Exposure to Consistent Room Reverberation Facilitates Consonant Perception

Norbert Kopčo[1,4], Eleni Vlahou[1,2], Ueno Kanako[3] & Barbara Shinn-Cunningham[4]

[1]Institute of Computer Science, P. J. Šafárik University, Košice, Slovakia, [2]Department of Psychology, University of California, Riverside, [3]School of Science and Technology, Meiji University, Tokyo, Japan, [4]Center for Computational Neuroscience and Neural Technology, Boston University

1. BACKGROUND AND MOTIVATION

To achieve accurate speech communication in everyday conditions, the auditory system has to be able to adapt to different reverberant environments that distort the speech signal. Several recent studies showed that consistent exposure to a particular room facilitates speech perception both for a limited set of speech sounds (Beeston et al, 2014) and for sentences with rich lexical information (Srinivasan & Zahorik, 2013).

Current study:
We present the results of two experiments, investigating the effect of **room consistency** on phoneme perception
- using a **wide range** of consonants, representative of a language's phonetic repertoire,
- using nonsense vowel-consonant (VC) syllables, thereby factoring out **lexical influences** on perceptual compensation for reverberation.

Main questions:
- Does **prior exposure** to **consistent reverberant** environments **improve** phoneme perception?
- Is amount of perceptual **degradation** due to **different-room** carrier the same as that due to **anechoic** carrier?
- Which **phonetic features** are affected by adaptation to reverberation and how?

2. METHODS

Participants:
14 subjects participated in **Experiment 1**, 10 subjects in **Experiment 2** (6 participated in both). All participants were native speakers of American English.

Phonetic stimuli:
16 consonants (k, t, p, f, g, d, b, v, ð, m, n, ŋ, z, θ, s, and ʃ) were used, each preceded by vowel /a/. For each VC, 3 tokens were spoken by 3 talkers (two males, one female). In **Exp. 1**, all 16 consonants were used. In **Exp. 2** (and in phonetic feature analyses), 6 consonants, for which performance was at ceiling in Exp. 1, were excluded (k, t, n, s, ʃ and z). However, participants could still respond using all 16 consonants.

Simulated rooms:
BRIRs from two different large rooms were used, denoted as **R1** and **R2** (Figure 1). Both rooms exhibited **high levels of reverberation**. **R1** was measured in an elliptical church (distance from sound source 12 m). **R2** was measured in a large concert hall (distance from sound source 33 m). A 5-ms window was applied to the direct portion of the **R1** BRIR to remove most reverberant energy, generating "pseudo-anechoic" (**an**) BRIR. The resulting three BRIRs (**R1**, **R2**, and **an**) were equalized for overall energy. Because of its elliptic room shape, **R1** has a large echo around 60 ms after the direct sound, seen more prominently in the 500 Hz octave band (Fig. 1a). T_{60} values larger for **R1** than **R2** (Fig. 1b).

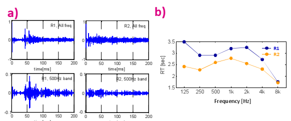


Figure 1. Acoustic characteristics of the BRIRs
a) Early time-domain portions of the responses in one ear. **b)** Reverberation time (T_{60}) obtained by the integrated impulse response method for each frequency band.

Experimental design:
On each trial, listeners were exposed to VC syllables and had to report the **final (target)** consonant (Figure 2). On most trials, a carrier consisting of **2-VC** or **4-VC** syllables preceded the target (in **Exp. 1**, control trials with no carrier were also included). VCs within a trial were separated by an inter-stimulus interval of 0.8 s. In **Exp. 1**, different length carriers (**no**, **2-VC** or **4-VC**) were randomly presented in different trials, whereas in **Exp. 2**, the **2-VC** or **4-VC** carriers were presented in separate blocks. In **same** trials, the carrier and target had the same reverberation (**R1** or **R2**). In **diff** trials, the carrier and the target contained different reverberation (Carrier:**R1**-Target:**R2** or Carrier:**R2**-Target:**R1**). In **an** trials, the carrier consisted of anechoic speech and the target was either **R1** or **R2**.

