Calibration of consonant perception to room reverberation

Eleni Vlahou[1,2,3], Kanako Ueno[4], Barbara G. Shinn-Cunningham[5] and Norbert Kopčo[2,3]

[1] Department of Computer Science and Biomedical Informatics, University of Thessaly, Greece

[2] Institute of Computer Science, Faculty of Science, P. J. Šafárik University, 04001 Košice, Slovakia

[3] Hearing Research Center and Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

[4] School of Science and Technology, Meiji University

[5] Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding

This work was supported by EU H2020-MSCA-RISE-2015 Grant No. 691229, VEGA 1/0355/20 and APVV-0452-12. EV was financed by Greece and the European Union (European Social Fund—ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project "Reinforcement of Postdoctoral Researchers—2nd Cycle" (MIS-5033021), implemented by the State Scholarships Foundation (IKY).

Abstract

Purpose: We examined how consonant perception is affected by a preceding speech carrier simulated in the same or a different room, for different classes of consonants. Carrier room, carrier length, and carrier length/target room uncertainty were manipulated. A phonetic feature analysis tested which phonetic categories are influenced by the manipulations in acoustic context of the carrier.

Method: Two experiments were performed, each with 9 participants. Targets consisted of 10 or 16 vowel-consonant (VC) syllables presented in one of 2 strongly reverberant rooms, preceded by a multiple-VC carrier presented either in the same room, a different reverberant room, or an anechoic room. In Experiment 1 the carrier length and the target room randomly varied from trial to trial while in Experiment 2 they were fixed within a block of trials.

Results: Overall, a consistent carrier provided an advantage for consonant perception compared to inconsistent carriers, whether in anechoic or differently reverberant rooms. Phonetic analysis showed that carrier inconsistency significantly degraded identification of the manner of articulation, especially for stop consonants, and, in one of the rooms, also of voicing. Carrier length and carrier/target uncertainty did not affect adaptation to reverberation for individual phonetic features. The detrimental effects of anechoic and different-reverberant carriers on target perception were similar.

Conclusions: The strength of calibration varies across different phonetic features, as well as across rooms with different levels of reverberation. Even though place of articulation is the feature that is affected by reverberation the most, it is the manner of articulation and, partially, voicing, for which room adaptation is observed.

Introduction

Reverberation is ubiquitous in everyday settings. It has a pervasive influence on the acoustic signals reaching a listener, affecting their temporal structure, spectral content, and interaural differences (Shinn-Cunningham, 2003). Numerous studies show that reverberation can impair spatial hearing and speech perception. For example, it negatively affects sound localization in the horizontal plane (Hartmann, 1983), selective auditory attention to a speech source in the presence of competing sources (Ruggles & Shinn-Cunningham, 2011), and speech intelligibility, particularly for children and older adults, nonnative listeners and hearing-impaired individuals (Assman & Summerfield, 2004; Lecumberri et al., 2010; Nábělek & Donahue, 1984; Takata & Nábělek, 1990). On the other hand, there is strong evidence that adult listeners can quickly adapt to and take advantage of reverberation in many situations (Helfer, 1994; Shinn-Cunningham, 2003). For instance, listeners are sensitive to the statistical regularities that are present in everyday reverberation and exploit these regularities to separate the contributions of sound sources and environmental filters (Traer & McDermott, 2016). Reverberation can facilitate distance perception (e.g., Zahorik, Brungart & Bronkhorst, 2005). Furthermore, exposure to different rooms during phonetic training can enhance implicit phonetic learning (Vlahou et al., 2019). Collectively, these results demonstrate that reverberation can both disrupt and enhance auditory perception, and that listeners use various adaptation mechanisms to mitigate the negative impacts of reverberation and to improve auditory and speech perception.

Different researchers have postulated monaural and binaural adaptation mechanisms that use information from the preceding context to modify and to improve speech perception in reverberation (Beeston et al., 2014; Brandewie & Zahorik, 2010; Srinivasan & Zahorik, 2013; Watkins, 2005). While the specific underlying mechanisms are not fully understood, two primary mechanisms that have been hypothesized include sensitivity to temporal envelope information and to stable spectrotemporal properties in the environment (Zahorik & Anderson, 2013; Srinivasan & Zahorik, 2014; Stilp et al., 2016;

Watkins et al., 2011). Several studies suggest that listeners have high sensitivity to distortions to the signal amplitude envelope that are caused by room reverberation (Zahorik & Anderson, 2013) that appears to be specific to the reverberant-envelope, but not reverberant-fine structure signal (Srinivasan and Zahorik, 2014; Watkins et al., 2011). Related but distinct work has explored listeners' perceptual adjustment to stable spectrotemporal patterns in the acoustic environment (Stilp et al., 2016). Although this type of perceptual compensation is not due to reverberation or speech per se, it appears to help listeners to handle reverberation by decreasing perceptual weight for non-varying spectral cues and assigning larger perceptual weights for changing, and thus more informative, spectral cues.

Regardless of the exact contributions and complementarity of the underlying mechanisms, recent behavioral research has elucidated how adaptation to reverberation affects speech processing. In a seminal study, Watkins (2005) exposed listeners to different levels of reverberation, using monaural speech tokens from the continuum from "sir" to "stir". He showed that, for the same amount of reverberation imposed on the same speech token, listeners shifted their responses towards "sir" or "stir" depending on the level of reverberation in the preceding carrier phrase. Later studies replicated this finding with other speech sounds (Beeston et al., 2014) and non-speech contexts (Watkins & Makin, 2007). Zahorik and colleagues used binaural tasks with speech stimuli presented in reverberation and noise. They demonstrated that prior exposure to a consistent room significantly improved performance for stimuli taken from the Coordinate Response Measure corpus (Bolia et al., 2000; used in Brandewie & Zahorik, 2010) and for sentences with rich phonetic and lexical content taken from the TIMIT database (Garofolo et al., 1993; used in Srinivasan and Zahorik, 2013). These results provide robust evidence that exposure to consistent rooms improves subsequent speech processing, but also raise important new questions.

First, it is not clear whether adaptation to reverberation generalizes across speech sounds and phonetic features with different acoustic properties. Adaptation does generalize across stimuli with diverse lexical content, and, thus, is ecologically beneficial for real-world listening (Srinivasan and Zahorik, 2013). However, the use of lexical items does not enable a precise examination of adaptation processes at the segmental phoneme level, factoring out the contribution of higher order linguistic cues. A previously mentioned early study showed that monaural compensation mechanisms affect perception of the 'sir'-'stir' contrast (Watkins, 2005). A later study showed that adaptation extends to other stops differing in place of articulation, especially /p/ and /b/ (Beeston et al., 2014). Stop consonants are popular candidates for studies investigating speech under adverse conditions, as they are particularly susceptible to masking by noise and temporal smearing by reverberation (e.g., Assman & Summerfield, 2004). Less is known about whether consistent room exposure improves the perception of other features that are also susceptible to room distortions (e.g., non-sibilant fricatives, place contrasts; Gelfand & Silman, 1979). A more detailed investigation of adaptation patterns across different speech sounds in different rooms can better inform theories and models of speech intelligibility in everyday listening environments.

While there is strong evidence that speech perception can be dramatically improved after exposure to consistent reverberation, less is known about how different inconsistent environments affect performance. Brandewie and Zahorik (2018, Experiment 1) replicated the finding of improved speech-in-noise perception after exposure to a consistent room, compared to a baseline condition where no prior room context was given. Examining the effects of inconsistent carriers, they found that when there was a switch from one reverberant room to a room with different reverberation, performance was significantly worse than in the consistent condition, and that the amount of degradation depended on the relative strength of reverberation in the carrier vs. target rooms. Specifically, the disruption was larger when the switch was from a more reverberant carrier room to a less reverberant target room, compared to when the carrier room was less reverberant than the target room. The authors suggested that this might occur because some of the adaptation to the less-reverberant carrier transferred to the new room, improving performance and reducing the difference from the consistent condition. These results motivate further examination of how the acoustic properties of a preceding and new environment interact, especially when the speech is not masked by noise and the system has an opportunity to estimate the room characteristics from the unmasked signals.

Another important issue is the duration of the preceding acoustic context needed for the perceptual system to calibrate. For spatial hearing, there is evidence that localization performance in a weakly reverberant room can continue to improve after hours of exposure (Shinn-Cunningham, 2000). For speech perception, evidence from recent studies suggests more rapid adaptation timescales. For instance, monaural compensation occurs in under a second, with exposure to a consistent previous context producing adaptation that builds up over, at least up to 500 ms of exposure (Beeston et al., 2014). On the other hand, binaural compensation can result in improvement over tens of seconds: adaptation to a room continues to improve with exposure to as many as 10 sentences in the room (Longworth-Reed et al., 2009). Yet, other studies found no increase in adaptation across multiple sentences (Srinivasan & Zahorik, 2014) and no evidence for long-term improvements over many trials (Brandewie & Zahorik, 2010). The exposure duration at which intelligibility improvement asymptotes was also observed to increase with SNR, from 850 ms for lower SNRs to 2.7 s for higher SNRs (Brandewie & Zahorik, 2013). These results suggest that the buildup of adaptation to reverberation for speech perception occurs on a time scales that range widely across conditions. Here we examine whether longer exposure to a preceding consistent vs. inconsistent environment is more beneficial for individual phonetic features of VC syllables in challenging listening environments without noise masking distorting the acoustic properties of the room.

Finally, while past work explored effects of the acoustic properties and the duration of the carrier, less emphasis has been given to non-acoustic factors, such as the ability to direct selective attention to the target speech. For example, knowing when target speech will appear might affect the ability either to benefit from a preceding consistent carrier or to overcome the disruption caused by an inconsistent carrier. Research on speech perception in complex auditory scenes suggests that prior knowledge of the spatial position and voice of a target speech can reduce attentional load and improve selective auditory attention and speech intelligibility in reverberation (e.g., Best et al., 2008; Shinn-Cunningham & Best, 2008). The importance of various aspects of cognitive function (including attention, working memory, speech of processing etc.) on speech perception in adverse listening environments (e.g., with noise or multiple talkers) has been documented in numerous studies (see Dryden et al., 2017 for a review). However, more research is needed to determine whether these top-down factors can also enhance adaptation to reverberation.

Here, we performed two behavioral experiments that studied adaptation to room reverberation for consonant perception. In both experiments listeners were exposed to vowel-consonant (VC) syllables from a carrier phrase, followed by a target VC syllable simulated as being presented in one of two rooms, R1 or R2. The task was to identify the consonant in the final, target syllable. The carrier room was R1, R2, or anechoic space. The length of the carrier varied, containing either 2 or 4-VC syllables. Finally, the carrier/target uncertainty varied across the experiments. In Experiment 1 both the carrier length and the target room randomly varied from trial to trial, i.e., participants could not predict when and from which simulated room the target would appear. In Experiment 2 the carrier length and the target room were fixed, i.e., participants knew in advance when and from which simulated room they would hear the target.

The two reverberant rooms simulated in this study, R1 and R2, had broadband T_{60}'s of approximately 2.5 s and 3 s, respectively, and differed both in room volume and in the distance from source to listener. This strong reverberation was chosen to avoid performance ceiling effects that would preclude us from observing any benefits of adaptation. Previous studies have tackled the ceiling issue by using noise maskers (e.g., Zahorik and Brandewie, 2010) or by lowpass-filtering the stimuli (Beeston et al, 2014). Although adding noise makes the task more difficult, the unique effects of reverberation and the listeners' compensation mechanisms might differ in multiple ways between masked and unmasked conditions. First, if masking noise of levels comparable to the speech is continuously present, then the masking energy always dominates at least over a part of the speech sound's reverberant tail, particularly after the word offset. Thus, the noise may mask the reverberant tail, preventing the auditory system from, e.g., directly estimating frequency-dependent values of T_{60} for which the system might otherwise compensate. Second, the noise has intrinsic, random temporal modulations that are independent of the target speech sound (while, in contrast, the reverberant tail of a sound is deterministically related to the direct sound via the BRIR). These independent modulations are likely to interfere with the system's ability to estimate the temporal modulations like dips in the target stimulus envelope critical for distinguishing certain consonants (e.g., for the "sir-stir" contrast; Watkins, 2005; Beeston et al., 2014). Third, the constant noise energy is likely to dominate the overall signal energy, particularly at the temporal dips of target signal. Since such dips have different depths depending on the consonant (e.g., for the "sir-stir" contrast) and on the room, if the dips are filled in by the same amount of noise energy, the resulting modulation depth becomes more similar across the rooms and consonants, making it difficult to distinguish the consonants or to compensate for/tune to the distinct reverberation effects of each room. Fourth, the target's reverberant tail is binaurally decorrelated, especially in its later portions, as determined by the BRIR. On the other hand, the constantly present masking noise has an approximately constant, relatively high, interaural correlation. And finally, since the target and masker were at different locations in the studies using noise, the mechanisms of spatial release from masking (SRM) are likely to have contributed to target speech identification (Bronkhorst, 2000), possibly interacting with any reverberation compensation mechanism. Specifically, since the amount of SRM decreases with reverberation (LeClere et al., 2015), SRM can differentially influence the observed effects in different rooms in the noise-masking studies. We therefore expected differences in performance between the current and previous studies, which typically used less reverberant rooms and additive noise maskers. Also, the two rooms used here differed in multiple acoustic characteristics, summarized in Fig. 1, and in the speaker/listener locations. Since adaptation to reverberation drops significantly at high levels of reverberation (Zahorik & Brandewie, 2016), we expected differences in performance between the two rooms especially when comparing the effect of inconsistent reverberant carrier and the anechoic carrier.

