

Manuscript Details

Manuscript number	BRLN_2018_116_R1
Title	Written sentence context effects on acoustic-phonetic perception: fMRI reveals cross-modal semantic-perceptual interactions
Article type	Short Communication

Abstract

This study examines cross-modality effects of a semantically-biased written sentence context on the perception of an acoustically-ambiguous word target identifying neural areas sensitive to interactions between sentential bias and phonetic ambiguity. Of interest is whether the locus or nature of the interactions resembles those previously demonstrated for auditory-only effects. fMRI results show significant interaction effects in right mid-middle temporal gyrus (RmMTG) and bilateral anterior superior temporal gyri (aSTG), regions along the ventral language comprehension stream that map sound onto meaning. These regions are more anterior than those previously identified for auditory-only effects; however, the same cross-over interaction pattern emerged implying similar underlying computations at play. The findings suggest that the mechanisms that integrate information across modality and across sentence and phonetic levels of processing recruit amodal areas where reading and spoken lexical and semantic access converge. Taken together, results support interactive accounts of speech and language processing.

Keywords speech perception; crossmodal integration; semantic context; functional magnetic resonance imaging

Corresponding Author Sara Guediche

Corresponding Author's Institution Brown University

Order of Authors Sara Guediche, Yuli Zhu, Domenic Minicucci, Sheila Blumstein

Suggested reviewers David Gow, Nienke van Atteveldt, Milene Bonte, Richardson Fiona, Colin Phillips, Ksenija Marinkovic

Submission Files Included in this PDF

File Name [File Type]

CoverLetter_BL_July8.docx [Cover Letter]

reviewerResponseBL_July8.docx [Response to Reviewers]

SCW_changes_marked.docx [Revised Manuscript with Changes Marked]

Highlights.docx [Highlights]

SCW_2019_July8.docx [Manuscript File]

BrainAndLang_Fig.tiff [Figure]

Table1_BL.docx [Table]

Table2_BL.docx [Table]

SupplementaryMaterials_BL_July8.pdf [e-Component]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Research Data Related to this Submission

There are no linked research data sets for this submission. The following reason is given:
Data will be made available on request

Ref: BRLN_2018_116

Title: Written sentence context effects on acoustic-phonetic perception: fMRI reveals cross-modal semantic-perceptual interactions

Journal: Brain and Language

Dear Professor Li.,

Thank you for the opportunity to revise our manuscript, BRLN_2018_116 entitled “Written sentence context effects on acoustic-phonetic perception: fMRI reveals cross-modal semantic-perceptual interactions”. We have provided a detailed response to each reviewer comment in the uploaded document designated as “response to reviewers”. Each response is provided in italics below the reviewer comment, which is provided in regular typeface.

We believe the suggested revisions have significantly improved the manuscript and hope that the changes better convey the novelty and overall contribution of these findings to understanding interactions between reading and speech processing and to neural models of language processing, more generally.

As requested by the journal, we submitted the revised manuscript with the changes marked. To facilitate the revision process, we have also provided an unmarked “clean” version.

Thank you for your time and consideration.

Sincerely,

Sara Guediche
Basque Center on Cognition, Brain and Language
Postdoctoral Researcher
Marie Curie Fellow

Ref: BRLN_2018_116

Title: Written sentence context effects on acoustic-phonetic perception: fMRI reveals cross-modal semantic-perceptual interactions

Journal: Brain and Language

Dear Professor Li.,

Thank you for the opportunity to revise our manuscript, BRLN_2018_116 entitled "Written sentence context effects on acoustic-phonetic perception: fMRI reveals cross-modal semantic-perceptual interactions". We have addressed reviewer comments below. Comments of the reviewers are in regular typeface and our response is in italics.

Dear Dr. Guediche,

I'm writing with regard to your submission to Brain and Language (BRLN). Enclosed are the comments on your manuscript from reviewers whose expertise falls within the area of your research.

As you can see from the attached, while the reviewers comment that your study has the potential to make a significant contribution to the literature, they have raised a number of theoretical and methodological issues with your study. Most clearly is the concern, from both reviewers, that the theoretical motivation and significance are not clearly spelt out, and as a consequence, the theoretical novelty/contribution of your work is uncertain. I would ask, along with the reviewers, that you consider carefully whether this study conveys to our BRLN readership truly significant new work on top of your previous work, or, whether it's only a small incremental step with a slightly new context/modality. The latter case will not be the favored approach for a publication in our journal.

With these comments in mind, I encourage you to carefully consider the reviewers' comments and submit a revised version of your manuscript for further consideration. Please provide a response letter that discusses in detail how you have addressed the reviewers' concerns, and in places where you fail to address their concerns, why.

We believe the suggested revisions have significantly improved the manuscript and hope that the changes better convey the novelty and overall contribution of these findings to understanding interactions between reading and speech processing and to neural models of language processing, more generally. In particular, the results provide evidence for cross-modal (visual/auditory), cross-hierarchical (sentence context meaning/phonetic ambiguity) effects on perception, and identify the locus and pattern of cross-modal sentential effects on phonetic perception. By leveraging a similar design and analysis approach to that previously used in a study conducted in the auditory modality, the findings allowed us to assess potential similarities or differences in the locus and computations demonstrated for auditory-only effects shown in prior research. Thus, the findings call for further specification in neural and computational models of both spoken and written language processing.

Please clearly mark the changed text or updated material in the manuscript. I will likely send your revised submission to the same reviewers, or to new reviewers where appropriate. The outcome of further reviews will determine whether your revised manuscript will be accepted or rejected.

The paper has been extensively rewritten. However, as requested by the journal, we submitted the revised manuscript with the changes marked. In addition, to facilitate the revision process, we have also provided an unmarked "clean" version.

Thank you for your time and consideration,

Sincerely,

*Sara Guediche
Basque Center on Cognition, Brain and Language
Postdoctoral Researcher
Marie Curie Fellow*

-Reviewer 1

Summary:

The authors present an fMRI experiment investigating the integration of sentential context from a written sentence with spoken target words. They find that areas hypothesized to be involved in semantic-perceptual interactions are activated in this task, similar to previous findings using an auditory context and auditory target (Guediche et al. 2013). The results are discussed relative to models of cross-modal integration and top-down processing in language comprehension.

General comments:

Overall, I found the study to be straightforward with a clear set of results. The paradigm is similar to a previous study by two of the authors that I am familiar with, and it does a nice job taking that approach and applying it to a cross-modal situation. There are a few places where the authors could be clearer in the theoretical motivation for the study and interpretation of the results, but I believe these issues could easily be addressed with some revision. I will outline the main issues below, and list more specific, minor points after.

1. Motivation for the study

The questions posed in the study have to do with the extent to which top-down semantic information can influence acoustic-phonetic processing. This is a classic question in speech perception, and it is a question that there continues to be considerable debate over. The methods used here (like those in Guediche et al.) are excellent for addressing this debate, particularly with regard to the neuroanatomical location of these effects.

However, I think the motivation for the current study could be stronger. The introduction makes it clear that there is evidence for auditory-auditory interactions of this type, from the authors' previous work, as well as others. So, the question being addressed in the current study is not whether top-down processing occurs, but whether it occurs in a cross-modal situation with a written context and auditory target word.

We thank the reviewer for this comment. We have reorganized and revised the text describing the motivation for this study, which is to investigate cross-modal integration of information in language by examining the interaction between top-down processing of sentential context and bottom-up processing of acoustic phonetic properties of speech. Here, we identify the neural locus and nature of potential integration mechanisms that optimally combines two different sources (levels) of information extracted from two different modalities (visual and auditory).

The introduction also establishes that cross-modal interactions have been shown to affect phonological processing but not acoustic-phonetic processing (p. 3). The distinction between these levels of processing is a little unclear to me in this context, as both would seem to be evidence of cross-modal interactions; some clarification would be helpful.

Indeed both phonological and acoustic-phonetic effects reflect cross-modal interactions. Our point was that there is a difference between more abstract phonological processing and lower-level acoustic-phonetic processing. Whereas phonological effects in reading reflect an abstraction of the auditory input, the acoustic-phonetic effects in reading found here reflect the influence of a visually presented semantic context on the perception of sub-phonemic, low-level acoustic-phonetic properties. The distinction between phonetic and phonological processing in this context was made to highlight the need for increased specification in detailing the interactions between reading and speech processing in neuro-anatomical models of language processing (going beyond the abstract phonological processes currently incorporated in neuroanatomical models of reading). We have added text to clarify the distinction at the bottom of page 3 and at the bottom of page 14. Nevertheless, this was not primary to the goal of the study and so we have, at the same time, shortened the text that highlights this distinction so that it would not detract from the main question regarding the locus and nature of cross-modal sentential effects on acoustic-phonetic perceptual ambiguity.

I am left with the sense that the specific question being addressed here is not whether top-down effects occur, or whether cross-modal interactions occur, but whether cross-modal interactions between semantic and lower-level perceptual representations are mediated by the same mechanisms as those involved in a purely auditory domain. I think the manuscript would be better motivated based on why this is an important question to address, above and beyond the questions of whether top-down processing or cross-modal interactions occur.

In point of fact, we intended to address all of the questions you enumerated in the previous paragraph. While fMRI, at present, cannot address questions about the details of underlying neurobiological mechanisms of perception, the locus and pattern of these interactions can be used to make inferences about the potential underlying computations and processes that give rise to flexible perception. The extent to which there are differences and/or similarities in locus/pattern in a cross-modal compared to a unimodal context further elucidates the potential mechanisms that support the integration of information facilitating comprehension and speaks to the nature of current neuroanatomical and computational models of language. We have attempted to clarify the importance of this point, both in the introduction and in the discussion.

One factor that may be important is that the task seems different from the auditory-only experiment of Guediche et al. in an important way. In the current experiment, subjects are reading a sentence and hearing a spoken target word. Thus, subjects' task is not to integrate the target word into the running spoken discourse (as it is in Guediche et al., 2013). Here, it is more like the written context acts as a semantic prime, which then exerts an influence on how the perceptual input is processed. It's not clear to me that this necessarily involves same underlying processes as integrating meaning with bottom-up phonetic input during online spoken language processing. Some clarification about this would be useful. This would help better frame the study relative to previous work and provide a clearer motivation for the questions being addressed.

The task for the subjects is the same in both the Guediche et al. (2013) and the current experiment, which is to identify whether an auditorily presented target word was 'goat' or 'coat'. Thus, we do not fully agree that the two tasks differ. Although it is true that the stimulus input was different – i.e. it was auditory within modality in Guediche et al. 2013, and it is cross-modality in the current experiment, it is not totally clear to us that there would be an a priori reason to think that the subject is not integrating the meaning of the read phrase with the meaning of the heard target as they did when they heard the phrase. It is also possible that both the within and cross-modality instances reflect semantic priming. However, it is also possible that the meaning of a written sentence context may not activate the predicted auditory target word form to the same level as in the auditory task, which may contribute to some of the differences that emerged between the cross-modal and auditory-only tasks. In either case, some meaning of the phrase (providing the sentence context) needs to be extracted to influence access to the auditory target word. That said, the results of the current study suggest that the integration of sentential context with phonetic ambiguity across modalities, may be supported by a similar underlying mechanism as that for the auditory-only context. In contrast, although both rely on the ventral stream, the specific brain regions differ. As suggested, we have clarified this issue in the discussion, at the bottom of page 12 and the bottom of page 13.

2. Mechanisms driving these effects (particularly w.r.t. top-down effects)

I also have some questions about the precise mechanisms involved, and whether they are the same as or similar to those observed in Guediche et al. (2013). Do the effects found in the current study really reflect top-down processing? The strongest predictions from interactive models would seem to be that the semantic context (whether driven by auditory or visual input) can influence low-level perceptual representations. If the authors had observed these interactions in areas argued to be directly involved in speech perception (e.g. pSTG), that would provide more convincing evidence of a top-down effect. However, the interactions are found in areas suggested to reflect amodal integration of semantic and perceptual information (p. 9).

In our view, the reported interaction effect provides the strongest evidence for feedforward/feedback interactive models (and the influence of (top-down processing on an acoustic manipulation that affects perception)) in the current study and, as well, in Guediche et al. (2013). In particular, the cross-over pattern reflected an interdependent effect sensitive to both the quality of the acoustic input and the predictive bias of the sentence context. Namely, decreased activation for acoustically ambiguous targets in a semantically biased context is consistent with top-down modulation, and increased activation for unambiguous targets in a semantically biased context reflects enhancement activation due to congruency. This BOLD response crossover interaction result is more consistent with an interactive (feedforward/feedback) than a feedforward computational account of neural information processing. The neural locus of the reported interaction effect is meant to identify those regions where top-down and bottom-up effects may mediate flexibility in perception. Even though the areas identified are amodal and thus not involved in low-level speech perception per se, they involve temporal lobe areas involved in lexical and semantic access.

Importantly, effects were **not** observed in decision-making areas, which would have been predicted by completely feedforward models. Thus, the results are inconsistent with feedforward models (as the authors state), but they also do not seem to provide the strongest support for interactive models. Compare this with the results of Guediche et al., where the interaction effects were observed in both MTG and posterior portions of STG. The locations of the effects in the current study appear to be more anterior than those effects, with no interaction effects in posterior STG.

As you indicate, the interaction does not emerge in frontal areas, which would be predicted by feedforward models. Nor does the pattern of the interaction track with condition difficulty which would also be predicted by feedforward models, as effects of context in feedforward models are attributed to decision-making processes.

Perhaps the differences between the two experiments can provide sufficient explanation for why the effects are not observed in more low-level speech perception areas. To be clear, I find the results to be convincing evidence that cross-model interactions occur, and

the demonstration of interactions between written semantic context and acoustic-phonetic processing is noteworthy on its own. But I think the authors could clarify the extent to which the results provide evidence of top-down processing in the way that Guediche et al. and other studies have shown. It may help to have more discussion of the similarities and differences between the current results and those of Guediche et al. (2013).

