Voice familiarity via training affects listening effort during a voice cue sensitivity task with vocoder degraded speech

Ada Biçer[1,2]*, Thomas Koelewijn[1,2], Deniz Başkent[1,2]

1 Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
2 Research School of Behavioral and Cognitive Neuroscience, Graduate School of Medical Sciences, University of Groningen, Groningen, The Netherlands
** Contact author: a.bicer@rug.nl*

Abstract

Understanding speech in real-life can be challenging and effortful when multiple people speak at the same time. In speech-on-speech (SoS) perception, normal hearing (NH) listeners can use fundamental frequency (F0) and vocal-tract length (VTL) voice cues to separate speech streams, spoken by different talkers. However, such voice segregation can be challenging for cochlear implant (CI) users, as CI users have a reduced sensitivity to F0 and VTL voice cues. Additionally, vocoder studies show that listening effort is increased with increased spectral degradation in the speech signal. In SoS listening, familiarity with a talker's voice can improve speech intelligibility for NH listeners. However, it is unknown if voice familiarity improves sensitivity to F0 and VTL voice cues and affects listening effort, especially when the speech signal is vocoder degraded.

In this study, we aimed to provide voice familiarity by implicit short-term voice training. During training, participants listened to an audiobook segment of approximately 30 minutes that contained 13 chapters, and after each chapter, they answered a context related question. Voice sensitivity, namely just-noticeable-differences (JNDs) for F0 and VTL voice cues combined (F0+VTL), was measured with an odd-one-out task in a 3 alternative forced choice adaptive paradigm. Simultaneously, listening effort was measured via pupillometry.

Our results showed that voice training did not improve sensitivity to small F0+VTL voice cue differences measured at the threshold level for both non-vocoded and vocoded conditions. However, according to Generalized Additive Mixed Models (GAMM) analysis results, effort while listening to vocoded speech was less for trained (familiar) compared to untrained voices. These findings suggest that voice familiarity through implicit voice training can be of benefit for voice cue perception through reducing listening effort for vocoded speech, even in the absence of a behavioral effect.

Introduction

When listening to multiple talkers at the same time, listening to familiar voices can provide a speech intelligibility benefit. A familiar voice that belongs to a partner or close friend can improve speech-on-speech (SoS) listening in normal hearing (NH) listeners (Holmes et al., 2018; Johnsrude et al., 2013). Additionally, it is possible to make a naïve listener familiar with a

previously unheard voice, through voice training. Voice training can be done explicitly, by training listeners with a speaker recognition task (Nygaard & Pisoni, 1998) or implicitly, through voice exposure over a period of time (Kreitewolf et al., 2017), similar to familiarizing with voices in real-life. It might be that the implicit knowledge we acquire about voices through exposure is related to talker-specific voice cues. Two main voice cues are fundamental frequency (F0) and vocal-tract length (VTL). While F0 is related to glottal pulse rate and perceived as the voice pitch, VTL is related to formant frequencies, and changes with the speaker's height. In SoS situations when there are clear differences in F0 and VTL between target and masker voices, intelligibility of the target speech is high (Darwin et al., 2003). However, when speech signals are inherently degraded via cochlear implants (CIs), F0 and VTL voice cues are not optimally perceived by the CI users, which contributes to the challenge of understanding speech in SoS situations (El Boghdady et al., 2019). Previous research showed that perceptual discriminability of F0 and VTL voice cues were limited with vocoder manipulations (Gaudrain & Başkent, 2015) and for CI users (Gaudrain & Başkent, 2018). While voice familiarity might improve speech intelligibility for NH listeners in SoS situations, where voice information is used to segregate target and masker speech, , it is relatively unknown whether sensitivity to F0+VTL voice cues reduced by means of vocoding, would improve with voice training.

SoS perception is shown to be effortful in studies using pupillometry as an objective measure for listening effort (Koelewijn et al., 2012). Additionally, when speech is spectrally degraded, such as in CIs, listening also becomes more effortful (Winn et al., 2015), which contributes to the challenge of listening to speech in adverse conditions. As cognitive resources are limited, a reduction in listening effort would spare cognitive resources that can then be used for other listening related processes, which in turn might improve speech intelligibility (Rabbitt, 1966). Even though voice familiarity may show a benefit at behavioral level by improving speech intelligibility, our knowledge on how voice familiarity may affect listening effort remains scarce.

