Perceptual flexibility at an advanced age: Training seniors to perceive a nonnative voicing contrast

Sidsel Holm Rasmussen[1*], Jonas Albæk Villumsen[1], Birgitte Poulsen[1], Ocke-Schwen Bohn[1*]

[1]Aarhus University, Denmark

Abstract. This paper presents data from a perception study in which the effects of nonnative phonetic training in listener groups of juniors and seniors are compared. Specifically, the study includes L1 Danish participants (aged 20–30 and 60–76) and two age-matched control groups who were tested on their ability to identify the L2 English sibilants /s/ and /z/ in 2AFC tasks. The junior (n = 14) and senior (n = 15) experimental groups were asked to complete 10 online sessions which trained, with immediate feedback, their perception of the contrast syllable-initially. Unlike the control group, both training groups had significantly improved identification scores on the trained identification task at post-test, and further analyses suggested little performance difference between the trained juniors and seniors over time. Furthermore, their learning gains appeared to be partially retained as neither age group showed significant decline in identification accuracy eight weeks after training had finished. In conclusion, we find that participants above the age of 60 demonstrate largely the same capabilities for phonetic learning as do the younger participants, and we suggest that an advanced age does not necessarily compromise speech learning ability.

Keywords. aging research, language learning, phonetic training, L2 speech perception, voicing contrast

A peer-reviewed contribution to *Journal for Language and Aging Research* (JLAR).
Submitted: 2025-05-19
Accepted: 2025-11-10
Published: 2025-12-29
DOI: 10.15460/jlar.2025.3.2.1806

*. Corresponding authors, ocke.bohn@cc.au.dk, sidsras@gmail.com

Please cite as Rasmussen, Sidsel Holm, Villumsen, Jonas Albæk, Poulsen, Birgitte & Bohn, Ocke-Schwen. 2025. “Perceptual flexibility at an advanced age: Training seniors to perceive a nonnative voicing contrast.” *Journal of Language and Aging Research* 3(2): 161–180. 10.15460/JLAR.2025.3.2.1806.

1 Introduction

The most widely used models of second language speech, PAM-L2 (Best and Tyler 2007) and SLM-r (Flege and Bohn 2021) claim that the ability to restructure one's phonetic system extends over the whole life span. Evidence supporting this claim comes from three sources: (1) As demonstrated by, e. g. Werker and colleagues (Werker and Tees 1984; Werker and Logan 1985), apparent decline in speech perception abilities with age reflects a shift in attention, not neural atrophy. (2) An individual's language environment causes shifts in attention that may be quite subtle (often with respect to the native language, see Harrington, Palethorpe, and Watson (2000) and Chang (2012) or dramatic (when this shift leads to native- or near-native like production and perception of new phonetic categories after long-term immersion in the nonnative language environment, see Flege, Takagi, and Mann (1995) and Garibaldi and Bohn (2019). (3) Shifts in attention can also be induced in lab settings through training regimes (e. g. Bradlow et al. 1997; Sereno and Wang 2007). Phonetic training studies typically use versions of the High Variability Perceptual Training (HVPT) paradigm in attempts to emulate naturalistic experience by exposing trainees to members of to-be-trained categories produced by multiple speakers in multiple phonetic environments (Logan and Pruitt 1995). The evidence for the malleability of adults' phonetic abilities from these three sources is massive (as reviewed by Bohn 2018, but incomplete in a way that is quite problematic for any claim of life-long perceptual flexibility. As pointed out by Henrich, Heine, and Norenzayan (2010, 33) in their critique of the behavioral sciences near-exclusive focus on participants from WEIRD[2] societies, "undergraduates . . . form the bulk of the database in the experimental branches of the behavioral sciences".

