Adaptation to Room Reverberation in Nonnative Phonetic Training

MGH/HST Athinoula A. Martinos Center for Biomedical Imaging

Eleni L Vlahou[1], Aaron Seitz[1] and Norbert Kopčo[2,3]

[1]University of California, Riverside, [2]Šafárik University, Košice, Slovakia, [3]Harvard Med. School – Martinos Center, MGH, Charlestown, MA, USA

1. BACKGROUND AND MOTIVATION

Speech communication often occurs in adverse listening conditions, such as noisy and reverberant environments. Room reverberation distorts the speech signal and hampers intelligibility, especially for nonnative listeners (Nábělek and Donahue, 1984). Native listeners benefit from prior exposure to consistent reverberation (Brandewie & Zahorik, 2010, Ueno et al. 2005), but less is known about the patterns of interference and adaptation to room reflections for nonnative listeners, during the acquisition of novel phonetic categories.

Current study: Here, we address these issues by training different groups of adults on a difficult nonnative phonetic contrast in a virtual acoustic environment

- using speech stimuli presented in **anechoic** space or in anechoic and in **simulated room** environments,
- crossed with **explicit** and **implicit** training.

Main questions:

- Does exposure to **room** environments facilitate or interfere with learning of new speech sounds?
- Will subjects trained with a single **anechoic** environment be able to differentiate sounds presented in **reverberant** environments?
- Will there be different patterns of interference/adaptation to room reverberation for **explicit** and **implicit** training?
- Will learning generalize to new speech tokens, new talkers, and/or new rooms?

2. METHODS

Subjects and experimental conditions: Four groups (5 subjects each) were trained and tested on a difficult nonnative phonetic contrast. Training was performed **explicity** or **implicitly**, with sounds presented in **anechoic** or in **reverberant** environments. 5 more subjects were tested and re-tested with the same material over a period of 1 week, without training in between (**no-training control group**).

Phonetic stimuli and simulated room reflections: We used the Hindi **dental-retroflex** phonetic distinction (Werker & Tees, 1984). Phonetic stimuli were CV syllables, starting with dental/retroflex sounds and followed by the long [:i], from two native Hindi speakers. There were 10 different tokens/phonetic category/speaker. Each token was convolved with 4 different room reverberations termed "bathroom" (ba), "ping-pong" (pg), "cafeteria" (ca), "office" (of) and an anechoic environment (an; details in Kopčo and Shinn-Cunningham, 2011 and Kayser et al., 2009).

Training: Experimental groups were trained with **5 tokens/phonetic category** ("**Trained tokens**") from **1 speaker only** ("**Trained voice**", counterbalanced across participants) in 4 daily sessions, 45 min/session. In each session:

- Groups trained with **1 room** (**Explicit-1-Room** and **Implicit-1-Room**) were trained with sounds presented in **anechoic** space (600 trials/session).
- Groups trained with **3 rooms** (**Explicit-1-Room** and **Implicit-3-Room**) were trained with sounds presented in anechoic space and two reverberant environments ("bathroom" (ba) and "ping-pong" (pg), 200 trials/room in 40-trial randomly interspersed blocks). Implicit and explicit training paradigms are illustrated in Figures 1 and 2, respectively.