The present study investigated the effect of voice familiarity via implicit voice training, on F0+VTL voice cue discrimination and listening effort, when speech signal was non-vocoded and vocoded. Additionally, acoustic/linguistic variability of the stimuli was manipulated. After a short voice training, sensitivity to F0+VTL voice cues was measured by means of an adaptive procedure with just-noticeable-differences (JNDs) as the outcome measure, while in parallel, listening effort was assessed using pupillometry.

Methods

Participants

Sixteen NH native Dutch speaking adults between the ages of 21 and 39 years participated in the experiment. Participants with normal hearing audiometric thresholds were included in the study. At the beginning of the experimental session, a written informed consent was obtained from the participants. The ethical approval was obtained from the Ethics Committee of the University Medical Center Groningen and participants received hourly compensation for their participation.

Stimuli

Stimuli for voice training

Stimuli used for voice training were auditory recordings of the first 13 chapters of "The Twits" (de Griezels) by Roald Dahl in Dutch language. A female native Dutch speaker took part in the recordings of the audiobook. From these recordings a male voice version was created by shifting the F0 -12 semitones (st) and shifting the VTL +3.8 st, as male voice was defined as having a difference of -12 F0 and +3.8 VTL difference from the original female voice (Gaudrain & Başkent,

2015, 2018). Female voice was also processed without shifting the F0 and VTL to keep the possible sound processing artifacts constant.

Stimuli for voice sensitivity and listening effort measurements

The same female Dutch speaker who took part in recordings of the audiobook for voice training, also took part in recordings of stimuli that were used during voice cue sensitivity measurements. We recorded meaningful Dutch CVC words from the Nederlandse Vereniging voor Audiologie (NVA) corpus (Bosman & Smoorenburg, 1995). The stimuli used during the voice sensitivity measurements were consonant-vowel triplets (CVCVCV) created from individual CVs that were extracted from the recorded CVC words. The procedure of extracting the CVs was similar to Gaudrain & Başkent (2015, 2018). During the adaptive task, while F0+VTL shifted versions of individual CVs were processed, CV durations were also equalized to 200 ms. CV triplets presented in one condition were either acoustic/linguistically the same (fixed item) or different (variable item), depending on the item variability condition. Stimuli were presented as either non-vocoded or vocoded signals. Vocoding manipulations were in line with Gaudrain & Başkent (2015), using settings that produce voice identification sensitivity performance matching that of highly performing CI users. A noise vocoder with 12- band, 12th order zero-phase bandpass filters between 150 and 7000 Hz was implemented, using Greenwood frequency-place mapping function for partitioning the cutoff frequencies of the filters. The temporal envelope was extracted with half-wave rectification and later low-pass filtered in each frequency band with a cutoff frequency of 300 Hz. Analysis and synthesis filters used the same settings.

Procedure

The experimental session started with voice training. Half of the participants listened to the first 13 chapters of "The Twits" uttered by the original female voice and the other half uttered by the simulated male voice. At the end of each chapter, a content related question was presented to ensure engagement. During voice training, participants did not receive any instructions to attend to the voice and training lasted approximately 30 minutes. Therefore, the implemented voice training was implicit and short term. Voice training was followed by the voice discrimination task, which was similar to the procedure of Gaudrain & Başkent (2015, 2018). JNDs were obtained using an odd-one-out task in a 3 Alternative Forced Choice (AFC) adaptive paradigm, during which listeners chose the stimuli that sounded different (the target voice) than the other two stimuli that sounded identical to each other (the reference voice).

At the start of each trial, there was a 3 second silence, which was followed by the stimuli presentation. After the stimuli presentation, there was another 3 second silence. From the start of the trial until the end of 3 second silence after stimuli presentation, a fixation dot was presented in the center of the screen while x- and y- gaze data and pupil diameters of both eyes were recorded. Following pupillometry measurements, selection boxes numbered 1 to 3, referring to the order of the three CV triplets presented, was shown on the screen.