We tested these hypotheses in a series of analyses. First, to assure overall comparability of the current and previous studies, we examined the effects of a consistent carrier relative to the inconsistent carrier and to the baseline no-carrier condition of Exp. 1, using the overall percent correct consonant identification as the performance measure. Based on Brandewie & Zahorik (2018), we expected performance to be better for the same carrier, compared to both the no carrier and to inconsistent carriers. Subsequently, all comparisons were across different types of carrier; therefore, Exp. 2 did not include no-carrier trials.

A central goal of this study was to examine how consistent and inconsistent carriers affect performance across speech sounds with diverse spectrotemporal properties. To this end, in the next part of the analysis the consonants were grouped based on their distinctive features of manner of articulation, place of articulation, and voicing. Previously, Beeston et al. (2014) focused on the place of articulation feature and found reverberation adaptation effects when only 3 stop consonants were considered. Here, we used a broader set of consonants and used information theory to examine which of the features were affected the most by adaptation to reverberation.

In this analysis, we contrasted performance across the two experiments to examine in detail how carrier length and target room uncertainty affects speech intelligibility across the different classes of speech sounds. Specifically, both the target's temporal position and room was chosen randomly on each trial in Exp. 1 while they were fixed within a block in Exp. 2. Knowing the temporal configuration of the carrier and target syllables as well as the target room in advance might allow listeners to ignore the carrier, reducing attentional load and improving selective auditory attention to the target syllable. On the other hand, it is possible that if participants know when the target occurs, they may simply ignore the carrier, reducing any adaptation to the carrier's reverberation characteristics. This in turn is likely to reduce the effect of both consistent and inconsistent carriers. Such effects of temporal and contextual expectation across the two experiments are expected to interact with carrier room, and to be greater for the longer carrier length.

Note that Exp. 1 was performed using a larger set of 16 consonants as stimuli. Since performance was at ceiling for 6 of those consonants, Exp. 2 presented only the remaining 10 consonants, although the participants could still use all 16 consonants when responding.

Methods

Participants

Nine young male and female listeners participated in Experiment 1 (21-35 years old) and nine different male and female listeners in Experiment 2 (21-35 years old). Four participants (2 in Exp. 1) had previous experience with psychophysics procedures. All participants had normal hearing, as confirmed by an audiometric screening (set at 20 dB HL for frequencies ≤ 8 kHz for both ears), and spoke English as their first language. All procedures were approved by the Boston University Institutional Review Board.

Speech Material

Sixteen consonants (k, t, p, f, g, d, b, v, ð, m, n, ŋ, z, θ, s, and ʃ) were used, each preceded by the vowel /a/. We used vowel-consonant (VC), rather than consonant-vowel syllables, as preliminary listening indicated that reverberation effects were greatest for final consonants (see also Gelfan & Silman, 1979). Stimuli were produced by three speakers, with one male recording taken from CUNY-NST corpus (Resnick et al., 1975) and one male and one female recording from the corpus described by Yund and Buckles (1995). For each VC, three tokens were spoken by each of three talkers. This resulted in a total of 144 unique speech tokens (16 VCs x 3 talkers x 3 tokens). Overall level differences across talkers were removed by equalizing the root-mean-square (RMS) energy levels of all tokens.

In Experiment 1 participants performed at ceiling for six of the consonants (k, t, n, z, s and ʃ), with correct identification responses exceeding 90% in all tested conditions. Trials containing these consonants as target stimuli were removed from the analyses in Exp. 1 and these consonants were not included as targets in Exp. 2. However, in both experiments these stimuli were included in the carrier syllables and participants could still respond that they heard one of these consonants as targets. Analyses including these consonants as targets in Exp. 1 are presented in Appendix D.

For the feature-based analysis, the consonants were grouped by their manner of articulation, place of articulation, and voicing. Table 1 shows the feature classification used in this study.

Table 1. Phonetic feature classification. Consonants not used as target stimuli are in bold. All consonants were available as responses in both experiments.

Room Simulation

To simulate the presentation of stimuli in different rooms, the VC tokens were convolved with binaural room impulse responses (BRIRs). The BRIRs were recorded using a setup consisting of an omni-directional (up to 2 kHz) dodecahedral loudspeaker system and a manikin head (Head Acoustics, HMM2) that faced the speaker system. The choice of omnidirectional loudspeaker system was made, even though this type of loudspeaker system has different directional characteristics than a human talker, as the results obtained with an omnidirectional loudspeaker can be thought of as approximating the average of different speaker orientations. BRIRs from two different large rooms were used, denoted as R1 and R2. The R1 response was measured in a large concert hall (room volume 22,776 m^3, 2020 seats) with the manikin located on the second balcony, 33 m from the speaker system located on the stage. The R2 response was measured in an elliptical church (room volume 13.333 m^3) with the manikin relatively close (12 m) to the sound source, which was located beside the altar. The impulse responses were measured using the swept-sine method (Suzuki et al., 1995), for a time-stretched pulse of 1.35 seconds duration and with synchronous averaging (Satoh et al., 2004). An anechoic BRIR (AN) was derived from the R2 BRIR by time-windowing the first 5 ms of the response using a rectangular window to remove most of the reverberant energy. The resulting three BRIRs (R1, R2, and AN) were equalized for overall RMS energy. This equalization made the direct sound energy of R1, R2 and anechoic rooms quite different. However, the perceived loudness of speech stimuli convolved with the three BRIRs was comparable, as confirmed by informal listening.

Figure 1 shows the acoustic properties of the BRIRs. Early time-domain portions of the responses in one ear are shown in Fig. 1A. R2 has a large echo around 50 ms after the direct sound, likely due to its elliptic room shape. Fig. 1B shows reverberation times (T_{60}) as a function of frequency. R2 has a larger T_{60} than R1 at all frequencies. Fig. 1C shows the Clarity Index C_{50}, i.e., the ratio of the early energy (0-50 ms) to the late energy (beyond 50 ms) in the impulse response as a function of frequency. C_{50} is lower in R2 than in R1, especially in the mid-frequency bands (250-1000 Hz). This analysis suggests that R2 should be more disruptive to speech perception than R1, while AN can be considered an ideal environment for speech perception, without any acoustic distortion.

The stimuli consisted of sequences of 0, 2, or 4 carrier VCs convolved with one BRIR followed by a target VC convolved with the same or a different BRIR. The stimulus onset asynchrony between individual VCs in sequences was always 0.8 s. Due to the long reverberation times of the BRIRs, the reverberant Carrier’ tails overlapped with the Target signals in the current stimuli. This might have caused energetic masking and interaural de-correlation of the Target stimuli by the Carrier energy, affecting their intelligibility. Appendix A contains acoustic analysis that shows that these effects were relatively small, especially compared to the intrinsic masking by the vowel in the target VC (Beeston et al., 2014) and especially towards the end of the direct portion of the Target stimulus containing the consonant that the listeners need to identify.

-------------------------- INSERT FIGURE 1 ABOUT HERE -------------------------------

Setup

The experiments were performed in an experimental laboratory in the Boston University Hearing Research Center. In both experiments, participants were seated in front of an experimental computer inside a double-walled sound-proof booth. The experiments were implemented in MATLAB software (Mathworks Inc.). Stimuli were presented through a D/A converter (TDT RP2) and headphone amplifier (TDT HB7) driving insert headphones (Etymotic Research, ER1) at a comfortable listening level (adjusted by the experimenter). Participants responded using a graphical user interface (GUI) with 16 graphical buttons labeled with the 16 VCs (‘ok’, ‘ot’, ‘op’, ‘of’, ‘og’, ‘od’, ‘ob’, ‘ov’, ‘oth ð‘, ‘om’, ‘on’, ‘ong’, ‘oz’, ‘oth θ‘, ‘os’, ‘osh’), clicking with a computer mouse the button corresponding to the perceived target VC.

Procedure

Prior to each experiment, a short training session was conducted to familiarize participants with the connection between the response GUI and the corresponding VC sounds. Participants were instructed to click on graphical buttons to produce the corresponding sounds, in a self-paced manner, until they felt confident about the relationship between sound and response. Upon clicking one of the buttons, a VC spoken by one male talker in an anechoic room was presented. There were no time constraints in this practice session, which typically took several minutes.

Next, there was a short warm-up phase in which participants completed a session of 10 sample trials, identical to the test sessions described below. Participants were instructed to listen to the sounds and report the consonant in the final syllable. No feedback was provided. In Experiment 2 this warm-up phase was conducted each time the carrier length changed (described below).

The warm-up was followed by the experimental runs. On each experimental trial, participants heard an initial carrier, consisting of 2- or 4-VC syllables, followed by a target VC syllable. On each trial, each of the carrier syllables was randomly selected. In Experiment 1 there was an additional control condition, in which participants only heard the target VC without a preceding carrier. The task was to report the consonant in the final target VC by mouse-clicking on the corresponding button in the GUI. The reverberation of the syllables in the carrier was randomly selected on each trial to be either R1, R2, or AN. With the exception of the no-carrier trials in Exp. 1, the reverberation of the target syllable was R1 on half the trials and R2 on the other half (in Exp. 1 no-carrier trials, R1, R2, and AN trials were presented with equal probability). The length of the preceding carrier (0, 2, or 4 VCs) varied randomly from trial to trial in Exp. 1, whereas it was blocked (2 or 4 VCs) in Exp. 2. Similarly, the target room varied randomly in Exp. 1 and was blocked in Exp. 2. A random voice was selected for each trial and was consistent for all VC syllables within the trial. All three voices and three tokens per target VC were presented an equal number of times.

Each of the two experiments contained 720 trials in total. In Experiment 1 the trials were distributed across three sessions of 240 trials each. Each session contained (a) each of the 16 consonants in the target VC for each carrier length (2 and 4 VCs), carrier room (AN, R1, R2) and target reverberation (R1, R2) and (b) 48 control trials without a preceding carrier (No Carrier), with each of the 16 consonants as targets, for each room (R1, R2, AN). In Experiment 2, trials were distributed across two daily sessions of 360 trials. Each session contained two repetitions of each of the 10 consonants in the target VC for each of the three talkers and carrier reverberation (AN, R1, R2), in separate blocks for each carrier length (2 and 4 VCs). In each session the target reverberation was fixed (R1 or R2), with the order counterbalanced across participants.

Statistical Analyses

For overall consonant identification, participants' percent correct scores were logit transformed and entered into ANOVA tests. All figures show untransformed values and all error bars in the figures indicate standard error of the means. To quantify phonetic feature identification, we used the Information Transfer Rate (ITR) score, an information-theory derived measure (Shannon, 1948) commonly used for phonetic feature perception analyses (e.g., Miller & Niceley, 1955; Beeston et al., 2014; Sagi & Svirsky, 2008). The ITR is obtained by normalizing the mutual information between the speech stimuli and the participants' responses by the stimulus entropy. A score of 1 indicates no confusions, whereas a score of 0 indicates random guessing. Unlike percent correct scores, this measure takes into account unbalanced categories and response biases (Sagi & Svirsky, 2008).

Results

The results presentation is divided into two main parts. First, we present the across-consonant average percent-correct identification data from Experiment 1 and 2, to confirm that the overall pattern of build-up and breakdown of adaptation is similar to the previous studies. Then, the main analysis focuses on the phonetic feature identification performance from both experiments, to examine the effects of carrier/target uncertainty, carrier room and carrier length on consonant identification in the two target rooms. Finally, we present a brief analysis of confusion matrices showing the error patterns for individual consonants. Additionally, Appendix B contains an analysis of the overall performance for the three talkers used in the study, showing that intelligibility was above chance for all three, without reaching ceiling levels.

Across-Consonant Average Identification Performance

Three analyses were performed on the across-consonant average percent correct identification data, mainly to confirm that, overall, identification of final consonants in VCs improves with consistent-room carriers while it is made worse with inconsistent ones. Figure 2 plots the across-participant-averaged percent correct responses for different target rooms as a function of the carrier room for both experiments.

-------------------------- INSERT FIGURE 2 ABOUT HERE --------------------------------

Experiment 1: No-carrier baseline performance

First, we examined the participants' baseline performance for targets not preceded by any carrier. The goals were to confirm that the reverberant rooms used in this study can degrade speech intelligibility significantly, and to establish baseline performance against which we could directly evaluate the effect of preceding carriers on consonant identification accuracy.

Results

The leftmost data points in Fig. 2a, corresponding to the no-carrier (NC) baseline condition of Exp. 1, show the mean identification accuracy for target stimuli simulated from all 3 rooms used in this study. The presence of reverberation had a dramatic effect on consonant intelligibility. While identification accuracy in the anechoic room reached 88% (*SE* = 3.2), in the two reverberant rooms it fell by about 30%. Further, the results show that intelligibility was higher for target room R1 (62%, *SE* = 4.5) than R2 (56%, *SE* = 5). Confirming these observations, a one-way repeated measures ANOVA with Target Room (R1, R2, anechoic) as the within-participants factor showed a significant effect ($F(2,16) = 65.44$, $p < 0.0001$).