The database in second language speech research is not quite as narrow in that many studies include participants up to the age of ca. 35 years, but as the authors of a recent perceptual training study point out, "attention to older adults over 60 years is entirely lacking" (Zhang, Liao, and Truong 2024). A thorough review of the literature reveals that this statement is almost accurate. In a recent review of the effects of aging on bilingual language, Reifegerste (2021) notes that "to date, there has been relatively little research on the effects of aging on . . . aspects . . . of bilingual language such as phonetics, phonology . . . ". Among the very few studies of older adults' speech learning ability is a series of studies conducted by Kubo and colleagues. Kubo, Akahane-Yamada, and Yamada (2000) conducted a training study in which L1 Japanese listeners of various ages (in their 40s, 50s and 60s) were trained to identify American English /r/ and /l/. The authors reported significant improvements in accuracy for the three age groups, however, their results and that of an earlier study (Yamada 1993) suggested that the efficacy of training decreases by age. Kubo, Akahane-Yamada, and Yamada (2000) concluded that aging interferes with the learning of new sounds, but not completely. Likewise, a large-scale training study with 1000 L1 Japanese speakers between the ages of 10 and 70 concluded that amount of improvement decreases with age, but the elderly trainees can improve significantly (Akahane-Yamada, Takada, and Kubo 2002).

A more recent study by Kubo and colleagues (Kubo, Akahane-Yamada, and Akagi 2012; Kubo, Akagi, and Akahane-Yamada 2015) provided interesting additional

2. The WEIRD acronym stands for Western, Educated, Industrialized, Rich, and Democratic (Henrich, Heine, and Norenzayan 2010, 33).

information by examining training efficacy across age groups for American English /r/ and /l/ in different phonetic environments. Kubo, Akahane-Yamada, and Akagi (2012) and Kubo, Akagi, and Akahane-Yamada (2015) reported improvement for all age groups, but this had to be qualified for the positions in which this contrast was tested: All age groups showed comparable identification improvements for /r/-/l/ in final position, but the older adults improved less than the younger participants when tested on /r/-/l/ in initial clusters (Kubo, Akahane-Yamada, and Akagi 2012) and initial singletons (Kubo, Akagi, and Akahane-Yamada 2015). This observation fits in well with a study by Lively, Logan, and Pisoni (1993), who found that L1 Japanese listeners (no age specified) identified postvocalic /r/ and /l/ quite accurately but were less accurate at identifying these liquids in prevocalic position (both as singletons and in clusters).

Jähi, Alku, and Peltola (2015) compared two groups of L1 Finnish speaking seniors (aged 62–73 years) on the effect of listen-and-repeat training for the production of two nonnative vowel contrasts. The groups differed in that one had demonstrated an interest in language learning (i. e. had studied a foreign language at least once a week for 2–10 years), whereas the other had not. The authors concluded that learning to produce a non-native phoneme can be easier for elderly learners who show a general interest in languages, but that study did not examine whether the effect of interest in language learning would be different for different age groups.

Tamminen et al. (2021) trained L1 Finnish speaking seniors aged 61–71 years on the nonnative /f-v/ contrast and reported that training altered identification similarly in the elderly participants and in young adults who were studied earlier (Tamminen et al. 2015). Neither of the Tamminen et al. (2015) and Tamminen et al. (2021) studies included untrained age-matched control groups.

Felker et al. compared two L1 Dutch age groups (aged 18–31 and 65–84) for the effectiveness of explicit instruction regarding the English /æ-ɛ/ and word-final /t-d/ contrasts. They summarized their findings by stating that “a brief phonetic instruction can improve phonological awareness and perception of L2 contrasts in younger and older adult listeners” Felker et al. (2023, 27).

Even more recently, Zhang, Liao, and Truong (2024) studied how L1 English-speaking adults (age 60–64 years) benefitted from perceptual training of Mandarin lexical tones and whether training efficacy was related to the participants’ perceptual acuity (as determined by an adaptive pitch discrimination task) and their ability to respond correctly to melodies consisting of five notes differing in fundamental frequency by indicating the height of these notes by clicking on screen displays. The training efficacy of the trainees in the Zhang, Liao, and Truong (2024) study was compared to an aged-matched control group which was not trained.

All participants took a 4AFC identification task and a discrimination task before training, immediately after training, and at a delayed post-test (two months after training had finished). Training was conducted over four weeks in eight sessions each lasting 20 minutes. Zhang, Liao, and Truong (2024) reported that the trainees exhibited a more pronounced improvement in tone categorization than the control group, and that individual differences in perceptual acuity were significantly related to gains from training. While the Zhang, Liao, and Truong (2024) study suggests that seniors exhibit perceptual flexibility in that they benefit from perceptual training, that leaves open the question of how the training efficacy for seniors compares to that for juniors.