The reviewer makes an important point and we have discussed this issue in more detail in the discussion. We agree that the loci of interactions in the cross-modal tasks suggest the involvement of more anterior (more abstract), amodal areas involved in semantic processing compared to those found in Guediche et al. (2013). At the neural level, this may be a consequence of when and where information from two modalities converges such that a predictive context differentially modulates ambiguous vs. unambiguous acoustic properties of a target word. We also find effects in aSTG and mid-MTG, in the right hemisphere (effects in Guediche et al. 2013 were left-lateralized). The right-hemisphere bias fits with recent papers showing greater right-hemisphere recruitment for reading compared to listening. We have added text, in the discussion, on page 12 and page 13 to further elaborate potential similarities and differences.

-That paper also included a discussion of the differences between different types of top-down accounts (e.g. predictive coding vs. TRACE), which may help clarify this issue in the current study as well.

We have edited the discussion, on page 14, to further clarify this issue and cite the argument presented in Guediche et al. (2013).

3. Lack of effect on one side of the /g/-/k/ continuum

I was somewhat surprised by the lack of an interaction for the coat-biasing sentences, though this is consistent with the results of Guediche et al. I think more discussion of this is warranted. The pattern of results may be due to the fact that the ambiguous stimulus is /g/-biased (i.e. 62% /g/ responses in the neutral context). This perhaps makes it more difficult for the /k/-biasing context to "push around" the perceptual representation. Is that the correct interpretation?

We agree with the interpretation provided by Reviewer 1. Given that the ambiguous stimulus in the current study, across this set of participants, was perceived as 'goat' 60 percent of the time, this may reflect the overall bias towards more 'goat' responses in goat-biasing sentence contexts but not for more 'coat responses' in more coat-biasing sentence contexts.

However, we felt it was beyond the scope of this paper to provide a detailed discussion of the asymmetry beyond what we included in the text. It is worth noting that other studies (one cited in the paper by Burton and Blumstein, 1995) have also found asymmetries despite using a boundary that was more "symmetrical", i.e. 50%.

Note: In checking the behavioral results, we realized there was a typo in reporting the p-value for this interaction. This has been corrected and the full R output table is now included in Supplementary materials.

Related to this, I would ask if the authors could provide some additional analyses of the behavioral data. Specifically, since a Context x Target Type interaction is observed for the goat-biased vs. neutral contexts, it would be good to have a follow-up analysis confirming that responses to the ambiguous stimulus are different in the goat-biased vs. neutral contexts (i.e. the simple main effect of Context for the ambiguous Target Types).

As requested by the reviewer, we have conducted these additional analyses and incorporated them into the manuscript.

Also, RTs are provided in Table 1, but there are no statistical analyses of them. Perhaps the authors simply have no predictions to make for the RTs, but I would expect that, at minimum, the RTs would be longer for ambiguous stimuli than the other two stimulus types; this appears to be the case given the data in Table 1.

The reaction time analyses only show main effects of ambiguity. We now include them in Supplementary materials.

Specific comments:

p. 3 last paragraph: It took me a few times to read this paragraph to understand what the authors were trying to say. My understanding is that the first sentence is meant to establish cross-modal interactions between meaning and phonological representations. The second sentence is meant to establish these interactions for letters (i.e. no meaning) and acoustic-phonetic representations. This leaves the question of whether there are interactions between meaning and acoustic-phonetic representations (last sentence). Is that interpretation correct? I would suggest revising this paragraph to make these points more clearly.

This text has been rewritten.

p. 6: The manuscript is organized with the Method section at the end, which is fine. However, I think there needs to be more of a transition into the Results section. A summary of the task at the start of the Results section or end of the Introduction would be helpful.

The guidelines require that for a short communications article, the methods be presented at the end of the article. We agree that there needs to be a transition into the results and thus have added text at the end of the introduction that summarizes the task and hypotheses.

p. 10 line 8, "In this case": Please clarify. Do you mean in the case of the models being referenced here, or in the case of the current study?

Thank you for pointing out the lack of clarity. The intended meaning was “in those models”. Much of the introduction has been rewritten and we have taken care to remove such ambiguities.

p. 10 second paragraph: Please clarify this paragraph. Do you mean to say that existing models do not account for interactions between cross-modal semantic processing and lower-level acoustic phonetic processing, but that given the current data, they should?

In the rewritten text, we have clarified what we intended to say. In particular, should the results of the study confirm such cross-modal sentential-phonetic ambiguity interactions, neuro-anatomical models of reading would need to be modified to predict such interactions and their locus.

p. 13 Procedure section: Was the same random order of trials used for all participants?

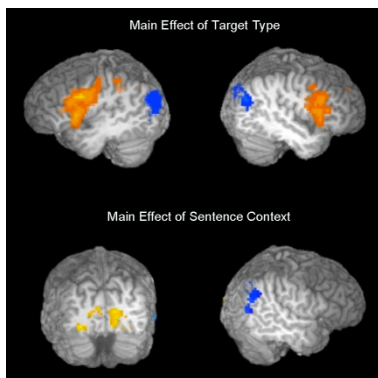
The same random presentation within a run was used for all participants. This information was present, though perhaps not clear and so the wording has been clarified in the methods section.

p. 15 line 1, "justified by the data": What was the procedure used to determine the random effects structure? Was it a backward-stepping procedure (e.g. start with a model containing all random effect terms and remove higher-order terms until a significant decrease in model fit is observed)?

We used a forward-stepping model. This information has been added.

Figures: The one figure shows the areas involved in the Context x Target Type interactions, but I think it may be helpful to have figures for the main effects also.

The suggested figures (below) have been added to the supplementary materials.



Top Panel: Regions showing differences in activity between Ambiguous and Unambiguous targets. The left hemisphere is on the left side. Sagittal slices at X= -43, 43, corrected at a voxelwise threshold of $p < .01$ (top panel). **Bottom Panel:** Regions showing differences in activity between Goat-biased and Neutral sentence contexts. Yellow-scale reflects greater activity for Neutral compared to Goat and blue scale reflects greater activity for Goat compared to Neutral. Coronal Slice at Y = 73, corrected at a voxelwise threshold of $p < .01$. Sagittal slices at X= 50, corrected at a voxelwise threshold of $p < .05$ (see Table 2 of the manuscript).

-Reviewer 2

The current paper extends the prior work of the authors on the effects of semantic context on phonetic processing. The authors use a cross-modal (visual context – auditory target) paradigm to determine whether semantic context priming across modality has the same effect on the phonetic processing of voicing on words (GOAT vs. COAT) as shown previously in their behavioral paradigms as well as their prior imaging study (Guediche et al., 2013). Their results show that they do show a similar interaction pattern in anterior MTG and STG as their previous studies suggesting that semantic priming across modalities has an impact on the processing of ambiguously voiced stimuli. While the authors' findings are compelling, there are some nuances in their findings regarding the specific regions of STG and MTG (anterior rather than posterior) that render some of their conclusions regarding post-lexical processing less valid than suggested.

The use of the terminology “post-lexical” is based on models of speech processing which to date only incorporate the lexical level. We agree that the use of this terminology raises unnecessary confusions and have replaced it with “decision-making” processes, a term also used by feedforward-only models.

I do not have specific itemized comments rather than suggestions for making the argument structure in the paper clearer for the audience and the movement from hypotheses, predictions, results and conclusions more facile and valid.

We thank the reviewer for this observation. We have restructured the paper. Additional text has been added to the discussion to elaborate the logic of the conclusions.

Given the structure of the manuscript format with the Methodology appearing last and the Results appearing after the Introduction, it would be prudent to state clearly your manipulation and hypotheses in a brief paragraph at the transition between sections 1 and 2. As it stands now, your Introduction ends with a discussion of general theory regarding feedforward and feedback models of activation.

We have extensively rewritten the introduction. We now include our hypotheses and describe the stimuli and task.

As one reads through the Results, even after having read the Methods first, it is unclear what the predictions are given the particular manipulations especially considering that both the semantic contexts and the VOT of the Target auditory words are being manipulated.

In rewriting the introduction, the manipulations, analyses, and hypotheses have been further elaborated and clarified. In particular, we have described the predictions about the specific interaction of interest between the two manipulated factors (semantic bias of the sentence context, and phonetic ambiguity).

In the conclusion on page 10, the authors state: “Interactions were not found in other regions typically associated with post-perceptual decision-making processes or executive control (e.g., pars opercularis and dorsolateral prefrontal cortex), conflict resolution (e.g. the anterior cingulate), or auditory sensory processing (e.g., Heschl’s gyrus).” It is important to point out that there were main effects of ambiguity in these same regions as shown in Table 2.

We thank the reviewer for this thoughtful comment. We have added a few sentences, on page 13 (first full paragraph) of the discussion, to remind the reader that the main effects of ambiguity do emerge in these regions. With the predictions now clarified, we have tried to

further highlight the important difference between regions that are sensitive to effects of ambiguity compared to those that are sensitive to the interaction between ambiguity and the semantic bias of the sentence context.

The findings from Guediche et al., 2013 are more posterior ($Y=-31$ and $Y=-51$) on the MTG and STG, whereas the findings from this study are far more anterior nearing the temporal pole. From most accounts, these regions are far closer to the primary and secondary auditory belt regions and could be considered less “amodal” compared to the ATL regions. Because the authors reference both their previous findings as well as those of Rueckl et al 2015, it would be important to note when speaking about the STG/MTG whether the authors are referring to Anterior, Mid, or Posterior portions of these regions. This is of greater importance in the Introduction where the authors refer broadly to findings in the literature. The authors are more careful in the conclusion section (but should still be clear) to point out the findings are more anterior.

We have clarified these differences in the introduction and re-labeled the regions that emerged in our analyses accordingly.

Lastly, the discussion of feedforward and feedback activation should be tempered here because of the nature of the BOLD signal. The author’s refer to the work of Marinkovic et al 2003. Their MEG findings from 2002 show that when reading a single word activity spreads forward between 0-300msec but at around 400msec the activity begins to move from anterior to posterior suggesting both feedforward and feedback activation can be seen within a second. The BOLD signal is not that precise and the activation that is seen in ATL may reflect the confirmation of predictive inferences from the sentence context (see Ferstl & Neumann, 2008; Kuperberg, Lakshmanan, Caplan, & Holcomb, 2006; Mar, 2004, 2011; Prat, Mason, & Just, 2012).

Reviewer 2 brings up a very important point about differences between MEG studies that report the time course of spreading activation. We hope that the changes made to the introduction have further clarified that we are not investigating the time course of changes in activation due to feedforward and feedback processing. Rather, we are investigating the locus and nature of the interactions of the two factors of interest (sentential context and ambiguity), which may reflect the combination of feedforward and feedback modulation. While the observed interaction suggests that both feedforward and feedback signals are contributing to the changes in the BOLD signal, as Reviewer 2 points out, we cannot directly measure the time course of feedback or feedforward processing in real-time.

Regarding the reviewer’s second point, we hope that the added text describing how each of the simultaneously manipulated factors modulates activity in these regions in an interdependent (crossover interaction) manner (activity to ambiguous target changes in opposite direction depending on context) clarifies why it is unlikely that these regions are merely reflecting the confirmation of predictive inferences, as those reported in the studies discussed above. The current study is also distinguished from the references provided above in that phonetic ambiguity is also being manipulated and shows that a low level acoustic manipulation impacts activity in those regions differentially, depending on context.

The current interaction findings do mimic those of Guediche et al 2013 and cannot be interpreted in a straightforward fashion. One question is whether such findings would extend to the conflict situation (goat context – unambiguous ‘coat’).

In order to address this question, we conducted two additional analyses:

The first was parallel to the reported analysis but with the unambiguous ‘coat’ stimulus- as suggested by the reviewer. A 2x2 ANOVA was conducted with Sentence Type (goat-biased, neutral) and Target Type (Boundary, Unambiguous ‘coat’) as within-subject factors. At a p-value of .01, only the anterior cingulate was significant, Therefore, the table below reports all regions at a p-value of .05 that are greater than 100 voxels. As can be observed, the regions reported below do not include regions in temporal cortex. Rather, they are all regions that have been associated with decision-making processes and conflict monitoring, error detection/correction.

	x	y	z	size	t-value
LIFG	-28	8	26	373	17.52
Right Cerebellum (Crus I)	41	-58	-25	301	19.47
Left Cingulate	-7	-4	26	228	28.31
Right Thalamus	5	-31	-4	158	22.03

In addition, to determine regions that were specifically sensitive to the congruency of the sentence context and targets, we compared the biased sentence contexts (goat-biased and coat-biased) when they were paired either with the congruent unambiguous target or with the incongruent targets. The only regions showing enhanced activation for the congruent condition was the right angular gyrus. Both the cingulate and left insula show greater activation for the incongruent condition. The table below includes all clusters above 100 voxels at $p = .05$. Again, there were not significant clusters that emerged in temporal cortex.