Discussion

The comparison of no-carrier intelligibility in the anechoic room vs. the strongly reverberant rooms shows that the reverberation associated with the utterance of a vowel in a VC pair distorts perception of the subsequent consonant signal, interfering with identification (also see Appendix A). Note that performance degradation due to reverberation is likely to be much smaller for initial consonants (i.e., if the stimuli were CVs instead of VCs), as these consonants would not be affected by the vowel-related reverberation as much (e.g., Gelfand & Silman, 1979); specifically, the additional energy due to reverberation from a vowel will overlap the energy of a subsequent consonant, but not a preceding consonant, as used, e.g., in Beeston et al. (2014). Informal piloting prior to the current study supported this prediction. It is also important to note that, as described in the Introduction section, the masking effect of strong reverberation might be different from the masking effect of noise, which was used to limit the baseline performance in several previous studies performed in less reverberant rooms (e.g., Brandewie & Zahorik, 2018). Finally, the detrimental effect of reverberation was larger in room R2 than R1 in the current study. This is consistent with acoustic analysis showing a higher T_{60} and a lower C_{50} for this environment (see Fig. 1B-C).

Experiment 1: Effect of a preceding carrier relative to the no-carrier baseline performance

Next, we examined the effect of a preceding carrier relative to the no-carrier baseline. Specifically, we tested whether an inconsistent carrier (different reverberation or anechoic room) degrades performance relative to baseline and/or whether a consistent carrier causes an improvement. Note that, in order to compare the different carrier rooms with the no carrier condition, data were averaged across carrier length.

Results

Fig. 2a shows the across-participant average consonant identification accuracy (%) for Experiment 1, as a function of carrier room and target room. A repeated-measures ANOVA with the factors of target room (R1, R2) and carrier room (No Carrier, Same, Different, Anechoic) showed a main effect of target room ($F(1,8) = 10.81$, $p = 0.0111$, $\eta_p^2 = 0.57$), owing to improved overall performance for the less reverberant target room R1. There was also a main effect of carrier room ($F(3,24) = 3.08$, $p = 0.046$, $\eta_p^2 = 0.28$), and no interaction ($F(3,24) = 1.96$, $p = 0.148$, $\eta_p^2 = 0.196$). Following the significant effect of carrier room, we performed post-hoc pairwise comparisons, corrected by the Holm-Bonferroni method. To minimize the number of comparisons, the two inconsistent rooms (anechoic and different) were pooled [0.5*(anechoic+different)] and treated as one contrast. First, we performed a directional pairwise comparison between the no-carrier and same carrier, based on our hypothesis that performance would be improved after exposure to a consistent carrier compared to the no-carrier (*buildup*; Brandewie &

Zahorik, 2018). However, there was no significant difference between the two conditions (t(17) = -0.93, *p* = 0.18). On the other hand, performance was significantly worse in the inconsistent carriers compared to the same carrier (t(17) = 2.40, df = 17, *p* = 0.014). Finally, performance did not differ between the no-carrier and the inconsistent carriers (t(17) = 1.21, *p* = 0.121).

Discussion

Contrary to previous reports (e.g., Brandewie & Zahorik, 2018) we did not find a significant improvement in performance after exposure to a consistent carrier, relative to a no-carrier baseline. Several important differences between the two studies can account for this discrepancy, including the use of different rooms, speech materials and tasks, as well as the absence of noise-masking. On the other hand, compared to the consistent carrier, performance was significantly worse when the target was preceded by an inconsistent carrier. These results suggest that, in strongly reverberant environments, listeners are less able to benefit from a consistent preceding context, compared to a no carrier baseline condition, while at the same time, tuning speech perception to the acoustics of an inconsistent (reverberant or anechoic) room can be detrimental. Specifically, the improvement in consonant identification in consistent vs. inconsistent rooms was on average 5% in target room R1 and only 1% in target room R2.

Experiment 2: Effect of consistent vs. inconsistent carriers

In Exp. 2 the no-carrier baseline condition was not included. The analysis of consonant identification performance therefore focused on establishing that the improvement in performance is observed for the consistent vs. inconsistent carrier conditions, similar to the results of Exp. 1 and of previous studies. In this analysis, data were again averaged across carrier length.

Results

Fig. 2b shows the across-participant average consonant identification accuracy (%) for Experiment 2, as a function of carrier room and target room. A repeated-measures ANOVA with the factors of target room

(R1, R2) and carrier room (Same, Different, Anechoic) showed a main effect of carrier room ($F(2,16)$ = 8.84, p = 0.006, η_p^2 = 0.52), and a carrier X target room interaction (F(2,16) = 9.077, p = 0.009, η_p^2 = 0.53). One-sided post-hoc pairwise comparisons showed that performance was significantly worse in the inconsistent carriers compared to the same carrier for target room R1 (t(8) = 5.59, p = 0.0003) whereas no significant differences were observed for target room R2 (t(8) = 1.59, p = 0.08).

Discussion

The pattern of results for the Same, Different, and Anechoic conditions is similar to that of Exp. 1, while the overall performance in Exp. 2 is better, presumably due to lower carrier length/target room uncertainty. Specifically, contrary to Exp. 1, in Exp. 2 carrier length and target room were fixed within a block, and thus participants knew in advance when and from which room the target syllable would appear. This might have helped participants to ignore the carrier and focus attention to the target speech, thus reducing attentional load and improving overall identification performance. Again, compared to the consistent carrier, performance was significantly worse with inconsistent carriers. This time, the effect was observed for target room R1 but not R2. However, the results are very similar to Exp. 1, with the consistent-vs-inconsistent performance difference of 7% in target room R1 and 2% in R2.

Effects of carrier room, carrier length, and carrier/target uncertainty on phonetic features

A central goal of this study was to investigate the effects of different carrier and target characteristics on major classes of speech sounds. To this end, in the next part of the analysis the consonants were grouped into phonetic features based on manner of articulation, place of articulation, and voicing. Specifically, consonants were grouped into different categories according to their features (see Table 1 and Fig. 5, which shows confusion matrices for individual consonants discussed below, and which also shows the category labels along the x-axis and y-axis). Then, confusion matrices were derived, separately for place, manner, and voicing. In these confusion matrices, the stimulus-response pairs were only considered at the feature level, i.e., identifying the voiced labial stop of /b/ as an unvoiced labial stop of /p/ would increase the number of voiced-unvoiced (i.e., incorrect) responses in the voicing feature category, labial-labial (i.e., correct) responses in the place category, and stop-stop (i.e., correct) responses in the manner category. Based on these new matrices, we computed for each individual participant the information transfer rate (ITR) for each feature across the different combinations of carrier and target rooms. We expected that the benefits of consistent carrier would differ across the phonetic features as different features are affected differently by reverberation. Specifically, place of articulation has been shown to be particularly sensitive to reverberation (Gelfand and Silman, 1979). Therefore, it is possible that this feature will benefit the most from a consistent carrier, e.g., if tuning to the carrier allows the system to overcome some of the negative effects of reverberation (as observed for initial consonants in Beeston et al., 2014). On the other hand, if the reverberation distorts the place of articulation cues such that they cannot be recovered, no adaptation to this feature is expected.

Three characteristics of the carriers and targets were systematically manipulated across the two experiments: the carrier room (same, different, anechoic), the carrier length (2 or 4 VCs), and the carrier/target uncertainty (in Exp. 1 the carrier length and the target room varied randomly from trial to trial, and thus listeners could not predict the target onset or its room; in Exp. 2 these parameters were fixed within a block; however, note that there were other differences between Exp. 1 and 2 as well). The main prediction regarding the carrier room was that performance would be better after exposure to a consistent carrier compared to either of the inconsistent carriers. Considering the two inconsistent carriers, Brandewie & Zahorik (2018) observed that a carrier with reverberation larger than the target was more disruptive than vice versa. Thus, a potential outcome was that the disruptive effect of the anechoic carrier would be smaller than that of either of the reverberant-room carriers. Alternatively, the anechoic carrier might be the most disruptive as the anechoic room was very dissimilar from both of the reverberant rooms, while the two reverberant rooms were relatively similar to each other. These effects of the carrier room were predicted to grow with carrier length, as it was expected that the tuning to each carrier room would get stronger over time, resulting in a larger improvement for the longer consistent carrier and a larger degradation for the longer inconsistent carriers. Regarding carrier/target uncertainty, it was expected that knowing when and from which room to expect the target might allow listeners to ignore the carrier altogether and focus attention exclusively on the target. This, in turn, would result in reduced interference from inconsistent carriers. Finally, while the two target rooms were both similar in that they were strongly reverberant, it was expected that the effects of carrier would be more visible in the less reverberant target room R1 than in R2, consistent with Zahorik & Brandewie (2016).

-------------------------- INSERT FIGURE 3 ABOUT HERE -------------------------------

Results

To preface our results, we did not find evidence that carrier length or uncertainty affect adaptation to reverberation for any feature (no interaction involving these factors with carrier room; see statistical analyses below). Thus, Figure 3 shows the across-participant average ITR score, as a function of carrier room, separately for each phonetic feature (separate panels) and each target reverberation (different colors within each panel), with results pooled across carrier length and experiment.

Consistent with our average consonant identification results, overall performance tended to be higher for target room R1 (blue) and for the same carrier condition. For both rooms, manner of articulation was the feature with the highest transmission (ITRs ranging between about 0.7-0.8), followed by voicing (ITRs ranging from 0.6-0.75) and place of articulation (ITRs from about 0.4-0.5). The particularly low performance for place is consistent with previous work on phonetic confusions in noise and reverberation, showing that place is negatively affected, especially for consonants in the final position (e.g., Gelfand & Silman, 1979; Miller & Nicely, 1955).

For each class of features (manner of articulation, voicing, and place of articulation), a mixed ANOVA was performed on participants' ITR values, with carrier/target uncertainty (experiment) as a between-participants factor and with carrier room (Same, Different, Anechoic), carrier length (2 vs. 4 VCs) and target room (R1, R2) as within-participant factors.

For manner of articulation, the mixed ANOVA yielded a significant main effect of Carrier Room ($F(2,32) = 4.71$, $p = 0.023$, $\eta_p^2 = 0.23$), while Experiment, Target Room, and Carrier Length were not significant either as main effects nor as interactions (Experiment: $F(1,16) = 4.47$, $p = 0.051$, $\eta_p^2 = 0.22$; Target Room: $F(1,16) = 3.21$, $p = 0.092$, $\eta_p^2 = 0.17$; Carrier Length: $F < 1$, ns; all other p's > 0.16). Given this, the black line in Figure 3 collapses across Target Room to better visualize the significant effects. Based on our directed hypothesis that adaptation to reverberation would be stronger for the Same carrier, we performed one-sided post-hoc comparisons between Same vs. Different and Same vs. Anechoic carriers. The pairwise comparisons (adjusted with the Holm-Bonferroni correction for multiple comparisons) showed a significant difference between Same and Different ($t(71) = 2.74$, $p = 0.026$) and Same and Anechoic ($t(71) = 2.69$, $p = 0.009$).

For voicing, the ANOVA showed a significant interaction between carrier room and target room ($F(2,32) = 5.49$, $p = 0.01$, $\eta_p^2 = 0.26$). No other main effect or interaction came out as significant (carrier room: $F(2,32) = 3.19$, $p = 0.054$, $\eta_p^2 = 0.17$; Carrier length X Carrier Room: $F(2,32) = 2.45$, $p = 0.102$; Experiment X carrier length X carrier room: $F(2,32) = 3.07$, $p = 0.073$; Experiment X target room X carrier room: $F(2,32) = 2.71$, $p = 0.082$; Experiment X target room X carrier length X carrier room: $F(2,32) = 2.95$, $p = 0.0718$; all other p's > 0.10). Post-hoc pairwise comparisons (adjusted with the Holm-Bonferroni correction) found a significant difference for the Same vs. Different carrier ($t(35) = 3.12$, p =

0.0018) and Same vs. Anechoic carrier (t(35) = 3.17, p = 0.002) for room R1, but not for room R2 (Same vs. Different: t(35) = 0.03, p = 0.49; Same vs. Anechoic: t(35) = -1.79, p = 0.95).

For place, there was a significant main effect of target room ($F(1,16) = 14.98$, $p = 0.001$, η_p^2 = 0.4836), owing to better overall performance in target room R1 compared to target room R2. There were no other main effects or interactions (carrier type: $F(2,32) = 1.76$, $p = 0.188$; all other p's > 0.20). As the overall improved performance on the less reverberant target room was expected and was not a main interest of this study, we did not follow up on this result further.

Because considerable across-subject differences in performance were observed, Appendix C provides information about individual subject performance for the significant effects found in the ANOVAs above.

To further examine the significant ITR improvements with consistent vs. inconsistent carriers, we attempted to identify which individual phonetic features corresponding to manner of articulation and voice drive the effects shown in Fig. 3. While ITR cannot be computed when a single feature is considered in isolation, it is possible to compute what proportion of the responses for an individual feature were correct. And, while the measures of ITR and percent correct are not equivalent (e.g., in the extreme, if a subject consistently reverses the responses in a two-alternative task, the ITR is 1 while the percent correct is 0), analysis of individual feature's percent correct might identify some factors that also influenced the effects in terms of ITR. With that caveat in mind, Fig. 4 plots the percent correct identification of individual phonetic features corresponding to manner of articulation (left-hand panel, shown for both target rooms) and voice (right-hand panel, only for target room R1). The left-hand panel shows that the percent correct performance is only influenced by carrier consistency for the stop consonants, suggesting that some of the improvement in the ITR comes from better identification of stop consonants. Similarly, the right-hand panel shows that the percent correct performance is influenced by carrier consistency more for the voiced than for the unvoiced consonants which might suggest that the ITR improvement in Fig. 3 was driven more by the voiced consonants. Paired t-tests performed on these percent correct data showed significant improvements for stop consonants ($t(17) = 3.58$, $p = 0.0023$) and voiced consonants ($t(17) = 3.48$, $p = 0.0029$), while for the fricative, nasal, and unvoiced consonants the difference was not significant ($p > 0.5$).