If the ability to restructure one's phonetic system extends over the entire life span, as claimed by current models, then this claim needs more support from studies with participants who are quite a bit older than what is almost exclusively found in speech learning studies. The present study is motivated by the obvious gaps in the literature: First, few studies have directly compared training efficacy across age groups. Second, to the best of our knowledge, no study has ever compared the training trajectory across age groups. Are there differences between age groups regarding the effect of training across training sessions? Third, nothing is known about age differences regarding the robustness of training effects. If training is effective, does the training effect last or does it decline after the last training session, and are there age differences? Fourth, does training a new contrast in one syllable position extend to improved perceptual accuracy of the same contrast in a different position and if so, are there age differences?

The present study addressed these questions in a pretest-training-posttest-delayed posttest design with two groups of native Danish speakers, a group of juniors aged 20–30 years, and a group of seniors aged 60–76 years. With respect to the senior group, the aim of the study is not to be representative of seniors' speech learning ability in general but to determine whether chronological age per se affects the ability to restructure speech perception for the contrast chosen for our study. We chose the English /s/-/z/ contrast because Danish does not have voiced fricatives like /z/, and because previous studies have shown that native Danish listeners with little English-language experience have problems identifying tokens from this contrast correctly (Trapp and Bohn 2002; Romano-Hvid 2003; Bohn and Ellegaard 2019).[3] The present study examined the identification of syllables exemplifying the /s/-/z/ contrast to address the general question of how an advanced chronological age affects speech learning ability. The specific research questions are:

- RQ1: Does the effect of training (difference in perceptual accuracy before and after training) differ between the two age groups?
- RQ2: Do the two age groups differ with respect to their learning trajectories? That is, does training lead to improved accuracy at different rates throughout the training period?
- RQ3: Do the two age groups differ with respect to any lasting effect of the training?
- RQ4: Does training a new contrast in one syllable position extend to improved perceptual accuracy of the same contrast in a different position and if so, are there age differences?

Addressing these questions will, we expect, shed light on the general issue of seniors' cognitive abilities. Traditionally, an advanced age was often considered to be a period

3. Trapp and Bohn (2002) trained L1 Danish adolescents to identify coda /s/ and /z/ from minimal pairs and reported a mean identification accuracy before training of 59 %. After four training sessions of 30 minutes duration each, perceptual accuracy increased to 73 %. While the trainees' mean identification scores had improved after training, there was no significant improvement in their measured production accuracy. In the Bohn and Ellegaard (2019) study, inexperience native Danish listeners identified the initial consonant in [s] tokens as /s/ at a rate of 63.6 % (and as /z/at 31.3 %), and the initial consonant in [z] tokens as /z/ at a rate of 54.2 % (and as /s/ at 35.4 %).

of inevitable cognitive decline (Deary et al. 2009). However, the validity of this view is challenged by more recent studies which suggest that "more people are living to older ages with better overall functioning" (Christensen et al. 2013, 1507) and which has led Ramscar et al. (2014) to speak of a "myth of cognitive decline". With respect to the cognitive challenges involved in learning additional languages, few studies have examined grammatical language learning abilities in seniors,[4] and those who have report that "learning a new language in old adulthood is certainly feasible" (Kliesch et al. 2018, 63). The present study focuses specifically on nonnative speech learning ability in seniors.

2 Methods

2.1 Participants

We recruited two groups of participants: Juniors between the ages of 18 and 30 (which is the age group typically found in training studies) and seniors between the ages of 60 and 78. The junior participants were recruited, among others, from a webpage seeking research participants among Aarhus University students, from postings at the university library, and by word of mouth. The senior participants were recruited mainly through postings at the local libraries, through our visit to a lecture at *Aarhus Folkeuniversitet*, which provides lectures and seminars open to the public, and by word of mouth. Interested individuals were asked to fill out a brief online questionnaire on the projects' home page which elicited information about the native language, age, nonnative language experience, and hearing ability. 71 individuals who were either between 18 and 30 years old, or between 60 and 78 years, who had not spent more than six months in an English-speaking country or had experience with a language that the /s/-/z/ contrast, and who reported normal hearing (including normal hearing with a hearing aid), were invited to visit our lab and participate in the study.