JNDs were measured using an adaptive procedure where target voice became closer to the reference voice by sifting the F0+VTL voice cues together using STRAIGHT software (Kawahara & Irino, 2005) in steps of 2 st, in a 2-down 1-up staircase procedure, that corresponded to 70.7% correct on the psychometric function. Step size was updated after 15 trials or after the voice difference became smaller than twice the step size by $\sqrt{2}$. JNDs were measured in two directions: from a target male voice towards a female reference voice and from a target female voice towards a male reference voice. Target voice either belonged to the familiarized voice (female or simulated male voice) or to an unfamiliar voice, which was the voice unused during training, either female or simulated male voice. At the end, participants completed the JND test that consisted of 8 conditions, involving conditions with familiar or novel voice used as reference voice, where stimuli were either variable or fixed, presented as non-vocoded and vocoded signals.

Analysis

Voice sensitivity data analysis

F0+VTL JNDs were analyzed with a 2x2x2 repeated-measure analysis of variance (ANOVA) with voice training (untrained, trained voice), vocoder (non-vocoded, vocoded) and item variability (fixed, variable) as within-subject factors on the log-transformed JNDs, using R. Type I errors were corrected with False Discovery Rate (FDR).

Listening effort data analysis

Pupillometry measurements were first preprocessed, in line with Koelewijn et al. (2021). Trials containing 20% or more eye blinks were removed from further analysis. For the remaining trials, eye blinks were corrected with linear interpolation. Trials with eye movements outside of the visual display area were also excluded from further analysis. In order to remove high frequency artifacts, an 11-point moving average smoothing filter was used. A baseline value was calculated by taking the mean of the 1 s time window before stimulus onset. Baseline correction was performed by subtracting the trace's baseline value from the value for each time point within that trace. These preprocessed pupil traces were then averaged for each participant and each condition and Peak Pupil Dilation (PPD), Mean Pupil Dilation (MPD), Peak Pupil Dilation Latency (PPDL) and averaged baseline pupil diameter values were calculated. Between the time window of stimulus onset until response prompt, the maximum pupil dilation value in mm compared to baseline was calculated as PPD. MPD was calculated, within the same time window, as the average pupil dilation in mm relative to baseline and PPDL denotes the time in ms from the onset of stimulus until the PPD.

Separate 2x2x2 repeated measures ANOVAs were performed with voice training, vocoder and item variability as within-subject factors on each of the pupil outcome measures (PPD, MPD, PPDL and pupil baseline). In addition, a non-linear regression analysis was performed using Generalized Additive Mixed Models (GAMMs) to estimate the change in pupil dilation over time (van Rij et al., 2019), using R. Preprocessing of the data for GAMM analysis was performed the same way as described above with the exception of down-sampling data from 120 Hz to 30 Hz and not applying a smoothing filter. For GAMM analysis we pooled data over item variability conditions and performed a planned comparison between vocoder and voice training conditions. With GAMM analysis we looked at the change in pupil diameter over time with contrasts between the trained and untrained voices for both vocoded and non-vocoded speech signals. Contrasts between vocoded and non-vocoded speech for trained and untrained voices were also analyzed.

Results

Voice training

The answers to 13 multiple choice questions showed that the overall accuracy score was high (M= 88.94%, SD = 9.45%). Individual scores from participants also showed that no participant scored lower than 76.92% correct. These high accuracy scores demonstrate that listeners were attentive during voice training.

Voice sensitivity

Figure 1 displays the log transformed F0+VTL JNDs for levels of different conditions, including voice training, vocoder and item variability.

