Some People are "More Lexical" than Others

Mako Ishida [a,d] , Arthur G. Samuel [b,c,d] , and Takayuki Arai [a]

[a] *Sophia University, Tokyo, Japan.*

[b] *Basque Center on Cognition, Brain and Language, Donostia, Spain.*

[c] *IKERBASQUE, Basque Foundation for Science.*

[d] *Stony Brook University, Dept. of Psychology, Stony Brook, NY, the United States of America.*

Corresponding author:

Arthur Samuel
Department of Psychology
Stony Brook University
Stony Brook, NY 11794-2500
USA
Tel: +01 631-632-7792
Email: Arthur.Samuel@stonybrook.edu

Word Count: 4809 (main text + abstract)

Abstract

People can understand speech under poor conditions, even when successive pieces of the waveform are flipped in time. Using a new method to measure perception of such stimuli, we show that words with sounds based on rapid spectral changes (stop consonants) are much more impaired by reversing speech segments than words with fewer such sounds, and that words are much more resistant to disruption than pseudowords. We then demonstrate that this lexical advantage is more characteristic of some people than others. Participants listened to speech that was degraded in two very different ways, and we measured each person's reliance on lexical support for each task. Listeners who relied on the lexicon for help in perceiving one kind of degraded speech also relied on the lexicon when dealing with a quite different kind of degraded speech. Thus, people differ in their relative reliance on the speech signal versus their pre-existing knowledge.

Keywords: Degraded speech; perceptual units; lexical support for speech perception

People can understand speech under an impressively wide range of listening conditions. In fact, one difficulty that speech researchers face is that the process is so good that it is difficult to examine the system's operation because it works so quickly and accurately. For that reason, researchers have used various techniques to stress the system in order to be able to probe what it is doing. These techniques include presenting the speech in noise (e.g. Miller & Isard, 1963), filtering away different parts of the spectrum (e.g., Wilson, Zizz, Shanks, & Causey, 1990), vocoding the speech (e.g., Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005), compressing the speech in time (e.g., Dupoux & Green, 1997), dropping out pieces of the signal (e.g., Huggins, 1964), and other manipulations that impair perception enough to see when errors occur.

In the current study, we use a method that Saberi and Perrott (1999) introduced (see also Steffen & Werani, 1994) that we will call Locally Time-Reversed Speech, or LTRS. With LTRS, an utterance is first segmented into pieces of a fixed size (e.g., every 50 msec, or every 100 msec), each such segment is then reversed along the time axis (i.e., played backwards), and the segments are put back together. This is quite different than simply playing the whole utterance backwards, as it breaks up the speech every N msec. Now, if each segment were tiny, e.g., only 1 msec long, this manipulation would not harm the signal very much because there is not much change in the waveform on such a short scale. The surprising result reported by Saberi and Perrot was that the segments could be quite long before listeners thought that intelligibility was impaired – with segments as long as 130 msec, listeners only rated the speech as having lost half its intelligibility.

Saberi and Perrot's (1999) report has spawned about a half dozen other studies of

LTRS, and has been cited in arguments about whether speech is decoded into syllable-sized units (suggested by the long segments that are tolerated) versus phoneme-sized units. The basic effect has now been shown in French (Magrin-Chagnolleau, Barkat, and Meunier, 2002) and in German (Kiss et al., 2008), in addition to English (Greenberg & Arai, 2001; Remez et al., 2013). In all cases, researchers have shown that with very small segments (e.g., 10 msec) performance is quite good, and as the segments get longer, performance declines. There has been some variation across studies in how quickly the curve falls as a function of segment size, but this variation presumably mostly traces to differences in how the measurements were done. For example, Remez et al. pointed out that some studies, including the original Saberi and Perrot paper, presented a very small number of stimuli repeatedly (e.g., only a single sentence in the original paper), and asked subjects to rate intelligibility, whereas in other studies listeners were required to report the words that were presented. Remez et al. included both measures, and as expected, found that subjective intelligibility yielded much longer estimates of the tolerated segment size than actual intelligibility measures.

Having listeners transcribe what they hear is clearly a better measure of perception than getting subjective intelligibility ratings, but even the transcription approach has some significant limitations. Because transcription of sentences is often slow and difficult, in most studies researchers have had to present each stimulus repeatedly to allow listeners to write or type what they hear. This repetition interacts with the use of sentences in which the listeners clearly are using the context to guess quite a bit, making the reports less informative about what is actually being perceived. In the current study, we apply a technique based on signal detection procedures to measure what listeners perceive when they hear LTRS stimuli. This approach overcomes the existing problems, and allows us to address two important issues.

In Experiment 1 we use this method to examine whether the "critical" time window depends on the phonetic properties of the speech. As Remez et al. (2013) pointed out, some speech sounds (e.g., vowels and fricatives) have relatively static spectral properties, and on these grounds one might expect that perception of such sounds would be more robust across longer time windows than speech sounds like stop consonants that change rapidly over time. The first experiment thus assesses how perception of degraded speech depends on the details of the phonetic signal.