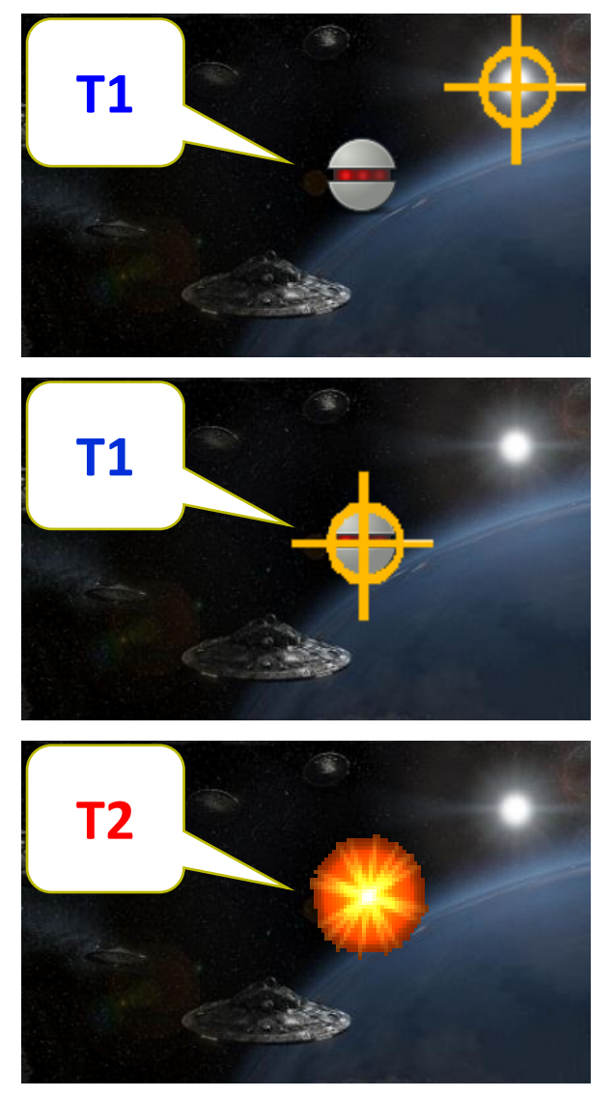

Figure 1. Schematic representation of the **implicit training** paradigm. Implicit training employed a videogame which promoted stimulus-reward contingencies (Seitz & Watanabe, 2005). In each trial, a character appeared on the screen and produced two identical Hindi sounds from one category ("**T1**"; retroflex for half participants). If the player managed to shoot the character, it produced two identical sounds from the other category ("**T2**"). As the player got better, characters were moving faster.

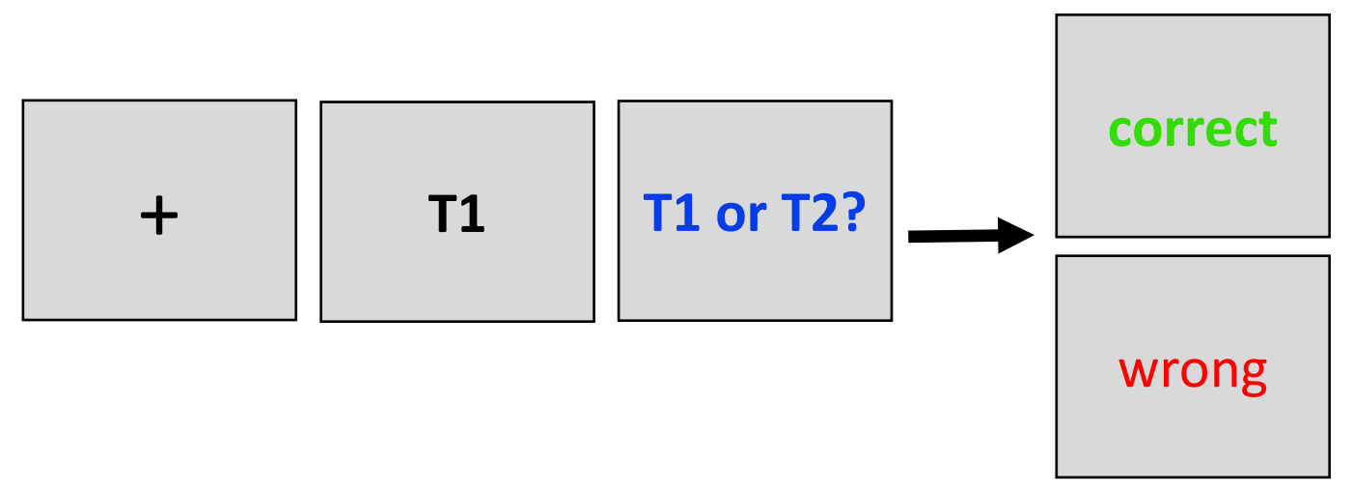

Figure 2. Schematic representation of the **explicit training** paradigm. Explicit training employed a 2I-2AFC categorization task. In each trial, a sound was heard, coming from T1 or T2. Subjects had to listen carefully to the sound and decide whether it belonged to T1 or T2. Immediate feedback was provided after each response.