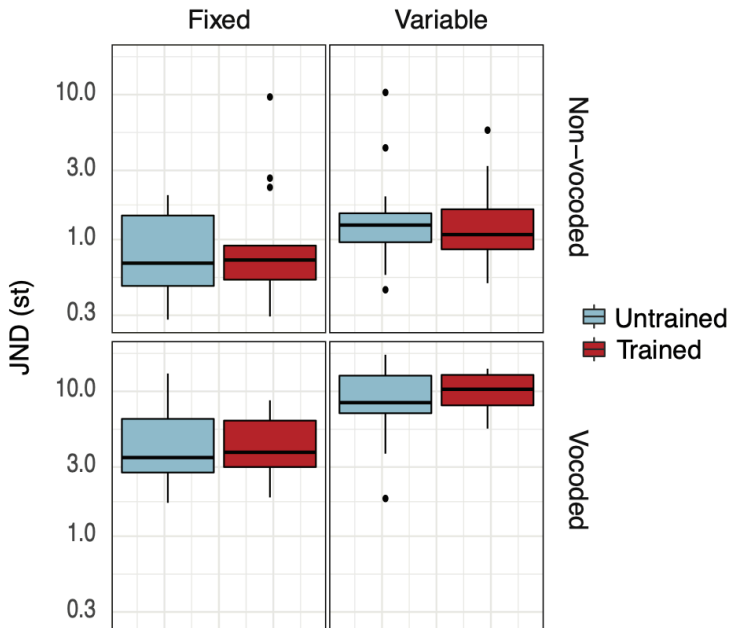

Figure 1 — Voice cue sensitivity in F0+VTL JNDs for conditions of voice training (untrained, trained), item variability (fixed, variable) and vocoder (non-vocoded, vocoded) is shown. JNDs shown in the y-axis are plotted on a logarithmic scale. The boxes illustrate the upper and lower quartiles, and the median is represented as the midline of the boxes.

The results show that vocoder manipulations had the largest main effect on the JNDs [$F_{(1,15)}$ = 346.74, p < 0.001, η^2_g = 0.69]. When vocoded signals were presented, on average, JNDs were significantly larger (7.33 st) than when non-vocoded signals were presented (1.45 st). The second largest main effect was observed with item variability manipulations [$F_{(1,15)}$ = 65.03, p < 0.001, η^2_g = 0.21]. When variable items were presented, average JNDs were significantly larger (5.80 st) than when fixed items were presented (2.99 st). Contrary to expectation, voice training did not have a significant main effect on the JNDs [$F_{(1,15)}$ = 1.16, p = 0.298, η^2_g = 0.00].

There was a significant interaction between vocoder and item variability conditions on the JNDs, [$F_{(1,15)}$ = 8.82, p < 0.05, η^2_g = 0.01]. A pairwise t-test was conducted for multiple comparisons and results showed that item variability conditions (variable vs fixed) were significantly different from each other in non-vocoded conditions [$t_{(15)}$ = 5.38 , p_{FDR} < 0.001], and in vocoded conditions [$t_{(15)}$ = 8.43, p_{FDR} < 0.001]. There were no other significant interactions.

Listening effort

ANOVA results

Figure 2 displays the pupil dilation response (mm) over time, where pupil traces were averaged over all participants and per conditions.

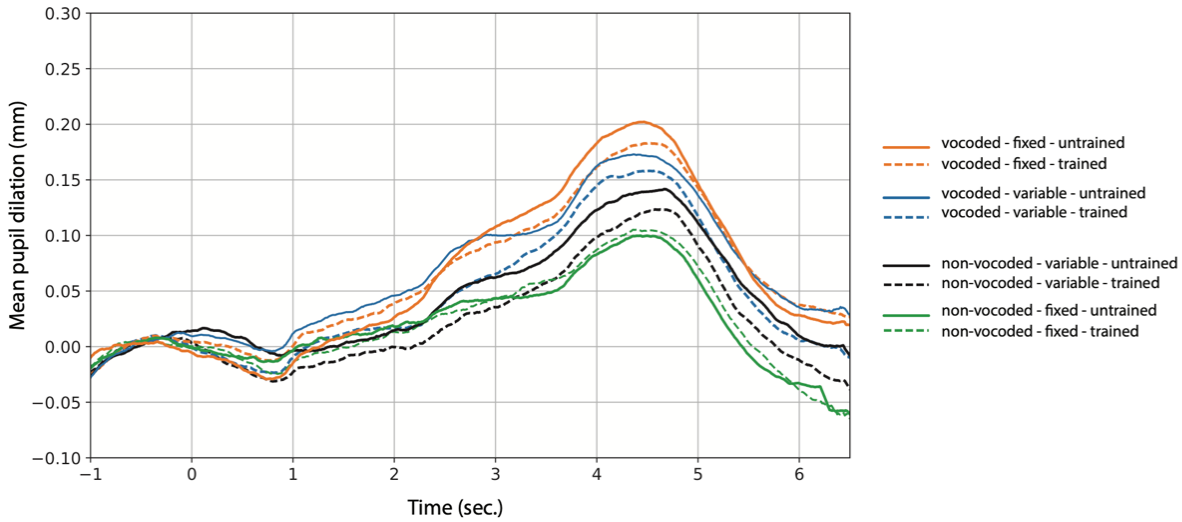

Figure 2 — Mean pupil dilation (mm) over time (sec) for all conditions is shown. Time window between -1 and 0 seconds is denoted as the baseline and pupil dilation response to stimuli presentation is illustrated from seconds 0 to 6.5. Solid lines represent untrained voices and dashed lines represent trained voices.

