirical discrepancy between the results of Samuel (2001) and Norris et al. (2003). To this end, we plan to conduct further experiments to more closely investigate the relative time-course of the perceptual learning and adaptation effects. This empirical effort will be critical for the development of computational models of these phenomena.

Finally, because our experiments have indicated that the adaptation paradigm does not provide the necessary means of assessing either lexical or conceptual feedback in speech perception, we are exploring different methodologies. One approach that appears promising is a matching task in which listeners are asked to adjust a sample segment until it seems to match a corresponding sound embedded in a lexical (or conceptual) context. The question is whether this context will bias the matching process (e.g., whether /?/ is matched to a stimulus further along the /s/ continuum when embedded in the context "bronchiti?" compared to when embedded in the context "replenish?"). This technique promises to be useful for investigating the possible effects of both lexical and conceptual feedback.

Davis, Colin (2007). Is speech perception influenced by top-down feedback?: Full Research Report.
ESRC End of Award Report, RES-000-22-1656. Swindon: ESRC

References

Bowers, J.S., & Davis, C.J. (2004). Is speech perception modular or interactive? Trends in Cognitive Science, 8, 3-5.

Eimas, PD, & Corbitt, JD (1973) Selective adaptation of linguistic feature detectors. Cognitive Psychology, 4, 99-109.

Fodor, J.A. (1983). Modularity of mind: An essay on faculty psychology. MIT Press.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. Behavior Research Methods, Instruments & Computers, 35, 116-124.

Ganong, W.F. (1980). Phonetic categorization in auditory word perception. Journal of Experimental Psychology: Human Perception and Performance, 6, 110–125.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? Cognitive Psychology, 51, 141-178.

McClelland, J.L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? Trends in Cognitive Science, 10, 363–369.

McQueen, J. M., Norris, D., & Cutler, A. (2006). Are there really interactive speech processes in speech perception? Trends in Cognitive Science, 10, 533.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. Behavioral and Brain Sciences, 23, 299-370.

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

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. Psychological Science, 12, 348–351.

Samuel, A.G., 1986. Red herring detectors and speech perception: In defence of selective adaptation. Cognitive Psychology, 18, 452–499.

Samuel, A.G., & Kat, D. (1996). Early levels of analysis of speech. Journal of Experimental Psychology: Human Perception and Performance, 22, 676–694.

Vroomen, J., Linden, S. van, Gelder, B. de, & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory–visual speech perception: contrasting build-up courses. Neuropsychologia, 45, 572-577.

Watt, R.J., & Andrews, D.P. (1981). APE: Adaptive probit estimates of psychometric functions. Current Psychological Reviews, 1, 205-214.

---

[1] We gratefully acknowledge the assistance of James Twist, who collected these data during the course of his third-year undergraduate project that was supervised by Professor Bowers.