Testing:

- Before and after training, all groups were tested with sounds coming from **both** Hindi speakers.
- **All 10 tokens/phonetic category** were used from each speaker, presented in **all 5 different** simulated rooms.
- The order of rooms was fixed (1st: ping-pong, 2nd: bathroom, 3rd: cafeteria, 4th: anechoic, 5th: office), whereas tokens within each room were presented in random order
- The "trained voice" was presented first, followed by the "untrained voice"

Analyses: Proportion (percentage) correct responses from **training** (panel **4**) and **testing** (panels **3—5**) **sessions** were arcsine-square root transformed and entered into ANOVA analyses. In all figures rooms that were used during training are shaded in grey, error bars are SEMs and the dashed red line indicates chance performance.

Setup: Experiments were run in a small quiet room at UCR, using an Apple Mac Mini computer. The sounds were presented binaurally over Senheiser 650 headphones, at an individually adjusted comfortable listening level.

3. BASELINE PERFORMANCE AND ASSESSMENT OF TEST-RETEST EFFECTS

Results (Figure 3):

3A: Mean **baseline** performance was assessed. Data were averaged across groups.

- Initial performance in all rooms was above chance, suggesting that listeners were able to differentiate the sounds to some extent prior to training, without showing ceiling effects (error rates above 35%).
- Initial performance was comparable across rooms and voices ("trained" (1st) vs. "untrained" (2nd)).

3B: Pre-post test performance of Control group was compared. Data were averaged across "trained" (1st) and "untrained" (2nd) voice.

- Initial performance was comparable across rooms.
- No evidence of learning from pre- to post-test.

Pre-test performance was comparable across rooms and across "trained" and "untrained" voice intervals. No evidence for performance improvements without training