At the first visit, three seniors did were excluded based on the results from a short hearing test (see below). The remaining 68 participantsthen took part in a 2AFC identification task with /sV/ and /zV/ syllables (see below, Procedure and Stimuli). One purpose of this first visit was to only include potential participants in the training study whose identification accuracy allowed for improvement in the training regime, which is why we set a threshold of less than 90 % identification accuracy for this identification task (following the procedures of previous training studies, (e. g. Saito et al. 2022). Of the original 71 candidates, 10 junior and 10 senior candidates tested above the inclusion threshold and were not invited to participate in the experiment. Of the remaining 48 participants, two juniors had not trained as agreed and were excluded, so the results of the present study are based on 46 participants: 23 juniors (16f), age range 20-30 years, mean age = 23.2 years (SD = 2.1) and 23 seniors (15f), age range 60–76 years, mean age 65.2 years, SD = 4.1). Participants were randomly assigned to an experimental group which took part in the training, or a control group which was not trained. The junior groups consisted of 14 trainees (10 f) and 9 controls (6 f), the senior groups of 15 trainees (9 f) 8 controls (6 f).

4. An extensive overview is provided by the contributions to Gabrys-Barker 2018.

2.2 Procedure and Stimuli

Table 1: Timeline of sessions

Pre-training	Training	Post-training	Delayed Post-training
Preparation			
Informed consent, hearing thresholds, instructions using PowerPoint: Voicing in /s-z/			
Production tasks			
/s-z/ delayed repetition, initial and final	10 training sessions, ca. 15 minutes each, ca. 3 days apart	/s-z/ delayed repetition, initial and final	/s-z/ delayed repetition, initial and final
Identification tasks			
Identification task I (initial /f-v/) Identification task II (final /s-z/) Identification task III (initial /s-z/)		Identification task II (final /s-z/) Identification task III (initial /s-z/)	Identification task II (final /s-z/) Identification task III (initial /s-z/)

Table 1 presents the timeline of the study. In individual sessions, each participant visited our lab when first joining the experiment. They were informed about the purpose and design of the experiment and signed an informed consent form. We then assessed their hearing thresholds, which for most participants was at least 35 dB HL at 500, 1000, 2000, and 4000 Hz on one ear.

Participants were then shown a PowerPoint presentation on a laptop computer with instructions in Danish. The presentation also contained links to the different perception tasks. The participant controlled the presentation via mouse clicks. An experimenter was present at every meeting, and participants were given the opportunity to ask clarifying questions throughout the experiment.

The procedure for each participant's first meeting is outlined here below:

1. Introduction to the topic of the experiment, including an explanation of /s/ and /z/ being contrastive in English, and /z/ being pronounced with voicing.

2. Instructions on how the fricative voicing contrast is created including a prompt to make the participant feel the presence of vocal fold vibration with their hand on their throat while making a /z/ sound (like a buzzing bee) and the absence of voicing when making an /s/ sound (like a hissing snake).

3. Instructionsto adjust the volume to a comfortable hearing level.

4. How to carry out the production task.[5]

5. The present report focuses on the perception aspects of the training, the results from the production task will be reported elsewhere.

JLAR 3 (2025)
10.15460/jlar.2025.3.2.1806

5. 46 slides containing the delayed-repetition task and sound files.

6. Instructions on how to conduct the perception tasks (I, II, III), which were built using the web-tool PERCY (Draxler 2011)

7. Perception Task I: Familiarization task, identification of /fV/-/vV/. Based on the results of earlier studies (Bohn and Ellegaard 2019; Horslund and Bohn 2022) we assumed that this would be an easy task for our participants, whose L1 has a /f-/ contrast. The stimuli were tokens from the Shannon et al. (1999) corpus with V = /i, , u/, produced by the male talker who Shannon et al. (1999) reported to be one of the most intelligible of their 10 talkers.[6] Participants were presented with a randomized set of 30 stimuli (2 initial consonants x3 vowels x 5 tokens) which had been normalized for peak intensity and to which participants responded by clicking on one of two buttons marked <F> and <V>.

8. Perception Task II: Identification of syllable-final /Vs/-/Vz/. This task was included to examine whether training of the /s/-/z/ contrast in initial position would have an effect on the perception of the contrast in syllable-final position. The tokens, with V = /i, , u/, were obtained from recordings of two native English speakers (1f, 1m), normalized for peak intensity, and validated by three native English speakers. Participants were presented with a randomized set of 60 stimuli (2 final consonants x 3 vowels x 5 tokens x 2 talkers) to which they responded by clicking one of two buttons marked <S> and <Z>.