In Experiment 2, we use our technique to determine how much perception of such degraded speech relies on support from lexical representations. Prior research with speech degraded in other ways has demonstrated that listeners do use lexical context to guide perception of phonetic input (e.g., Grataloup et al., 2009; Samuel 1981, 1996). In addition to looking for lexical influences on perception of LTRS stimuli, in Experiment 2 we ask whether the degree to which a given listener relies on the lexicon is a general property of that listener: If an individual brings to bear lexical information when trying to understand speech degraded via LTRS, does that same individual also tend to rely on the lexicon when confronting speech with a very different kind of challenge? For this test, we had the same group of listeners listen to LTRS stimuli, and to stimuli in which a single phoneme could be replaced by white noise. Warren (1970) discovered that listeners are generally unable to detect that a phoneme is missing when it is replaced by a loud noise, an effect he termed phonemic restoration. Samuel (1981, 1996) demonstrated that a significant source of the restoration is lexical knowledge – listeners restore missing phonemes more in words that they know than in matched pseudowords. In Experiment 2, we test whether the lexical influence on phonemic restoration correlates with any lexical influence on how listeners perceive LTRS stimuli.

In our new LTRS paradigm, on each trial listeners hear two items. Each item is either a word (e.g., "academic") or a pseudoword that differs by one phoneme (e.g., "acabemic"). The first item is a locally time-reversed stimulus in a male voice and the second item is normally produced, in a female voice. The task is to judge if the two items were the same (e.g. "academic" followed by "academic", or "acabemic" followed by "acabemic"), or different (e.g., "academic" – "acabemic", or "acabemic" – "academic"). If perception of the LTRS-modified first item is good, judging its phonetic similarity to the normal second item should be accurate, but if LTRS modification disrupts perception, making the same-different judgment will be difficult. As will be seen, this method provides a very sensitive signal detection based measure of perception of the degraded speech. Because this method uses a discrimination task in which listeners are never asked to report what word (or pseudoword) they hear, the results cannot be traced to any post-perceptual decision stage.

Experiment 1

Experiment 1 had three goals. First, we sought to establish that our signal detection test was an excellent way to assess perception of LTRS stimuli. Second, we wanted to determine whether lexical support played a significant role in perceiving such stimuli. Finally, we wished to see whether the critical time-window differed for words that differed in their relative proportions of more static sounds (fricatives) versus more dynamic ones (stops).

Methods

Participants

A total of 60 undergraduate students from Stony Brook University (12 males, 48 females) participated in Experiment 1. All were native American English speakers, age 18 or older, and had no known hearing problems. They received research credit for their participation.

Stimuli

Two sets of words were selected from the MRC Psycholinguistic database: 72 fricative-dominant stimuli and 72 stop-dominant stimuli. All words were 3, 4, or 5 syllables long, had a Kucera-Francis written frequency greater than 1/million, and a familiarity rating of 500-700 in the MRC database. Fricative-dominant words were chosen to have relatively many fricatives ($M = 1.72$) as compared to stops ($M = 0.72$), while stop-dominant words contained more stops ($M = 3.18$) than fricatives ($M = 0.79$). Working within the constraints of the overall occurrence rates of stops and fricatives in English words of this sort, these two sets provide a strong test of whether perceptual degradation via LTRS depends on the degree to which the speech has rapid spectral changes versus relatively steady-state segments: There are over twice as many steady-state fricatives in the fricative-dominant set as in the stop-dominant set, and over four times as many stops in the stop-dominant set than in the fricative-dominant set. The mean frequency of words in the two sets was matched (fricative-dominant words: 16.45; stop-dominant words: 18.32), based on the CLEARPOND database (Marian, Bartlotti, Chabal, and Shook, 2012).

After selecting the fricative- and stop-dominant words, matched pseudowords were created by changing the place of articulation of one phoneme in each word (e.g., “academic” – “acabemic”). This degree of change was designed to be small enough that if perception were to be impaired, errors would be likely. With this procedure, there were 72 fricative-dominant words, 72 fricative-dominant pseudowords, 72 stop-dominant words, and 72 stop-dominant pseudowords. All items were spoken by one male and one female native American English speaker, and recorded in a sound shielded room. The sounds were first digitized at 48 kHz (16 bit), and downsampled to 16 kHz (16 bit). The amplitude of stimuli was also normalized with GoldWave digital audio software, and each token was stored as a WAV file (16Hz, 16bits, mono). Words and pseudowords from the male voice were locally time-reversed by MATLAB with reversal window lengths of 10, 30, 50, 70, 90, and 110 ms. The MATLAB script imposed 5-msec linear onset and offset amplitude shaping at the points where successive segments were joined to prevent clicks from being generated at those points. The sets of words and pseudowords are shown in the Appendix.

Procedure

Participants were randomly assigned to either the fricative-dominant or stop-dominant stimuli group. Each trial included a locally time-reversed stimulus in the male voice followed by a non-reversed stimulus in the female voice. The inter-stimulus-interval was 400 ms. By presenting the two items of a trial in two very different voices, we forced the judgments to be based on phonetic encoding (what we wished to measure), rather than any acoustic differences that a listener might detect.

We will refer to the first item – the one that was subjected to the local time reversal manipulation – as the target. We will refer to the second item (presented without any time reversal) in a trial as the standard. The standard was either the same as the target (the same word or the same pseudoword presented twice), or different (a word followed by its matched pseudoword, or vice versa). Listeners were instructed to judge if the first speaker and the second speaker produced the same items, and to respond "same" or "different" using two labeled buttons on a response pad. "Same" and "Different" were specified as either containing the same set of vowels and consonants or having one segment changed (i.e., the phonetic coding that is the focus here).

Each subject did 288 trials, representing the four possible trial structures (word-word; pseudoword-pseudoword; word-pseudoword; and pseudoword-word) for each of the 72 items. For each set (fricative dominant or stop dominant) the 72 items were divided into 6 subsets of 12 items, with each subset assigned to one of the 6 segment durations. Six groups of participants, in a Latin Square, were used to counterbalance the 6 subsets of items across the 6 segment durations. Stimuli were presented through SONY MDR-V900HD headphones in a sound shielded room, and the experiment lasted about 30 minutes. The 288 trials for each subject were individually randomized.