Congruent-Incongruent

	x	y	z	Cluster	t-value
Cingulate/Corpus collosum	-1	-4	20	504	-4.74
Left Insula	-27	17	-7	223	-3.64
Right Angular gyrus	35	-73	35	111	3.31

Written sentence context and phonetic perception

Written sentence context effects on acoustic-phonetic perception: fMRI reveals cross-modal semantic-perceptual interactions

Sara Guediche^{1,4}, Yuli Zhu², Domenic Minicucci¹, and Sheila E. Blumstein^{1,3}

¹*Department of Cognitive, Linguistic & Psychological Sciences, Brown University*

²*Neuroscience Department, Brown University*

³*Brown Institute for Brain Science, Brown University*

⁴*BCBL. Basque Center on Cognition, Brain and Language, Donostia-San Sebastian, Spain*

Corresponding author: Sara Guediche

Sara Guediche is now at the BCBL. Basque Center on Cognition, Brain and Language, San Sebastian, Spain
s.guediche@bcbl.eu

Yuli Zhu is now a resident Physician ~~in Chicago, Illinois~~ at Langone Medical Center, NY

~~technical staff in the pain imaging lab~~ at Massachusetts General Hospital
System Analyst at The Athinoula A. Martinos Center for Biomedical Imaging

Abstract

~~The current~~This study ~~investigates the neural areas that are sensitive to~~examines cross-modality effects of a semantically-biased *written* sentence context on the perception of ~~phonetically~~an acoustically-ambiguous stimuli. ~~The goal is to determine whether cross-modality effects engage similar brain areas as those previously shown for within-modality auditory sentence processing. Behavioral results revealed that the meaning of a read sentence context influenced the auditory perception of an ambiguous stimulus associated with two possible real-word targets distinguished by the acoustic property voice-onset time. FMRI results show that activation in temporal areas (right middle temporal gyrus and bilateral anterior middle temporal gyri) was modulated by~~target identifying neural areas sensitive to interactions between ~~the semantiesentential~~ bias of the visually presented written sentence context and the quality of the spoken acoustic and phonetic input. ambiguity. Of interest is whether the locus or nature of the interactions resembles those previously demonstrated for auditory-only effects. FMRI results show significant interaction effects in right mid-middle temporal gyrus (RmMTG) and bilateral anterior superior temporal gyri (aSTG), regions along the ventral language comprehension stream that map sound onto meaning. These regions are situated along a ventral stream consisting of neural areas commonly activated~~more anterior than those previously identified for both~~auditory-only effects; however, the same cross-over interaction pattern emerged implying similar underlying computations at play. The findings suggest that the mechanisms that integrate information across modality and across sentence and phonetic levels of processing recruit amodal areas where reading and spoken lexical and semantic access converge. Taken together, results support interactive

accounts of speech and language processing. ~~These cross-modality effects seem to rely on interactions between semantic and perceptual processes at points of overlap between spoken and written language processing networks facilitating semantic-phonetic integration.~~

1. Introduction

~~Sensory~~ Seemingly rudimentary perceptual processing of sensory input ~~is~~ can be influenced by information from a different ~~modalities~~ modality. For example, change in the flutter rate of an auditory stimulus affects perception of flicker rate of a visual light stimulus (Shipley, 1964). ~~Similarly, auditory perception of a complex auditory speech stimulus is affected~~ Such perceptual flexibility is fundamental for the accurate mapping of more complex and highly variable sensory input such as the auditory speech signal. The goal of this fMRI study is to examine cross-modality, visual-auditory, effects of reading a sentence context on the auditory perception of speech input in order to determine the neural locus and nature of potential cross-modal interactions that support flexibility in speech perception. To this end, we manipulate the semantic bias of a written sentence context and the acoustic-phonetic ambiguity of two possible target words differing in the acoustic dimension, voice-onset time.

Experimental research has shown that the perception of speech can be influenced by visual information from multiple sources including articulatory gestures (McGurk and MacDonald, 1976; Bertelson et al., 1997), congruent orthography (van Atteveldt et al., 2004; Keetels et al., 2016; Bonte et al., 2017), and ~~other~~ written ~~context~~ text (Gagnepain et al., 2012; Sohoglu et al., 2012; Wild et al., 2012; Clos et al., 2014). ~~The current study focuses on cross-modality effects across different levels of linguistic processing by investigating potential interactions between the semantic bias of a written sentence context and subsequent acoustic phonetic perception, and the influence of semantic-phonetic interactions on brain activity.~~

~~————Prior studies examining interactions~~2012). Interactions between reading and speech perception have ~~focused~~emerged in studies focusing on the influence of written information (matching text and meaning contexts) on phonological processing. ~~Results show that orthographic processing of letters (from text input) can also induce changes in, i.e.~~ the perception of ~~lower-level acoustic phonetic properties of speech that distinguish between two phonetic categories and drive changes in activity in auditory brain regions associated with phonetic perception (Keetels et al., phonemes.2016; Bonte et al., 2017).~~ To our knowledge, there is no empirical evidence showing the influence of a written semanticsentence context on perception of lower level, acoustic–phonetic (“sub-phonemic”) properties of speech nor is there information about the neural areas that support potential interactions between the reading sentence comprehension network and lower-level acoustic phonetic processing-acoustically ambiguous speech stimuli.

~~————Within the auditory modality, the influence of semantic bias of a spoken sentence context on effects differentially influence~~ the perception of ~~an acoustic dimension, voice-onset time (VOT), distinguishing voiced from voiceless stop consonants, has been examined~~phonetically ambiguous compared to unambiguous words (Borsky et al., 1998; Guediche et al., 2013; Gow and Olson, 2016). FMRI results ~~showed~~show interactions between the semantic bias of a sentence and acoustic–phonetic ~~manipulation~~manipulations in a region that encompasses parts of the left middle and superior temporal gyrus (LMTG/STG) (Guediche et al., 2013). ~~Neural~~Importantly, neural activity was differentially modulated by the semantic bias of the context depending on the quality of the acoustic phonetic information, ~~suggesting that this area appears to integrate semantic and acoustic phonetic information.—~~ In particular, there was a cross-

over interaction: an unambiguous target stimulus showed *increased* activation in a semantically biased context whereas an acoustically ambiguous target showed *decreased* activation. This interaction pattern reflected an interdependent effect sensitive to both the quality of the acoustic input and the predictive bias of the sentence context consistent with an interactive (feedforward/feedback) rather than a feedforward computational account of neural information processing, which would predict neither the observed interaction pattern nor the ventral stream locus of the interaction (e.g., Ahissar, 2004; Nelken and Ahissar, 2008; Friston, 2010).

——Information about the meaning of a sentence context can also be conveyed through written input and may similarly impact acoustic–phonetic processing engaging similar neural pathways as those shown for within auditory modality stimuli. ~~However, an additional consideration for neural models of reading is that multiple routes have been proposed for accessing semantic information (Coltheart et al., 1993; Binder et al., 2005; Plaut and McClelland, 1996; Jobard et al., 2003; Binder et al., 2005). This may impact the brain regions that support the integration of semantic and phonetic information.~~

——~~Indeed, the~~What is not known is the locus of such effects. The initial processing of written versus spoken input relies on different brain regions recruiting visual cortex (including ventral occipital temporal cortex, vOTC) for written input (Dehane et al., 2002; Price and Devlin, 2011; Richardson et al., 2011) and auditory cortex (including Heschl’s gyrus and surrounding areas) for spoken input (Davis and Johnsrude, 2003; Peelle et al. 2010). However, later stages ~~of phonological and semantic processing~~ appear to ~~engage~~converge on to common areas. For example, manipulations of orthographic-phonological congruency modulate activation, producing a supramodal

response, in brain regions associated with abstract phonological processes including the superior temporal gyrus (van Atteveldt et al., 2004). At even later stages of semantic processing, reading and spoken language share many brain areas ~~associated with the semantic network~~ including the anterior superior and middle temporal gyri, and inferior parietal cortex, areas associated with lexical, semantic, and sentence processing (e.g., Fiez and Petersen, 1998; Vandenberghe et al., 2002; Jobard et al., 2003; Binder et al., 2005; Lau and Poeppel, 2008; Binder et al., 2009). A study that specifically investigated the overlap between the written and spoken word recognition networks, across different languages, identified sites of convergence in bilateral IFG, bilateral MTG, STG, and the left SMG (Rueckl et al., 2015). A number of other neuroimaging studies and meta-analyses also point to amodal involvement of the anterior temporal lobes in sentence comprehension and semantic processing, more generally (e.g., Marinkovic and Dhond, 2003; also see Richardson et al., 2011; Price, 2012). ~~While it is widely accepted that visual orthographic input may be converted to phonological representations, the proposed direct route from written input to semantic information (Coltheart et al. 1993) also allows the possibility for bypassing phonological and lower level acoustic phonetic processing. Therefore, it is unclear which points of speech-reading convergence, if any, serve as a site (or sites) for integrating acoustic phonetic and semantic information across modalities;~~ Lau and Poeppel, 2008; Bemis and Pykkänen, 2012).

Thus, it is an open question which points of speech-reading convergence areas serve as sites for integrating acoustic-phonetic and semantic information across modalities and whether and how the integration mechanisms across the two modalities might differ with respect to locus and activation pattern from those in the auditory

modality. Across models of reading, multiple routes have been proposed for accessing semantic information including routes that may bypass phonological and lower level acoustic-phonetic processing (Coltheart et al., 1993; Binder et al., 2005; Plaut and McClelland, 1996; Jobard et al., 2003; Cohen et al., 2004, Binder et al., 2005), potentially impacting the brain networks involved in integrating semantic and phonetic information.

Interactive accounts of speech and language processing (modeled after other perceptual domains, where top-down effects influence ventral stream processing (e.g., Ahissar, 2004; Bar et al., 2006 Friston, 2010)) posit that context effects on perception occur through “feedforward/feedback” interactions (top-down and bottom-up processes) (e.g., Nelken and Ahissar, 2008), and predict that multiple brain regions would be sensitive to the interaction between semantic and acoustic/phonetic manipulations. In particular, cross-modal interactions between semantic and phonetic manipulations are likely to emerge along the ventral speech/language processing stream, specifically. Thus, in the current study, of the superior/middle temporal gyri, regions implicated in lexical access and word recognition may be involved, as has been shown for within-auditory modality interactions between semantic bias and phonetic manipulations (Guediche et al., 2013; Davis et al., 2011; see also Gow and Olson, 2016). ~~Additionally, integrating semantic and phonetic information, across the two modalities may lead to interactions in amodal areas that are common to reading and spoken language comprehension such as the anterior superior and middle temporal gyri (aSTG, aMTG, SMG, IFG). Thus, the~~ The locus of semantic-phonetic interactions may depend on the regions that provide semantic feedback signals (amodal and/or modality independent regions), which interact with subsequent feedforward processes activated by phonetic information (see Gow and

Olson, 2016). Thus, integrating semantic and phonetic information across the two modalities may lead to interactions in amodal areas that are common to reading and spoken language comprehension such as the anterior superior and middle temporal gyri (aSTG, aMTG, SMG, IFG). If the underlying computations that integrate semantic and phonetic information are similar to those previously found, in the auditory domain, then the same crossover interaction pattern should be observed here.

FeedforwardIn contrast, feedforward (bottom-up) models of speech perception, which hypothesize that context effects on perception result from post-perceptualother higher level cognitive processes, and involve would predict interactions in areas associated with decision-making (e.g., Davis et al., 2011; although see Sohoglu et al., 2012; Norris et al. 2000; McQueen et al. 2006). Recent research 2006) such as in the pars opercularis (BA45) and dorsolateral prefrontal cortex. In this case, these regions should show increased activity in the more difficult conditions. In sum, the current study will address the following questions: will reading a semantically biased context interact with and influence the perception of an acoustically ambiguous speech signal, as has focused on the potential signals that feed back from frontal areas to been shown in the auditory cortex (e.g. Sohoglu and Davis 2016). In such a view, modality (Borsky et al. 1998; Guediche et al., 2013; Gow and Olson, 2016), and if so, what will be the locus and pattern of this interaction? Will the locus of such cross-modality effects be similar to or different from those found within the auditory modality; and will the interaction pattern mirror those found within the auditory modality?

We predict that cross-modal interactions between semantic and aeoustiephonetic manipulations should emerge in areas often referred to in the models as “post-perceptual”

~~or decision-making such as the pars opercularis (BA45) and dorsolateral prefrontal cortex are likely to emerge along the ventral speech/language processing stream, as has been shown for within auditory modality interactions (Guediche et al., 2013; Davis et al., 2011; see also Gow and Olson, 2016), as well as in the MTG and additional amodal sentence processing areas that are common to reading and spoken sentence processing such as the anterior superior and middle temporal gyri (aSTG) and angular gyri (AG) (Bemis and Pylkkanen, 2013, Regev et al., 2013). We also hypothesize that the underlying computations that integrate information will be similar to those previously found in the auditory domain, resulting in a similar crossover interaction pattern.~~

~~In this study, participants read sentence fragments (neutral – ‘he saw the ___’ or biased – ‘he milked the ___’) followed by an auditorily presented target stimulus that was acoustically ambiguous - ‘g/coat’ or it was unambiguous – ‘goat’ and ‘coat’. They were asked to press one button to indicate they heard the word ‘goat’ and another button to indicate they heard the word ‘coat’. Behavioral and fMRI results are presented below.~~

2. Results

2.1 Behavioral Results

Behavioral results for the 18 participants included in the imaging analysis are shown in Table 1.

~~The results of the A~~ mixed effects regression analysis model on the accuracy data was conducted. The results showed a significant Context Bias x Target Type interaction for the goat-Bias versus neutral contrast code, $p = .014$, whereas the interaction for the coat-biased versus neutral code was not significant, $p = .7$ (see Supplementary

Materials for full details of model output). The follow-up t-test on the boundary stimulus supports this asymmetry, neutral versus coat-biased, $p = .12$, and neutral vs. goat-biased, $p = .058$ (FDR-corrected for 36 comparisons). Thus, compared to neutral sentence contexts, goat-biasing sentences (e.g. ‘he fed the’) had a significant effect on ‘goat’ proportion responses for the ambiguous stimulus compared to the unambiguous ‘goat’ stimulus, whereas the coat-biasing sentences (e.g. ‘he buttoned the’) did not affect responses for the ambiguous stimulus compared to the unambiguous ‘coat’ stimulus. These findings are consistent with the results reported in Guediche et al. (2013), which showed a similar asymmetric bias and with the literature showing perceptual asymmetries for velar stop consonants (e.g., Burton and Blumstein 2005; Myers and Blumstein, 2008). Analyses on the reaction time data only show a significant effect of ambiguity, $p < .001$ (see Supplementary Materials).