9. Perception Task III: Identification of syllable-initial /sV/-/zV/. This was the main focus of the study, in which we compared the effect of perceptual training of syllable-initial /s/ and /z/ on the identification of the contrast in initial position. The stimuli were tokens from the Shannon et al. (1999) corpus with V = /i, , u/, produced by the female and male talker who the authors reported to be the most intelligible of their 10 talkers. Participants were presented with a randomized set of 60 stimuli (2 initial consonants x 3 vowels x 5 tokens x 2 talkers) which had been normalized for peak intensity and to which participants responded by clicking on one of two buttons marked <S> and <Z>.

Steps 4. and 5. are irrelevant to the present study because the results of the production tasks from before and after training will be reported elsewhere. All perception tasks (I–III) were initiated by clicking on a PERCY link which started the experiment, which in all instances was conducted as a 2AFC identification task.[7] Results from perception task I (identification of /fV/ vs. /vV/) will not be reported here; as expected, all participants performed at or near ceiling.

The participants' first visit to the lab ended with the first training session. Training consisted of the presentation of 120 trails of /sV/ and /zV/ syllables (i. e. two randomizations of the tokens presented in Perception Task III), to which the participants responded by clicking on one of two boxes labeled "S" and "Z". If the response was correct, the box lit up in green, if it was incorrect, the box lit up in red, and the

6. Shannon et al. (1999) reported intelligibility scores ranging from 96.1 % to 98.9 % for the ten talkers. The tokens used for the present study were from talkers whose intelligibility score was 98.9 %.

7. Unlike discrimination tasks in which listeners are likely to attend to within-category differences, identification tasks have been reported to promote more robust perceptual learning (Logan and Pruitt 1995).

stimulus was played again. This direct feedback was provided because, as Roediger and Butler (2011, 20) pointed out, "general (right or wrong) feedback is not very helpful if the correct answer is not provided. Correct answer feedback usually produces robust gains". At the end of each training session, participants were informed about the proportion of correct responses. Training sessions lasted ca. 15 minutes.

After the first meeting, we made an appointment for a second visit with the control participants, which was scheduled for three weeks after the first meeting, i. e. the time it took the trainees to complete the training sessions. The trainees borrowed headphones (Sennheiser HD 560S) to be used during the training sessions at home and a training schedule for the next three weeks, according to which the training sessions were spaced three days apart. We also arranged dates for a second meeting after training had ended (post-test), as well as an optional third meeting for the delayed post-training test two months after training had ended which 75 % of the trainees were available for. Trainees received the link to the experiment via e-mail, and they were encouraged to conduct 9 training sessions at home in a quiet room with minimal background noise. At the beginning of each session, the software asked participants about the environment in which they conducted the training session, the device used to run the session, (e. g. laptop or tablet), and the sound source, and the participants then adjusted the volume to a comfortable hearing level.

All participants received vouchers at the second meeting (worth 200 DKK for the controls and 400 DKK for the trainees). An additional voucher worth 100 DKK was offered to trainees after their third meeting (Delayed Post-Test).

2.3 Results

Data from the perception tasks II, III, and training sessions were downloaded from PERCY (Draxler 2014) and merged in R studio (R Core Team 2023). Trials in which English /s/ and /z/ were correctly identified received a score of 1, and incorrect responses received a score of 0. Before training all participants identified syllable final and syllable initial /s/ and /z/ in two separate tasks (II and III). The threshold for inclusion in the present study was a maximum score of 90% in the identification task targeting the initial English sibilants (Task III). The participants included had a mean identification score of 69.5 % in task II (final sibilants), and a mean score of 71.8 % in task III (initial sibilants). Figure 1 shows the sample means for the four separate groups at Pre-Test. Whereas group means were quite similar for Task III (initial; due to pre-established inclusion threshold), Task II accuracy means vary more across the individual groups, most notably with the Seniors Training group scoring higher than the Junior Training group, which has implications for an analysis that implements Task II (final fricative identification).