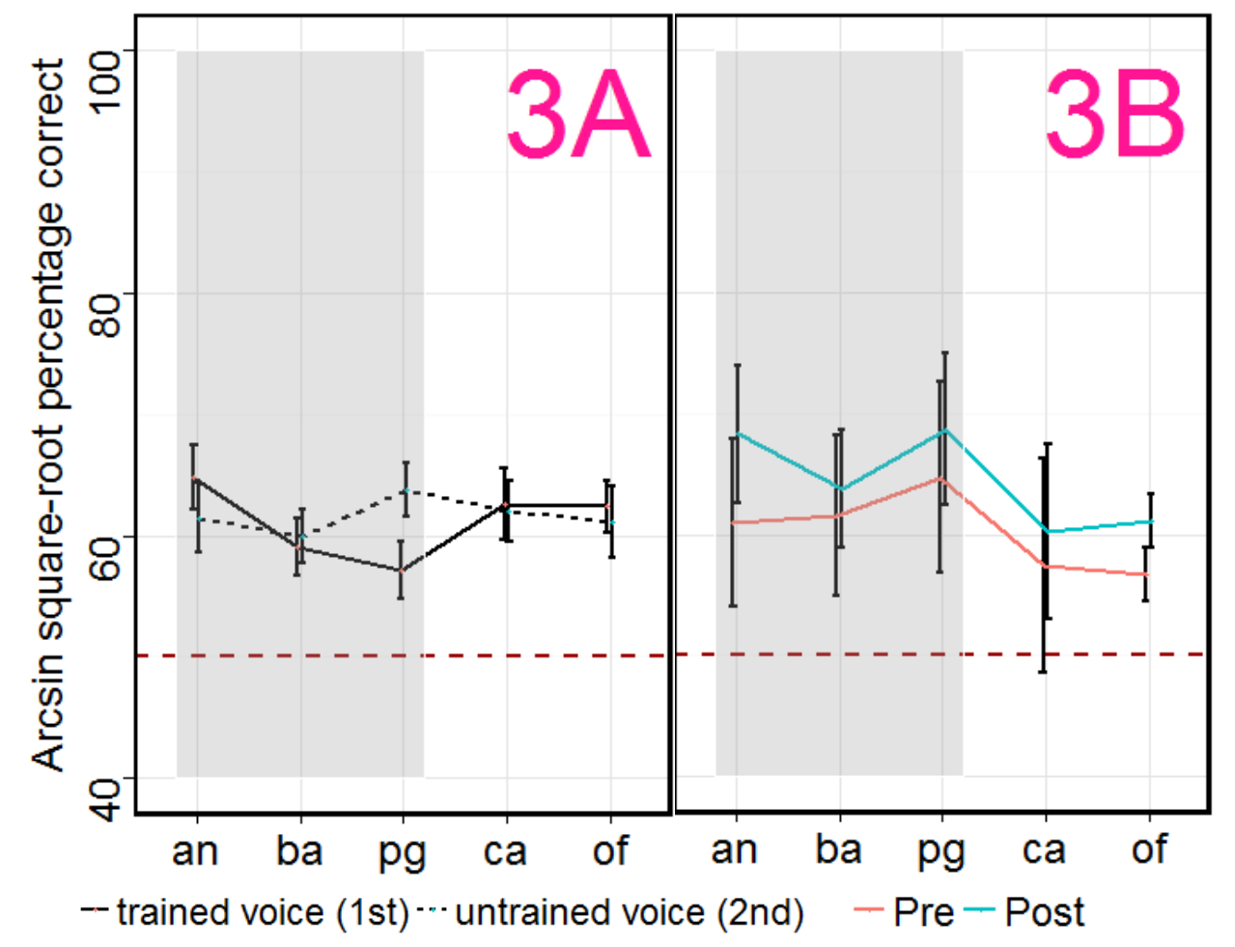

Figure 3. A: Mean baseline performance of **all groups** as a function of testing room. **B:** Mean pre- and post-test performance of **control group** as a function of testing room

