Nonnative Phonetic Category Training in Varying Acoustic Environments

Eleni Vlahou[1,2], Aaron Seitz[2] & Norbert Kopčo[1]

[1]Institute of Computer Science, P. J. Šafárik University, Košice, Slovakia, [2]Department of Psychology, University of California, Riverside

1. BACKGROUND AND MOTIVATION

Past research has shown that reverberation has a pronounced detrimental effect on speech intelligibility (e.g., Nábělek & Donahue, 1984), but no previous studies investigated how it affects the **acquisition** of **novel phonetic categories**. In a previous study (Vlahou et al., 2014; see Methods) we used virtual acoustics to train listeners on a difficult nonnative phonetic distinction. During training, stimuli were presented to different subject groups a) in a **single** (**anechoic**) room or b) in **multiple** (**anechoic** and **two reverberant**) rooms. Learning effects were assessed by comparing **pre**-test vs. **post**-test performance of each group. **Figure 1** summarizes our main findings:

For **explicit** training:

- No difference between groups trained in multiple or in a single environment (Fig. 1A vs. 1B).
- Large improvement for trained voice and trained rooms (shaded areas, Fig. 1A-B).
- Strong generalization to untrained rooms (unshaded areas, Fig. 1A-B) and weaker generalization to an untrained voice (Fig. 1E-F).

For **implicit** training:

- Effective for participants trained in **multiple rooms** but not in a **single anechoic** environment (Fig. 1C vs. 1D). This difference between the two groups might have been caused by lack of exposure to (a) **reverberation** or, (b), **room-to-room variation** during training for the implicit-1-Room group.
- For the **3-Room** group (Fig. 1D), learning observed for the anechoic environment (an) but little improvement shown for the trained rooms (pg, ba). The pg and ba improvement was less than or equal to the two untrained rooms (ca and of). However, a possible confound was the **fixed** order of rooms during testing, favoring rooms presented later (ca: 3rd, an: 4th, of:5th) compared to rooms presented earlier (pg: 1st, ba: 2nd).
- No generalization to an untrained voice for either 1- or 3-Room groups (Fig. 1G-H).

Main questions in current study:

Here, we attempted to separate the alternative explanations for the effects observed in the previous study. Different groups of participants were trained implicitly:

- In a **single reverberant** environment. If exposure to **reverberation** during training facilitates implicit learning, then this group should show improved posttest performance. Alternatively, if **room variation** is a critical factor, then no learning should be observed.
- In **multiple** rooms, but with **random** room order during testing, thereby removing the potential confound of fixed room order.