As shown in Figure 2, pupil dilation responses were larger when speech was vocoded, compared to non-vocoded, which was also shown by a significant main effect of vocoder on the PPD [$F_{(1,15)} = 14.82$, $p < 0.01$], MPD [$F_{(1,15)} = 10.90$, $p = < 0.01$] and on PPDL [$F_{(1,15)} = 7.50$, $p < 0.05$]. According to ANOVA results, there was no significant main effect of voice training on the PPD [$F_{(1,15)} = 2.38$, $p = 0.144$], MDP [$F_{(1,15)} = 1.31$, $p = 0.270$] or PPDL [$F_{(1,15)} = 0.72$, $p = 0.409$]. There was also no significant main effect of item variability on PPD [$F_{(1,15)} = 0.00$, $p = 0.976$], MPD [$F_{(1,15)} = 0.06$, $p = 0.812$] or PPDL [$F_{(1,15)} = 0.01$, $p = 0.945$]. However, item variability had a significant effect on the baseline pupil dilation [$F_{(1,15)} = 12.67$, $p < 0.01$]. During baseline period, pupil dilation responses were significantly larger when variable items were presented compared to fixed items. There was also a significant interaction between item variability and vocoder conditions on the baseline [$F_{(1,15)} = 4.88$, $p < 0.05$]. Multiple comparisons showed that variable and fixed items were significantly different from each other when the signal was vocoded [$t_{(28,7)} = 3.95$, $p_{FDR} < 0.001$], and no significant difference between levels of item variability condition was observed when the signal was non-vocoded [$t_{(28,7)} = 0.51$, $p_{FDR} = 0.616$]. There were no other significant interactions on the pupil outcome measures.

GAMM analysis results

We carried out GAMM analysis to assess pupil dilation response over time, where smooth functions are fitted to the data and the relationship between conditions can be non-linear. Details of performing GAMM analysis with pupillometry data can be found at van Rij et al. (2019).

We looked at the pupil dilation response over time between voice training and vocoder conditions as a planned comparison, by pooling over data from item variability, since item variability had no effect on the pupil dilation response during the stimuli presentation.