Results

The design of the same-different task allowed us to use signal detection's *d'* parameter to index how well listeners perceived the target. [1] If the local time-reversal prevented accurate phonetic encoding then listeners should have had difficulty judging whether the target matched the intact standard. The two possible error types provide the miss rates and false alarm rates needed for computing *d'*: One error type is responding "different" when the standard in fact matched the segmental structure of the target, and the other occurs if a listener responds "same" when in fact there was a segmental change from the target to the standard. We use the resulting *d'* scores to investigate the two central questions in Experiment 1: (1) Does the acoustic-phonetic structure of a word affect its susceptibility to the local time reversal manipulation?, and (2) Does the lexical status of the target affect its susceptibility to local time reversal?

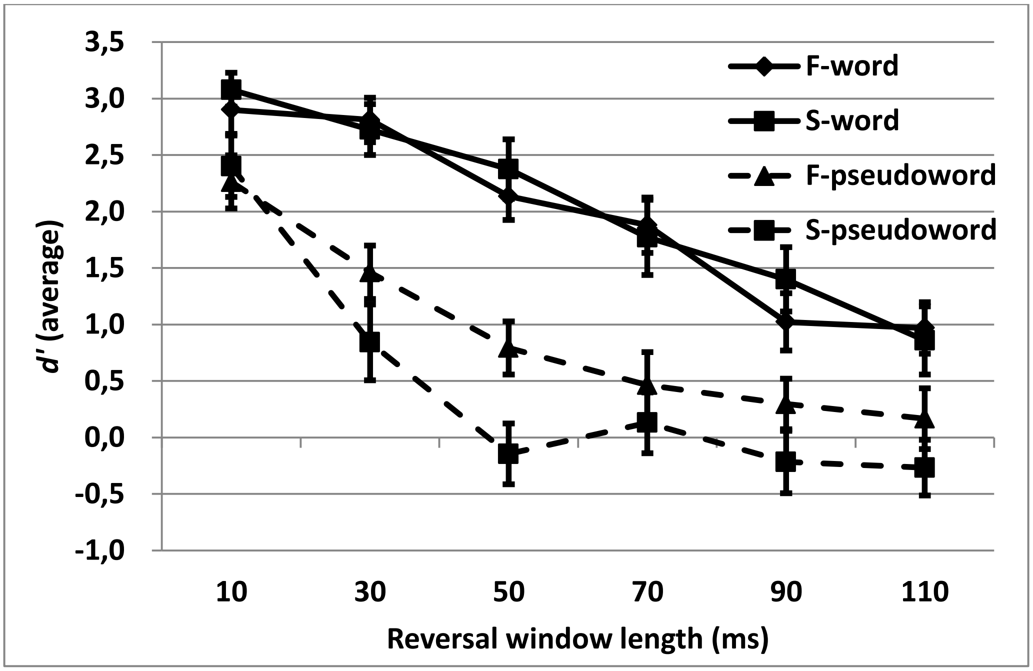

Figure 1 : Intelligibility of Locally Time-reversed Words and Pseudowords (F: Fricative dominant, S: Stop dominant). Error bars here and in other figures represent 95% confidence intervals.

For each subject, *d'* scores were computed for each cell of the experimental design. In cases in which a miss or false alarm rate was at ceiling or floor, it was replaced with the value of 1/2N (floor) or 1 – 1/2N (ceiling), with N equal to the number of items. An ANOVA was then performed on these *d'* scores with two Phoneme types (fricative-dominant vs. stop-dominant) as a between-subject factor, and with two Lexical status types (words vs. pseudowords) and six Reversal Window lengths (10, 30, 50, 70, 90, 110 ms) as within-subject factors. All data for this experiment and for the following experiment are available on Cognition's data archive.

Figure 1 presents the mean *d'* scores for fricative-dominant versus stop-dominant targets, with words versus pseudowords as the LTRS target stimuli, as a function of the local time reversal window size. The reduction in performance as the size of each reversed segment increased (F (5, 290) = 256.302, $p < .001$) is consistent with the general form of prior work on LTRS. Importantly, when the speech was essentially undistorted – at the 10 msec segment size – performance was very accurate, with essentially identical accuracy on the fricative-dominant words and on the stop-dominant words, confirming that the stimulus selection yielded well-matched sets. This matching at the 10-msec segment size generated a three-way interaction of Lexical status, Reversal Window lengths, and Phoneme types, F (5, 290) = 3.641, $p = .003$.

In unpacking this three-way interaction, we first note that when the LTRS target was a pseudoword (shown with dashed lines), once the window size exceeded 10 msec there was a robust disadvantage for the targets with many stops, compared to those with many fricatives; this difference was absent when the target was a real word. This produced a significant interaction between Lexical status and Phoneme type, *F* (1, 58) = 9.436, p = .003. Overall, fricative-dominant words were significantly more tolerant of local time distortion than stop-dominant words, *F* (1, 58) = 5.051, p = .028. The rapid spectral changes of stops appear to be quite vulnerable to local time distortion, whereas the relatively stable spectral patterns of fricatives are less disrupted.