-------------------------- INSERT FIGURE 4 ABOUT HERE -------------------------------

Discussion

Manner of articulation (left-most panel in Fig. 3) is the feature for which a consistent carrier yields the strongest benefit in terms if ITR compared to a carrier from an inconsistent room. This improvement might be partially explained by the observation that, in both rooms, the percent correct identification of the stop consonants has improved, whereas for fricatives and nasals there was no evidence of improvement (Fig. 4, left panel). On the other hand, the feature of voicing (center panel in Fig. 4) showed a strong same-vs-different carrier improvement in ITR for target room R1, but no such effect for R2. This improvement might be related to an improvement in the percent correct identification of the voiced consonants (Fig. 4, right panel). The room specificity of this effect suggests that tuning to the voicing characteristics in the more reverberant R2 carrier has a negative effect on consonant identification in the less reverberant R1 target, but not vice versa. Finally, there was no ITR improvement for the same-vs-different carrier for the place feature (right panel in Fig. 3), even though there was a trend for same-carrier improvement in target room R1. Overall, these results show that stop consonants are affected the most by adaptation to room reverberation. Similarly, Beeston et al. (2014) observed adaptive effects for stops preceding vowel and differing in their place of articulation. Thus, it is possible that these adaptive effects have different strengths depending on the position of the consonant within the word. The current results also show that voiced consonants can be affected by adaptation in certain rooms.

No effects of carrier length or carrier/target uncertainty were observed in this analysis. This suggests that at least at the level of the phonetic features, the immediately preceding carrier is the main driver of the adaptation changes and that this adaptation is fast enough to build up across two syllables of the short carrier used here.

Consonant confusion matrices for Individual phonemes

-------------------------- INSERT FIGURE 5 ABOUT HERE --------------------------------

Finally, we present an exploratory analysis of the consonant confusions across the different carrier and target rooms. This analysis is only exploratory because very few measurements per consonant were done, although given the results of the phonetic analysis (previous section) which did not find any significant effect of carrier length and uncertainty, the data were collapsed across these two factors, to partially alleviate this shortcoming. Fig. 5 plots the across-subject average confusion matrices for individual consonants separately for all combinations of carrier and target rooms. Note that the matrices are not square as more responses were allowed than the number of presented consonants considered.

As can be seen from Fig. 5, performance varied considerably across phonemes. The two consonants most severely affected by reverberation were /m/ and /θ/, with overall identification accuracy less than 45%. At the other extreme, /g/ and /d/ were perceived much more accurately, with average performance exceeding 85% correct. A closer examination of the participants’ errors reveals that, for each stimulus, confusions clustered around one or two dominant responses that tended to be consistent across carrier and target rooms. Specifically, for 8 out of the 10 target consonants, the primary confusion was consistent across all, or almost all, carrier and target rooms (6/6 conditions for 5 consonants /v/→/ð/, /m/→/ŋ/, /ð/→/d/, /f/→/ð/, /ð/→/ f/, and 5/6 conditions for 3 consonants /g/→/ŋ/, /ŋ/0-→/n/, /p/→/k/) and accounted, on average, for 52 % of all errors. Further examination of the participants' errors revealed that a few phonemes were mutually confusable, with the clearest cases being /θ/-/f/ (for both target rooms) and /d/-/ð/ (for the R1 target room). However, in most cases the phonemes were not equally confusable with each other, but rather showed a response bias. Specifically, certain nasals tended to be confused for specific other nasals: e.g., /m/ was systematically confused with /ŋ/, while for /ŋ/ the primary confusion was /n/ (which remained a response option even when not presented as a target consonant) and /m/ was the secondary confusion. For /b/ the primary confusion was /v/ (in 4/6 conditions), whereas /v/ was consistently confused with /ð/. In summary, these examples suggest that reverberation created a complex pattern of consonant confusion groups that were mostly asymmetrical.

General Discussion

This study investigated how final-consonant perception in a highly reverberant room is influenced by a preceding carrier phrase simulated from either the same or a different room. The effects of various combinations of carrier and target rooms were examined using natural reverberation, without adding noise or introducing other manipulations, such as abrupt cut-offs, that have been used in previous work (e.g., Brandewie & Zahorik, 2018; Zahorik & Brandewie, 2016; Srinivasan & Zahorik, 2013; Beeston et al., 2014). Here, for two reverberant target rooms, we examined different aspects of the preceding carrier and target: the carrier room (i.e., the preceding carrier either had the same room reverberation as the target, a different room reverberation, or was anechoic), the carrier length (either 2 or 4-VC syllables), and the carrier/target uncertainty (the carrier length and target room were either fixed or varied randomly from trial to trial). The main results were obtained after grouping consonants along three phonetic features (manner of articulation, place of articulation, and voicing), while secondary analysis was performed on percent correct consonant identification data averaged across the consonants.

Without a preceding carrier the simulated reverberant rooms degraded perception of some consonants while having a negligible effect on others. Specifically, the target consonants /z/, /n/, /t/, /s/, /k/, and /ʃ/ included in Exp. 1 were removed from further analysis because their perception was largely unaffected by the room reverberation. The participants' ITR performance using the full set of consonants is presented in Appendix D. As expected, including the six consonants raises overall performance, whereas the magnitude of adaptation to reverberation remains very similar across the two datasets. Not surprisingly, 3 of these 6 consonants were sibilants, with strong energy at higher frequencies, which have been shown to be resistant to both noise and reverberation (e.g., Gelfand & Silman, 1979; Danhauer & Johnson, 1991; Miller & Nicelly, 1955). Performance was also unaffected by the room acoustics for the unvoiced stop consonants /t/ and /k/, while it dropped significantly for the unvoiced stop /p/. While it is outside the scope of this study to determine why reverberation affects some consonants more than others, it is possible that the strong high- and mid-frequency bursts that are critical, respectively, for the perception of /k/ and /t/ survived reverberation, in contrast to /p/, which is instead characterized by a soft wide-band click that diminishes to a low frequency burst (Li & Allen, 2011; Li et al., 2010), making it more susceptible to temporal smearing by reverberation. Overall, in agreement with previous studies our results show that there is considerable variability in how reverberation affects different speech sounds, ranging from negligible, to moderate, to strong disruptions in perception (e.g., Danhauer & Johnson, 1991; Gelfand & Silman, 1979).

Averaged across the remaining 10 consonants, in Exp. 1 we expected to find a significant improvement in speech perception after exposure to a consistent carrier, relative to a no-carrier baseline condition (e.g., Brandewie & Zahorik, 2010; Beeston et al., 2014; Srinivasan & Zahorik, 2013, etc.).

However, we only found a weak improvement in overall identification accuracy. On the other hand, the inconsistent carriers, on average, impaired performance compared to the no-carrier baseline. Thus, overall, the negative effect of inconsistent carriers *re.* baseline was stronger than the positive effect of consistent carriers, while in previous reports by Brandewie & Zahorik (2018) an inconsistent carrier never led to worse performance than the no-carrier baseline. However, the inconsistent-carrier performance was observed to fall below the “silent” baseline in Exp. 2 of Beeston et al. (2014) which did not use noise masking. These results suggest that when the effect of carrier adaptation is measured without noise masking, listeners are less able to take advantage of a consistent preceding context to improve perception, while, at the same time, they are very susceptible to the disruptive effects of an inconsistent context. Overall, the effect of consistent-vs-inconsistent carrier was fairly small in the current study. For targets in room R1, the benefit was on average 5-7% in both Exp. 1 and 2, while for targets in room R2, the effect was negligible, on the order of 1-2%.

The remaining analyses considered data from both experiments evaluated by considering information transmitted for three classes of phonetic features. The major goal of this analysis was to examine whether the relative benefit of consistent vs. inconsistent carrier phrases for consonant perception was specific to certain phonetic features, e.g., stop consonants differing by their place of articulation (Beeston et al., 2014), or whether it also affects other features that are representative of the acoustic-phonetic diversity of everyday listening.

Phonetic feature analysis showed that the highest information transmission was observed for manner of articulation, followed by voicing and place of articulation. This is consistent with previous work on phonetic confusions in noise and reverberation (e.g., Miller & Nicely, 1955; Gelfand & Silman, 1979). The larger number of place errors is also consistent with previous reports (e.g., Miller & Nicely, 1955; Benki, 2004).

Manner was the feature that showed the most robust improvement in performance for the same-room carrier. This robust improvement was observed in both reverberant target rooms, but was restricted to the stop consonants. The strong adaptation to room reverberation that we found for stop consonants (see Fig. 4, left-most panel) is in line with previous studies which report strong monaural compensation for stop consonants (e.g., Beeston et al., 2014; Watkins et al., 2005). It appears that, even though stops can be substantially degraded by reverberation, they are the class of phonemes that can benefit the most from prior exposure to a consistent room.

For voicing there was a large consistent-vs-inconsistent carrier difference for the R1 target room, but no difference for the more reverberant R2 target room. This asymmetry might help explain the above-mentioned asymmetry in how much degradation was caused by the inconsistent-room carrier for target room R1 vs. R2 in the across-consonant average data. Also, it might be the cause of the previous report that there is a greater disruption in speech identification caused by a more reverberant carrier than by a less reverberant carrier (Brandewie & Zahorik, 2016). Specifically, the current results suggest that this asymmetry is driven primarily by specific disruptions in the identification of voicing, and specifically for the voiced consonants, for which the detrimental effect was significant.

Finally, for place of articulation there was only a weak trend for an improved performance on the same carrier in the R1 target room that did not reach significance. Thus, the place of articulation seems to be the feature that is the least affected by the specific characteristics of any given reverberant room and/or the characteristic to which the auditory system is tuning the least when adapting to a specific reverberant room.

The two different types of inconsistent carriers used in this study were expected to affect performance differently. On the one hand, the anechoic carrier might be more disruptive than the different-room reverberant carrier, as it has substantially different acoustic characteristics than both reverberant target rooms. On the other hand, the anechoic carrier does not distort the stimuli, giving the listeners a chance to have a good "look" at the "clean" version of each phoneme. Such looks may be beneficial when identifying target speech distorted by reverberant energy. Contrary to our predictions our results showed that there was a similar drop in performance for the anechoic and different-room reverberant carriers. This result might mean that the critical parameter of the carrier and target rooms is whether they are the same or different (as opposed to the specific amount of reverberation by which they differ). Alternatively, it may be that the two above-mentioned contradicting predictions about the effects of the anechoic carrier tend to cancel.

Our results failed to show an effect of carrier/target uncertainty on the across-phoneme averaged data. This is in line with previous reports that uncertainty about the temporal location of the target stimulus does not reduce the magnitude of adaptation to reverberation (Beeston et al., 2014). However, the current results preclude coming to a definitive conclusion, as in addition to the uncertainty the experimental design changed in other minor aspects between the two experiments reported here.

A consistent finding in our study was that the effect of the different carrier rooms was much smaller for the targets in room R2. This is likely due to the larger broadband T_{60} and C_{50} of R2, which resulted in a marked decrease in performance in the no-carrier condition. This strong reverberation not only made the baseline R2 performance worse, it also made it more difficult for listeners to benefit from prior exposure to this room for all the phonetic features. There was also only a modest negative effect of inconsistent carriers on R2 targets, consistent with a previous report that the more reverberant inconsistent carriers have a more negative effect on the less reverberant target than vice versa (Brandewie and Zahorik, 2016). However, importantly, R2 also differed from R1 in other aspects such as its elliptical shape and prominent low frequency resonances (Fig. 1a). In future studies it is important to examine whether such aspects are also important, e.g., by controlling the level of reverberation while using more rooms, with different wall materials and layouts. Independent of the exact cause, our results suggest that the magnitude of facilitation or disruption due to adaptation to reverberation can vary considerably depending on the acoustic properties of the target room.

An important question not addressed directly in this study is what mechanism supports adaptation to reverberation and to what reverberant characteristics of rooms listeners can adapt. Previous studies have shown that the adaptation operates both monaurally (e.g., Beeston et al., 2014) and binaurally (e.g., Longworth-Reed *et al*., 2009, Brandewie & Zahorik, 2010). Overall, the dominant effect of reverberation is that it changes the amplitude modulation structure of the signal by acting as a lowpass filter (Houtgast & Steeneken, 1973), smearing the spectral peaks and filling in spectral dips. Stilp et al. (2016) suggested that the adaptation operates by increasing cue weight of the cues, like spectral features, that are robust to reverberation while disregarding the cues that are rendered uninformative in a given context. This mechanism could explain why we observed strong effects for manner perception of stop consonants in both rooms, but less consistent effects for voicing. In terms of reverberant characteristics to which the system might be tuned, it is notable that some previous studies of adaptation to room reverberation manipulated the source-listener distance while keeping the room constant (Watkins et al., 2005, Beeston et al., 2014) while others actually varied the rooms (Brandewie & Zahorik, 2010). Likely, different adaptive processes need to be activated to compensate for the effect of speaker-listener distance within the same room and different ones for the stable distance-independent characteristics of a room. For example, when listening to a conversation of two speakers at different distances within one room, the compensation mechanism needs to adapt within seconds, or faster, as the speakers take turns in a conversation. On the other hand, in real environments people do not switch rooms frequently, thus a room compensation mechanism can easily tune to stable features of the room, like its T_{60}, over tens of seconds and minutes, as observed for speech perception in Longworth-Reed *et al*., 2009), or even on the scales of hours or days for sound localization (Shinn-Cunningham, 2000; Kopco et al., 2004).