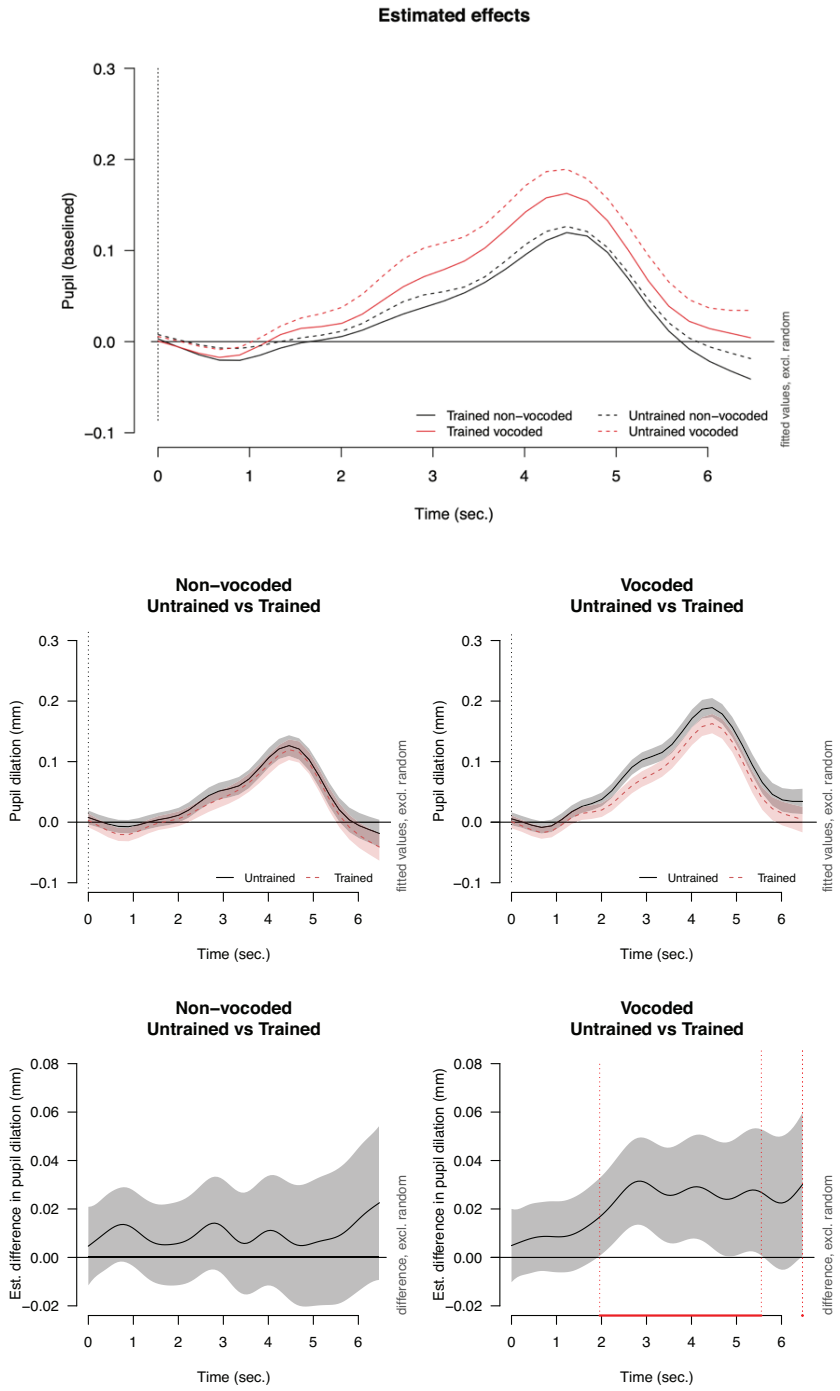

Figure 3 — GAMM analysis results for untrained and trained voices within non-vocoded and vocoded conditions are shown. Estimated effects of trained and untrained voices in vocoded and non-vocoded conditions are plotted in the estimated effects plot. Estimated differences in pupil dilation responses across time are plotted for contrasts of untrained and trained voices when speech is non-vocoded and vocoded. The time period where the estimated difference in pupil dilation between the untrained and trained voices significantly differs from the 0 line is indicated by the red line in the lower right panel x-axis.

Figure 3 shows the estimated differences in the pupil dilation between untrained and trained voices for non-vocoded and vocoded conditions. The red line on the below panel of Figure 3 illustrates the deviation from the 0 line. Visual inspection of Figure 3 shows that when the speech signal was vocoded, pupil dilation was significantly larger when listening to untrained voice compared to trained voice. There was no significant difference between pupil dilation response

to untrained and trained voices when speech signal was non-vocoded. This difference was confirmed by Ordered Factor Difference (OFD) statistics (p = 0.034).