Our second major question was whether performance would differ as a function of whether the LTRS target was a word versus a pseudoword. If lexical encoding produces a more stable percept for words than is available for pseudowords then word target trials should produce better levels of performance than pseudoword target trials. This word advantage was very clear: word targets (solid lines) were much more tolerant of local time distortion than pseudoword targets, *F* (1, 58) = 229.249, p < .001. If we take a *d'* of approximately 0.5 as the minimum needed to reflect performance above chance, the data indicate that the intelligibility of target words was above chance even at the 110-ms reversal window length, whereas the intelligibility of target pseudowords dropped to chance at the 50-ms reversal window length. Thus, there are very strong lexical effects in perception of locally time-reversed speech.

In Experiment 2, we focus on this lexical effect. In particular, we test whether the lexical influence found in Experiment 1 reflects a relatively stable perceptual style for a given individual. Put another way, are there people who rely on such top-down lexical guidance in speech perception more than other people?

Experiment 2

Experiment 1 established that when confronted with speech degraded by local time reversals, listeners relied heavily on the lexicon to encode the target for comparison to the standard. Previous work on phonemic restoration (Samuel, 1981, 1996) has shown that listeners also use the lexicon to help restore speech segments that have been replaced by noise. In Experiment 2, listeners were given both the new LTRS discrimination task, and a phonemic restoration task. We computed a measure of lexical influence for each listener for each task, and used these two measures to ask whether individuals vary in their reliance on the lexicon when dealing with degraded speech, regardless of the type of degradation: Are some people more reliant on top-down lexical support than others?

Methods

Participants

A total of 52 undergraduate students from Stony Brook University (23 males, 29 females) participated in Experiment 2. All were native American English speakers at least 18 years old, and had no known hearing problems. None had participated in Experiment 1. They received research credit for their participation.

Stimuli

There were two sets of stimuli: locally time-reversed stimuli (words and pseudowords), and phonemic restoration stimuli (words and pseudowords).

The LTRS stimuli were 60 stop-dominant words and 60 matching pseudowords, a subset of the items from Experiment 1. We focused on the stop-dominant items because they had produced stronger LTRS effects than the fricative-dominant items. The words contained many stops (*M* = 3.22) as compared to fricatives (*M* = 0.53). The mean frequency of words was 21.24 in the CLEARPOND database (Marian, Bartlotti, Chabal, and Shook, 2012). All words were all 3, 4, or 5 syllables long. New target stimuli in the male voice were made by locally time-reversing the words and pseudowords using 20, 40, 60, and 80 ms time windows; the (normal) items in the female voice were as in Experiment 1.

A phonemic restoration task used the procedure developed by Samuel (1996), with stimuli constructed by Mattys, Barden and Samuel (2014). The task is based on presenting listeners with two different types of stimuli. One stimulus type ("replaced") is made by identifying the location of a phoneme and its transitions into adjacent phonemes, and replacing that portion of the waveform with signal correlated white noise. The second stimulus type ("added") uses the same boundaries for the critical speech segment, but the speech is left intact and the signal correlated white noise is simply added to the segment. The motivation for the two stimulus types is that if listeners perceptually restore a missing speech sound (in a "replaced" item), the resulting percept should be of intact speech that has noise superimposed (an "added" item). Samuel (1981) first created these two stimulus types, and using a forced-choice task (was the item "added" or "replaced"?) with signal detection analyses, he showed that lexical support does make "replaced" items sound more like "added" items. Samuel (1996) developed an improved version of this method that provides richer data and also avoids some of the statistical assumptions underlying signal detection computations. On each trial the listener first hears either an "added" item or a "replaced" item, and then hears the speech token in its original form, with no noise. The task is to rate the similarity of the first item to the unmanipulated version. Any difference in rating the “replaced” versions and the “added” versions is evidence for incomplete perceptual restoration. As shown in previous studies (Mattys, Barden, & Samuel, 2014; Samuel, 1996), the two versions have more similar ratings when the stimuli are real words than when they are matched pseudowords because the stronger lexical influence in a real word increases the ratings of the “replaced” items, bringing them closer to the ratings for the “added” stimuli. This corresponds to the lower discriminability of the “added” and “replaced” stimuli for words than pseudowords observed by Samuel (1981) – the use of lexical information actually reduces performance on the task by making the two versions sound more similar. The directionality of this effect – more lexical use causes poorer discrimination – will be relevant when we look at the relationship of performance on the phonemic restoration task and performance with LTRS stimuli.

The critical phonemic restoration stimuli included 60 words and 60 matching pseudowords (see Mattys, Barden & Samuel, 2014, for more detail). They were all 4 or 5 syllables long, and white noise was either “added” to, or “replaced”, the first phoneme of the final syllable. The target phoneme was either a liquid (/l/, or /r/) or nasal (/m/, or /n/), sounds chosen on the basis of prior studies. Each word – pseudoword pair had the same stress pattern, and the same final two syllables. There were an additional 30 filler words and 30 filler pseudowords, in which the “added” or “replaced” phoneme was either in the first syllable or in the onset of the second syllable, to prevent listeners from only focusing on word endings. All words and pseudowords had been recorded by a female phonetician who spoke Standard Southern British English.

Procedure

Each participant did a set of phonemic restoration trials first, then the LTRS same-different task, and finally the remaining phonemic restoration trials. Half of the subjects did the phonemic restoration task with all of the word stimuli first, and with all of the pseudoword stimuli last. For the other half of the subjects, the pseudoword restoration stimuli preceded the LTRS task, and the word restoration stimuli followed it. For the restoration tasks, participations used 8 buttons on a response pad to rate the similarity of the "added" or "replaced" version to the intact version (1 = not similar, 8 = very similar). The reversed speech task was run using exactly the same procedures as in Experiment 1, with subjects instructed to respond "same" or "different" using labeled buttons, with the judgment being based on the phonemic match of the LTRS male item to the intact female item.