The current study has some limitations. First, the number of participants was relatively small. Second, the two rooms used here, while natural and realistic, have higher levels of reverberation than environments in which typical listeners spend the majority of their daily lives. This choice was motivated by our goal of directly examining the effects of reverberation on consonant perception by testing difficult conditions without combining it with the effect of noise masking. However, as has been shown by previous research, the benefit of a consistent carrier is diminished in very strongly reverberant target rooms (Brandewie & Zahorik, 2016). This was the case in our study, where very little improvement in consistent-carrier performance was observed even after removing six of the original consonants that were unaffected by room reverberation. Additionally, the disruptive effects of inconsistent carriers were also very small in the current study when overall percent correct performance is considered. Therefore, it should be noted that our results are likely to generalize to challenging environments such as churches, large lecture halls or concerts halls, but may not explain effects in modestly reverberant environments. Second, although we tested a number of phonetic units and we examined three carrier rooms, we included only two target rooms, with particular acoustic characteristics. Future studies should include additional strongly reverberant environments with different geometry and reverberation time. Similarly, the current study only analyzed 10 consonants preceded by a single vowel. While beyond the scope of this study, we believe that these different sources of variability need to be addressed in future studies in order to obtain more generalizable findings for adaptation of speech perception to reverberation. Finally, the reverberant tails of the Carrier stimuli in the current study extended to the target VCs in the current study. Thus, theoretically, the reverberant carrier VCs might have energetically affected the target VCs. Previous studies artificially removed a portion of the carrier reverberant tail to avoid any artifacts caused by the overlap of reverberation from a preceding VC during target VC presentation. Here, no such modifications were made, as the signal during the presentation of the consonant in the target VC was dominated by the intrinsic masking immediately preceding vowel and its reverberation, not by the preceding VC’s tail, and thus the energetic and binaural de-correlation effects of the carrier VC tails were minimal (Appendix A). An indirect confirmation of this argument is that the detrimental effects of different-reverberant and anechoic carriers are similar in the current study, even though the reverberant tail was only present in the former case.

In sum, the current results partially confirm the results of previous work while, at the same time, pointing to a more complicated picture for the non-initial consonant perception in reverberation (Helfer, 1994; Beeston et al., 2014). We found that for final consonants presented in particularly challenging rooms without masking noise, the effects of a preceding acoustic context on speech perception manifest as a disruption by inconsistent carriers that is as large or larger than an improvement by a consistent carrier. The effects of the preceding carrier affect certain phonemes and phonetic features more than others. Performance for manner of articulation and, partially, for voicing is improved after exposure to a consistent relative to an inconsistent carrier, while place of articulation is not affected. Although previous research has revealed important insights about adaptation to reverberation for speech perception, to our knowledge, this study is the first to show the patterns of improvement and disruption for high level of real-room reverberation without masking, while examining a large set of consonants that represent much of a language’s phonetic repertoire. When considering phonetic features, the current study did not find any effect of carrier duration or carrier/target uncertainty on the adaptive processes studied here. More research is needed to determine how listeners are able to overcome the disruptive effects of inconsistent carriers to understand speech in very challenging listening environments and when moving from one environment to another. Such understanding might also be useful for the development of prosthetic devices for the hearing impaired (Mason & Kokkinakis, 2014; Reinhart et al., 2015).

References

Assmann P., & Summerfield, Q. (2004). The Perception of Speech Under Adverse Conditions. In: Speech Processing in the Auditory System. Springer Handbook of Auditory Research, vol 18. Springer, New York, NY.

Beeston, A. V., Brown, G. J., & Watkins, A. J. (2014). Perceptual compensation for the effects of reverberation on consonant identification: Evidence from studies with monaural stimuli. *Journal of the Acoustical Society of America, 136,* 3072–3084.

Best, V., Ozmeral, E. J., Kopčo, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences, 105,* 13174-13178.

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *Journal of the Acoustical Society of America, 107*, 1065–1066.

Brandewie, E., & Zahorik, P. (2018). Speech intelligibility in rooms: Disrupting the effect of prior listening exposure. *Journal of the Acoustical Society of America, 143,* 3068–3078.

Brandewie, E., & Zahorik, P. (2013). Time course of a perceptual enhancement effect for noise-masked speech in reverberant environments. *Journal of the Acoustical Society of America,* 134, EL265–EL270. doi: 10.1121/1.4816263.

Brandewie, E., & Zahorik, P. (2010). Prior listening in rooms improves speech intelligibility. *Journal of the Acoustical Society of America, 128,* 291–299.

Bronkhorst, A. W. (2000). The cocktail party phenomenon: a review of speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica*, *86*, 117–128.

Danhauer, J. L., & Johnson, C. E. (1991). Perceptual features for normal listeners' phoneme recognition in a reverberant lecture hall. *Journal of the American Academy of Audiology*, 2, 91-8.

Dryden, A., Allen, H., Henshaw, H., & Heinrich, A. (2017). The association between cognitive performance and Speech-in-Noise perception for adult listeners: A systematic literature review and meta-analysis. *Trends in Hearing, 21*.

Garofolo, J., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., & Zue, V. (1993). DARPA TIMIT acoustic-phonetic continuous speech corpus. Linguistic Data Consortium, Philadelphia, PA.

Gelfand, S.A., and Silman, S. (1979). Effects of small room reverberation upon the recognition of some consonant features. *Journal of the Acoustical Society of America,* 66, 22–29.

Hartmann, W. M. (1983). Localization of sound in rooms. *Journal of the Acoustical Society of America, 74,* 1380‑1391.

Helfer, K. S. (1994). Binaural cues and consonant perception in reverberation and noise. *Journal of Speech and Hearing Research, 37,* 429-438. doi: 10.1044/jshr.3702.429.

Houtgast, T., & Steeneken, H.J.M. (1973). The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acustica, 28,* 66-73.

ISO3382 (1997). Acoustics -Measurement of the reverberation time of rooms with reference to other acoustical parameters.

Kopčo N., Schoolmaster, M., & Shinn-Cunningham, B. G. (2004). Learning to Judge Distance of Nearby Sounds in Reverberant and Anechoic Environments. In: Proc. Joint congress CFA/DAGA ’04 22.-25.03.2004 in Strasbourg, France.

Leclère, T., Lavandier, M., & Culling, J. F. (2015). Speech intelligibility prediction in reverberation: Towards an integrated model of speech transmission, spatial unmasking, and binaural de-reverberation. *Journal of the Acoustical Society of America, 137*, 3335–3345.

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication, 52,* 864‑886.

Li, F., Menon, A., and Allen, J. B. (2010). A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *Journal of the Acoustical Society of America, 127*, 2599–2610. doi.org/10.1121/1.3295689.

Li, F., and Allen, J. B. (2011). Manipulation of consonants in natural speech. *IEEE Transactions on Audio, Speech & Language Processing, 19*, 496–504. doi.org/10.1109/TASL.2010.2050731.

Longworth-Reed, L., Brandewie, E., & Zahorik, P. (2009). Time-forward speech intelligibility in time-reversed rooms. *The Journal of the Acoustical Society of America*, 125, EL13-9.

Mason, M., & Kokkinakis, K. (2014). The perception of consonants in reverberation and noise by adults fitted with bimodal devices. *Journal of Speech, Language, and Hearing Research, 57,* 1512–1520. doi:10.1044/2014_JSLHR-H-13-0127.

Mershon, D. H., Ballenger, W. L., Little, A. D., McMurtry, P. L., and Buchanan, J. L. (1989). Effects of room reflectance and background noise on perceived auditory distance. *Perception, 18,* 403-416.

Miller, G. A., & Nicely, P. A. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27,* 338–352.

Nábělek, A. K., & Donahue, A. M. (1984). Perception of consonants in reverberation by native and nonnative listeners. *Journal of the Acoustical Society of America, 75,* 632‑634.

Reinhart, P. N, Souza, P., Srinivasan, N. K., and Gallun, F. (2015). Effects of Reverberation and Compression on Consonant Identification in Individuals with Hearing Impairment. *Ear and hearing, 37(2)*, 144-152.

Resnick, S. B., Dubno, J. R., Hoffnung, S., & Levitt, H. (1975). Phoneme errors on a nonsense syllable test. *Journal of the Acoustical Society of America, 58,* 114.

Ruggles, D., & Shinn-Cunningham, B. (2011). Spatial selective auditory attention in the presence of reverberant energy: Individual differences in normal-hearing listeners. *Journal of the Association for Research in Otolaryngology, 12*, 395–405.

Sagi, E., & Svirsky, M. A. (2008). Information transfer analysis: A first look at estimation bias. *Journal of the Acoustical Society of America, 123,* 2848–2857.

Satoh, F., Sakamoto, S., and Tachibana, H. (2004). "Comparison between the MLS and TSP methods for room impulse response measurement under time-varying condition," in RADS (A satellite symposium on ICA 2004), p. B05.

Schroeder, M. R. (1981). Modulation Transfer Functions: Definition and Measurement. *Acustica, 49*, 179-182.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27,* 379-423.

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification, 12*, 283–299.

Shinn-Cunningham, B. G. (2003). Acoustics and perception of sound in everyday environments. In *Proceedings of the 3rd International Workshop on Spatial Media*, Aizu-Wakamatsu, Japan.

Shinn-Cunningham, B. G. (2000). Learning reverberation: Considerations for spatial auditory displays. In P. Cook (Ed.), *Proceedings of the International Conference on Auditory Display* (pp. 126‑134). Atlanta, GA: Georgia Institute of Technology.

Srinivasan, N. K., & Zahorik, P. (2014). Enhancement of speech intelligibility in reverberant rooms: Role of amplitude envelope and temporal fine structure. *Journal of the Acoustical Society of America, 135*, EL239–EL245.

Srinivasan, N. K., & Zahorik, P. (2013). Prior listening exposure to a reverberant room improves open-set intelligibility of high-variability sentences. *Journal of the Acoustical Society of America, 133*, EL33–EL39.

Suzuki, Y., Asano, F., Kim, H., and Sone, T. (1995). An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses. *Journal of the Acoustical Society of America, 97*, 1119—1123.

Takata, Y., & Nábělek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *Journal of the Acoustical Society of America, 88,* 663‑666.

Traer, J., & McDermott, J. H. (2016). Statistics of natural reverberation enable perceptual separation of sound and space. *Proceedings of the National Academy of Sciences, 113,* E7856–E7865.

Vlahou, E., Seitz, A. R., & Kopčo, N. (2019). Nonnative implicit phonetic training in multiple reverberant environments. *Attention, Perception, & Psychophysics 81,* 935–947.

Watkins, A.J., Raimond, A.P., & Makin, S.J., (2011). Temporal-envelope constancy of speech in rooms and the perceptual weighting of frequency bands. *Journal of the Acoustical Society of America*, *130*, 2777-2788.

Watkins, A. J. (2005). Perceptual compensation for effects of reverberation on in speech identification. *Journal of the Acoustical Society of America, 118*, 249-262.

Watkins, A. J., & Makin, S. J. (2007). Steady-spectrum contexts and perceptual compensation for reverberation in speech identification. *Journal of the Acoustical Society of America, 121,* 257–266.

Yund, E. W., and Buckles, K. M. (1995). Multichannel compression hearing aids: Effect of number of channels on speech discrimination in noise. *Journal of the Acoustical Society of America, 97*, 1206-1223.

Zahorik, P., & Anderson, P.W. (2013). Amplitude modulation detection by human listeners in reverberant sound fields: effects of prior listening exposure. In *Proceedings of Meeting on Acoustics, 19,* 050139.

Zahorik, P., & Brandewie, E. (2016). Speech intelligibility in rooms: Effect of prior listening exposure interacts with room acoustics. *Journal of the Acoustical Society of America, 140*, 74–86.

Zahorik, P. & Brungart, D. & Bronkhorst, A. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica, 91*, 409-420.

Table 1. Phonetic feature classification. Consonants not used as target stimuli are underlined and in bold. All consonants were available as responses in both experiments.

FIGURE CAPTIONS

Figure 1. Acoustic properties of the BRIRs used in the Experiments. Blue and orange symbols are used for rooms R1 and R2, respectively. (A) Time-domain impulse responses from the left ear. (B) Reverberation time (T_{60}), and (C), Clarity index (C_{50}) as a function of frequency.

Figure 2. Across-participant average consonant identification accuracy (%) for (a) Experiment 1 and (b) Experiment 2, plotted as a function of carrier room. Data are averaged across carrier length. Color represents target room. Error bars show standard errors of the mean.