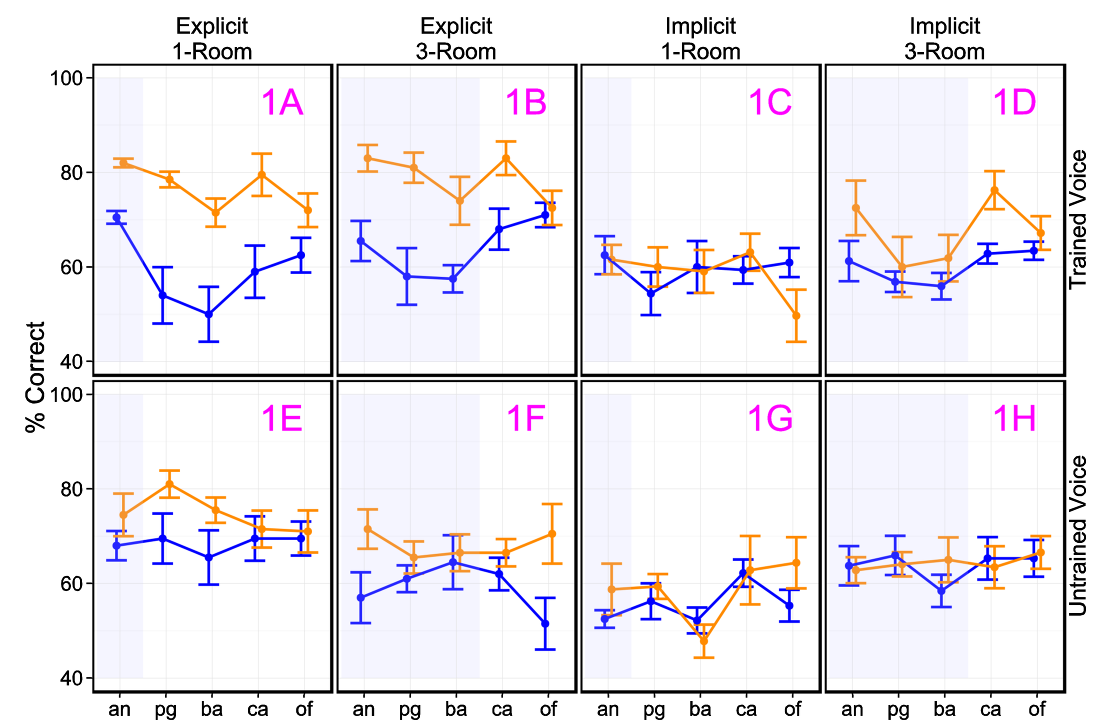


Figure 1. Mean **pre**-test and **post**-test performance of each group as a function of room, separately for the trained and untrained voice.

2. METHODS

Subjects and experimental conditions: Two groups (5 subjects each) were trained **explicitly** (Fig.1A-B, E-F) with sounds presented in **anechoic** (**1-Room**) or in an **anechoic** and 2 **reverberant** environments (**3-Room**). Fifteen participants were trained **implicitly** in **3 rooms**. During **testing**, rooms were presented in a **random** (7 people; current study) or **fixed** order (8 people, Fig.1D, 1H; 1st: pg, 2nd: ba, 3rd: ca, 4th: an, 5th: of, Vlahou et al., 2014). 14 participants were trained **implicitly** in **1 room** (an: 8, Fig.1C, 1G; ba: 2, pg: 4). 13 more subjects were tested and re-tested with the same material over a period of 10 days, without training in between (**no-training control group**).

Phonetic stimuli and simulated room reverberation: We used Hindi **dental-retroflex** CV syllables (Werker & Tees, 1984) from two native speakers. Each syllable was convolved with Binaural Room Impulse Responses (BRIRs) of 4 different rooms, termed "ping-pong" (**pg**), "bathroom" (**ba**), "cafeteria" (**ca**), "office" (**of**) and an anechoic environment (**an**; see Vlahou et al., 2014).

Training: One of the two Hindi speakers was used during training ("**Trained Voice**", counterbalanced across participants) in 4 daily sessions, 45 min/session. In each session:

- 1-Room groups were trained with sounds presented in one room (anechoic, bathroom or ping-pong, 600 trials/session).
- 3-Room groups 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).

Explicit training consisted of a 1I-2AFC test. Feedback was provided after each response. The implicit training paradigm is illustrated in Fig. 2.

Figure 2. 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.

Testing:

- Before and after training, all groups were tested with sounds coming from **both** Hindi voices in **all 5 different** simulated rooms.
- The "trained voice" was presented first, followed by the "untrained voice"

Analyses: Proportion (percentage) correct responses were arcsine-square root transformed and entered into ANOVA analyses. In all figures, shaded areas show rooms used during training and error bars are SEMs, corrected for within-subject designs.

3. IMPLICIT TRAINING IN MULTIPLE ROOMS

Figure 3. Mean **pre**-test and **post**-test performance of the **implicit-3-Room group** for participants tested in **random** room order (3A-B), for all participants, averaged over random and fixed room order (3C-D), and the **control** group (3E-F), as a function of room, separately for the trained and untrained voice (for the control group, for Hindi voice presented 1st and 2nd).