The same equipment was used as in Experiment 1. Each phonemic restoration block (word or pseudoword) consisted of 180 trials, and the LTRS degraded speech task included 240 trials. In each case, stimuli were presented in a randomized order. The entire experiment lasted about 50 minutes.

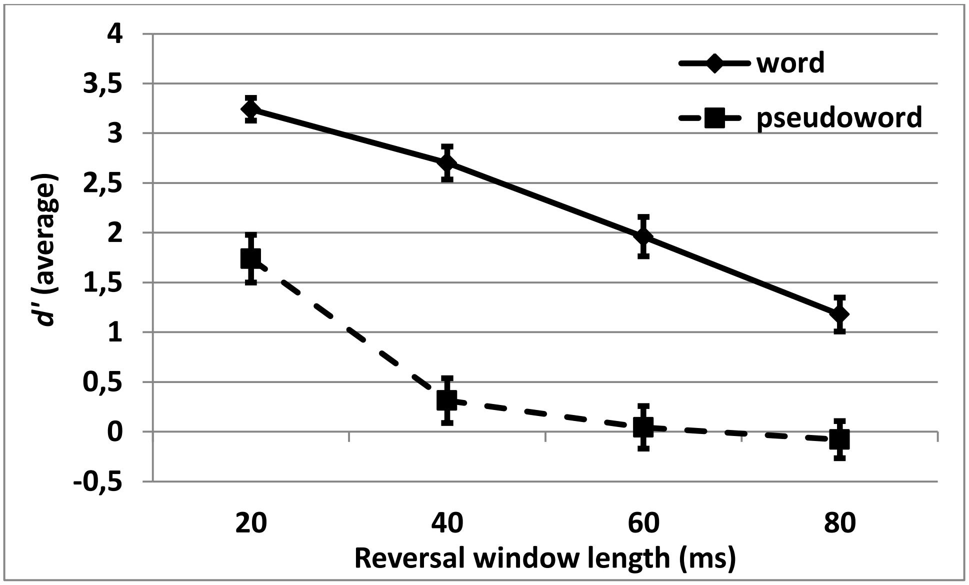

Figure 2 : Intelligibility of Locally Time-reversed Words and Pseudowords (Experiment 2). F: Fricative dominant, S: Stop dominant. Error bars represent 95% confidence intervals.

Results

For the locally time-reversed stimuli, *d'* scores were computed as in Experiment 1, and were submitted to an ANOVA with Lexical status of the target (word or pseudoword) and Reversal Window lengths (20, 40, 60, and 80 ms) as within-subject factors. The results, shown in Figure 2, are remarkably similar to what we observed in Experiment 1. The intelligibility of locally time-reversed stimuli declined as the reversal window lengths increased, *F* (3, 153) = 330.507, p < .001, with a steeper decline for pseudowords than for words, *F* (3, 153) = 22.248, p < .001. The intelligibility of words was well above chance at the longest 80-ms reversal window length, while performance on pseudowords was at chance with segment lengths as short as 40 ms. Critical to the main purpose of Experiment 2, we replicated the strong lexical effect on perceiving locally time-reversed stimuli, *F* (1, 51) = 228.407, p < .001. This is essential for our goal of testing whether lexical influences are stronger in some individuals than in others.

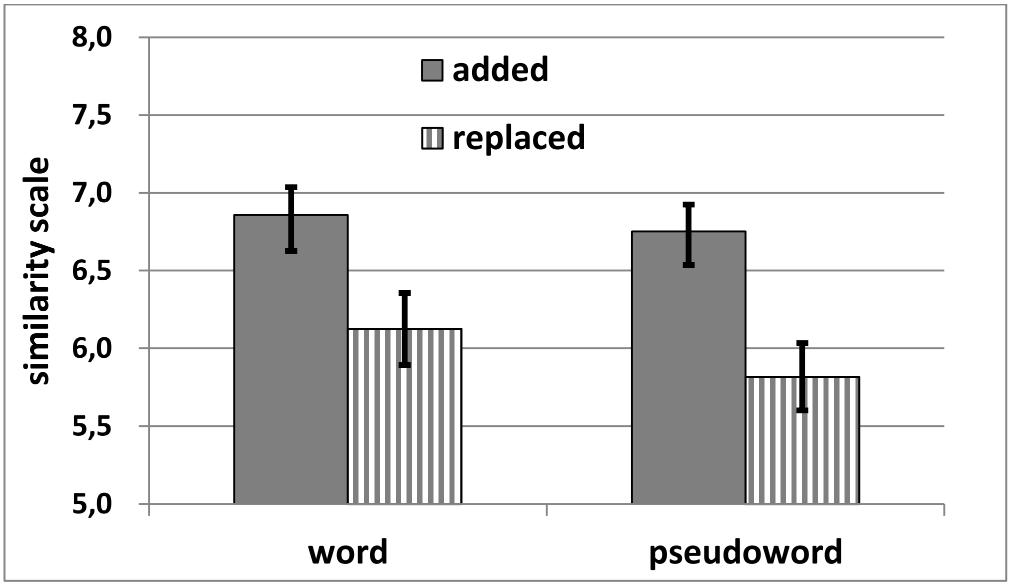

Figure 3: Similarity judgments for words and pseudowords with either an "added" or "replaced" phoneme