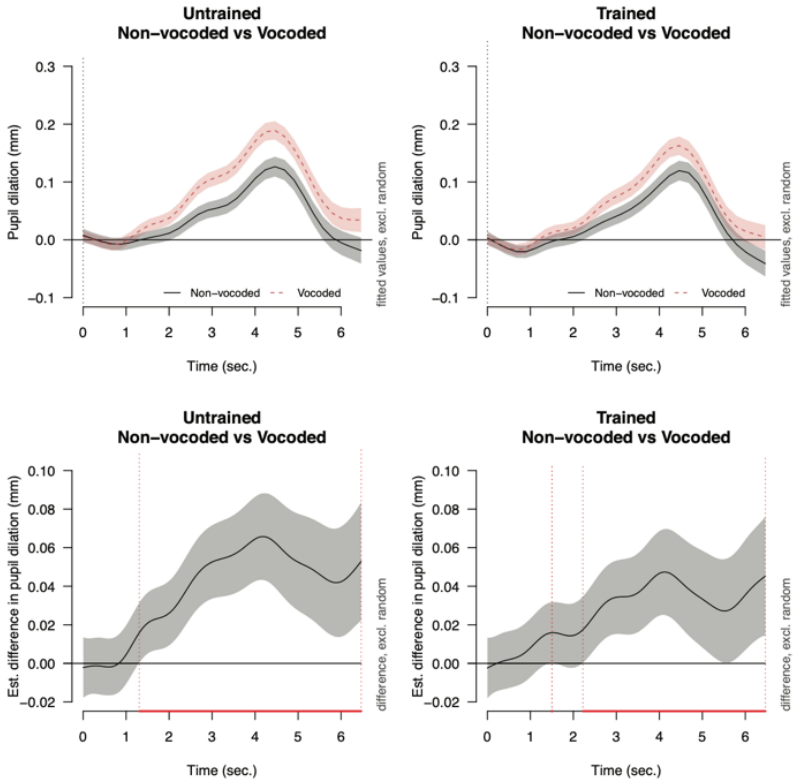

Figure 4 — GAMM analysis results showing differences in pupil dilation responses across time for contrast between non-vocoded and vocoded speech for untrained and trained voices is shown. The time period where the estimated difference in pupil dilation between the non-vocoded and vocoded speech significantly differs from the 0 line is indicated by the red line in the lower panel x-axis of the plot.

Figure 4 shows the contrast between non-vocoded and vocoded speech for untrained and trained voices. Visual inspection of Figure 4 and the OFD statistics shows that pupil dilation responses to vocoded and non-vocoded voices were significantly different when voices were both untrained and trained. However, OFD statistics also showed an interaction between training and vocoding (p = 0.031), which indicates that estimated differences (contrasts) for non-vocoding and vocoding in pupil dilations were smaller for trained voices than for untrained voices. Deviance explained from the final model was 77.1%.

Discussion

The present study examined the effect of voice training, as a way of providing voice familiarity, on voice sensitivity through F0+VTL voice cue discrimination and listening effort. Acoustic linguistic variability of the stimuli was manipulated with the item variability condition by presenting variable or fixed items. Additionally, stimuli were either vocoded or non-vocoded.

Our results showed that voice training did not influence voice cue discriminability. Sensitivity to small differences in F0+VTL, at threshold level, did not improve with voice familiarity. It is possible that other covarying voice cues such as pitch contour and speaking rate might have additionally been in use during voice discrimination. Therefore, voice discrimination based only

on using F0+VTL voice cues might not be sufficient to observe a familiarity benefit. GAMM analysis performed as a planned comparison showed that voice training had an effect on the pupil dilation response only when speech was vocoded. Pupil dilations were larger, indicating more effort, when listening to untrained vocoded voices than listening to trained vocoded voices. This effect lasted almost the entire duration of the stimulus presentation, from 2 seconds until 5.5 seconds. However, we did not see the effect of voice training on the pupil dilation response, when speech was non-vocoded. This suggests that, when speech is vocoded and input demands were high, voice familiarity might have acted like a compensatory mechanism for perception of degraded speech (Başkent et al., 2016). The results from both contrasts shown in Figures 3 and 4 indicated that trained voices were less affected by vocoding. Listening effort decreased for trained voices compared to untrained voices when speech was vocoded, while the contrast between vocoded and non-vocoded speech was also smaller for trained voices, compared to untrained voices.

As expected, vocoded speech resulted in significantly poorer discrimination of F0+VTL voice cues. Additionally, increased acoustic linguistic variability through variable item presentation also hindered voice cue discrimination performance. There was also an interaction between item variability and vocoder conditions. JNDs were significantly larger (performance was worse) when variable items were presented compared to fixed items when speech was both vocoded and non-vocoded.