4. TRAINING PERFORMANCE

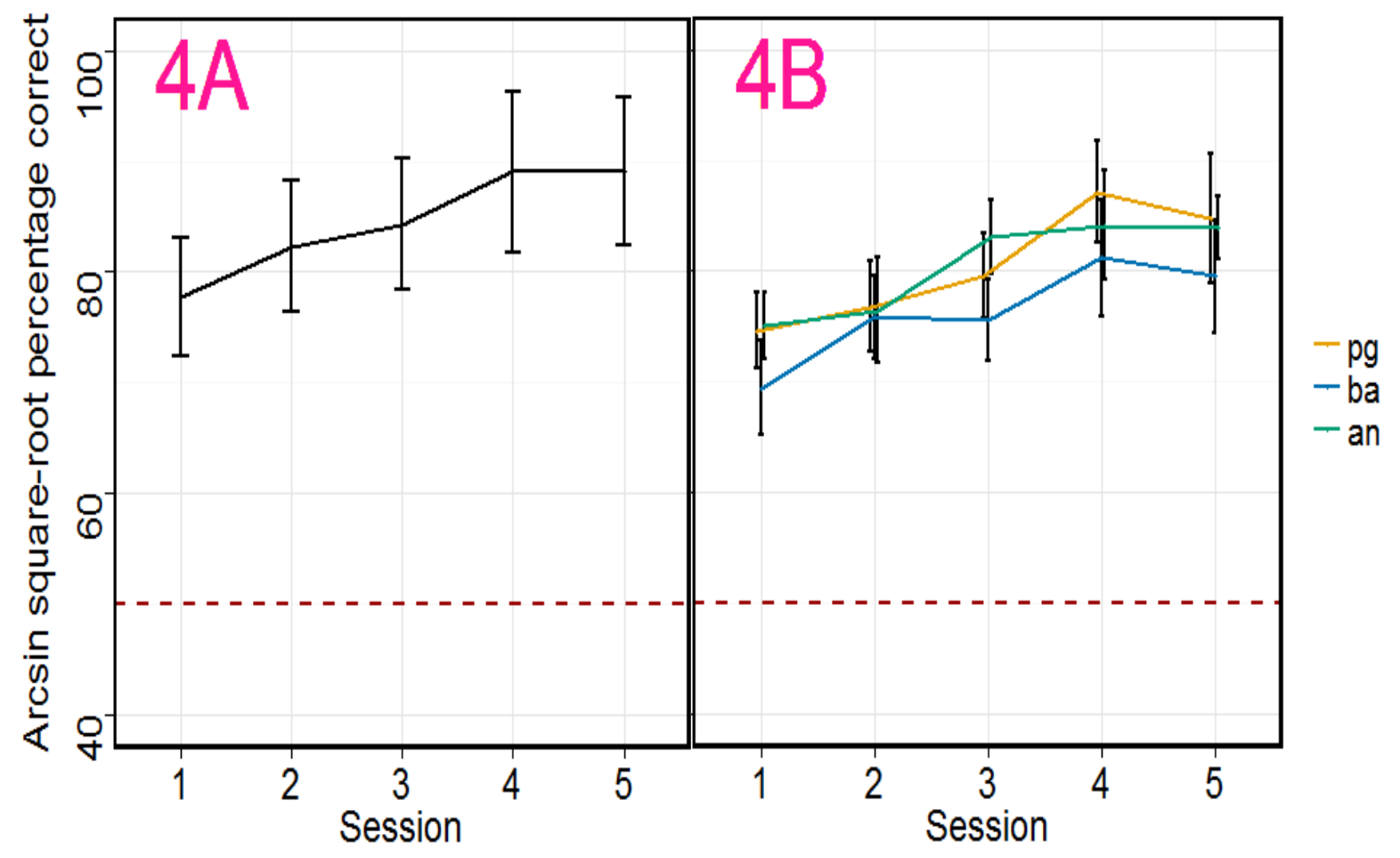

Figure 4. Mean training performance of **Explicit-1-Room** (A) and **Explicit-3-Room** (B) groups, as a function of daily session (1-5), separately for each training room (B).

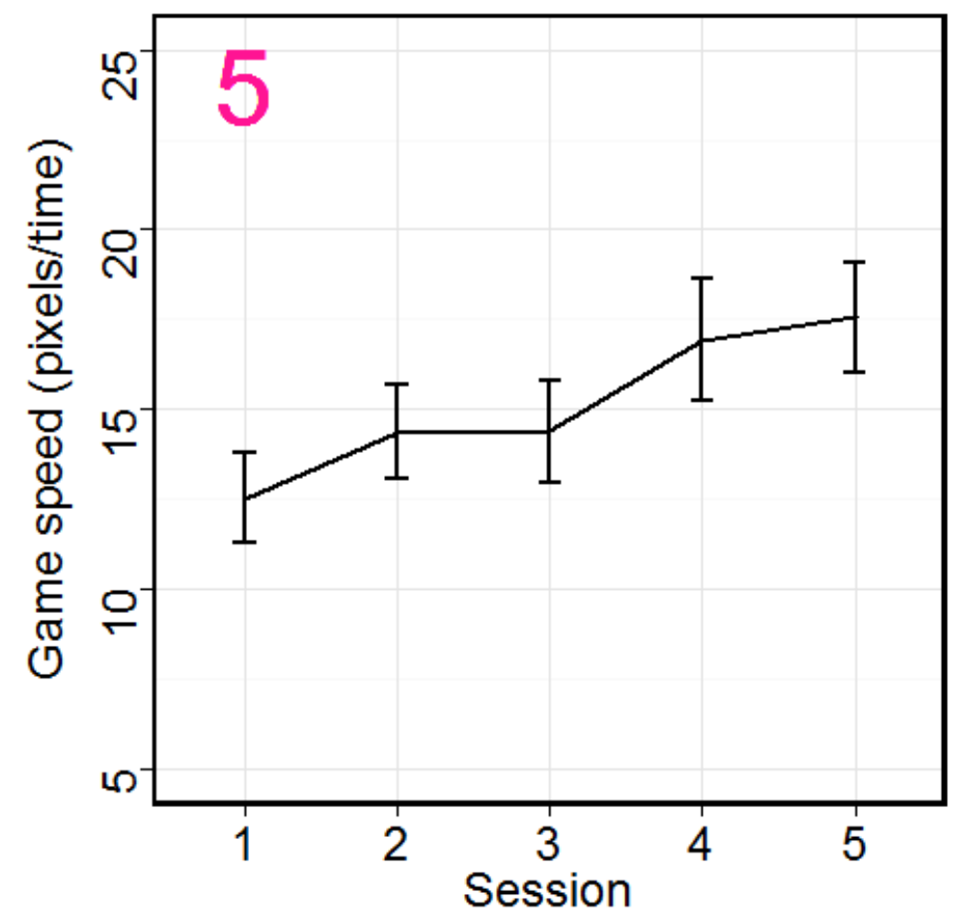

Figure 5. Mean game speed of both **Implicit groups**, as a function of daily session (1-5).

Results (Figure 4):

Average **training** performance of Explicit-1-Room (A) and Explicit-3-Room groups (B) was assessed.

- High performance for both groups (Fig. 4A-B).
- Comparable performance across different rooms (Fig. 4B).
- performance in anechoic room slightly better in 1-Room than in 3-Room training.