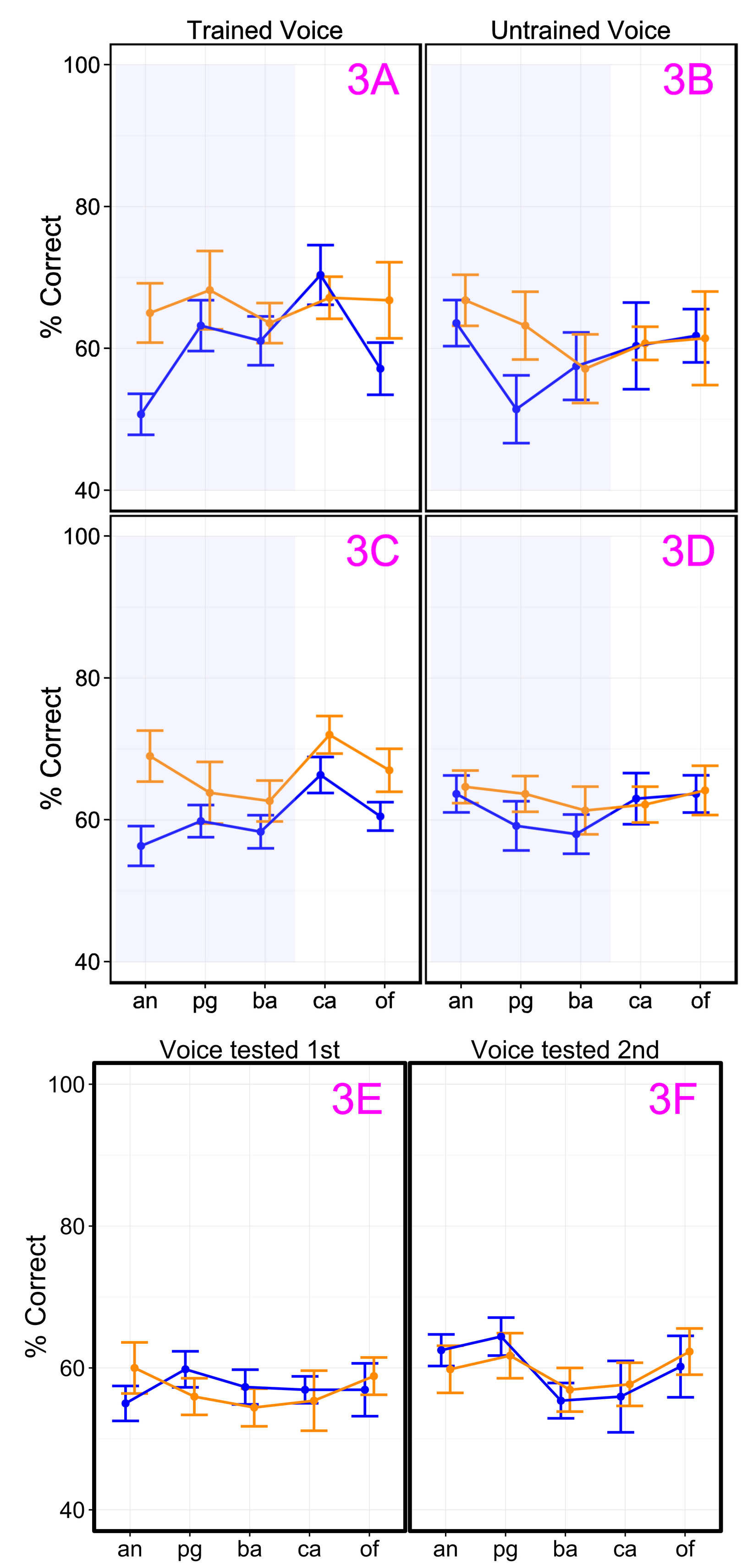
Data from participants tested in random room order (3A-B):

- Support previous findings of pre-post test **improvement** for speech stimuli coming from the **trained** voice (compare Fig. 3A to Fig. 1D).
- Do not confirm larger gains for the untrained reverberant rooms compared to the two trained reverberant rooms (compare Fig. 3A to Fig. 1D).
- Confirm previous findings of **no learning** for the **untrained** voice (compare Fig. 3B to Fig. 1H).

Averaged data over all participants trained implicitly in multiple rooms (3C-D) show:

- **Learning** for the implicit-3-room group for sounds coming from the **trained** voice (3C-D; Interaction time x voice, $F_{(1,14)} = 6.52$, $p = .0229$).
- Good **generalization** to untrained rooms coming from the trained voice (3C). The anechoic environment tends to show larger improvement.
- **No (or little) generalization** of learning to the **untrained** voice (3D).

No learning for the **control** group (3E-F) from **pre**-test to **post**-test , for either room/voice tested 1st or 2nd.