Pupillometry data showed that listening to vocoded signals resulted in significantly larger pupil dilation response, reflected by PPD, MPD and PPDL. Winn et al. (2015) showed that while understanding speech, listening effort increased with decreased spectral bands in a vocoder which increases signal degradation. In addition to findings of Winn et al. (2015), our results show that pupil dilation responses were larger with vocoder manipulations during voice discrimination. Item variability was shown to only affect the pupil baseline, which, according to the adaptive gain theory (Aston-Jones & Cohen, 2005), can be interpreted as an increase in anticipated task difficulty, when variable items were presented compared to fixed items. The interaction between vocoder and item variability in the pupil baseline shows that anticipation of increased task difficulty specifically happened when speech was vocoded and variable items were presented, which was indeed the hardest condition in the experiment.

Together our findings suggest that voice familiarity through implicit voice training, which is similar to becoming familiar with specific voices in real life, can be of benefit at the level of voice cue perception through reducing listening effort for vocoded speech. This way CI users might benefit from spared cognitive resources which can be used for other listening related tasks such as understanding speech, remembering what was being said, and voice discrimination.

Acknowledgements

We would like to thank Etienne Gaudrain for experimental software, data analysis and helpful discussions on study design; Terrin Tamati for helpful discussions on study design. This study was funded by the VICI grant 918-17-603 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw), Heinsius Houbolt Foundation, and the Rosalind Franklin Fellowship.

References

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci., 28*(1), 403-450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., Sarampalis, A., Wagner, A., & Gaudrain, E. (2016). Cognitive compensation of speech perception With Hearing Impairment, Cochlear Implants, and Aging: How and to What Degree Can It Be Achieved? *Trends in Hearing*, *20,* 233121651667027. https://doi.org/10.1177/2331216516670279

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC Syllables and Sentences for Listeners with Normal Hearing and with Three Types of Hearing Impairment. *Audiology*, *34*(5), 260-284. https://www.tandfonline.com/doi/abs/10.3109/00206099509071918

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *114*(5), 2913-2922. https://doi.org/10.1121/1.1616924

El Boghdady, N., Gaudrain, E., & Başkent, D. (2019). Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users? *The Journal of the Acoustical Society of America*, *145*(1), 417–439. https://doi.org/10.1121/1.5087693

Gaudrain, E., & Başkent, D. (2015). Factors limiting vocal-tract length discrimination in cochlear implant simulations. *The Journal of the Acoustical Society of America*, *137*(3), 1298–1308. https://doi.org/10.1121/1.4908235

Gaudrain, E., & Başkent, D. (2018). Discrimination of Voice Pitch and Vocal-Tract Length in Cochlear Implant Users. *Ear and Hearing*, *39*(2), 226–237. https://doi.org/10.1097/AUD.0000000000000480

Holmes, E., Domingo, Y., & Johnsrude, I. S. (2018). Familiar Voices Are More Intelligible, Even if They Are Not Recognized as Familiar. *Psychological Science*, *29*(10), 1575–1583. https://doi.org/10.1177/0956797618779083

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

Kawahara, H., & Irino, T. (2005). Underlying Principles of a High-quality Speech Manipulation System STRAIGHT and Its Application to Speech Segregation. In: Divenyi P. (Eds.), *Speech Separation by Humans and Machines* (pp. 167–180). Springer, Boston, MA. https://doi.org/10.1007/0-387-22794-6_11

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil Dilation Uncovers Extra Listening Effort in the Presence of a Single-Talker Masker. *Ear and Hearing*, *33*(2), 291–300. https://doi.org/10.1097/AUD.0b013e3182310019

Koelewijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E. (2021). The effect of monetary reward on listening effort and sentence recognition. *Hearing Research*, *406*, 108255. https://doi.org/10.1016/j.heares.2021.108255

Kreitewolf, J., Mathias, S. R., & von Kriegstein, K. (2017). Implicit Talker Training Improves Comprehension of Auditory Speech in Noise. *Frontiers in Psychology*, *8*, 1584. https://doi.org/10.3389/fpsyg.2017.01584

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–376. https://doi.org/10.3758/BF03206860

Rabbitt, P. (1966). Recognition: Memory for words correctly heard in noise. *Psychonomic Science*, *6*(8), 383–384. https://doi.org/10.3758/BF03330948

Van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the Time Course of Pupillometric Data. *Trends in Hearing*, *23*, 233121651983248. https://doi.org/10.1177/2331216519832483

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear and Hearing*, *36*(4), e153–e165. https://doi.org/10.1097/AUD.0000000000000145