Results (Figure 5):

Mean **training** performance (game speed) of Implicit groups (averaged over 1-Room and 3-Room groups) was assessed.

- Players get better from first to last training session.

5. TESTING PERFORMANCE

Results (Figure 6):

Learning effects were assessed by comparing pre-post test performance of training groups. Data were averaged across trained and untrained tokens of the trained voice, and across bathroom-ping-pong (**ba-pg**) and cafeteria-office (**ca-of**) room pairs.

Explicit training (Fig. 6A-D):

- No difference between 1-Room vs. 3-Room training (Fig. 6A,B vs. 6C,D).
- Performance comparable in all 5 rooms within a condition.
- Large improvement for trained voice and trained rooms (shaded areas in Fig. 6A,C).
- Strong generalization to untrained rooms (non-shaded areas in Fig. 6A,C).
- Slightly weaker generalization to an untrained voice in all rooms (Fig. 6B,D).

Implicit training (Fig. 6E-H):

- No learning with 1-Room training (Fig. 6E,F).
- Some learning with 3-Room training for trained voice (Fig. 6G), but only in one of the trained rooms (an).
- Learning generalizes to untrained rooms (ca-of in Fig. 6G).
- No generalization of learning to an untrained voice (Fig. 6H).

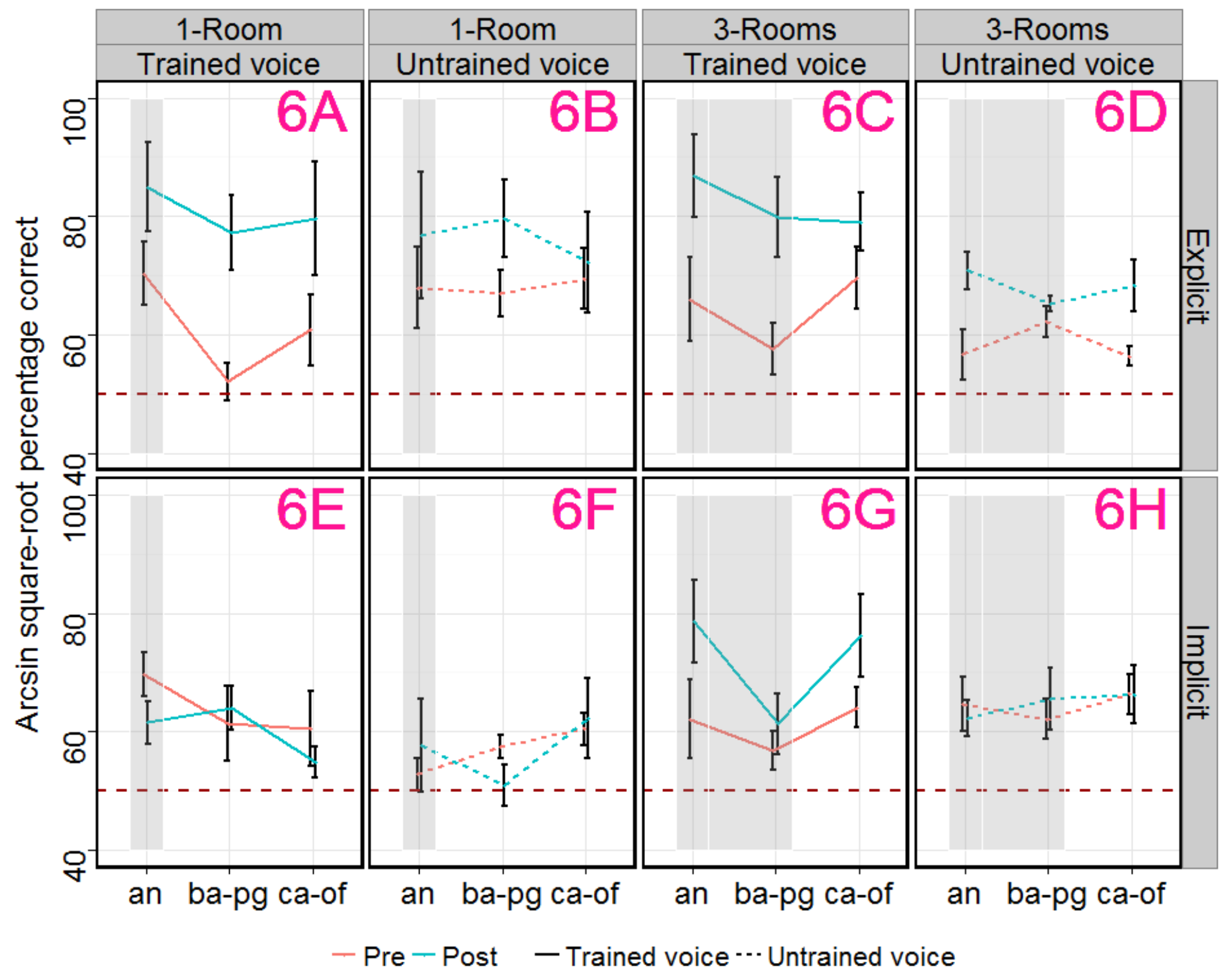

Figure 6. Mean pre- and post-test performance of **all groups** as a function of testing room.

Training in multiple rooms can enhance implicit phonetic category learning, but not explicit learning.

