The Effect of Artificial Voice Training on Speech Perception and Listening Effort

Carlos Eduardo Rangel Outeda, BASc Arts & Sciences

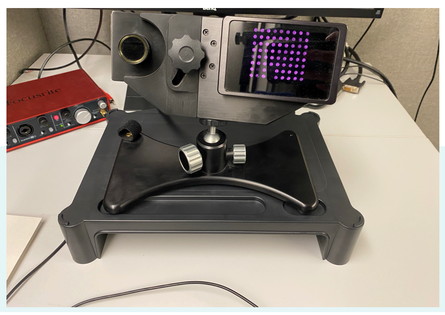
Supervised by Dr. Emma Holmes, UCL Psychology and Language Sciences

Introduction

- It is more difficult to identify and understand a speaker when there is background noise and speakers, ie. Restaurant
- However, people can recognise voices and understand them better if they belong to someone they know, ie. Family, friends
- People have been trained in lab conditions to be familiar with new voices. The result is recognition and understanding as if family/friend
- As little as 10 minutes of training can yield an effect on speech perception - voice better understood and recognised.
- Rather than focus on the effect of performance (recognition, understanding of listener), this project focuses on *Listening Effort*
- Listening Effort is how hard listener's had to work in order to understand/recognise a voice. It is not necessarily linked to performance.
- Those with hearing loss regularly report higher effort than those with regular hearing in the same task. Can voice training help reduce effort?

Objectives

- Determine whether or not certain auditory conditions can make listening tasks less effortful
  - Listening Effort is measured through pupillometry & self reporting effort after trials
- Observe the interplay between signal to noise ratio (how loud the target voice is compared to background noise) and familiarity with a voice, seeing which is more important
- Validate previous experiments by confirming if training voices in a lab for a limited time yields improvements in listening tasks



Methodology

There was a pool of the same five voices for all participants: EPS01, EPS02, EPS04, EPS05, and EPS06. However, the conditions each voice was assigned differed for each participant. In other words, for one participant EPS01 could be most exposed and assigned Fam1, for another participant EPS01 is least exposed Fam3, instead EPS04 is Fam1, and so on.

Participant Eligibility:

- Accustomed to British English
- No physical or mental circumstances which may Impede eye tracking, pupillometry, or attention
- Good to normal Hearing

Training Section:

Brief (10 trial) exposure to each voice with matching name, followed by task.

Task: Identify which name each audio clip belongs to.

Participant is exposed to each voice for different quantities of trials. Most trials = most exposure = Fam1. Then Fam2, then least exposure = Fam3

Goal: Participant artificially trained on three voices.

Recognition Section:

Introduction of two new unfamiliar voices: Unfam1, Unfam2

Task: Decide whether the voice in audio clip is familiar (fam1,2,3) or unfamiliar (new).
If familiar, they must say which voice it is.

Throughout this section, there was background Babble noise - unintelligible overlap of talking - to replicate noisy setting.

Noise to Sound Ratio (volume of babble noise vs Audio sound) = SNR. -3 SNR = more noise, 6 SNR = more audio sound

Goal: Determine if participants able to recognise familiar voices better. Listening effort also measured in this section.

Adaptability Section

Task: Identify words said by one of two voices playing simultaneously. Voices can be any of the five from the pool.

TMR = Target voice volume to Masker-voice volume Ratio. TMR changed depending if correct or not.

Goal: Calibrate the background noise for each voice condition to achieve an accuracy of 70.7% . Attempt to isolate listening effort.

Intelligibility Section.

Task: Same as Adaptability.

TMR = already calibrated for 70.7% accuracy across voice conditions. Each pair of target/ masker voice have their own TMR

Goal: Measure listening effort at different voice conditions whilst accuracy is controlled.

Crux of study:

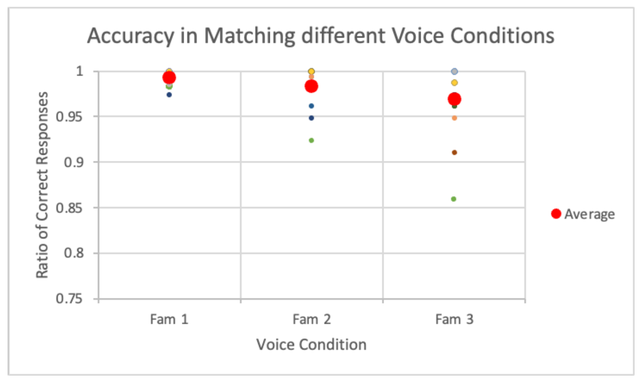
Effects on effort > effects on accuracy.

Participants were briefed on what listening effort means: How hard they worked to get an answer - not whether they think they got it right.

Results & Conclusions

The average effort of participants in the Intelligibility section for different voice conditions was statistically insignificant, ie. this study did not find that voice training yields a noticeable difference in *self reported* listening effort. It is relevant to note that this is still an ongoing study, and the pupillometry data has yet to be completed, and so is not included in these results. Moreover, not all the data was gathered by the time the placement was complete, resulting in the scholar not having the complete data at his disposal for the outputs.