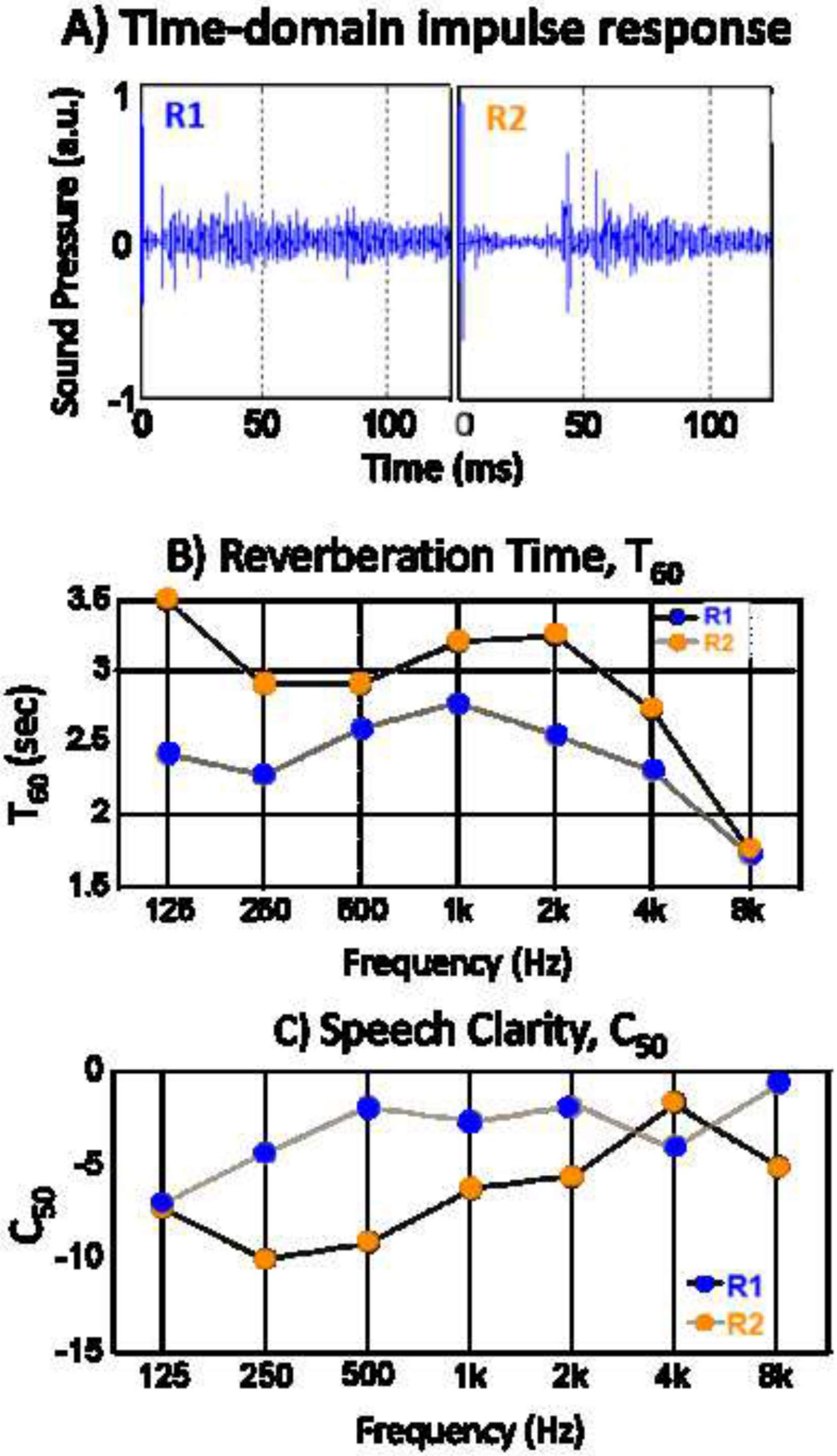
Figure 3. Across-participant average Information Transfer Rate (ITR) as a function of carrier room for manner or articulation, place of articulation and voicing, separately for target room R1 and R2. Asterisks denote significance of difference between Same and Different and Same and Anechoic carrier rooms (* $p < 0.05$; ** $p < 0.01$, one-sided t test).

Figure 4. Across-participant average percent correct feature identification as a function of carrier room. Symbols denote the individual phonetic features. Manner or articulation data (left-hand panel) are plotted separately for target room R1 (blue) and R2 (orange), and as an average across target rooms (avg, black). Voicing data (right-hand panel) show performance for target room R1. Error bars show SEMs. Asterisks denote significance of difference between Consistent and Inconsistent carrier rooms (** $p < 0.01$, one-sided t test).

Figure 5. Consonant confusion matrices. Across-participant average confusion matrices, pooled across the two experiments and carrier lengths. Separate matrices are shown for the Same (left-most panel), Different (middle) and Anechoic (right-most panel) carriers, and for each Target Room (R2, top; R1

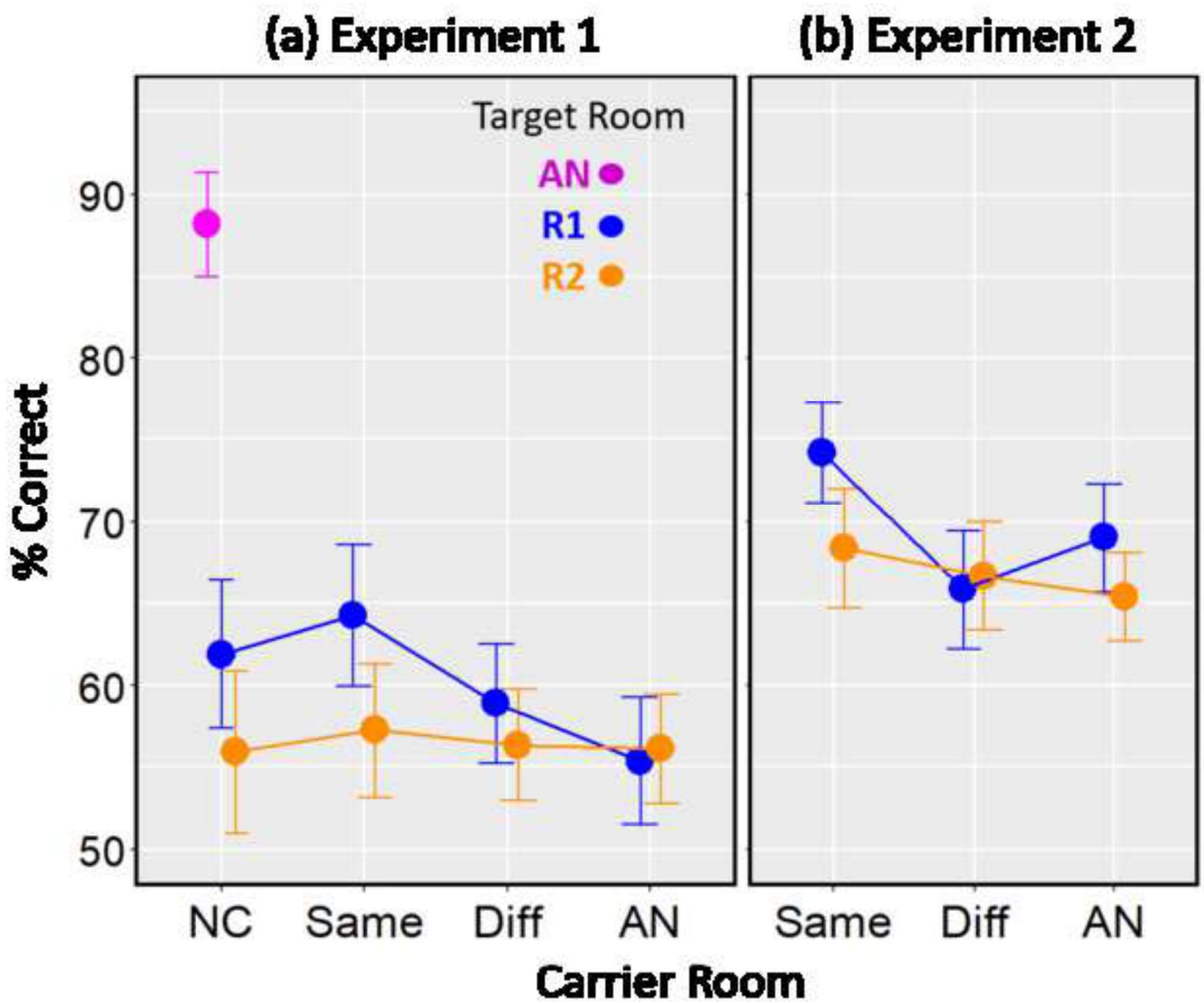
bottom). Columns show the actual speech stimuli that were presented and rows show the response options. Each cell [i, j] shows the percentage of times the consonant in column j was identified as the consonant in row i (empty cells denote 0 percentage). White colors in the tiles represent lower and red colors higher stimulus-response percentage. The legends shown along the vertical and horizontal axes denote consonant classification according to voicing (bold letters for voiced and plain for unvoiced), manner of articulation and place of articulation. Blue frame highlights the cells that represent correct responses.

Figure 1



A) Time-domain impulse response
R1
R2
Sound Pressure (a.u.)
1
0
-1
0
50
100
Time (ms)
B) Reverberation Time, T60
3.5
3
2.5
2
1.5
T60 (sec)
125
250
500
1k
2k
4k
8k
Frequency (Hz)
R1
R2
C) Speech Clarity, C50
0
-5
-10
-15
C50
Frequency (Hz)

Figure 2



(a) Experiment 1
(b) Experiment 2
Target Room
AN
R1
R2
% Correct
90
80
70
60
50
NC
Same
Diff
AN
Same
Diff
AN
Carrier Room