A total of 29 participants were assigned to the two training groups (14 Juniors and 15 Seniors), but only 14 of the senior trainees completed the post-training test of initial /s/ and /z/ due to a procedural error in the second time of testing. Data from trainee (P36) for whom we lack a crucial Task III Post-Test data pointwere excluded from the statistical analyses. Five trainees (2 Juniors and 3 Seniors) had completed one training session more than planned, and one (Junior, P25) had only completed six training sessions. In analyzing the training data we opted to remove the 11th training session of these five trainees and keep the six training sessions for P25. Since not all trainees were available to participate in the delayed post-test, the analysis related to learning retention includes 21 participants (9 Juniors and 12 Seniors) from whom we collected Task III test results at all three times of testing.

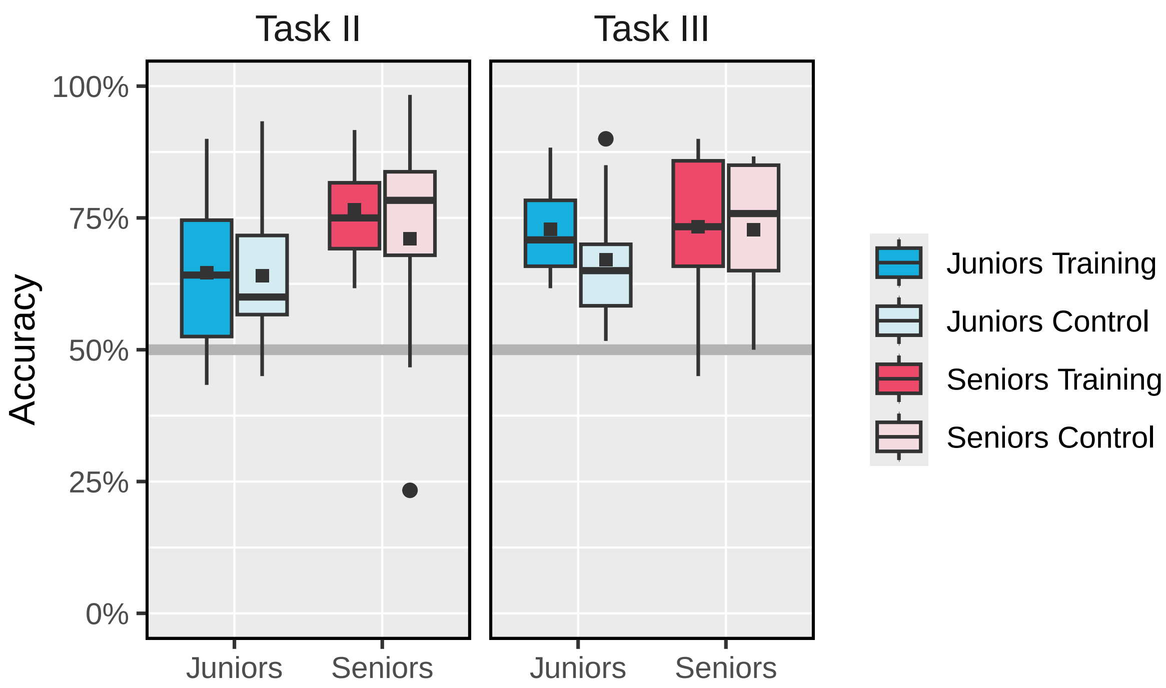

Figure 1: Boxplots of the distribution of accuracy scores for identification of English /s/ and /z/ at pre-test (implemented as the simple mean of hit/miss trial outcomes by each participant), stratified by group and task (task II: syllable-final and task III: syllable initial).

Addressing the specific research questions (RQ1–4) we applied mixed-effects logistic regression models using the *lme4* package (Bates et al. 2015) in R (R Core Team 2023). In the following we address each of our research questions separately.

2.4 Training effects on /s/ and /z/ identification in initial (trained) position

For RQ1, we modeled log-odds accuracy as a function of *Time* (Pre-test*, Post-test), *Age Group* (Juniors*, Seniors), and *Treatment* (Training*, Control), including all interactions. All predictors were treatment coded with (*) marking the reference level. Random effects included intercepts for participants and fricative (/s/, /z/), with by-participant slopes for Time.