However there were some relevant observations:

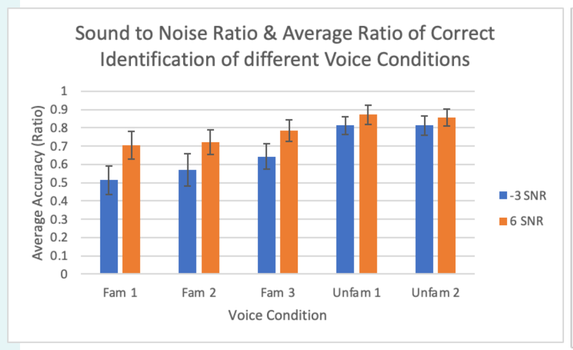


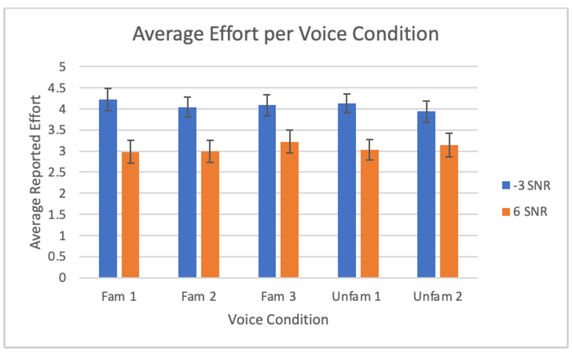
<u>Corroborates results of previous studies for artificial voice training: Voice training improves accurate identification.</u>

In the Training section, there was a statistically significant difference between the average accuracies for each condition.

Minimal amounts of artificial training (Under 1 hour for most familiar) with a voice was enough to improve listeners' ability to identify the voice correctly.

The variance amongst accuracies also decreased with increasing familiarity, which implies that voice training is more useful to certain participants than to others, since some participants had high accuracies regardless of familiarity.





<u>Corroborates previous studies - the louder the speaker's volume, the higher the accuracy.</u>

There seems to be no statistically significant difference between the three familiar conditions, which may mean that after a certain level of exposure - ie. the limited Fam3 training, voice training hits a ceiling on how much it can help with accurately recognising voices.

<u>Louder babble noise impedes recognition of familiar voices?</u>

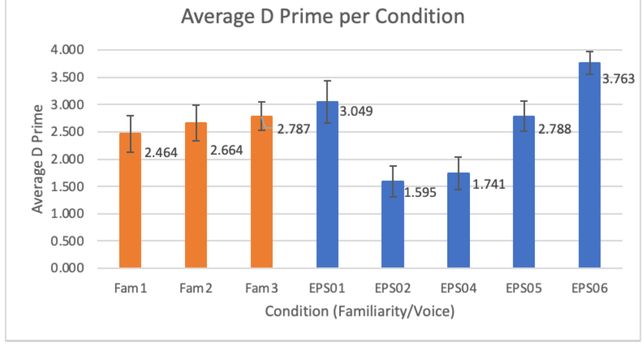
This graph may lead one to assume the accuracy of recognising unfamiliar voices is unhampered by louder background noise. However, this trend could be explained by the procedure - for Fam1,2,3 audio files participants had to identify a voice as familiar *and* which voice it was to be counted as accurate, which is more difficult than simply identifying a voice as unfamiliar as they did with Unfam1,2.

<u>Volume of target voice compared to masker voice affects effort to a larger degree than accuracy</u>

The novel observation in the data is that the difference in SNR in the recognition section led to a greater difference in effort than it did in accuracy.

In other words, accuracy and listening effort do not necessarily match up - which is an important datum for future study.

Regardless, louder voices can help listeners who are getting fatigued, even if it doesn't affect their ability to correctly recognise voices to the same extent.



<u>Sometimes the voice itself matters</u>

Participants found it considerably more difficult to distinguish EPS02 and EPS 04 from the rest. Perhaps they were similar, and participants confused them with each other. This was a rare instance in which the voice speaking was more relevant than the familiarity of said voice.

For Training and Intelligibility sections, which voice was the target voice did not have a statistically significant effect on the listeners performance nor effort. Yet here the voice condition matters whilst the familiarity does not.

References

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 25,* 975–979. https://doi.org/10.1121/1.1907229

Holmes, E., To, G., & Johnsrude, I. S. (2021). How Long Does It Take for a Voice to Become Familiar? Speech Intelligibility and Voice Recognition Are Differentially Sensitive to Voice Training. Psychological Science, 32(6), 903–915. https://doi.org/10.1177/0956797621991137

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a Cocktail Party: Voice Familiarity Aids Speech Perception in the Presence of a Competing Voice. Psychological Science, 24(10), 1995–2004. https://doi.org/10.1177/0956797613482467

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International journal of audiology*, *53*(7), 433–440. https://doi.org/10.3109/14992027.2014.890296

Stewart, R., & Wingfield, A. (2009). Hearing loss and cognitive effort in older adults' report accuracy for verbal materials. *Journal of the American Academy of Audiology*, *20*(2), 147–154. https://doi.org/10.3766/jaaa.20.2.7