To address that question, we first need to determine if we have a lexical effect for our listeners when they perceptually restore missing phonemes. For each listener, we computed the average similarity rating for: words/added, words/replaced, pseudowords/added, and pseudowords/replaced. A lexical influence would show up as closer ratings for the "added" versus "replaced" versions of words compared to that difference for pseudowords. As noted above, this corresponds to lower discriminability for words than for pseudowords, due to stronger restoration in words. Figure 3 shows the average ratings for these four cases, and this is in fact the observed pattern. An ANOVA was performed on the ratings with Lexical status (words vs. pseudowords) and Noise type (added vs. replaced) as within-subject factors. As one would expect, stimuli with an added phoneme (i.e., truly intact) were rated significantly more similar to normal than the stimuli with a replaced phoneme, *F* (1, 51) = 137.123, p < .001. In addition, words with a distorted phoneme were rated as significantly more similar to normal than pseudowords with a distorted phoneme, *F* (1, 51) = 5.487, p = .023. The critical question is whether there was a significant interaction between Lexical status and Noise type, as this is the test of whether the lexicon causes non-intact stimuli to sound more intact. The interaction was in fact reliable, *F* (1, 51) = 4.683, p = .035.

With a reliable lexical effect now established for each of our very different listening situations, we are in a position to address our central question: Is there a correlation between the strength of the lexical influence on phonemic restoration and the size of the lexical effect on perception of locally time-reversed stimuli? To compute this correlation, we derived a measure of lexical influence on each task, for each subject. For phonemic restoration, a value for each participant was computed as follows: First, we took the difference between the mean score of words with a noise-added phoneme and of words with a noise-replaced phoneme (words/added – words/replaced). Next, we did the same thing for pseudowords (pseudowords/added – pseudowords/replaced). Finally, we took the difference between the difference score for words and the difference score for pseudowords. For locally time-reversed stimuli, we computed the *d'* difference between words and pseudowords. Figure 4 shows the relationship between the lexical effect for each subject for phonemic restoration (Y-axis) versus the lexical influence on perception of LTRS degraded speech (X-axis). As is clear in the figure, there is a strong correlation, *r* =

0.43, $p < .01$. This robust correlation demonstrates that the same people who relied heavily on the lexicon to help decode locally time-reversed speech were the ones who showed the biggest difference between words and pseudowords in restoring missing phonemes: Individuals differ in their reliance on lexical support for phonemic encoding.

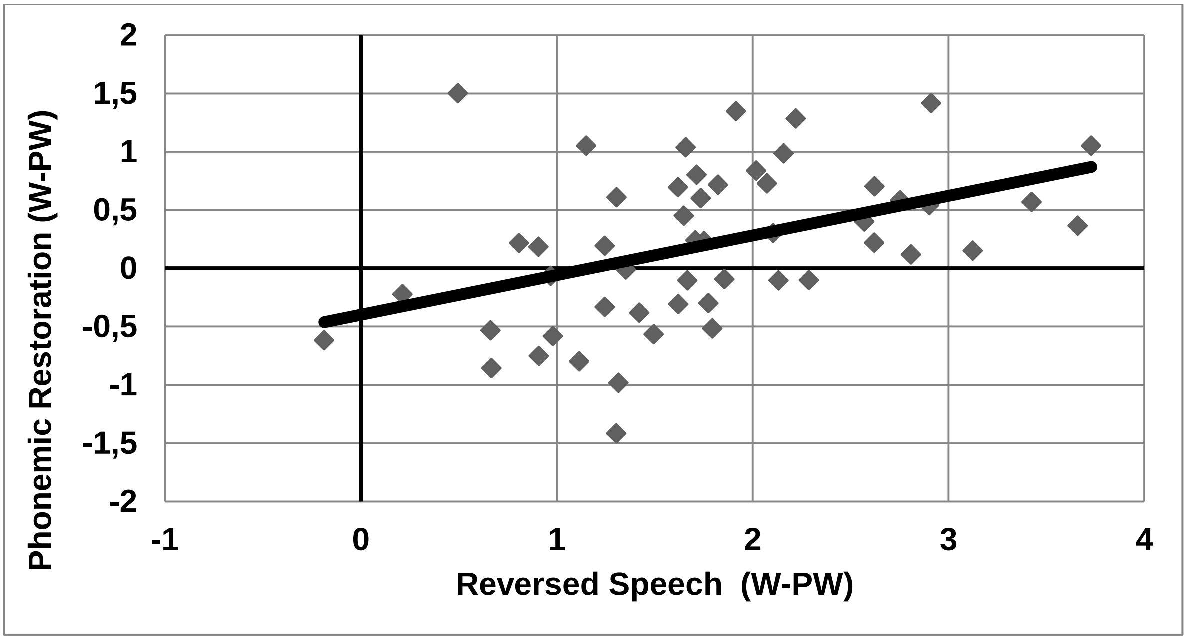

Figure 4: Correlation of the lexical influence on the phonemic restoration task with the lexical influence on the reversed speech task

General Discussion

Experiments 1 and 2 produced clear and consistent results. The intelligibility of locally time-reversed words was above chance with reversal windows over a hundred msec, whereas performance on pseudowords dropped to chance with reversal windows as short as 40 msec. In Experiment 1, the tolerance for time reversals also depended on the acoustic-phonetic properties of the speech, with perception failing quickly when words had many stop consonants.