~~—————These findings are consistent with the results reported in Guediche et al. (2013), which showed an asymmetric bias. Relative to the neutral context, more ‘goat’ responses were found in the goat-biased sentence context; however, no change was found for the coat-biased context. The magnitude of the perceptual effect as well as perceptual asymmetries for velar consonants have been reported in the literature~~

2.2. *fMRI Results*

The behavioral results showed that the semantic bias of a written sentence context influenced the perception of voice-onset time through the interaction between the ~~type of semantic context (goat-biased, vs. neutral) sentence contexts~~ and the ~~type of target type (ambiguous, vs. unambiguous).~~ Therefore, fMRI ‘goat’ target type. FMRI analyses

focused on the interaction between these two factors. ~~To that end, as~~ As in Guediche et al. (2013), a two-way ANOVA was conducted using Sentence Context (neutral, goat-biased) and target type (ambiguous, unambiguous 'goat') as fixed factors and participant as a random factor on the percent signal change values for each condition.

Table 2 shows the results. Regions that showed an interaction between Sentence Context and Target Type are corrected at a voxel-wise threshold of $p < .01$. Significant clusters, corrected for a cluster size threshold at an alpha of .05, were found in the right mid-middle temporal cortex (extending into the hippocampus), the right anterior superior temporal gyrus, and another cluster (uncorrected for cluster size) in the left anterior temporal lobe (see Table 2, Figure 1). As Figure 1 shows, there was a cross-over interaction with greater activity for unambiguous targets in the semantically biased compared to the neutral context and less activity for acoustically ambiguous targets in a semantically biased context compared to the neutral context.

The main effect of sentence context emerged ~~only~~ in one cluster at a $p < .01$ (corrected for cluster size at a threshold an alpha of .05), the right lingual gyrus extending into middle occipital gyrus (see Table 2). This cluster showed greater activity for the neutral compared to the goat-biased sentence context. At a reduced significance threshold, $p < .05$ (corrected for cluster size at an alpha of .05), increased activity for the goat-biased compared to neutral contrast was found in a cluster which included parts of the inferior parietal lobule and posterior MTG (see Table 2), and Figure in Supplementary Materials). This is consistent with prior studies showing that the angular gyrus is crucial for reading comprehension and enhances its coupling with anterior temporal cortex during combinatorial semantic processing (Molinaro et al., 2011).

In addition, the ANOVA revealed a main effect of Target Type (ambiguous, unambiguous ‘goat’) in several clusters with greater activity found for the ambiguous stimulus compared to the unambiguous stimulus in frontal, superior temporal, cingulate, and motor regions. There was more activity found for the unambiguous compared to the ambiguous stimulus in the right posterior middle temporal gyrus extending into the angular gyrus, occipital, posterior cingulate, and parahippocampal areas.

3. Discussion

~~Reading requires orthographic, phonological, and semantic processing and relies on a network of brain areas that overlap in part with spoken language processing. Thus, auditory speech processing may be subject to the same type of “top-down” influences of meaning from orthographic visual input as auditory spoken input. However, models of reading do not specify the underlying neural systems that support cross-modal integration of visually presented semantic information with acoustic-phonetic properties of speech. This study is the first to show such semantic-phonetic cross-modality effects on brain activity.~~

Three The goal of this study was to investigate the neural areas that support the integration of information across modalities and different levels of linguistic processing, and to assess potential similarities or differences in the locus or computations demonstrated for auditory-only effects in prior research (Guediche et al., 2013). To this end, we examined cross-modal, visual-auditory, effects between the meaning of a *written* sentence context and the acoustic-phonetic ambiguity of an auditory target word.

Significant interaction effects emerged behaviorally between the semantic bias of a sentence (goat-biased vs. neutral) and acoustic-phonetic manipulations (ambiguous vs. unambiguous 'goat'). The fMRI results showed three clusters emerged showing a Semantic Context by Target Type sensitive to this interaction: the right anterior and right middle portion of the middle temporal cortex (corrected for whole brain), and, in addition, the left anterior superior temporal cortex (uncorrected for whole brain). Thus, as predicted, interactions emerged in regions typically associated with amodal processing of lexical (mid-MTG) and semantic (aSTG) processing were sensitive to the sentential context/phonetic ambiguity interaction. As was the case for auditory sentence context effects, the areas showing this interaction effect were restricted to regions situated in the ventral language stream thought to support access to meaning (Hickok and Poeppel, 2007). Nevertheless, in contrast to auditory-only sentence context effects, the neural locus of the interaction was more anterior and more right-lateralized.

The right-hemispheric bias found in the current study is consistent with recent studies showing that comprehension networks for reading are more reliant on the right hemisphere (Horowitz-Kraus et al., 2015). The more anterior locus of the cross-modality interaction effect suggests that integration across the two modalities relies heavily on amodal semantic areas shown to be involved in processing the meaning of a sentence.

There are several possible reasons for the differences that emerged in the neural loci of the visual-auditory and auditory only interaction effects. The potential engagement of different processing routes for mapping visual and auditory inputs onto meaning is one possibility. Another possibility is that there may be inherent task differences involved in integrating information across two input streams compared to a single, auditory input

stream. Another consideration is that the meaning of a written sentence context may not activate the predicted auditory target word form to the same level as would an auditorily presented sentence context. Consequently, there could be greater reliance on anterior, semantic, areas than on middle temporal, lexical, areas as sentence meaning is integrated with an auditory word target.

Importantly, the neural areas showing the semantic bias x phonetic ambiguity interaction effect did not include any frontal regions, typically associated with decision-making processes or executive control (e.g., pars opercularis and dorsolateral prefrontal cortex). Feedforward (bottom-up) models would have predicted an interaction in such areas showing a pattern in which the most difficult condition would have the highest level of activation (neutral sentence paired with ambiguous target) and the easiest condition (biased sentence paired with unambiguous target), the lowest level of activation (c.f. Norris et al., 2000; McQueen et al., 2006; Myers and Blumstein, 2008; Davis et al., 2011; Sohoglu et al., 2012). Frontal areas were, however, sensitive to main effects of ambiguity, consistent with prior work, demonstrating their contribution to more general cognitive processes such as lexical selection, conflict resolution, decision-making, and the encoding of prediction errors (Guediche et al, 2013; Rogers and Davis, 2017).

The pattern of the observed interaction effects provides additional insight into the underlying computations that support cross-modal integration of sentential context with phonetic ambiguity to support comprehension. Similar to the results of Guediche et al. (2013), the changes of activity that emerged in all three areas showed a *crossover* interaction ~~with increased activation for unambiguous targets in a semantically biased context, and pattern with~~ decreased activation for acoustically ambiguous targets in a

semantically biased context. ~~Because the magnitude of percent change in the BOLD signal for the ambiguous target in the neutral sentence context was comparable to the percent change in the BOLD signal for the unambiguous target in the biased condition, the pattern of this interaction effect appears to reflect the integration of semantic and phonetic information and not perceptual (consistent with top-down modulation), and increased activation for unambiguous targets in a semantically biased context (reflecting enhancement activation due to congruency). The interdependent changes in brain activity observed in the current study are consistent with the idea that regions involved in flexible perception may be modulated by the differences between both bottom-up and top-down signals across conditions, rather than the differences in difficulty. Importantly, the interaction emerged in the ventral stream in amodal regions that may serve to facilitate the integration of meaning derived from the visual stream and, as would be predicted by feedforward-only accounts. A recent electrophysiological study also argues against a feedforward-only account of semantic context effects on acoustic-phonetic processing. Interactions were not found in other regions typically associated with post-perceptual decision-making processes or executive control (e.g., pars opercularis and dorsolateral prefrontal cortex), conflict resolution-phonetic perception; Getz and Toscano (2019) found that early auditory processing of an ambiguous acoustic-phonetic target (e.g. the anterior cingulate), or auditory sensory processing (e.g., b/pear) modulated N1 ERP responses, depending on the semantically associated predictions of the prime (e.g., Heschl's gyrus). teddy).~~

———A number of neuro-anatomical models of reading propose that semantic-phonological interactions may be mediated by connections between occipital, temporal,

parietal, and frontal areas (Jobard et al., 2003; Richardson et al., 2011; Carreiras et al., 2014). In this case, the phonological effects Taken together, the results of the current study are consistent with interactive models of language processing that allow for semantic influences on acoustic-phonetic perception via feedback and top-down modulation across different modalities. However, existing versions of interactive computational speech perception models such as TRACE and Predictive coding, do not include sentence level processing. Thus, it is not clear whether versions of these models that incorporate sentence meaning would predict the cross-over interaction pattern obtained in the current study (see Guediche et al., 2013 for further discussion).

The phonological processes represented by current neuroanatomical models of reading reflect access to more abstract sound representations in which the acoustic details of the input have been stripped away. ~~However, (e.g. Jobard et al., 2003; Richardson et al., 2011; Carreiras et al., 2014), despite recent evidence showing reading-induced phonetic recalibration (Bonte et al., 2017). The current results also suggest that neuroanatomical models of reading and speech perception have not yet incorporated~~ language should predict cross-modal interactions between semantic and interaction effects with lower level acoustic phonetic ~~speech~~ processing ~~where phonetic detail is a part of the sound representation, despite the fact that such interactions have been proposed theoretically~~ as those found in the current study and as expected by some theoretical accounts of reading (Morton 1969; Massaro & Cohen, 1991; Gaskell and Marslen-Wilson 1997; Harm and Seidenberg, 1994; 2004).

~~Rueckl et al. (2015) showed convergence between networks involved in phonological speech perception and reading in bilateral IFG, bilateral MTG to STG, and~~

left SMG. Their finding was consistent across languages of different levels of sound-letter transparency suggesting highly overlapping networks for speech and reading. Such overlap provides the opportunity for multiple sites of cross-talk between visual and spoken language input within and across both hemispheres. As described, the areas showing interactions in the current study were restricted to the temporal lobe and suggest that semantic/perceptual integration depends on ventral stream speech processing. Whether the influence of acoustic information on aMTG/STG activity Whether the influence of the acoustic-phonetic information on aMTG/STG activity as it interacts with the context is in the form of a trace of the acoustic details of speech or whether it is in the form of graded levels of activation of word representations (Chen and Davis, 2015; see Toscano et al. 2010) or sentence level information meaning (Vandenberghe et al., 2002; Spitsyna et al., 2006; Robson and Zahn, 2014) is still an open question.

ATL has been reported in a number of other studies and in meta-analyses examining reading

Summary

Together, these and other published Horowitz-Kraus et al (2015) showed that comprehension networks for reading, compared to listening, rely more on the right hemisphere, which may account for the bilateral effects found in the current study.

The main effects for Semantic Context and Target Type that emerged suggest that despite cross-modal semantic-phonetic interactions in areas common to reading and speech perception, modality-dependent areas are also at play. Thus, cross-modal semantic context effects on perception must rely on partially distinct and overlapping regions that

~~process written and spoken input, potentially integrating information in regions common to both networks.~~

~~———Feedforward (bottom-up only) accounts of speech processing (e.g., MERGE (Norris et al., 2000)) suggest that interaction effects should occur post-perceptually as a result of decision-making and executive control processes (e.f.2011; Sohoglu et al., 2012), typically attributed to frontal areas including the inferior frontal gyrus (BA 45/47) and dorsolateral prefrontal cortex (Badre and Wagner 2004; Binder et al., 2004). The findings of the current study are inconsistent with these models.~~

~~Our findings are better accountedcall for by a more integrated model of language processing that allows for cross-modal interactions between both feedback (top-down modulation) and feedforward (bottom-up) processes, producing bidirectional information flow amonginput from multiple modalities (reading and spoken input) and their interaction across multiple levels of processing. ~~These conclusions are based on the~~ The locus and nature of the interaction effect in the middle temporalcurrent study suggests that amodal mid- and anterior temporal cortex and are consistent with our original hypothesis, that areas showing an regions support cross-modal integration across sentence reading and speech processing. Additionally, the pattern of the interaction are common to reading and speech perception. effect supports models of language that allow for interactions between feedforward and feedback processing, challenging strictly feedforward accounts.~~

~~———Of interest, convergence between the neural systems involved in these processesreading and speech perception may be a crucial indicator of reading proficiency (Rueckl et al., 2015; Preston et al., 2016). In this view, reading is symbiotic on the speech~~

processing system. The results of the current study showing that high-level abstract semantic processing extracted from reading influences the perception of low-level acoustic phonetic properties of speech is consistent with this view. A consequence of such a functional architecture could be that the extent of cross-modal integration across different levels of processing (e.g., semantic-perceptual) has may also have an influence on reading proficiency.

4. Materials and Methods

4.1. Participants

Twenty-two right-handed native English speakers with normal hearing and no reported speech or language disorders, (14 male) with a mean age of 20.5 yrs (SD = 3.39) participated and were compensated \$25/hr. The data from two participants were eliminated for excessive movement in the scanner, one for being distracted while in the scanner (reported neuromuscular sensations), and one who failed to accurately perform the task (more than 40% responses were missing for the ambiguous target condition) leaving 18 participants for the data analysis.