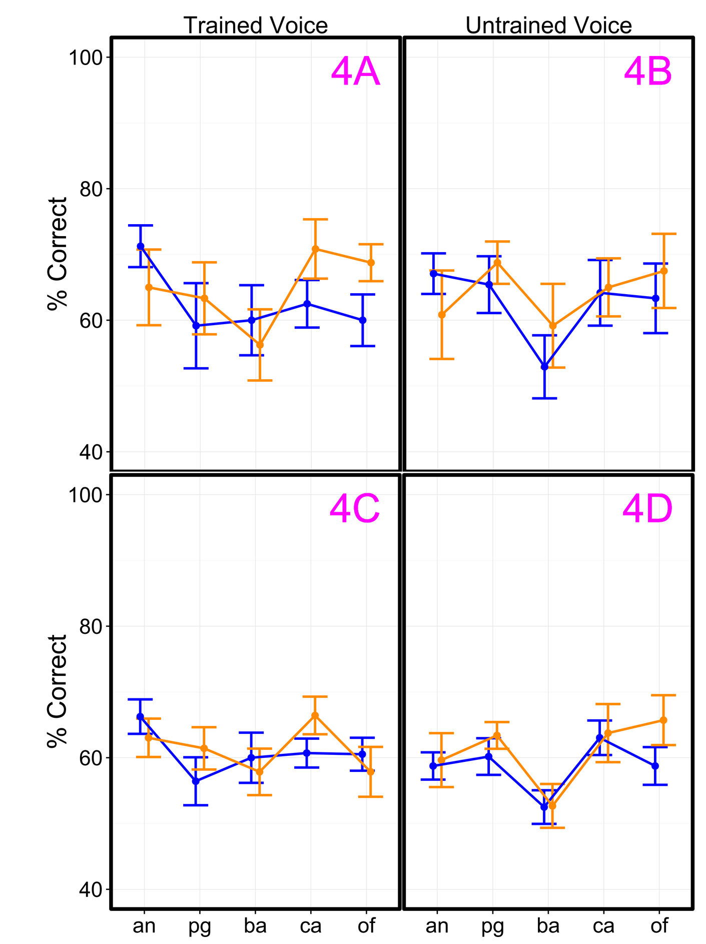


Implicit training is **effective** when speech sounds are presented in **multiple** rooms. Learning generalizes to untrained reverberant rooms when the stimuli come from the **trained** voice. However, **no generalization** is observed for speech coming from an **untrained** voice.

4. IMPLICIT TRAINING IN A SINGLE ENVIRONMENT

Figure 4. Mean **pre**-test and **post**-test performance of the **implicit 1-Room** group, for participants trained with a single reverberant environment (4A-B) and for all participants, averaged over single reverberant and anechoic room training (4C-D), as a function of room, separately for the trained and untrained voice.

- No evidence of improvement for participants trained in a single reverberant environment in either the trained or untrained voice (compare Fig. 4A, 4B to Fig. 1C, 1G).
- Averaged data over all participants trained implicitly in a single **anechoic** or **reverberant** room (4C-D) show **no pre**-test / **post**-test **improvement**.



No evidence for implicit learning without **room variation** during training.

5. CONCLUSIONS

- Implicit training in **varying** acoustic environments is effective whereas implicit training in a **single** room is not. For the **3-Room** group, significant improvement is observed for speech sounds coming from the **trained** voice, for both **trained** and **untrained** acoustic environments (Fig. 3C). For the **1-Room** group, no pre-test / post-test improvement is observed, independent of whether participants are trained in an **anechoic** (Fig. 1C) or **reverberant** environment (bathroom or ping-pong; Fig. 4A).
- This finding suggests that **room variation** during training is likely to be important for spontaneous phonetic learning. In line with previous studies (Lim & Holt, 2011), we show that increased variability along a **non-informative** dimension (room acoustics) shifts perceptual categorization towards more reliable cues, namely the invariant phonetic features that are robust against the variations that one experiences in different rooms. This leads to improved categorization performance for speech in reverberation but also to substantial improvement in the anechoic environment.
- This interpretation is consistent with the finding of **no generalization** of learning to an **untrained voice** for either 3- or 1-Room training groups (Figs. 3D & 4D). Since only one Hindi voice was used during training, participants were not able to ignore variations due to voice characteristics, possibly forming overspecified perceptual categories that included phonetically irrelevant talker-specific details.
- In sum, our results suggest that exposure to diverse room acoustics during the acquisition of novel phonetic categories facilitates nonnative phonetic learning.

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.
Lim, S.-J. & Holt, L. L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. Cognitive Science, 35, 1390—1405.
Seitz, A. R. & Watanabe, T. (2005). A unified model for perceptual learning. Trends in Cognitive Sciences, 9, 329–334.
Vlahou, E., Seitz, A. & Kopčo, N. (2014). ARO Abstract #806.
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.

ACKNOWLEDGMENTS

UCR research assistants: Ruchi Modi, Arjie Florentino, Rukayat Salaam, Surangi Jayasinghe & Catherine Nolasco.
Work supported by

- Slovak Research and Development Agency, SRDA, contract No. APVV-0452-12. NSF-BCS-1057625.
- This poster is the result of the project implementation: University Science Park TECHNICOM for Innovation Applications Supported by Knowledge Technology, ITMS: 26220220182, supported by the Research & Development Operational Programme funded by the ERDF.