These results suggest that it is at least premature, if not misguided, to use any critical segment length in LTRS stimuli to argue about perceptual units in speech perception. Remez et al. (2013) found relatively short critical windows when they removed a great deal of spectral information (via "sine wave" speech), and argued for phoneme-sized speech units. On the other side, Stilp et al. (2010) found that performance was linked to the speech rate in a way that appeared to be tied to the number of syllables presented per second (see Grataloup et al., 2009, for a related syllable-based argument). Poeppel and his colleagues (e.g., 2003; Hickock & Poeppel, 2007) have taken the apparently syllabic results as support for the view that one channel of speech analysis runs at an essentially syllabic rate. Stilp et al.'s approach shared many characteristics with an earlier research program that challenged the speech perception system by alternately presenting the speech to the left and right ears, at various rates. For example, Huggins (1975) found that intelligibility was most impaired at around 4 cycles per second, roughly the syllabic rate, and that if the speaking rate was increased, this minimum moved up linearly (keeping it aligned with the syllabic rate). These results were taken as support for the syllable in speech perception. However, Samuel (1991) demonstrated that the same pattern is observed for simple piano melodies, and that the performance degradation is unrelated to syllabic boundaries. Thus, as we suggested, it is not clear that pointing to particular segment lengths in LTRS stimuli will be a productive way to look for perceptual units in speech.

What is much more clear is that when listeners are confronted with speech that has been degraded they bring to bear sources of knowledge that can help. As shown in Experiment 2, one such source that is potentially quite useful is the existing knowledge that listeners have about words – lexical information. Previous research had shown that the lexicon is called upon to help restore phonemes that had been replaced by an extraneous sound (Samuel, 1981, 1996). In both Experiments 1 and 2, we observed a very large lexical effect on how well people could judge whether an LTRS-degraded target matched a clear standard: When the target was a word, much longer reversal windows were tolerated than when the target was a pseudoword. The most striking finding of the current study is that the relative reliance on lexical support in speech perception seems to vary across individuals in a stable way – the same people who showed strong lexical effects for LTRS stimuli did so for phonemic restoration stimuli. It is important to keep in mind that this correlation does not come from "better" subjects doing better on both tasks. On the contrary, as noted above, high lexical influences on the phonemic restoration task are associated with lower discrimination performance (Samuel, 1981, 1996). Thus, the results of Experiment 2 provide evidence for stable individual differences in the relative balance that different people strike between relying on the acoustic phonetic input, and relying on their pre-existing lexical knowledge.

References

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences" *Journal of Experimental Psychology: General*, 134(2), 222-241.

Dupoux , E., & Mehler , J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language, 29*(3), 316-335. doi: 10.1016/0749-596X(90)90003-I

Grataloup, C., Hoen, M., Veuillet, E., Collet, L., Pellegrino, F., and Meunier, F. (2009)., "Speech Processing: An Interactive Process", *Journal of Speech, Language, and Hearing Research*, 52, pp. 827-838

Greenberg, S. and Arai, T. (2001). “The relation between speech intelligibility and the complex modulation spectrum.” *Proc. 7th Eur. Conf. Speech Comm. Tech.*, pp. 473–476.

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402.

Huggins, A. W. F. (1964). Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America, 36,* 1055-1064.

Huggins, A.W.F. (1975). Temporally segmented speech. *Perception & Psychophysics*, 18, 149-157.

Kiss, M., Cristescu, T., Fink, M., and Wittmann, M. (2008). Auditory language comprehension of temporally reversed speech signals in native and non-native speakers. *Acta Neurobiol Exp (Wars).* 68, (2), 204-13.

Magrin-Chagnolleau, I., Barkat, M., and Meunier, F. (2002). Intelligibility of reverse speech in French: A perceptual study. *Proceedings of the 7th. International Conference on Spoken Language Processing, Interspeech 2002*, 3, p. 1669

Marian V, Bartolotti J, Chabal S, and Shook A. (2012.) Clearpond: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. PloS one 7 (8), e43230. 23.

Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 2, 217-228.

Mattys, S.L., Barden, K., & Samuel, A.G. (2014). Extrinsic cognitive load impairs low-level speech perception. *Psychonomic Bulletin and Review*, 21(3), 748-754.

Poeppel D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication* 41:245-255.

Remez, R. E., Thomas, E. F., Dubowski, K. R., Koinis, S. M., Porter, N. A. C., Paddu, N. U., Moskalenko, M., and Grossman, Y. S. (2013). Modulation sensitivity in the perceptual organization of speech. *Attention, Perception & Psychophysics*, 75, 1353-1358.

Saberi, K. and Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature*, 398, 760.

Samuel, A.G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of*

Experimental Psychology: General, 110, 474-494.

Samuel, A. G. (1991). Perceptual degradation due to signal alternation: Implications for auditory pattern processing. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 392-403.

Samuel, A.G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, 125, 28-51.

Steffen, A., &Werani, A. (1994). An experiment on temporal processing in language perception [In German]. In G. Kegel, T. Arnhold, K.Dahlmeier, G. Schmid,& B. Tischer (Eds.), Sprechwissenschaft undPsycholinguistik 6. Beiträge aus Forschung und Praxis [Speechscience and Psycholinguistics 6: Contributions from Research and Practice] (pp. 189–205). Opladen: Westdeutscher Verlag.

Stilp, C.E., Kiefte, M., Alexander, J.M., and Kluender, K.R. (2010). Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences. *Journal of the Acoustical Society of America*, 128(4), 2112-2126.

Warren, R. M. (1970). "Perceptual restoration of missing speech sounds," *Science*, 167, 392–393.

Wilson R.H., Zizz C.A., Shanks J.E., Causey G.D. (1990) . Normative data in quiet, broadband noise, and competing message for Northwestern University Auditory Test No. 6 by a female speaker. *Journal of Speech and Hearing Disorders*, 55, 771-778 .