Figure 3

Click here to access/download;Figure;Fig3.tif

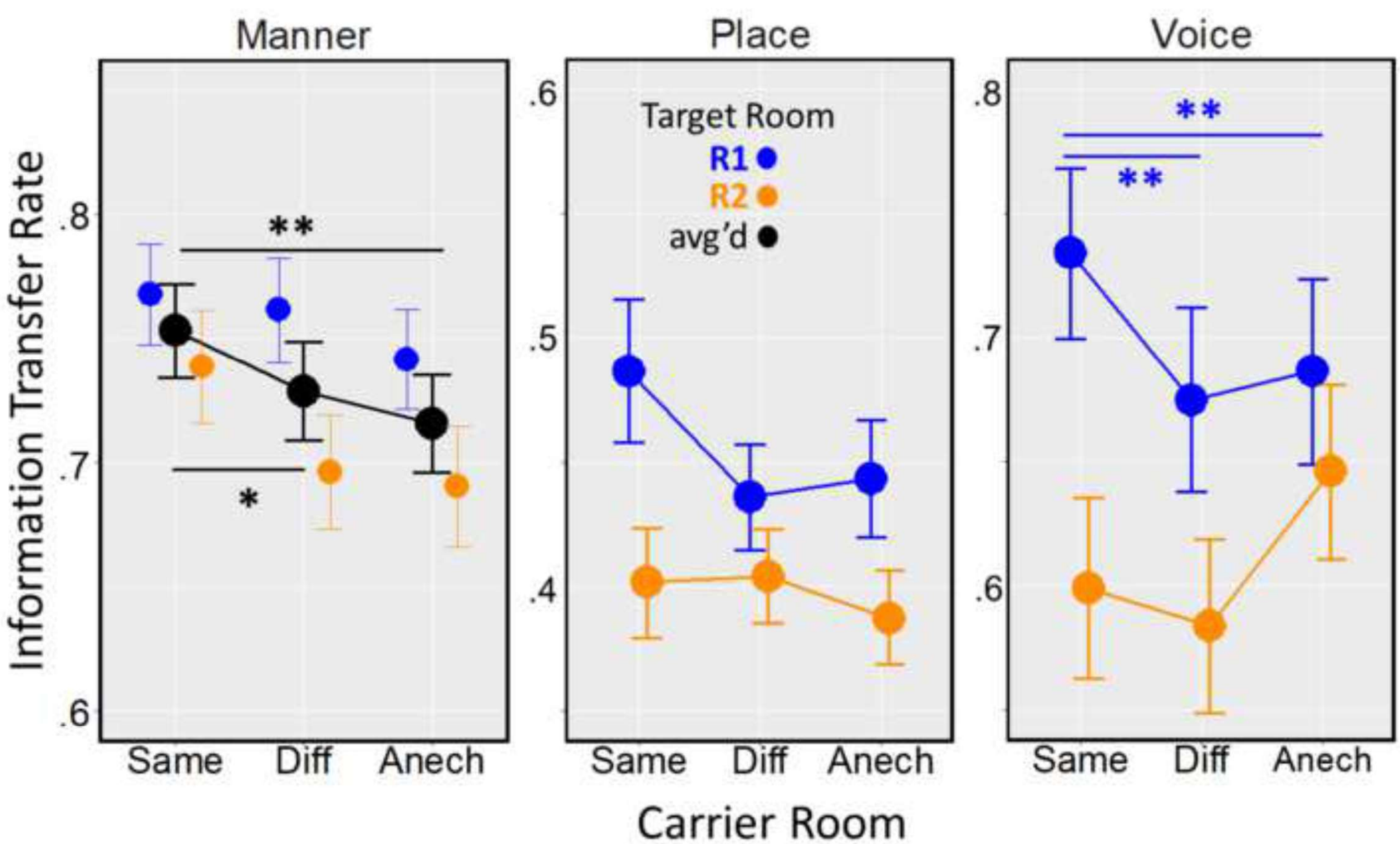


Figure 4

Click here to access/download;Figure;Fig4.tif

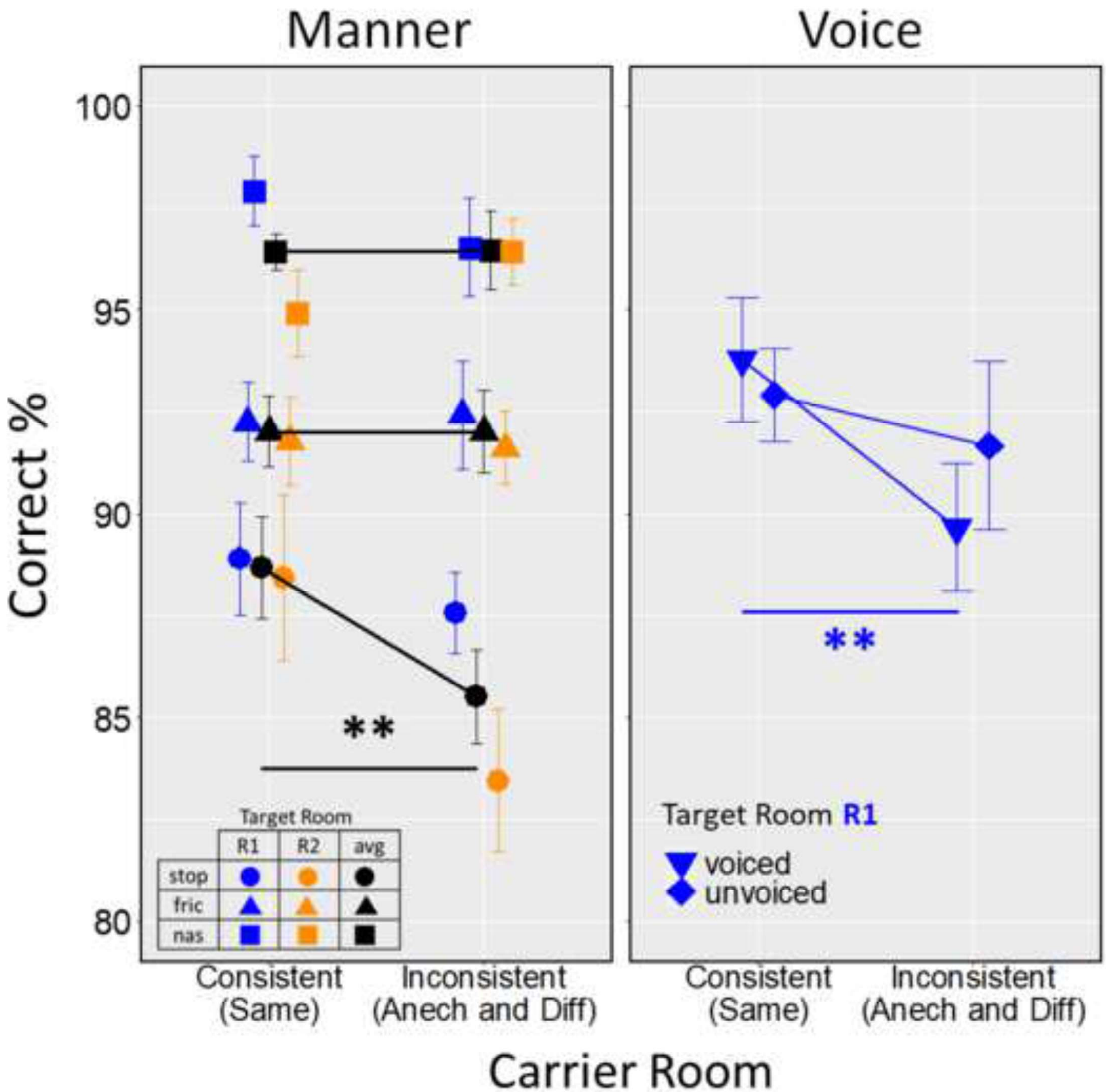


Figure 5

Click here to access/download;Figure;Fig5.tif

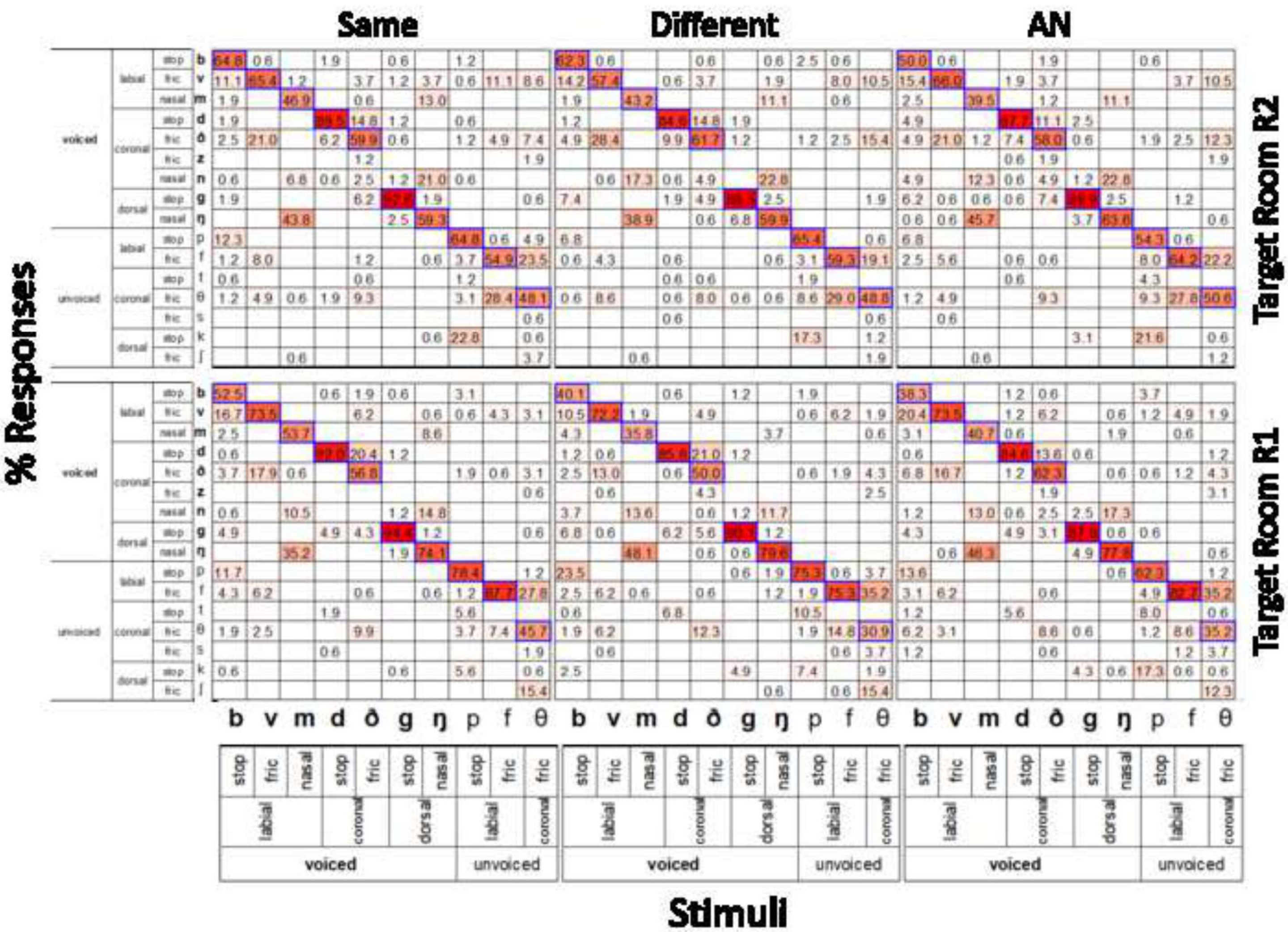


Table 1. Phonetic feature classification. Consonants not used as target stimuli are underlined and in bold. All consonants were available as responses in both experiments.

Feature		Consonants
Manner of articulation	Stop	**k**, **t**, p, g, d, b
	Fricative	f, v, ð, **z**, θ, **s**, **ʃ**
	Nasal	m, **n**, ŋ
Place of articulation	Labial	p, b, m, v, f
	Coronal	d, ð, θ, **t**, **n**, **s**, **z**
	Dorsal	**k**, g, ŋ, **ʃ** (post-alveolar)
Voicing	Voiced	g, d, b, v, ð, m, **n**, ŋ, **z**
	Unvoiced	**k**, **t**, p, f, θ, **s**, **ʃ**

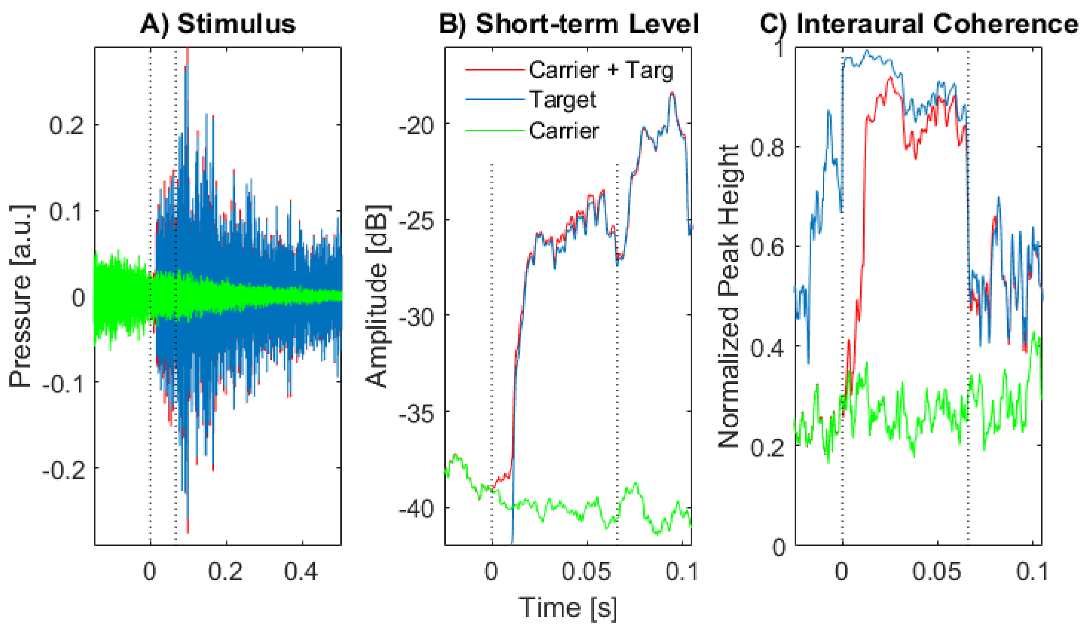


Figure A1: Illustration of acoustic overlap of the Carrier and Target stimuli in simulated reverberation. Target syllable 'ak' presented in room R2 was preceded by a Carrier from room R1. A) Time domain snippet of a left-ear stimulus showing separately a part of the Target (blue) and a reverberant tail of the preceding Carrier (green). Also, shown is the combined Carrier+Target stimulus (red). The direct-sound portion of the Target in this example was presented from time 0 to 0.06 s, indicated by the vertical dotted lines. B) Short-term stimulus level in running 10-ms time windows focused on the Target stimulus. C) Interaural crosscorrelation peak values focused on the Target.

Appendix A: Acoustic analysis of reverberant stimuli

The convolution of the CV tokens with the long BRIRs resulted in stimuli in which the reverberant tails of the Carriers overlapped with the Targets. This is illustrated in Fig. A1, the panel A of which shows an example 0.6-s segment of a stimulus including a Target (blue, starting at time 0 s), the reverberant tail of a preceding Carrier (green), as well as the combined Carrier+Target stimulus (red, identical with the green Carrier for time smaller than 0 s, and identical with, or slightly larger than, the blue Target for time larger than 0 s). The vertical dotted lines in Fig. A1 indicate the onset and offset of the direct portion of the Target stimulus. Two measures were computed to illustrate the effect of Carrier's reverberant tails on the Target. First, panel B shows the RMS power computed in a 10-ms running window for each of the stimuli to assess how much energetic masking form the Carrier reverberation might influence the Target signal. Second, panel C shows the height of the peak of normalized interaural cross-correlation to assess whether binaural properties of the Target signal were influenced by the Carrier reverberation. The effect of Carrier reverberation can be seen by comparing the blue and red lines in panels B and C. It is expected that the carrier reverberant energy would increase the overall level of the signal in panel B, and that it would decrease the correlation in panel C. And these effects are likely to be particularly important for the later portions of the Target which contain the consonants of the VC syllables. Panel B shows that the energetic masking effect was very small for the current stimuli, as the red and blue lines never differ by more than 1 dB, with the exception of the first couple of milliseconds after the onset of the Target, which only contain the vowel portion of the VC. Similarly, Panel C shows a relatively small effect of Carrier reverberation on the Target interaural correlation. First, the blue line shows that the reverberation related to the Target itself causes de-correlation, as the correlation in the second half of the stimulus is approximately 0.05 lower than in the first half, where it is equal to nearly 1. The red line is lower than the blue line particularly in the first half of the Target presentation. In the second half, which contains the consonant, the decrease is much smaller, less than 0.02, i.e., smaller than the decrease due to intrinsic Target-related reverberation.

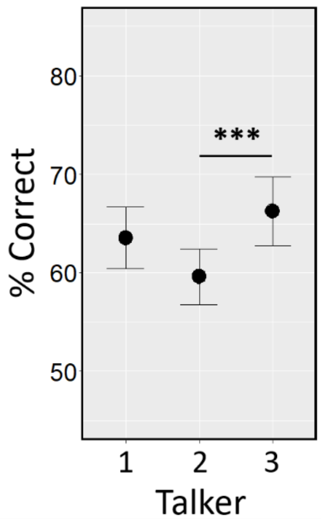


Figure B1. Across-participant average consonant identification accuracy (%), plotted as a function of the three different talkers used in the experiments reported below. Data are averaged across carrier length, carrier and target rooms and experiments. Error bars show standard error of the mean. Asterisks denote significance of difference; *** $p < 0.001$, two-sided t-test).

Appendix B: Talker effects

Prior to running our main analysis, we examined performance across the different talkers to ensure that there were no idiosyncratic voice characteristics that would substantially affect speech intelligibility (i.e., making the task too easy or too difficult).