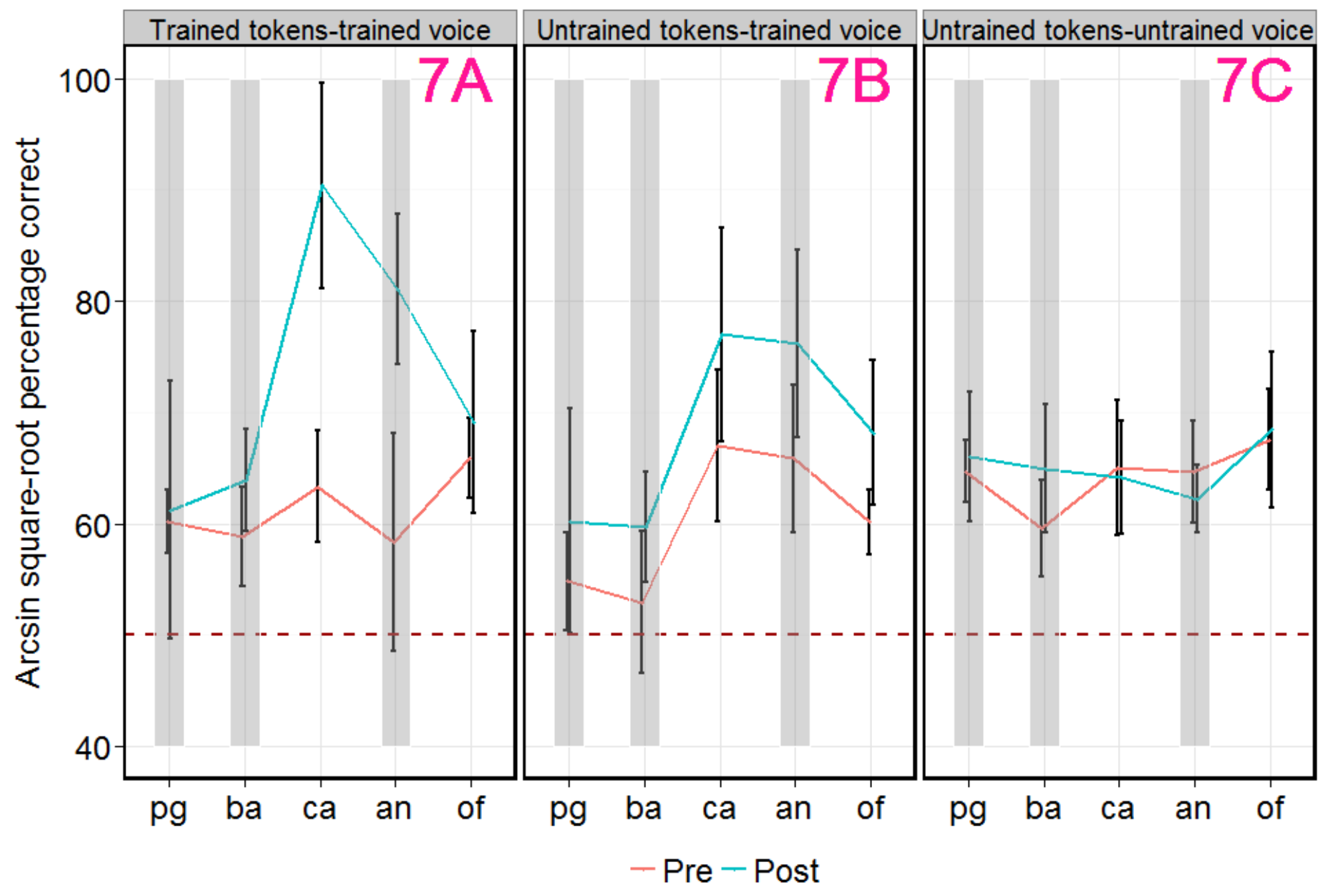

Figure 6. Mean pre- and post-test performance of **Implicit-3-room** group, plotted separately for each room. Room ordering corresponds to the order in which rooms were presented during testing.

Results (Figure 7):

Pre-post test performance of Implicit-3-R group was compared.

- Learning of stimuli presented in the trained anechoic environment, coming from the trained voice (Fig. 7A-B), but not from an untrained voice (Fig. 7C).
- Evidence of transfer of learning
  - to untrained tokens from the trained voice (7B)
  - to untrained reverberant environments, especially ca (7B).
- Little improvement for two of the trained reverberant environments (ba-pg; Fig. 7A-B).
- No transfer of learning to an untrained voice (Fig. 7C) in any of the testing rooms.
- Ordering of room presentation during testing may affect observed learning patterns.

Specificity of implicit learning might be related to the presentation order of rooms during testing.

6. CONCLUSIONS

- Varying the acoustic environment during training (1 Room vs. 3 Rooms) does not influence (positively or negatively) explicit training. Independent of whether participants are trained in anechoic or in anechoic and reverberant rooms (bathroom and ping-pong), there is approximately the same amount of learning for all rooms (i.e., near-perfect generalization). Generalization of learning to untrained tokens is near-perfect, while generalization to an untrained voice is slightly weaker.
- Varying the acoustic environment during training influences implicit training: when only one (anechoic) room is used during training, no implicit learning is observed. When listeners hear the phonetic contrast in three different rooms (anechoic, bathroom and ping-pong), voice-specific (i.e., non-generalizing to different voices) learning occurs for the trained voice in some rooms (anechoic, cafeteria, office) but not in the other rooms (bathroom, ping-pong).