Footnote

1. In addition to the perceptual index d’, signal detection theory also provides a parameter (Beta) that reflects the degree to which the listener is biased to report one answer or the other. We computed Beta scores for all of the conditions and observed that (1) there was not a very strong bias toward either response, and (2) there was little systematic change in Beta across the conditions of the experiments.

Appendix

The words and pseudowords used in Experiment 1. For the pseudowords, the sound whose place of articulation was changed to form the pseudoword is shown in bold font.

Stop-dominant Words	**Stop-dominant Pseudo-words**	**Fricative-dominant Words**	**Fricative-dominant Pseudo-words**
academic	aca**b**emic	additional	a**b**itional
acceptable	ac**sh**eptable	artificial	ar**p**ificial
activity	ac**p**ivity	associate	a**sh**ociate
anticipate	an**p**icipate	authority	authori**p**y
appointment	a**t**ointment	available	availa**d**le
architecture	archi**p**ecture	avenue	ave**m**ue
attitude	atti**p**ude	misery	mi**zh**ery
capable	ca**t**able	civilization	civili**zh**ation
capacity	ca**t**acity	confirmation	confir**n**ation
catastrophe	ca**p**astrophe	conventional	conven**s**onal
category	ca**p**egory	conversation	conver**sh**ation
certificate	cer**p**ificate	definition	defi**m**ition
comfortable	comfor**p**able	personally	per**sh**onally
commodity	co**n**odity	easily	ea**zh**ily

companionship	com**t**anionship	education	edu**t**ation
company	com**t**any	emotional	e**n**otional
competition	compe**p**ition	environment	environ**n**ent
consecutive	con**sh**ecutive	foundation	foun**b**ation
considerable	consi**b**erable	evolution	evolu**s**on
constitution	consti**p**ution	fascination	fasci**m**ation
contemporary	con**p**emporary	generosity	ge**m**erosity
correspondence	corres**t**ondence	hesitation	hesi**p**ation
decorated	de**t**orated	imitation	imi**p**ation
deliberate	deli**d**erate	information	infor**n**ation
department	de**t**artment	inheritance	inheri**p**ance
development	develop**n**ent	initially	i**m**itially
difficulty	diffi**t**ulty	insurance	in**ss**urance
disappointment	disa**t**ointment	intelligence	in**p**elligence
documentary	docu**n**entary	international	inter**m**ational
educated	edu**t**ated	intervention	in**p**ervention
electricity	electri**sh**ity	invitation	invi**p**ation
entertainment	enter**p**ainment	isolation	i**sh**olation
establishment	es**p**ablishment	machinery	machi**m**ery
executive	e**gzh**ecutive	magnificent	mag**m**ificent
expedition	expe**b**ition	mechanism	mecha**m**ism

experiment	ex**t**eriment	mercenary	merce**m**ary
extraordinary	extraordi**m**ary	miserable	mi**zh**erable
hospitality	hospi**p**ality	missionary	missio**m**ary
identical	i**b**entical	monastery	mo**m**astery
important	im**t**ortant	national	na**s**onal
incapable	in**t**apable	necessary	nece**sh**ary
incredible	incre**b**ible	negotiation	ne**d**otiation
independence	inde**t**endence	scientific	scien**p**ific
interpretation	interpre**p**ation	occasional	o**t**asional
interrupted	in**p**errupted	citizen	ci**p**izen
introduction	intro**b**uction	officer	offi**sh**er
remarkable	remar**t**able	official	offi**s**al
nobody	no**d**ody	operation	o**t**eration
opportunity	oppor**p**unity	opposition	o**t**osition
particular	par**p**icular	organization	orga**m**ization
photographic	pho**p**ographic	reception	re**sh**eption
political	poli**p**ical	philosophy	philo**sh**ophy
possibility	possi**d**ility	professional	profe**s**onal
predicament	pre**b**icament	psychology	psy**t**ology
prehistoric	prehis**p**oric	radiation	ra**b**iation
preposterous	pre**t**osterous	reasonable	reaso**m**able

profitable	profi**p**able	relationship	rela**s**onship
propaganda	propa**d**anda	relaxation	relaks**h**ation
property	pro**t**erty	revelation	revela**s**on
publication	publi**t**ation	revolution	revolu**s**on
publicity	publi**sh**ity	separation	se**t**aration
representative	represen**p**ative	sincerity	sin**sh**erity
republican	re**t**ublican	society	so**sh**iety
respectable	res**t**ectable	superficial	su**t**erficial
satisfactory	sa**p**isfactory	supervisor	su**t**ervisor
sophisticated	sophisti**t**ated	mysterious	my**sh**terious
spectacular	spec**p**acular	transformation	transfor**n**ation
stability	sta**d**ility	experience	e**ksh**perience
understandable	understan**b**able	universal	u**m**iversal
understatement	understate**n**ent	university	u**m**iversity
unexpected	unex**t**ected	voluntary	volun**p**ary
unpredictable	unpre**b**ictable	whatsoever	whats**h**oever

Author Contributions

MI and AGS jointly designed the study. MI and TA constructed the stimuli. MI collected the data and did the data analyses. MI wrote the first draft, TA provided feedback on the first draft, and AGS wrote the final draft. All authors approved the final version of the manuscript before submission.

Acknowledgments

We would like to thank April Pufahl, Julia Zhang, Yi Zheng, and Ai Mizoguchi for their help with this project. Support was provided by MINECO Grant PSI2014-53277 and by Centro de Excelencia Severo Ochoa SEV-2015-0490.