Fig. B1 shows across-participant-averaged percent correct responses as a function of the different talkers, collapsed across the two experiments, carrier lengths and the different carrier and target rooms. Overall, performance varied from approximately 60% (SE 2.86) for the most difficult-to-understand talker to 66% (SE 3.5) for the easiest-to-understand one. Intelligibility was well above chance for all three talkers ($p < 0.001$ in all cases), without reaching ceiling levels. A one-way repeated measures ANOVA showed a significant talker effect ($F(2,34) = 7.38$, p

$= 0.002$, $\eta_p^2 = 0.30$). Post-hoc pairwise tests, corrected for multiple comparisons, revealed that the third talker was significantly easier-to-understand than the second talker ($t(17) = -4.63$, $p = 0.0002$). Since the three talkers were evenly distributed across all the examined factors (carrier length, carrier room, target room and experiment), thus minimizing any potential bias, in the analyses in the main text the data are collapsed across the different talkers.

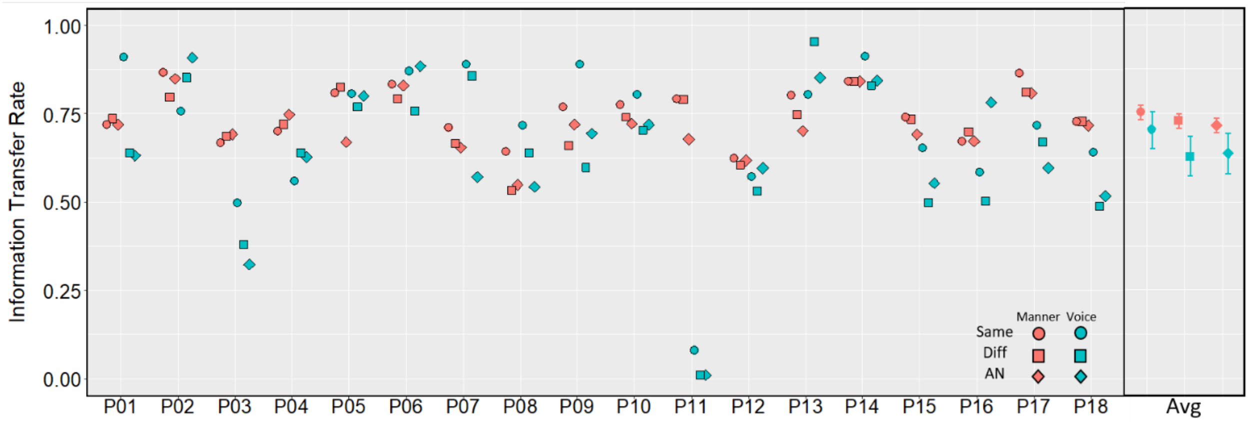


Figure C1. Individual participant (left) and across-participant-averaged (right) ITR scores for the significant results shown in Fig. 3. Data for manner of articulation are averaged across target rooms R1 & R2. Data for voice include the R1 target room. Carrier rooms and features are shown by different shapes and colors, respectively.

Appendix C: Across-subject variability in ITR performance

Following our main analysis on participants' ITR scores across manner of articulation, place of articulation and voicing (see section "Effects of carrier room, carrier length, and carrier/target uncertainty on phonetic features"), here we examine individual variability in performance for the two conditions where significant effects were observed. Figure C1 shows individual participants' performance for (a) Manner of articulation (averaged across target rooms R1 & R2), and, (b), Voicing, for target room R2. In agreement with previous studies (e.g., Brandewie & Zahorik, 2010; Brandewie & Zahorik, 2018), we found substantial individual differences in participants' performance. Overall, many (but not all) participants benefit from exposure to a consistent carrier compared to the two inconsistent carriers. Also, while both inconsistent carriers appear to degrade performance for most participants, there is significant variability in how different participants are affected and there is no clear pattern of differential effects across the Different and Anechoic carrier.

Appendix D: ITR performance with all 16 consonants included in Exp. 1

Here, we present the participants' ITR performance when all consonants used as target in Experiment 1 are considered. I.e., the 6 consonants (k, t, n, z, s and ʃ) for which performance was above 90% across all Exp. 1 tested conditions are included here, while they were left out in the section "Effects of carrier room, carrier length, and carrier/target uncertainty on phonetic features" (see "Speech Material" in Methods and Table 1). Note that, in Exp. 2, only 10 consonants were presented while the participants still could respond that they heard any of the 16 consonants of the full set.

Table D1 shows the across-participant average ITR values obtained with the two datasets (16 consonants or 10 consonants) for each class of features (manner of articulation, place of articulation, and voicing) as a function of carrier room (Same, Different and Anechoic carrier).

Table D1. Across-participant-average ITR obtained with data collapsed across the two experiments (upper section) and separately for Experiment 1 (lower section) if all 16 consonants were considered in Exp. 1 vs. if only 10 consonants were considered in Exp. 1 (the same 10 consonants were always considered in Exp. 2).

		Experiments 1 & 2		Experiment 1	
Feature	Carrier Room	16 consonants in Exp. 1 (Mean±SE)	10 consonants in Exp. 1 (Mean±SE)	16 consonants (Mean±SE)	10 consonants (Mean±SE)
Manner of Articulation	Same	0.79±0.08	0.76±0.10	0.79±0.07	0.72±0.08
	Diff	0.77±0.07	0.73±0.10	0.77±0.08	0.69±0.11
	Anech	0.76±0.08	0.72±0.10	0.77±0.09	0.69±0.11

Place of articulation	Same	0.51±0.13	0.44±0.14	0.55±0.13	0.42±0.15
	Diff	0.48±0.12	0.42±0.10	0.55±0.11	0.42±0.10
	Anech	0.48±0.11	0.42±0.11	0.52±0.10	0.39±0.10
Voicing	Same	0.71±0.19	0.65±0.25	0.68±0.22	0.57±0.29
	Diff	0.66±0.17	0.60±0.22	0.65±0.20	0.53±0.25
	Anech	0.69±0.18	0.63±0.23	0.67±0.20	0.56±0.26

In both analyses, performance for the Same carrier tended to be higher than for the two inconsistent carriers (on average, ITR improved by 0.03 for Experiments 1 & 2 and by 0.02 for Experiment 1). These results show that, as expected, including the 6 consonants in Experiment 1 raises the ITR values (by 0.04 for manner, 0.06 for place and 0.06 for voicing).

To examine whether the magnitude of adaptation to reverberation (improvement in the Same carrier compared to the two inconsistent carriers) is affected, we ran a mixed ANOVA on participants' ITR values, with carrier/target uncertainty (experiment) as a between-participants factor and with carrier room (Same, Different, Anechoic), carrier length (2 vs. 4 VCs) and target room (R1, R2) as within-participant factors, separately for each feature (manner of articulation, place of articulation, voicing; see section "Effects of carrier room, carrier length, and carrier/target uncertainty on phonetic features").

For manner of articulation (Fig. D1; left-hand panel), consistent with our previous analysis, we found a significant main effect of Carrier Room ($F(2,32) = 5.05$, $p = 0.018$, $\eta_p^2 = 0.23$), while Experiment, Target Room, and Carrier Length were not significant either as main effects nor as interactions (all p's > 0.10). One-sided post-hoc comparisons (adjusted with the Holm-Bonferroni correction) between Same vs. Different and Same vs. Anechoic carriers showed a significant difference between Same and Different ($t(71) = 1.97$, $p = 0.027$) and Same and Anechoic ($t(71) = 2.78$, $p = 0.003$).

For place (Fig. D1; right-hand panel), there was a significant main effect of experiment ($F(1,16) = 6.22$, $p = 0.024$, $\eta_p^2 = 0.28$), owing to improved performance in Experiment 1. There was also a main effect of target room ($F(1,16) = 17.86$, $p < 0.001$, $\eta_p^2 = 0.498$), owing to better overall performance in target room R1 compared to target room R2. There were no other main effects or interactions (carrier type: $F(2,32) = 3.13$, $p = 0.057$; all other p's > 0.19). The overall improved performance on Experiment 1 (which included the six consonants with high identification performance) and on the less reverberant target room was expected and was not a main interest of this study, thus we did not follow up on this result.

For voicing (Fig. D2), the results were somewhat less consistent with those reported in the main text. The ANOVA showed a significant 4-way interaction including all factors ($F(2,32) = 4.67$, $p = 0.0166$, $\eta_p^2 = 0.23$). Partial ANOVAs showed no significant main effects or interactions for Target Room R2 (Carrier Length: $F(1,16) = 3.97$, $p = 0.064$; all other p's > 0.19), in agreement with our previous results. For target room R1, there was a significant main effect of Carrier Room ($F(2,32) = 6.85$, $p = 0.0071$, $\eta_p^2 = 0.30$) and a significant Carrier Length X Experiment X Carrier Room interaction ($F(2,32) = 9.01$, $p = 0.0011$, $\eta_p^2 = 0.36$). To examine this more closely, we ran post-hoc pairwise comparisons (adjusted with the Holm-Bonferroni correction) for each Experiment and Carrier Length. To keep the number of comparisons low we compared performance on the Same carrier with performance on the inconsistent carriers (0.5*[Anechoic + Different]). T-tests showed a significant difference for the short carrier length in Experiment 1 ($t(8) = 3.77$, $p = 0.003$; top left panel in Fig. D2) and for the long carrier length in experiment 2 ($t(8) = 4.62$, $p = 0.0009$; bottom right panel in Fig. D2).

Overall, the results for manner of articulation and place of articulation are very similar across the two datasets (compare Fig. D1 with Figure 3 in main text). For voicing, however, contrary to our previous results, the benefit of the same carrier in target room R1 was not consistent across the different carrier lengths and experiments. However, due to the small number of participants for the complex study design, the significant higher-order interactions reported here should be examined with caution, especially given that they might be driven by the different number of consonants based used for the phonetic feature estimation in the two experiments.

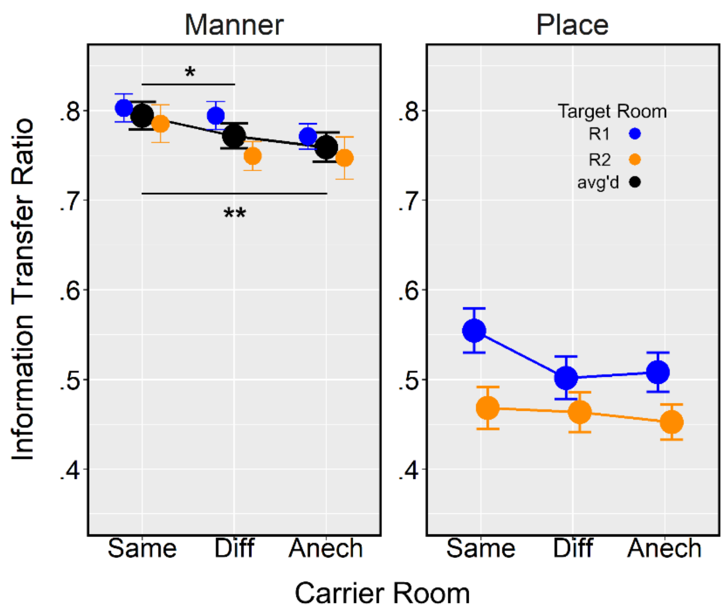


Figure D1. Across-participant average Information Transfer Rate (ITR) for manner of articulation (left) and place of articulation (right) as a function of Carrier Room, separately for target room R1 and R2, averaged across Carrier Length and Experiment. Error bars show SEM's. Exp. 1 data are based on 16 consonants, Exp. 2 data on 10 consonants. Asterisks denote significance of difference between Same and Different and Same and Anechoic carrier (*p < 0.05, and **p < 0.01, one-sided t test).

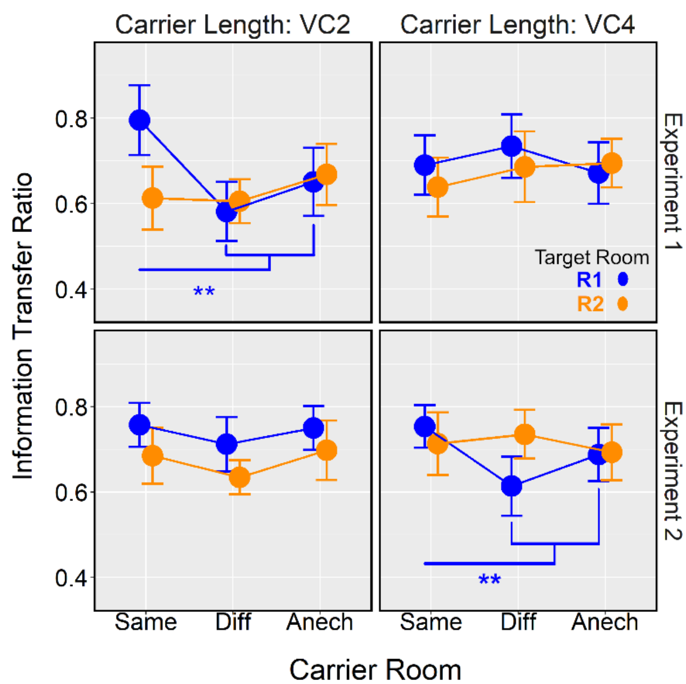


Figure D2. Across-participant average Information Transfer Rate (ITR) for voicing, as a function of Carrier Room, separately for target room R1 and R2, for each Carrier Length and Experiment. Exp. 1 data are based on 16 consonants, Exp. 2 data on 10 consonants. Error bars show SEM's. Asterisks denote significance of difference between Same and Inconsistent (Different and Anechoic) carrier (**p < 0.01, one-sided t test).