- The present results do not allow to draw clear conclusions on why little implicit learning is observed in the trained reverberant bathroom and ping-pong rooms. Potential explanations: 1) the acoustic characteristics of the phonetic features relevant for the trained discrimination are affected by the acoustics of these rooms; 2) Short-term adaptation within test-sessions affected results in rooms tested early (bathroom, ping-pong) vs. those tested late.
- These results suggest that when explicit feedback is provided, subjects can identify the critical characteristics for the phonetic distinction, without being disturbed by the stimuli being presented in varying acoustic environments, and show good generalization of learning. When explicit feedback is absent, the listeners only learn the contrast if it is presented in varying environments, possibly because this allows them to identify the invariant phonetic features that are important for the phonetic distinction.
- Phonetic learning through action videogames is a promising future direction. More research is needed to identify the training conditions that maximize learning of novel speech sounds and transfer of learning to novel settings.

7. REFERENCES

- 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.
- Brandewie, E. & Zahorik, P. (2010). Prior listening in rooms improves speech intelligibility. Journal of the Acoustical Society of America, 128, 291–299.
- Ueno, K., Kopčo, N. & Shinn-Cunningham, B. (2005). Calibration of speech perception to room reverberation. In Proceedings of the forum acusticum conference. Budapest.
- Werker, J. F. & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. Journal of the Acoustical Society of America, 75, 1866–1878.
- Kopčo, N. & Shinn-Cunningham, B. G. (2011). Effect of stimulus spectrum on distance perception for nearby sources. Journal of the Acoustical Society of America, 130, 1530–1541.
- Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V. & Kollmeier, B. (2009). Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. EURASIP Journal on Advances in Signal Processing.
- Seitz, A. R. & Watanabe, T. (2005). A unified model for perceptual learning. Trends in Cognitive Sciences, 9, 329–334.

8. ACKNOWLEDGMENTS

- Supported by APVV-0452-12 and NIH P41RR14075.
- Research assistants: Lisa Hong, Jennifer Rose, Ruchie Modi, Arjie Florentino, Eric Yoon, Rayan Karapet.