4.2. Stimuli

Sentence fragments and the target stimuli were taken from Guediche et al. (2013). As reported in Guediche et al. (2013), the voice-onset time (VOT) for the targets were as follows: unambiguous ‘coat’ (VOT, 70 ms), unambiguous ‘goat’ (VOT, 21 ms), and ambiguous ‘goat/coat’ stimulus (VOT, 40ms). The targets were preceded by three types of visually presented sentence contexts: ‘goat’-biased, e.g. “he milked the __”, ‘coat’-

biased, e.g. “he wore the ___”, or neutral, e.g. “he found the ___” (see Guediche et al., 2013 Appendix for full list).

4.3. Procedure

In a slow-event related sparse-sampling design, participants were presented with written sentence fragments followed by the acoustic word target. The total sentence fragment and ‘coat/goat’ presentation duration was 2 s. The sentence fragment appeared on the screen for 1.25 s in black-type text (size 24, Arial Bold font) centered on a computer monitor against a white background, then disappeared from the screen, followed in 50 ms by the acoustic target. The sentence context and the target were presented in a random order that was used across all the participants.

Participants followed the same instructions and trial procedure previously described in Guediche et al., (2013) pressing one button to indicate they heard the word ‘goat’ and another button to indicate they heard the word ‘coat’. A 12 s delay followed the end of each trial.

There were six experimental runs consisting of 36 pseudo-randomized trials. Across all runs, there were 24 trials in each of the nine conditions. Each run included four trials of each sentence context paired with each of the three target stimuli.

A practice run consisting of two of each of the different trial types was conducted during the anatomical scan.

4.4. Scanning Protocol

fMRI data were collected using a 3 Tesla Siemens Trio fMRI scanner equipped with a 32 receiver channel head coil. High-resolution T1 weighted structural images were acquired for each subject for the purpose of anatomical co-registration (TR = 1900 ms, TE = 2.98,

TI = 900 ms, FOV = 256, 1 mm³ isotropic voxels). Functional images were acquired using an echo-planar sequence (TR = 4000 ms, TE = 28 ms, FOV = 192 mm, 3 mm³ isotropic voxels) in thirty-three 3 mm³ thick slices. A sparse-sampling, slow-event related design was utilized in which the acquisition of each volume was followed by a 2 second silent gap during which the presentation of the stimuli occurred. Each of the six runs consisted of 146 EPI volumes collected over 9 min and 48 seconds, yielding a total of 876 EPI volumes.

A projector and MR compatible in-ear headphones using Bliss fMRI Runner (mertus.org) were used for the presentation of the stimuli.

4.5. Analysis

4.5.1. Behavioral Analysis

An analysis was conducted on the behavioral responses using a mixed effects model that included all conditions (implemented in R) with a logit-transformed proportion goat 'goat' responses as the dependent measure. The three context conditions were effect coded using two contrast codes: 1) goat-biased versus neutral (1, -1), and 2) coat-biased versus neutral (1, -1). The three target conditions were effect coded using two contrast codes 1) unambiguous 'goat' vs. ambiguous (1, -1), and 2) unambiguous 'coat' vs. ambiguous (1, -1). The maximum random effect structure justified by the data (~~using a forward stepping procedure~~ sentence context and ~~procedure~~) included random intercept and target slope for subject. Trials with no responses and outlier trials that were two standard deviations above the mean reaction time in each condition were removed from the analysis. The maximum number of excluded trials for any given participant in any given condition was 4 out of a total of 24 trials per condition.

4.5.2. *Imaging Analysis*

The imaging data were analyzed using The Analysis of Functional NeuroImages software (AFNI) (Cox, 1996). Functional images were corrected for head motion using a six-parameter rigid body transform. The structural images for each subject were normalized to Talairach stereotaxic space, and the functional images were aligned to structural images resampled to 3 mm³. Spatial smoothing was achieved using a 6-mm, full-width, half-maximum Gaussian kernel. Stimulus onset times for each of the 9 conditions were convolved with a gamma-variate hemodynamic function provided by AFNI and used in a general linear model (GLM) analysis on individual EPI data, including six motion parameters as covariates of no interest. Beta coefficients were converted to percent signal change units (baseline-corrected) and entered into an ANOVA analysis and masked using an 18-subject composite mask.

3.2.1. *fMRI ANOVA Analysis*

Given the perceptual asymmetry shown in the behavioral data, we used a 2x2 ANOVA analysis with Sentence Context (goat-biased, neutral) and Target Type (unambiguous 'goat'; ambiguous 'goat/coat') as within-subject factors and subject as a random factor using percent signal change as a dependent measure on the fMRI data. Monte Carlo simulations were performed in order to determine the significant cluster size (29 voxels) for a voxel-wise threshold of $p = .01$, at an alpha of .05.

Acknowledgments: This work was supported in part by NIDCD grant RO1 DC006220.

The authors acknowledge the contributions of Paul Allopenna, John Mertus, and Efthymia Kapnoula. We would also like to thank Professor Ping Li and reviewers Dr. Joe Toscano and an anonymous reviewer for their thoughtful feedback and comments on the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health.

References

- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8(10), 457-464.
- Badre, D., & Wagner, A. D. (2004). Selection, integration, and conflict monitoring: assessing and the nature generality of prefrontal cognitive control mechanisms. *Neuron*, 41(3), 473-487.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the national academy of sciences*, 103(2), 449-454.
- Bemis, D. K., & Pylkkänen, L. (2012). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, 23(8), 1859-1873.
- Bertelson, P., Vroomenti, J., & Gelderti, B.D. (1997). Auditory-visual interaction in voice localization and in bimodal speech recognition: The effects of desynchronization. *Audio-Visual Speech Processing: Computational & Cognitive Science Approaches*.
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7 (3), 295-301.
- Binder, J. R., Medler, D. A., Desai, R., Conant, L. L., & Liebenthal, E. (2005). Some neurophysiological constraints on models of word naming. *Neuroimage*, 27(3), 677-693.

- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex* 19(12), 2767-2796.
- Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: an fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, 17(9), 1353-1366.
- Borsky, S., B. Tuller, & Shapiro, L. P. (1998). "How to milk a coat:" the effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America*, 103(5), 2670-2676.
- Bonte, M., Correia, J. M., Keetels, M., Vroomen, J., & Formisano, E. (2017). Reading induced shifts of perceptual speech representations in auditory cortex. *Scientific Reports*, 7.
- Burton, M., & Blumstein, S. (1995). Lexical effects on phonetic categorization: The role of stimulus naturalness and stimulus quality. *Journal of Experimental Psychology: Human Perception and Performance*, 21(5), 1230–1235.
- Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, 18(2), 90-98.
- Chen, Y., Davis, M. H., Pulvermüller, F., & Hauk, O. (2015). Early visual word processing is flexible: evidence from spatiotemporal brain dynamics. *Journal of Cognitive Neuroscience*, 27(9), 1738-51
- Clos, M., Langner, R., Meyer, M., Oechslin, M. S., Zilles, K., & Eickhoff, S. B. (2014). Effects of prior information on decoding degraded speech: an fMRI study. *Human Brain Mapping*, 35(1), 61-74.

Cohen, L., Jobert, A., Le Bihan, D., & Dehaene, S. (2004). Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *Neuroimage*, 23(4), 1256-1270.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3), 162-173.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 23(8), 3423-3431.

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top-down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, 23(12), 3914-3932.

Fiez, J. A., & Petersen, S. E. (1998). Neuroimaging studies of word reading. *Proceedings of the National Academy of Sciences*, 95(3), 914-921.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Reviews Neuroscience*, 11(2), 127-138.

Gagnepain, P., R. N. Henson, & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology*, 22(7), 615-621.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12(5-6), 613-656.

Getz, L. M., & Toscano, J. C. (2019). Electrophysiological evidence for top-down lexical influences on early speech perception. *Psychological Science*, 30, 830-841.

- Gow Jr, D. W., & Olson, B. B. (2016). Sentential influences on acoustic-phonetic processing: A Granger causality analysis of multimodal imaging data. *Language, Cognition and Neuroscience*, 31(7), 841-855.
- Guediche, S., Salvata, C., & Blumstein, S. E. (2013). Temporal cortex reflects effects of sentence context on phonetic processing. *Journal of Cognitive Neuroscience*, 25(5), 706-718.
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review*, 111(3), 662.
- Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402.
- Horowitz-Kraus, T., Grainger, M., DiFrancesco, M., Vannest, J., Holland, S. K., & CMIND Authorship Consortium. (2015). Right is not always wrong: DTI and fMRI evidence for the reliance of reading comprehension on language-comprehension networks in the right hemisphere. *Brain Imaging and Behavior*, 9(1), 19-31.
- Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: a metaanalysis of 35 neuroimaging studies. *Neuroimage*, 20, 693-712.
- Keetels, M., Schakel, L., Bonte, M., & Vroomen, J. (2016). Phonetic recalibration of speech by text. *Attention, Perception, & Psychophysics*, 78(3), 938-945.
- Lau, E., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920-933.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, 38, 487-497.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology*, 23(4), 558-614.

McGurk, H., & MacDonald, J. (1976). Visual influences on speech perception processes. *Perception & Psychophysics*, 24(3), 253-257.

McQueen, J., D. Norris, et al. (2006). Are there really interactive processes in speech perception. *Trends in Cognitive Sciences*, 10(12), 533-533.

[Molinaro, N., Barber, H. A., & Carreiras, M. \(2011\). Grammatical agreement processing in reading: ERP findings and future directions. *cortex*, 47\(8\), 908-930.](#)

Morton, J. (1969). Interaction of information in word recognition. *Psychological review*, 76(2), 165.

Myers, E. and S. Blumstein (2008). The neural bases of the lexical effect: an fMRI investigation. *Cerebral Cortex*, 18(2), 278-288.

Nelken, I., & Ahissar, M. (2008). Low-level information and high-level perception: the case of speech in noise. *PLoS biology*, 6(5), e126.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences*, 23(3), 299-325.

Obleser, J. & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, 20, 633-640.

- Peelle, J. E., Johnsrude, I. S. & Davis, M. H. (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in Human Neuroscience*, 4, 1–3.
- Preston, J. L., Molfese, P. J., Frost, S. J., Mencl, W. E., Fulbright, R. K., Hoeft, F., ... & Pugh, K. R. (2016). Print-speech convergence predicts future reading outcomes in early readers. *Psychological science*, 27(1), 75-84.
- Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, 15(6), 246-253.
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, 62(2), 816-847.
- Richardson, F. M., Seghier, M. L., Leff, A. P., Thomas, M. S., & Price, C. J. (2011). Multiple routes from occipital to temporal cortices during reading. *Journal of Neuroscience*, 31(22), 8349-8247.
- Robson, H., Zahn, R., Keidel, J. L., Binney, R. J., Sage, K., & Lambon Ralph, M. A. (2014). The anterior temporal lobes support residual comprehension in Wernicke's aphasia. *Brain*, 137(3), 931-943.
- Rogers, J. C., & Davis, M. H. (2017). Inferior frontal cortex contributions to the recognition of spoken words and their constituent speech sounds. *Journal of cognitive neuroscience*, 29(5), 919-936.
- Rueckl, J. G., Paz-Alonso, P. M., Molfese, P. J., Kuo, W. J., Bick, A., Frost, S. J., ... & Lee, J. R. (2015). Universal brain signature of proficient reading: Evidence from four contrasting languages. *Proceedings of the National Academy of Sciences*, 112(50), 15510-15515.
- Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, 1328-1330.

- Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*(25), 8443-8453.
- Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences*, *113*(12), E1747-E1756.
- Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. (2006). Converging language streams in the human temporal lobe. *Journal of Neuroscience*, *26*(28), 7328-7336.
- Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, *21*(10), 1532-1540.
- Wild, C. J., M. H. Davis, & Johnsrude, I. S. (2012). Human auditory cortex is sensitive to the perceived clarity of speech. *Neuroimage*, *60*(2), 1490-1502.
- Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*(2), 271-282.
- Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*, *14*(4), 550-560.

Figure Caption

Figure 1. Regions showing interaction between Sentence Context (goat-biased, neutral) and Target Type (ambiguous, unambiguous 'goat'). The left hemisphere is on the left side. Sagittal slices at X=-47, and X=47. Bar graph shows percent signal change extracted from each region and standard error of the mean over subjects.

Highlights:

- Semantic bias of a written sentence context influences acoustic-phonetic perception
- Crossmodal semantic phonetic integration found in ventral language stream
- Crossmodal crossover interaction effect same as unimodal auditory interaction pattern
- Computations for uni- & crossmodal integration same despite distinct ventral locus
- Findings support interactive multimodal neural model of language processing

Written sentence context effects on acoustic-phonetic perception: fMRI reveals cross-modal semantic-perceptual interactions

Sara Guediche^{1,4}, Yuli Zhu², Domenic Minicucci¹, and Sheila E. Blumstein^{1,3}

¹*Department of Cognitive, Linguistic & Psychological Sciences, Brown University*

²*Neuroscience Department, Brown University*

³*Brown Institute for Brain Science, Brown University*

⁴*BCBL. Basque Center on Cognition, Brain and Language, Donostia-San Sebastian, Spain*

Corresponding author: Sara Guediche

Sara Guediche is now at the BCBL. Basque Center on Cognition, Brain and Language, San Sebastian, Spain
s.guediche@bcbl.eu

Yuli Zhu is now a resident Physician in Chicago, Illinois

Domenic Minicucci is now technical staff in the pain imaging lab at Massachusetts General Hospital

Abstract

This study examines cross-modality effects of a semantically-biased *written* sentence context on the perception of an acoustically-ambiguous word target identifying neural areas sensitive to interactions between sentential bias and phonetic ambiguity. Of interest is whether the locus or nature of the interactions resembles those previously demonstrated for auditory-only effects. fMRI results show significant interaction effects in right mid-middle temporal gyrus (RmMTG) and bilateral anterior superior temporal gyri (aSTG), regions along the ventral language comprehension stream that map sound onto meaning. These regions are more anterior than those previously identified for auditory-only effects; however, the same cross-over interaction pattern emerged implying similar underlying computations at play. The findings suggest that the mechanisms that integrate information across modality and across sentence and phonetic levels of processing recruit amodal areas where reading and spoken lexical and semantic access converge. Taken together, results support interactive accounts of speech and language processing.

