Many studies of sound perception often assumed that our auditory sensory processes are relatively static, rather than plastic. However, in everyday environments, we naturally and fluidly compensate for interfering effects of background noise and room reverberation. In order to investigate how listeners calibrate auditory perception to such acoustic interference, a listening experiment was performed to measure the effect of compensation on auditory perception.

Target words in sentences that provided strong contextual cues could be heard. The target word was one of a vowel pair embedded as the second word in one of three sentence types. The neutral sentence provided little context for the word. Sentences recorded with a Mid-western accent were convolved with different localization blur? A series of tests was conducted comparing the effect of reverberation on the identification of consonants. Test sounds were generated by convolving two types of binaural room impulse responses (BRIRs) measured in large real rooms with speech tokens. As a control condition, pseudo-anechoic BRIRs with negligible reverberation energy were used. Listeners were asked to identify the consonant in a vowel-consonant target. The target was preceded by a carrier phrase consisting of vowel-consonant pairs from the same talker. In some cases, the target and carrier phrase were processed by the same BRIRs, while in others the BRIR’s processing target and carrier differed. Consistent effect of calibration was observed in one of the simulated rooms, but not in the other, suggesting that the ability to compensate for the effects of reverberation depends on the specific pattern of reverberation produced in a given room. [Work supported by AFOSR and NSF.]

Sentences recorded with a Mid-western accent were convolved with head-related impulse responses that included different room reverberation conditions. The stimuli were presented binaurally through headphones in an echo-attenuated chamber and subjects (n = 23) typed the sentences they heard. The target word was one of a vowel pair (cattle/kettle, jam/gem, gas/guess, past/pest) embedded as the second word in one of three sentence types. The neutral sentence provided little context for the word. Target words in sentences that provided strong contextual cues could be congruent or incongruent with the expectations of the subject, for example, “The cattle/kettle grazed in the meadow.” Subjects made significantly more errors in the incongruent sentences compared to the neutral (Wilcoxon = 3.572 p < 0.05) or congruent sentences (Wilcoxon = 3.36 p < 0.05). When the target word was in a congruent sentence subjects performed equally well in reverberant or pseudo-anechoic conditions (Wilcoxon = 1.298) but they made more errors in the reverberant condition for both neutral (Wilcoxon = 3.359, p < 0.05) and incongruent sentences (Wilcoxon = 2.241, p < 0.05). Results suggest that reverberation may cause listeners to rely more heavily on linguistic context to determine word meaning. [Work supported by NOHR, AFOSR.]

Perceptual compensation for reverberation is observed when the reverberation is applied to a test word (from a “sir”-to-“stir” continuum) and its context (e.g., “next you’ll get to click on”) are varied independently. Increasing reverberation in test words decreases listeners’ “stir” responses, as reverberation fills the gap that cues the [t]. Compensation occurs when the context’s reverberation is commensurately increased, and “stir” responses increase back to the level found with minimal test-word reverberation. Compensation is strongest with speech contexts but also occurs with some noise-like contexts, including “signal-correlated noise” that has the wideband temporal envelope of the original speech. Also effective is a wideband noise that is given the temporal envelope seen at the output of a single auditory filter in response to speech. A narrow-band version of this audiatory-filtered noise is not effective, but when contexts are made by summing of three or five of these bands, their effectiveness increases correspondingly. Compensation appears to be informed by the “tails” that reverberation adds at offsets, so it merely requires contexts with suitable temporal-envelope fluctuations. However, effects seem confined to the context’s frequency region, so the crucial offsets need to be in a wide range of frequency bands. [Work supported by EPSRC.]
The intelligibility of speech transmitted from closed offices to adjacent spaces is strongly affected by the signal-to-noise ratio at the receiver position and the acoustical characteristics of the spaces involved. Previous studies have suggested that the effect of room acoustics on speech intelligibility in closed offices and rooms is negligible and can be ignored (as with intelligibility quantifiers such as the articulation index). The purpose of this study is to show that in conditions of very low signal-to-noise (i.e., when high speech privacy is a primary concern), the influence of room acoustics rises dramatically. To this end, multiple subjects were given tests of speech intelligibility in simulated sound fields. Speech samples were presented to subjects with seven levels of signal-to-noise and four different reverberation times. The results from these tests show that as reverberation time increases, speech intelligibility decreases much more sharply for very low (< 8 dB) signal-to-noise situations than in higher (+10 dB) signal-to-noise situations. This suggests an important relationship between room acoustics and speech privacy/security.


This study evaluates a preprocessing approach for reducing reverberation effects when the input signal is not ideal, dry speech, but rather a realistic speech signal picked up by a close microphone in a room. And this study shows how the situation affects the input and the further approach compared to a dry signal as the input. Steady-state suppression, as this study shows how the situation affects the input and the further application effects when the input signal is not ideal, dry speech, but rather a speech signal before it is radiated from loudspeakers. A lecture was simulated in two different halls (reverberation times of 1.2 and 1.8 s) in which public address systems were installed. The simulation software CATT-Acoustic was used and impulse responses were calculated for the input to the preprocessing approach and for a listener position. Stimuli for a syllable identification test were prepared by convolving speech signals with the calculated impulse responses. Speech signals were given with and without steady-state suppression. The inclusion of natural and electroacoustical impulse responses makes the study of steady-state suppression more realistic and tests its robustness. [Work supported by JSPS (1769111).]