	Carrier (2-VC or 4-VC)				Target
Same Carrier-Target Reverberation (**R1** or **R2**)			av	ag	ap
	at	ap	ab	ag	að
Different Carrier-Target Reverberation (Carrier: **R1**-Target:**R2**, Carrier:**R2**-Target:**R1**)			ab	ad	af
	av	ak	aŋ	ag	ag
Anechoic Carrier, Target Reverberation **R1** or **R2**			aθ	að	aŋ
	ap	ak	av	ag	aθ

What was the **final** consonant?

Figure 2. Experimental design

Analysis:
Percent correct responses were rau transformed and entered into repeated measures ANOVA. Data were averaged across talkers. In all figures, error bars are SEMs. Consonant confusions were calculated for different phonetic features (**manner**, **place**, and **voicing**), separately for the different carrier characteristics (**same**, **diff** and **an**). From each confusion matrix, Information Transfer Ratio was computed as ITR= H(X:Y)/H(X), where H(X:Y) is the mutual information of X and Y, and H(X) is the self-information (entropy) of X).

3. EXPERIMENT 1

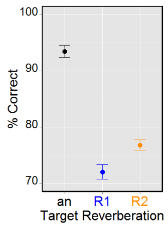


Figure 3. Baseline performance
Percent correct consonant identification as a function of target reverberation (**an**, **R1** and **R2**) for syllables presented *without* preceding carrier in **Exp. 1**.

In **anechoic** room, performance near ceiling.
Reverberation degraded intelligibility
- in both rooms,
- more in **R1** than in **R2**.

No-carrier intelligibility was dramatically **reduced** in strong **reverberation.**

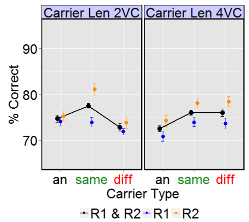


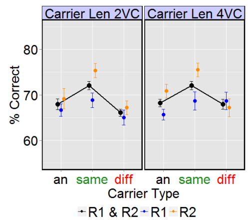
Figure 4. Effect of carrier type when VC length varies
Percent correct consonant identification as a function of carrier type, and carrier length. Data plotted either as an average across target reverberation (**R1&R2, black**), or separately for **R1** and **R2**.

Performance was
- better with **same** carrier than in baseline (compare to Fig. 3),
- better with **same** carrier than with non-matching carriers for all conditions except for **4-VC diff** carrier (significant interaction of carrier length and type, $F_{1.54,20.07} = 7.98$, $p = .0049$),
- similar across the two target reverberations (interaction between carrier type and target reverberation not significant, $F_{1,13} = 0.24$),

Improved performance with **diff 4-VC** carrier *re.* other non-match carriers might have been caused by lower uncertainty about the target temporal position. However, the lower uncertainty is also in **an 4-VC** condition.

Performance was better for targets preceded by carrier with **matching *re.* non-matching** reverberation, except for the **4-VC diff** carrier.

4. EXPERIMENT 2



Experiment 2 tested whether target temporal position uncertainty caused the lack of degradation for **diff 4-VC** in **Exp. 1**. **4-VC** and **2-VC** carriers were presented in separate blocks. Only 10 consonants were used.

Figure 5. Effect of carrier type when VC length is fixed
For legend see Figure 4.

Performance was
- better with **same** carrier than with **non-matching** carriers for all conditions (interaction of carrier length and type non-significant, $F_{2,18} = 0.72$),
- similar across the two target reverberations, even though effects stronger for target **R2** (significant interaction between carrier type and target reverberation, $F_{1.52,13.64} = 8.51$, $p = .0063$),

So, lack of degradation in **Exp. 1 diff 4-VC** might have been caused by uncertainty about the target temporal position. However, no significant difference between **Exp. 1 & 2** results was found.

Exposure to **matching** reverberation facilitates consonant perception for both short and long carriers, independent of carrier type, when target temporal position uncertainty is eliminated.