1. Introduction

Seemingly rudimentary perceptual processing of sensory input can be influenced by information from a different modality. For example, change in the flutter rate of an auditory stimulus affects perception of flicker rate of a visual light stimulus (Shipley, 1964). Such perceptual flexibility is fundamental for the accurate mapping of more complex and highly variable sensory input such as the auditory speech signal. The goal of this fMRI study is to examine cross-modality, visual-auditory, effects of reading a sentence context on the auditory perception of speech input in order to determine the neural locus and nature of potential cross-modal interactions that support flexibility in speech perception. To this end, we manipulate the semantic bias of a *written* sentence context and the acoustic-phonetic ambiguity of two possible target words differing in the acoustic dimension, voice-onset time.

Experimental research has shown that the perception of speech can be influenced by visual information from multiple sources including articulatory gestures (McGurk and MacDonald, 1976; Bertelson et al., 1997), congruent orthography (van Atteveldt et al., 2004; Keetels et al., 2016; Bonte et al., 2017), and written text (Gagnepain et al., 2012; Sohoglu et al., 2012; Wild et al., 2012; Clos et al., 2012). Interactions between reading and speech perception have emerged in studies focusing on the influence of written information (matching text and meaning contexts) on phonological processing, i.e. the perception of phonemes. To our knowledge, there is no empirical evidence showing the influence of a written sentence context on perception of lower level, acoustic-phonetic (“sub-phonemic”) properties of speech nor is there information about the neural areas that

support potential interactions between the reading sentence comprehension network and acoustically ambiguous speech stimuli.

Within the auditory modality, sentence context effects differentially influence the perception of phonetically ambiguous compared to unambiguous words (Borsky et al., 1998; Guediche et al., 2013; Gow and Olson, 2016). fMRI results show interactions between the semantic bias of a sentence and acoustic-phonetic manipulations in a region that encompasses parts of the left middle and superior temporal gyrus (LMTG/STG) (Guediche et al., 2013). Importantly, neural activity was differentially modulated by the semantic bias of the context depending on the quality of the acoustic phonetic information. In particular, there was a cross-over interaction: an unambiguous target stimulus showed *increased* activation in a semantically biased context whereas an acoustically ambiguous target showed *decreased* activation. This interaction pattern reflected an interdependent effect sensitive to both the quality of the acoustic input and the predictive bias of the sentence context consistent with an interactive (feedforward/feedback) rather than a feedforward computational account of neural information processing, which would predict neither the observed interaction pattern nor the ventral stream locus of the interaction (e.g., Ahissar, 2004; Nelken and Ahissar, 2008; Friston, 2010).

Information about the meaning of a sentence context can also be conveyed through written input and may similarly impact acoustic-phonetic processing engaging similar neural pathways as those shown for within auditory modality stimuli. What is not known is the locus of such effects. The initial processing of written versus spoken input relies on different brain regions recruiting visual cortex (including ventral occipital

temporal cortex, vOTC) for written input (Dehane et al., 2002; Price and Devlin, 2011; Richardson et al., 2011) and auditory cortex (including Heschl's gyrus and surrounding areas) for spoken input (Davis and Johnsrude, 2003; Peelle et al. 2010). However, later stages appear to converge on to common areas. For example, manipulations of orthographic-phonological congruency modulate activation, producing a supramodal response, in brain regions associated with abstract phonological processes including the superior temporal gyrus (van Atteveldt et al., 2004). At even later stages of processing, reading and spoken language share many brain areas including the anterior superior and middle temporal gyri, and inferior parietal cortex, areas associated with lexical, semantic, and sentence processing (e.g., Fiez and Petersen, 1998; Vandenberghe et al., 2002; Jobard et al., 2003; Binder et al., 2005; Lau and Poeppel, 2008; Binder et al., 2009). A study that specifically investigated the overlap between the written and spoken word recognition networks, across different languages, identified sites of convergence in bilateral IFG, bilateral MTG, STG, and the left SMG (Rueckl et al., 2015). A number of other neuroimaging studies and meta-analyses also point to amodal involvement of the anterior temporal lobes in sentence comprehension and semantic processing, more generally (e.g., Marinkovic and Dhond, 2003; also see Richardson et al., 2011; Price, 2012; Lau and Poeppel, 2008; Bemis and Pyllkanen, 2012).

Thus, it is an open question which points of speech-reading convergence areas serve as sites for integrating acoustic-phonetic and semantic information across modalities and whether and how the integration mechanisms across the two modalities might differ with respect to locus and activation pattern from those in the auditory modality. Across models of reading, multiple routes have been proposed for accessing

semantic information including routes that may bypass phonological and lower level acoustic-phonetic processing (Coltheart et al., 1993; Binder et al., 2005; Plaut and McClelland, 1996; Jobard et al., 2003; Cohen et al., 2004, Binder et al., 2005), potentially impacting the brain networks involved in integrating semantic and phonetic information.

Interactive accounts of speech and language processing (modeled after other perceptual domains, where top-down effects influence ventral stream processing (e.g., Ahissar, 2004; Bar et al., 2006 Friston, 2010)) posit that context effects on perception occur through “feedforward/feedback” interactions (top-down and bottom-up processes) (e.g., Nelken and Ahissar, 2008), and predict that multiple brain regions would be sensitive to the interaction between semantic and acoustic-phonetic manipulations. In particular, cross-modal interactions between semantic and phonetic manipulations are likely to emerge along the ventral speech/language processing stream. Thus, in the current study, of the superior/middle temporal gyri, regions implicated in lexical access and word recognition may be involved, as has been shown for within-auditory modality interactions between semantic bias and phonetic manipulations (Guediche et al., 2013; Davis et al., 2011; see also Gow and Olson, 2016). The locus of semantic-phonetic interactions may depend on the regions that provide semantic feedback signals (amodal and/or modality independent regions), which interact with subsequent feedforward processes activated by phonetic information (see Gow and Olson, 2016). Thus, integrating semantic and phonetic information across the two modalities may lead to interactions in amodal areas that are common to reading and spoken language comprehension such as the anterior superior and middle temporal gyri (aSTG, aMTG, SMG, IFG). If the underlying computations that integrate semantic and phonetic

information are similar to those previously found, in the auditory domain, then the same crossover interaction pattern should be observed here.

In contrast, feedforward (bottom-up) models of speech perception, which hypothesize that context effects on perception result from other higher level cognitive processes would predict interactions in areas associated with decision-making (e.g., Davis et al., 2011; although see Sohoglu et al., 2012; Norris et al. 2000; McQueen et al. 2006) such as in the pars opercularis (BA45) and dorsolateral prefrontal cortex. In this case, these regions should show increased activity in the more difficult conditions. In sum, the current study will address the following questions: will reading a semantically biased context interact with and influence the perception of an acoustically ambiguous speech signal, as has been shown in the auditory modality (Borsky et al. 1998; Guediche et al., 2013; Gow and Olson, 2016), and if so, what will be the locus and pattern of this interaction? Will the locus of such cross-modality effects be similar to or different from those found within the auditory modality; and will the interaction pattern mirror those found within the auditory modality?

We predict that *cross-modal* interactions between semantic and phonetic manipulations are likely to emerge along the ventral speech/language processing stream, as has been shown for within auditory modality interactions (Guediche et al., 2013; Davis et al., 2011; see also Gow and Olson, 2016), as well as in the MTG and additional amodal sentence processing areas that are common to reading and spoken sentence processing such as the anterior superior and middle temporal gyri (aSTG) and angular gyri (AG) (Bemis and Pylkkanen, 2013, Regev et al., 2013). We also hypothesize that the

underlying computations that integrate information will be similar to those previously found in the auditory domain, resulting in a similar crossover interaction pattern.

In this study, participants read sentence fragments (neutral – ‘he saw the ___’ or biased – ‘he milked the _’) followed by an auditorily presented target stimulus that was acoustically ambiguous - ‘g/coat’ or it was unambiguous – ‘goat’ and ‘coat’. They were asked to press one button to indicate they heard the word ‘goat’ and another button to indicate they heard the word ‘coat’. Behavioral and fMRI results are presented below.

2. Results

2.1 Behavioral Results

Behavioral results for the 18 participants included in the imaging analysis are shown in Table 1. A mixed effects model on the accuracy data was conducted. The results showed a significant Context Bias x Target Type interaction for the goat-Bias versus neutral contrast code, $p = .013$, whereas the interaction for the coat-biased versus neutral code was not significant, $p = .95$ (see Supplementary Materials for full details of model output). The follow-up t-test on the boundary stimulus supports this asymmetry, neutral versus coat-biased, $p = .12$, and neutral vs. goat-biased, $p = .058$ (FDR-corrected for 36 comparisons). Thus, compared to neutral sentence contexts, goat-biasing sentences (e.g. ‘he fed the’) had a significant effect on ‘goat’ proportion responses for the ambiguous stimulus compared to the unambiguous ‘goat’ stimulus, whereas the coat-biasing sentences (e.g. ‘he buttoned the’) did not affect responses for the ambiguous stimulus compared to the unambiguous ‘coat’ stimulus. These findings are consistent with the results reported in Guediche et al. (2013), which showed a similar asymmetric bias and

with the literature showing perceptual asymmetries for velar stop consonants (e.g., Burton and Blumstein 2005; Myers and Blumstein, 2008). Analyses on the reaction time data only show a significant effect of ambiguity, $p < .001$ (see Supplementary Materials).

2.2. fMRI Results

The behavioral results showed that the semantic bias of a written sentence context influenced the perception of voice-onset time through the interaction between the goat-biased vs. neutral sentence contexts and the ambiguous vs. unambiguous ‘goat’ target type. fMRI analyses focused on the interaction between these two factors. As in Guediche et al. (2013), a two-way ANOVA was conducted using Sentence Context (neutral, goat-biased) and target type (ambiguous, unambiguous ‘goat’) as fixed factors and participant as a random factor on the percent signal change values for each condition.

Table 2 shows the results. Regions that showed an interaction between Sentence Context and Target Type are corrected at a voxel-wise threshold of $p < .01$. Significant clusters, corrected for a cluster size threshold at an alpha of .05, were found in the right mid-middle temporal cortex (extending into the hippocampus), the right anterior superior temporal gyrus, and another cluster (uncorrected for cluster size) in the left anterior temporal lobe (see Table 2, Figure 1). As Figure 1 shows, there was a cross-over interaction with greater activity for unambiguous targets in the semantically biased compared to the neutral context and less activity for acoustically ambiguous targets in a semantically biased context compared to the neutral context.

The main effect of sentence context emerged in one cluster at a $p < .01$ (corrected for cluster size at an alpha of .05), the right lingual gyrus extending into middle occipital

gyrus (see Table 2). This cluster showed greater activity for the neutral compared to the goat-biased sentence context. At a reduced significance threshold, $p < .05$ (corrected for cluster size at an alpha of .05), increased activity for the goat-biased compared to neutral contrast was found in a cluster which included parts of the inferior parietal lobule and posterior MTG (see Table 2, and Figure in Supplementary Materials). This is consistent with prior studies showing that the angular gyrus is crucial for reading comprehension and enhances its coupling with anterior temporal cortex during combinatorial semantic processing (Molinaro et al., 2011).

In addition, the ANOVA revealed a main effect of Target Type (ambiguous, unambiguous ‘goat’) in several clusters with greater activity found for the ambiguous stimulus compared to the unambiguous stimulus in frontal, superior temporal, cingulate, and motor regions. There was more activity found for the unambiguous compared to the ambiguous stimulus in the right posterior middle temporal gyrus extending into the angular gyrus, occipital, posterior cingulate, and parahippocampal areas.

3. Discussion

The goal of this study was to investigate the neural areas that support the integration of information across modalities and different levels of linguistic processing, and to assess potential similarities or differences in the locus or computations demonstrated for auditory-only effects in prior research (Guediche et al., 2013). To this end, we examined cross-modal, visual-auditory, effects between the meaning of a *written* sentence context and the acoustic-phonetic ambiguity of an auditory target word.

Significant interaction effects emerged behaviorally between the semantic bias of a sentence (goat-biased vs. neutral) and acoustic-phonetic manipulations (ambiguous vs. unambiguous 'goat'). The fMRI results showed three clusters sensitive to this interaction: the right anterior and right middle portion of the middle temporal cortex (corrected for whole brain), and, in addition, the left anterior superior temporal cortex (uncorrected for whole brain). Thus, as predicted, regions typically associated with amodal processing of lexical (mid-MTG) and semantic (aSTG) processing were sensitive to the sentential context/phonetic ambiguity interaction. As was the case for auditory sentence context effects, the areas showing this interaction effect were restricted to regions situated in the ventral language stream thought to support access to meaning (Hickok and Poeppel, 2007). Nevertheless, in contrast to auditory-only sentence context effects, the neural locus of the interaction was more anterior and more right-lateralized.

The right-hemispheric bias found in the current study is consistent with recent studies showing that comprehension networks for reading are more reliant on the right hemisphere (Horowitz-Kraus et al., 2015). The more anterior locus of the cross-modality interaction effect suggests that integration across the two modalities relies heavily on amodal semantic areas shown to be involved in processing the meaning of a sentence.