Figure 2 shows the estimated marginal means derived from the RQ1 model, and here we summarize the modeled results: At pre-test, junior trainees performed at 73.9 % accuracy, improving to 94.2 % accuracy at post-test, $\beta = 1.74$, $z = 6.838$, $p < .001$. At pre-test, senior trainees performed at 74.9 % accuracy, improving to 90.3 % accuracy. Seniors did not perform statistically significantly differently from juniors at pre-test, $\beta = 0.05$, $z = 0.229$, $p > .05$.

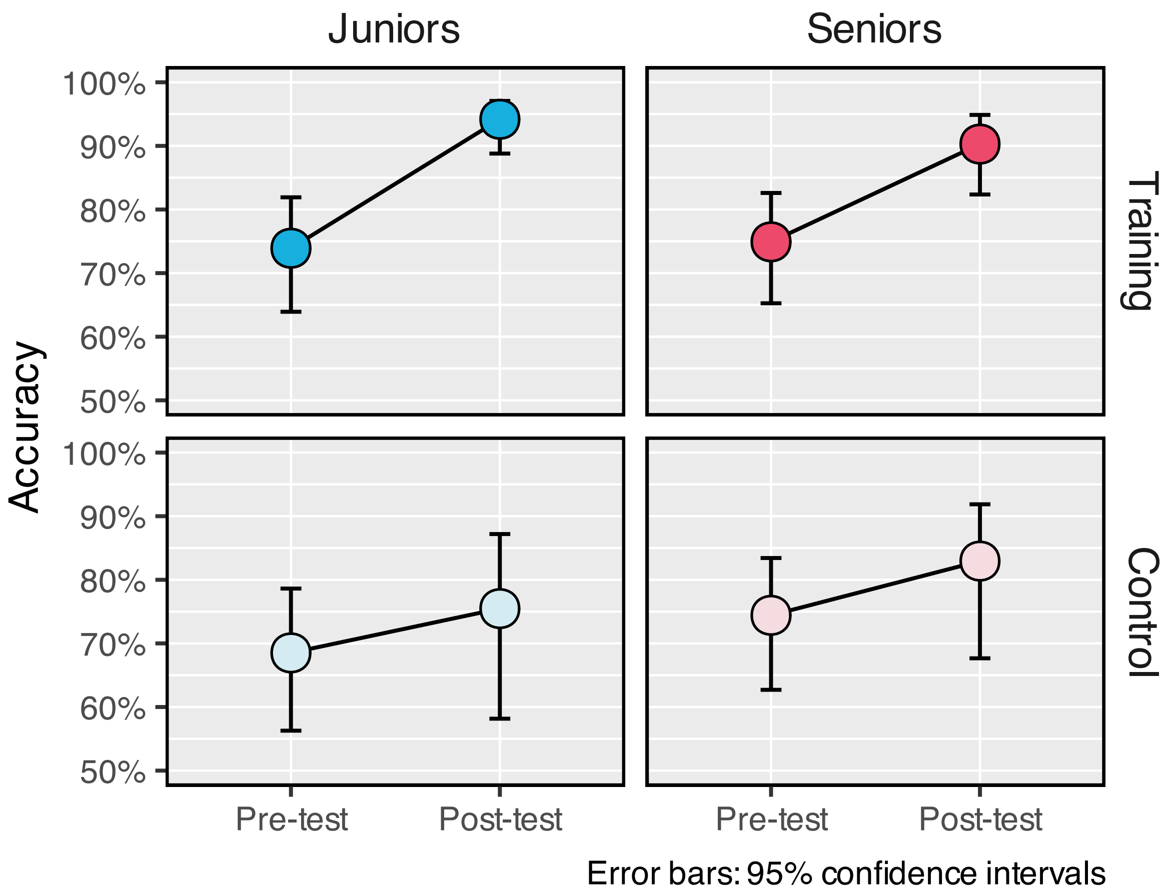

Figure 2: Estimated marginal means derived from the RQ1 model transposed onto the % correct scale