5. PHONETIC FEATURE ANALYSIS

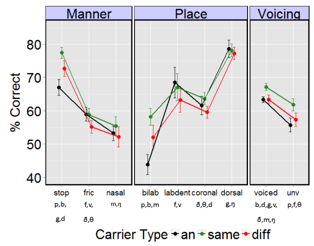


Figure 6. Phonetic feature analysis
Mean performance on the consonant identification task as a function of phonetic feature and carrier type, separately for **manner**, **place** of articulation and **voicing.** Data averaged across experiments, carrier length, and target reverberation.

Carrier type interacts with
- **manner** ($F_{4,68}$=7.47, p < .0001), strongest effect for stops,
- **place** ($F_{6,102}$=5.46, p = .0001), strongest effect for bilabials.

No interaction with **voicing**.
Same better than **diff** for all features. **An** performance less consistent with **same** or **diff**.

Degradation due to non-matching carrier:
- is **consistent** across phonetic categories for **diff**-room reverberant carrier,
- **varies** from category to category for **anechoic** room carrier.

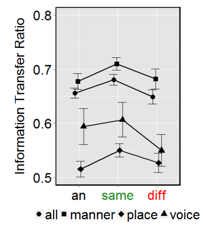


Figure 7. Information Transfer Ratio
ITF for all consonants (**all**), and for each phonetic feature (**manner**, **place** and **voice**), as a function of carrier type (**an**, **same**, **diff**).

For **all**, results similar to % correct (Figure 5).
Large differences between ITRs among features:
- **manner** > **voice** > **place**.

Non-matching carrier:
- degrades performance for all features,
  - for **diff**, largest degradation in **voice** (t-test, p<0.0005),
  - for **an**, largest degradation in **place** (t-test, n.s.).

Which feature is affected the most by non-matching carrier depends on carrier reverberation type. It is voicing (for **diff**) or place (for **anech,** n.s.).

6. CONCLUSIONS

- Strong reverberation degrades baseline speech intelligibility.
- Exposure to carrier with consistent (**same**) reverberation results in improved performance (*re.* baseline).
- Exposure to non-matching carrier causes degradation in performance (*re.* **same**). This degradation is strong and independent of VC-length, target reverberation, or target position uncertainty when the carrier is **anechoic**. It is also consistent across target reverberation and position uncertainty when carrier is reverberant and short (**2-VC**).
- When carrier is long (**4-VC**) and has **diff**-room reverberation, the degradation is less consistent across the experiments (observed in **Exp. 2** but not **Exp. 1**). Uncertainty about the temporal position of the target in **Exp. 1** may be the cause, as blocking runs by VC-length in **Exp. 2** resulted in degradation of performance even in this condition (however, statistical analysis comparing the two experiments did not show a significant difference).
- Phonetic feature analysis based on both **% correct** and **ITR**s showed improved performance for the **same** condition (*re.* **an** or **diff**). However, the amount of degradation due to non-matching room depends on carrier type and method of analysis: in terms of % correct, largest degradation was observed for bilabials and **anech** carrier; in terms of **ITR**, largest degradation was in voicing for **diff**-reverberant room.
- Additional experiments and analysis are needed to explain the differences in results.

REFERENCES

Beeston, A. V., Brown, G. J., & Watkins, A. J. (2005). 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.
Srinivasan, N. K., & Zahorik, P. (2013). Prior listening exposure to a reverberant room improves open-set intelligibility of high-variability sentences. *JASA, 133,* EL33—EL39.
Ueno, K., Kopčo, N., & Shinn-Cunningham, B. (2005). Calibration of speech perception to room reverberation. In *Proceedings of the forum acusticum conference*. Budapest.

ACKNOWLEDGMENTS

Work supported by
- CELEST, a National Science Foundation Science of Learning Center (NSF SMA-0835976).
- Slovak Research and Development Agency, SRDA, contract No. APVV-0452-12.
- SOFOS, ITMS: 26110230088, supported by the Research & Development Operational Programme funded by the ESF of the European Union.