There are several possible reasons for the differences that emerged in the neural loci of the visual-auditory and auditory only interaction effects. The potential engagement of different processing routes for mapping visual and auditory inputs onto meaning is one possibility. Another possibility is that there may be inherent task differences involved in integrating information across two input streams compared to a single, auditory input stream. Another consideration is that the meaning of a written sentence context may not

activate the predicted auditory target word form to the same level as would an auditorily presented sentence context. Consequently, there could be greater reliance on anterior, semantic, areas than on middle temporal, lexical, areas as sentence meaning is integrated with an auditory word target.

Importantly, the neural areas showing the semantic bias x phonetic ambiguity interaction effect did not include any frontal regions, typically associated with decision-making processes or executive control (e.g., pars opercularis and dorsolateral prefrontal cortex). Feedforward (bottom-up) models would have predicted an interaction in such areas showing a pattern in which the most difficult condition would have the highest level of activation (neutral sentence paired with ambiguous target) and the easiest condition (biased sentence paired with unambiguous target), the lowest level of activation (c.f. Norris et al., 2000; McQueen et al., 2006; Myers and Blumstein, 2008; Davis et al., 2011; Sohoglu et al., 2012). Frontal areas were, however, sensitive to main effects of ambiguity, consistent with prior work, demonstrating their contribution to more general cognitive processes such as lexical selection, conflict resolution, decision-making, and the encoding of prediction errors (Guediche et al, 2013; Rogers and Davis, 2017).

The pattern of the observed interaction effects provides additional insight into the underlying computations that support cross-modal integration of sentential context with phonetic ambiguity to support comprehension. Similar to the results of Guediche et al. (2013), the changes of activity that emerged in all three areas showed a *crossover* interaction pattern with decreased activation for acoustically ambiguous targets in a semantically biased context (consistent with top-down modulation), and increased activation for unambiguous targets in a semantically biased context (reflecting

enhancement activation due to congruency). The interdependent changes in brain activity observed in the current study are consistent with the idea that regions involved in flexible perception may be modulated by the differences between *both* bottom-up and top-down signals across conditions, rather than the differences in difficulty, as would be predicted by feedforward-only accounts. A recent electrophysiological study also argues against a feedforward-only account of semantic context effects on acoustic-phonetic perception; Getz and Toscano (2019) found that early auditory processing of an ambiguous acoustic-phonetic target (e.g., b/pear) modulated N1 ERP responses, depending on the semantically associated predictions of the prime (e.g. teddy).

Taken together, the results of the current study are consistent with interactive models of language processing that allow for semantic influences on acoustic-phonetic perception via feedback and top-down modulation across different modalities. However, existing versions of interactive computational speech perception models such as TRACE and Predictive coding, do not include sentence level processing. Thus, it is not clear whether versions of these models that incorporate sentence meaning would predict the cross-over interaction pattern obtained in the current study (see Guediche et al., 2013 for further discussion).

The phonological processes represented by current neuroanatomical models of reading reflect access to more abstract sound representations in which the acoustic details of the input have been stripped away (e.g. Jobard et al., 2003; Richardson et al., 2011; Carreiras et al., 2014), despite recent evidence showing reading-induced phonetic recalibration (Bonte et al., 2017). The current results also suggest that neuroanatomical models of language should predict cross-modal interaction effects with lower level

acoustic phonetic processing such as those found in the current study and as expected by some theoretical accounts of reading (Morton 1969; Massaro & Cohen, 1991; Gaskell and Marslen-Wilson 1997; Harm and Seidenberg, 1994; 2004).

Whether the influence of the acoustic-phonetic information on aMTG/STG activity as it interacts with the context is in the form of a trace of the acoustic details of speech or whether it is in the form of graded levels of activation of word representations (Chen and Davis, 2015; see Toscano et al. 2010) or sentence meaning (Vandenberghe et al., 2002; Spitsyna et al., 2006; Robson and Zahn, 2014) is still an open question.

Summary

Together, these and other published findings call for a more integrated model of language processing that allows for input from multiple modalities (reading and spoken input) and their interaction across multiple levels of processing. The locus of the interaction effect in the current study suggests that amodal mid- and anterior temporal regions support cross-modal integration across sentence reading and speech processing. Additionally, the pattern of the interaction effect supports models of language that allow for interactions between feedforward and feedback processing, challenging strictly feedforward accounts.

Of interest, convergence between the neural systems involved in reading and speech perception may be a crucial indicator of reading proficiency (Preston et al., 2016). In this view, reading is symbiotic on the speech processing system. The results of the current study showing that high-level abstract semantic processing extracted from reading influences the perception of low-level acoustic phonetic properties of speech is consistent

with this view. A consequence of such a functional architecture could be that the extent of cross-modal integration across different levels of processing (e.g., semantic-perceptual) may also have an influence on reading proficiency.

4. Materials and Methods

4.1. Participants

Twenty-two right-handed native English speakers with normal hearing and no reported speech or language disorders, (14 male) with a mean age of 20.5 yrs (SD = 3.39) participated and were compensated \$25/hr. The data from two participants were eliminated for excessive movement in the scanner, one for being distracted while in the scanner (reported neuromuscular sensations), and one who failed to accurately perform the task (more than 40% responses were missing for the ambiguous target condition) leaving 18 participants for the data analysis.

4.2. Stimuli

Sentence fragments and the target stimuli were taken from Guediche et al. (2013). As reported in Guediche et al. (2013), the voice-onset time (VOT) for the targets were as follows: unambiguous ‘coat’ (VOT, 70 ms), unambiguous ‘goat’ (VOT, 21 ms), and ambiguous ‘goat/coat’ stimulus (VOT, 40ms). The targets were preceded by three types of visually presented sentence contexts: ‘goat’-biased, e.g. “he milked the ___”, ‘coat’-biased, e.g. “he wore the ___”, or neutral, e.g. “he found the ___” (see Guediche et al., 2013 Appendix for full list).

4.3. Procedure

In a slow-event related sparse-sampling design, participants were presented with written sentence fragments followed by the acoustic word target. The total sentence fragment and ‘coat/goat’ presentation duration was 2 s. The sentence fragment appeared on the screen for 1.25 s in black-type text (size 24, Arial Bold font) centered on a computer monitor against a white background, then disappeared from the screen, followed in 50 ms by the acoustic target. The sentence context and the target were presented in a random order that was used across all the participants.

Participants followed the same instructions and trial procedure previously described in Guediche et al., (2013) pressing one button to indicate they heard the word ‘goat’ and another button to indicate they heard the word ‘coat’. A 12 s delay followed the end of each trial.

There were six experimental runs consisting of 36 pseudo-randomized trials. Across all runs, there were 24 trials in each of the nine conditions. Each run included four trials of each sentence context paired with each of the three target stimuli. A practice run consisting of two of each of the different trial types was conducted during the anatomical scan.

4.4. Scanning Protocol

fMRI data were collected using a 3 Tesla Siemens Trio fMRI scanner equipped with a 32 receiver channel head coil. High-resolution T1 weighted structural images were acquired for each subject for the purpose of anatomical co-registration (TR = 1900 ms, TE = 2.98, TI = 900 ms, FOV = 256, 1 mm³ isotropic voxels). Functional images were acquired using an echo-planar sequence (TR = 4000 ms, TE = 28 ms, FOV = 192 mm, 3 mm³ isotropic voxels) in thirty-three 3 mm³ thick slices. A sparse-sampling, slow-event

related design was utilized in which the acquisition of each volume was followed by a 2 second silent gap during which the presentation of the stimuli occurred. Each of the six runs consisted of 146 EPI volumes collected over 9 min and 48 seconds, yielding a total of 876 EPI volumes.

A projector and MR compatible in-ear headphones using Bliss fMRI Runner (mertus.org) were used for the presentation of the stimuli.

4.5. Analysis

4.5.1. Behavioral Analysis

An analysis was conducted on the behavioral responses using a mixed effects model that included all conditions (implemented in R) with a logit-transformed proportion ‘goat’ responses as the dependent measure. The three context conditions were effect coded using two contrast codes: 1) goat-biased versus neutral (1, -1), and 2) coat-biased versus neutral (1, -1). The three target conditions were effect coded using two contrast codes 1) unambiguous ‘goat’ vs. ambiguous (1, -1), and 2) unambiguous ‘coat’ vs. ambiguous (1, -1). The maximum random effect structure justified by the data (~~using a forward-stepping procedure~~ and sentence context procedure) included random intercept and target slope for subject. Trials with no responses and outlier trials that were two standard deviations above the mean reaction time in each condition were removed from the analysis. The maximum number of excluded trials for any given participant in any given condition was 4 out of a total of 24 trials per condition.

4.5.2. Imaging Analysis

The imaging data were analyzed using The Analysis of Functional NeuroImages software (AFNI) (Cox, 1996). Functional images were corrected for head motion using a six-

parameter rigid body transform. The structural images for each subject were normalized to Talairach stereotaxic space, and the functional images were aligned to structural images resampled to 3 mm³. Spatial smoothing was achieved using a 6-mm, full-width, half-maximum Gaussian kernel. Stimulus onset times for each of the 9 conditions were convolved with a gamma-variate hemodynamic function provided by AFNI and used in a general linear model (GLM) analysis on individual EPI data, including six motion parameters as covariates of no interest. Beta coefficients were converted to percent signal change units (baseline-corrected) and entered into an ANOVA analysis and masked using an 18-subject composite mask.

3.2.1. fMRI ANOVA Analysis

Given the perceptual asymmetry shown in the behavioral data, we used a 2x2 ANOVA analysis with Sentence Context (goat-biased, neutral) and Target Type (unambiguous ‘goat’; ambiguous ‘goat/coat’) as within-subject factors and subject as a random factor using percent signal change as a dependent measure on the fMRI data. Monte Carlo simulations were performed in order to determine the significant cluster size (29 voxels) for a voxel-wise threshold of $p = .01$ at an alpha of .05.

Acknowledgments: This work was supported in part by NIDCD grant RO1 DC006220.

The authors acknowledge the contributions of Paul Allopenna, John Mertus, and Efthymia Kapnoula. We would also like to thank Ping Li and reviewers Joe Toscano and an anonymous reviewer for their thoughtful feedback and comments on the manuscript.

The content is solely the responsibility of the authors and does not necessarily represent the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health.

References

- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8(10), 457-464.
- Badre, D., & Wagner, A. D. (2004). Selection, integration, and conflict monitoring: assessing and the nature generality of prefrontal cognitive control mechanisms. *Neuron*, 41(3), 473-487.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the national academy of sciences*, 103(2), 449-454.
- Bemis, D. K., & Pylkkänen, L. (2012). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, 23(8), 1859-1873.
- Bertelson, P., Vroomenti, J., & Gelderti, B.D. (1997). Auditory-visual interaction in voice localization and in bimodal speech recognition: The effects of desynchronization. *Audio-Visual Speech Processing: Computational & Cognitive Science Approaches*.
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7 (3), 295-301.
- Binder, J. R., Medler, D. A., Desai, R., Conant, L. L., & Liebenthal, E. (2005). Some neurophysiological constraints on models of word naming. *Neuroimage*, 27(3), 677-693.

- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex* 19(12), 2767-2796.
- Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: an fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, 17(9), 1353-1366.
- Borsky, S., B. Tuller, & Shapiro, L. P. (1998). "How to milk a coat:" the effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America*, 103(5), 2670-2676.
- Bonte, M., Correia, J. M., Keetels, M., Vroomen, J., & Formisano, E. (2017). Reading induced shifts of perceptual speech representations in auditory cortex. *Scientific Reports*, 7.
- Burton, M., & Blumstein, S. (1995). Lexical effects on phonetic categorization: The role of stimulus naturalness and stimulus quality. *Journal of Experimental Psychology: Human Perception and Performance*, 21(5), 1230–1235.
- Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, 18(2), 90-98.
- Chen, Y., Davis, M. H., Pulvermüller, F., & Hauk, O. (2015). Early visual word processing is flexible: evidence from spatiotemporal brain dynamics. *Journal of Cognitive Neuroscience*, 27(9), 1738-51
- Clos, M., Langner, R., Meyer, M., Oechslin, M. S., Zilles, K., & Eickhoff, S. B. (2014). Effects of prior information on decoding degraded speech: an fMRI study. *Human Brain Mapping*, 35(1), 61-74.

- Cohen, L., Jobert, A., Le Bihan, D., & Dehaene, S. (2004). Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *Neuroimage*, 23(4), 1256-1270.
- Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3), 162-173.
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 23(8), 3423-3431.
- Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top-down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, 23(12), 3914-3932.
- Fiez, J. A., & Petersen, S. E. (1998). Neuroimaging studies of word reading. *Proceedings of the National Academy of Sciences*, 95(3), 914-921.
- Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Reviews Neuroscience*, 11(2), 127-138.
- Gagnepain, P., R. N. Henson, & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology*, 22(7), 615-621.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12(5-6), 613-656.
- Getz, L., & Toscano, J. C. (2019). Semantic context influences early speech perception: Evidence from electrophysiology. *The Journal of the Acoustical Society of America*, 145(3), 1789-1789.

- Gow Jr, D. W., & Olson, B. B. (2016). Sentential influences on acoustic-phonetic processing: A Granger causality analysis of multimodal imaging data. *Language, Cognition and Neuroscience*, 31(7), 841-855.
- Guediche, S., Salvata, C., & Blumstein, S. E. (2013). Temporal cortex reflects effects of sentence context on phonetic processing. *Journal of Cognitive Neuroscience*, 25(5), 706-718.
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review*, 111(3), 662.
- Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402.
- Horowitz-Kraus, T., Grainger, M., DiFrancesco, M., Vannest, J., Holland, S. K., & CMIND Authorship Consortium. (2015). Right is not always wrong: DTI and fMRI evidence for the reliance of reading comprehension on language-comprehension networks in the right hemisphere. *Brain Imaging and Behavior*, 9(1), 19-31.
- Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: a metaanalysis of 35 neuroimaging studies. *Neuroimage*, 20, 693-712.
- Keetels, M., Schakel, L., Bonte, M., & Vroomen, J. (2016). Phonetic recalibration of speech by text. *Attention, Perception, & Psychophysics*, 78(3), 938-945.
- Lau, E., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920-933.

- Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, 38, 487-497.
- Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology*, 23(4), 558-614.
- McGurk, H., & MacDonald, J. (1976). Visual influences on speech perception processes. *Perception & Psychophysics*, 24(3), 253-257.
- McQueen, J., D. Norris, et al. (2006). Are there really interactive processes in speech perception. *Trends in Cognitive Sciences*, 10(12), 533-533.
- Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, 47(8), 908-930.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76(2), 165.
- Myers, E. and S. Blumstein (2008). The neural bases of the lexical effect: an fMRI investigation. *Cerebral Cortex*, 18(2), 278-288.
- Nelken, I., & Ahissar, M. (2008). Low-level information and high-level perception: the case of speech in noise. *PLoS biology*, 6(5), e126.
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences*, 23(3), 299-325.
- Obleser, J. & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, 20, 633-640.

- Peelle, J. E., Johnsrude, I. S. & Davis, M. H. (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in Human Neuroscience*, 4, 1–3.
- Preston, J. L., Molfese, P. J., Frost, S. J., Mencl, W. E., Fulbright, R. K., Hoeft, F., ... & Pugh, K. R. (2016). Print-speech convergence predicts future reading outcomes in early readers. *Psychological science*, 27(1), 75-84.
- Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, 15(6), 246-253.
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, 62(2), 816-847.
- Richardson, F. M., Seghier, M. L., Leff, A. P., Thomas, M. S., & Price, C. J. (2011). Multiple routes from occipital to temporal cortices during reading. *Journal of Neuroscience*, 31(22), 8349-8247.
- Robson, H., Zahn, R., Keidel, J. L., Binney, R. J., Sage, K., & Lambon Ralph, M. A. (2014). The anterior temporal lobes support residual comprehension in Wernicke's aphasia. *Brain*, 137(3), 931-943.
- Rogers, J. C., & Davis, M. H. (2017). Inferior frontal cortex contributions to the recognition of spoken words and their constituent speech sounds. *Journal of cognitive neuroscience*, 29(5), 919-936.
- Rueckl, J. G., Paz-Alonso, P. M., Molfese, P. J., Kuo, W. J., Bick, A., Frost, S. J., ... & Lee, J. R. (2015). Universal brain signature of proficient reading: Evidence from four contrasting languages. *Proceedings of the National Academy of Sciences*, 112(50), 15510-15515.
- Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, 1328-1330.

- Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*(25), 8443-8453.
- Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences*, *113*(12), E1747-E1756.
- Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. (2006). Converging language streams in the human temporal lobe. *Journal of Neuroscience*, *26*(28), 7328-7336.
- Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, *21*(10), 1532-1540.
- Wild, C. J., M. H. Davis, & Johnsrude, I. S. (2012). Human auditory cortex is sensitive to the perceived clarity of speech. *Neuroimage*, *60*(2), 1490-1502.
- Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*(2), 271-282.
- Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*, *14*(4), 550-560.

Figure Caption

Figure 1. Regions showing interaction between Sentence Context (goat-biased, neutral) and Target Type (ambiguous, unambiguous 'goat'). The left hemisphere is on the left side. Sagittal slices at $X=-47$, and $X=47$. Bar graph shows percent signal change extracted from each region and standard error of the mean over subjects.

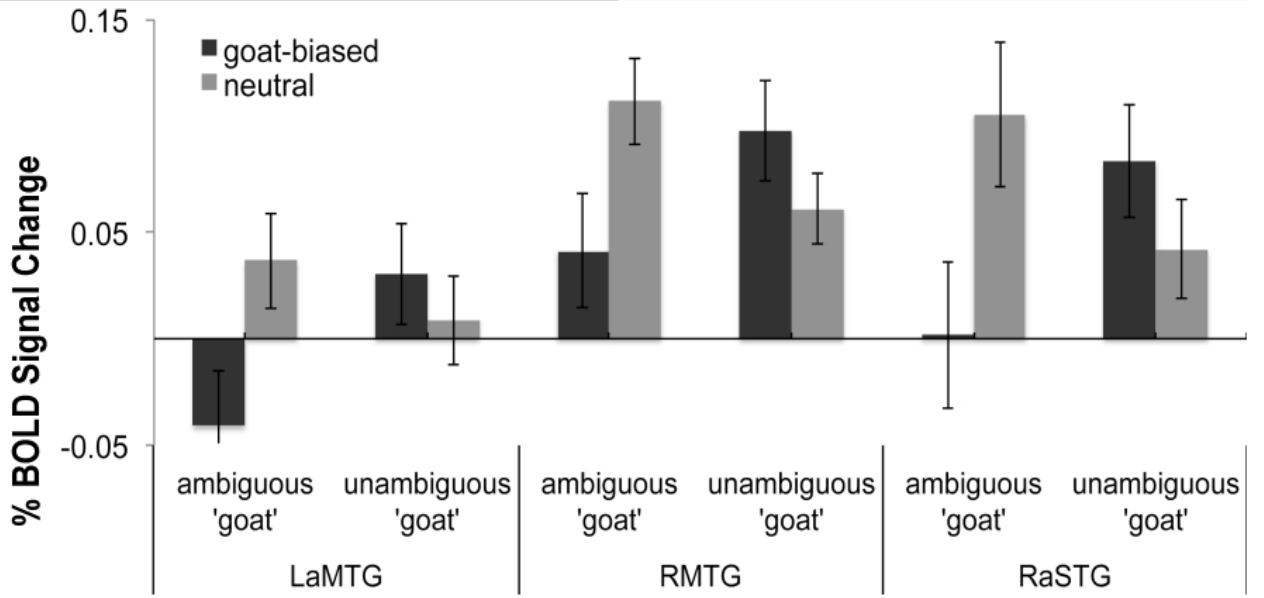
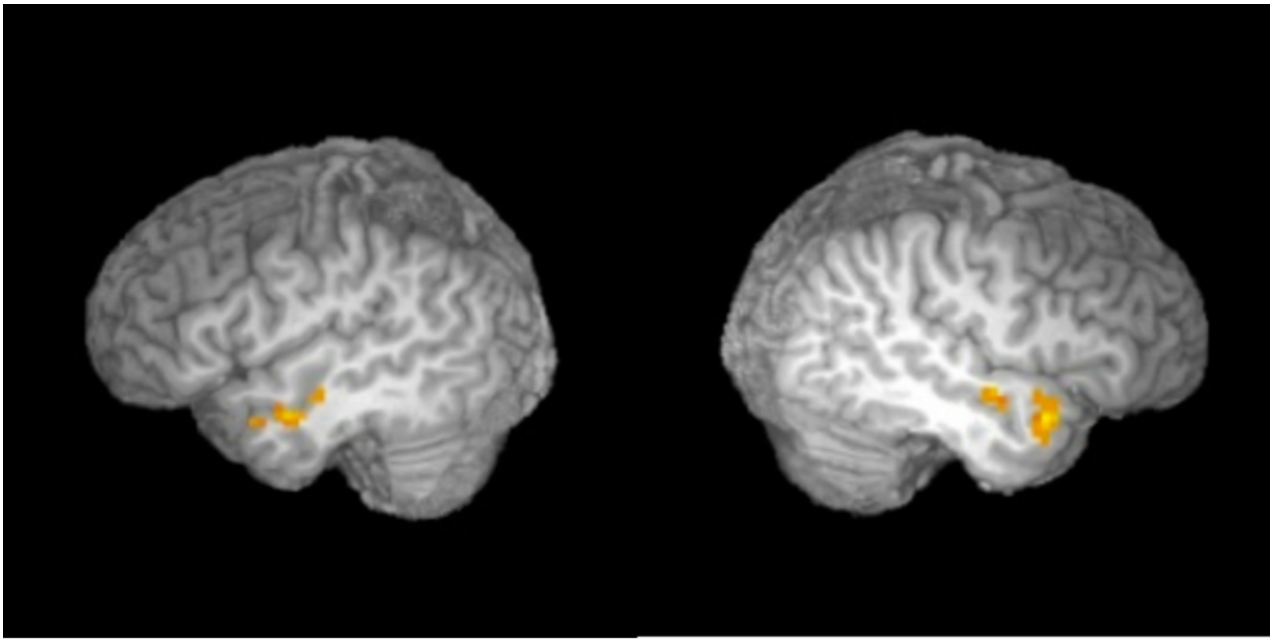


Table 1. Behavioral Results

	Unambiguous ‘goat’	Ambiguous	Unambiguous ‘coat’
	M (SEM)	M (SEM)	M (SEM)
<u>Percent Responses</u>			
Goat-biasing	99.1 (0.4)	68.1 (6.2)	1.9 (0.7)
Neutral	100 (0)	62.2 (6.6)	1.7 (0.7)
Coat-biasing	99.5 (0.5)	58.6 (5.8)	0.7 (0.5)
<u>Reaction-Time</u>			
Goat-biasing	648 (31)	920 (57)	685 (43)
Neutral	663 (39)	931 (57)	670 (32)
Coat-biasing	677 (39)	906 (52)	665 (33)

Note: Average percent ‘goat’ responses and reaction times in ms for each condition.

Standard error of the mean over subjects reported in parentheses.

Table 2. fMRI Results

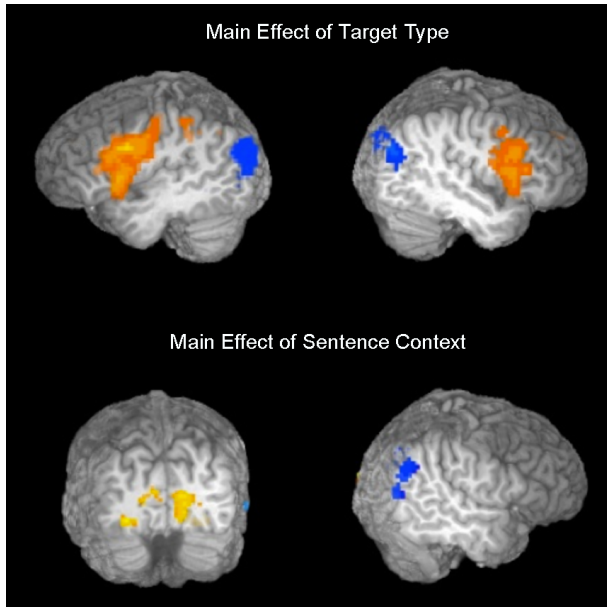
	Talairach (x, y, z)	Size	t-value
SentenceContext(goat-biased,neutral) x TargetType(ambiguous, unambiguous'goat')			
Right middle temporal gyrus/Insula	41, -13, -7	38	20.87
Right anterior superior temporal gyrus	44, 14, -16	36	22.59
*Left anterior superior/middle temporal gyrus	-43, -1, -16	25	20.87
Sentence Context			
<i>Neutral > Goat-biased</i>			
Right lingual, BA 18, BA19	14, -67, -1	517	5.17
<i>Goat-biased > Neutral</i>			
BA39 /right angular gyrus ¹	-52, -52, 41	219	3.49
Target Type			
<i>Ambiguous > Unambiguous</i>			
L insula and inferior frontal gyrus (IFG) (BA 13/45)	-28, 23, 8	1264	12.02
Left cingulate	-7, 26, 32	1177	10.65
Right Insula, IFG BA 47, BA 13	32, 11, 2	910	9.79
Left middle frontal and superior frontal gyrus 9	-22, 38, 23	79	5.17
Left superior temporal gyrus	-58, -25, 2	74	5.6
Left postcentral gyrus and inferior parietal lobule	-43, -31, 50	74	4.21
Left Precuneus/ angular gyrus	-28, -61, 38	56	4.49
<i>Unambiguous > Ambiguous</i>			
Right posterior middle temporal gyrus/BA39	41, -70, 33	377	5.65
Left middle occipital cortex	-31, -67, 5	244	4.83
Left posterior cingulate	-19, -52, 20	111	4.55
Left parahippocampal/fusiform gyrus	-34, -40, -7	83	5.50
Left cingulate Gyrus/BA31	-1, -46, 41	83	5.66

¹ This region only emerged at a $p = .05$, corrected for cluster size at an alpha of .05.

SUPPLEMENTARY MATERIALS

Additional fMRI Results

Main Effects:



Top Panel: Regions showing differences in activity between Ambiguous and Unambiguous targets. The left hemisphere is on the left side. Sagittal slices at X= -43, 43, corrected at a voxelwise threshold of $p < .01$ (top panel). **Bottom Panel:** Regions showing differences in activity between Goat-biased and Neutral sentence contexts. Yellow-scale reflects greater activity for Neutral compared to Goat and blue scale reflects greater activity for Goat compared to Neutral. Coronal Slice at Y = 73, corrected at a voxelwise threshold of $p < .01$. Sagittal slices at X= 50, corrected at a voxelwise threshold of $p < .05$ (see Table 2 of the manuscript).

Additional fMRI Analysis

A 2x2 ANOVA was conducted with Sentence Type (goat-biased, neutral) and Target Type (Boundary, Unambiguous 'coat') as within-subject factors. At a p -value of .01, only the anterior cingulate was significant, The table below reports all regions at a p -value of .05 that are greater than 100 voxels.

	x	y	z	size	t-value
LIFG	-28	8	26	373	17.52
Right Cerebellum (Crus I)	41	-58	-25	301	19.47
Left Cingulate	-7	-4	26	228	28.31
Right Thalamus	5	-31	-4	158	22.03