The interaction term between *Time: Post-test* and *Age Group: Seniors* does not indicate a difference between junior and senior trainees in their improvement from pre- to post-test, $\beta = 0.61$, $z = 1.76$, $p > .05$. The junior control group performed at 68.0 % accuracy on the pre-test and at 75.5 % accuracy at post-test. The difference between junior controls and trainees at pre-test was not statistically significant, $\beta = 0.26$, $z = 1.01$, $p > .05$. The interaction term between *Time: Post-test* and *Treatment: Control* suggests that the rate of improvement from pre- to post-test for the junior controls was lower than for the junior trainees, $\beta = 1.40$, $z = 3.76$, $p < .001$. The senior control group performed at 74.4 % accuracy at pre-test and at 82.9 % at post-test. The interaction term between *Age Group: Seniors* and *Treatment: Control* shows that, at pre-test, the performance of the senior control group was not statistically significantly different from the performance of the junior control group, $\beta = 0.24$, $z = 0.63$, $p > .05$. Finally, the interaction term between *Age Group: Seniors*, *Treatment: Control*, and *Time: Post-test* shows that the rate of improvement for senior controls from pre- to post-test was not statistically significantly different from junior controls, $\beta = 0.78$, $z = 1.46$, $p > .05$.

2.5 Learning trajectories for junior and senior trainees

For RQ2, we modeled log-odds accuracy as a function of *Training Session* (numeric; first session = 0, increments of +0.2 per session; sixth session = 1) and *Age Group*, with an interaction term. Random effects included participant intercepts, by-participant *Training Session* slopes, and fricative intercepts. *Training Session* was treated as numeric to model average improvement across the training span. Setting the comparison-level (1) to the sixth session estimates whether a difference in accuracy trajectories between juniors and seniors has emerged by the halfway point.

Figure 3 shows the estimated marginal means derived from the RQ2 model. At training session 1, junior trainees performed at 82.8% accuracy. At training session 6 (i. e. the approximate half-way point), they performed at 92.7% accuracy.

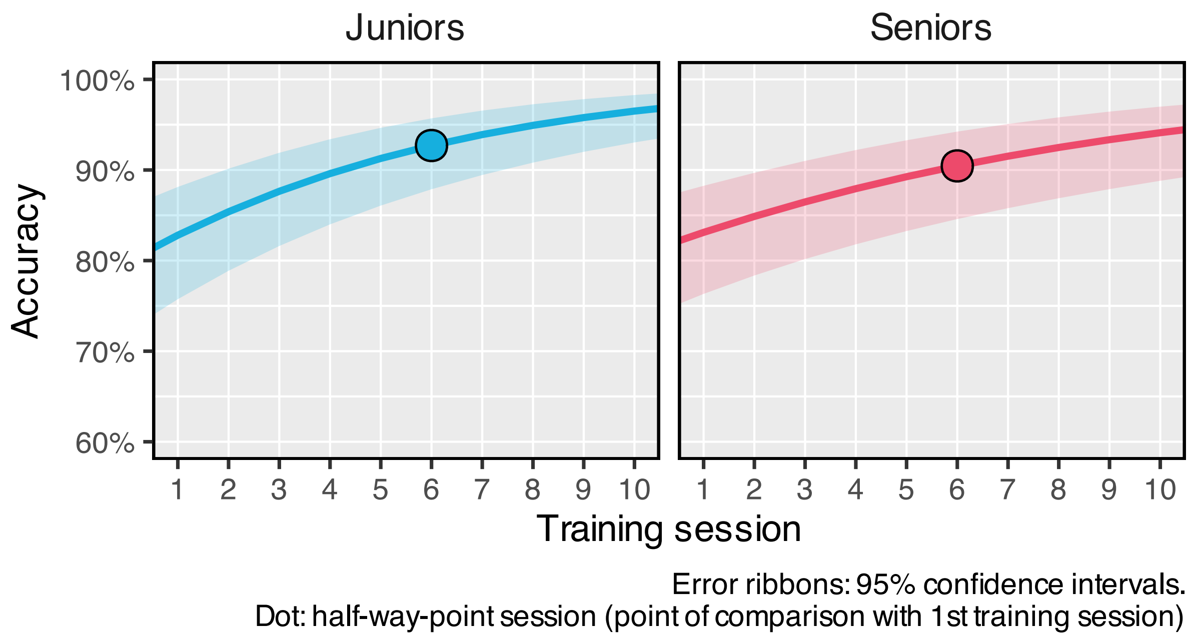

Figure 3: Estimated marginal means derived from the RQ2 model transposed onto the % correct scale

The difference is statistically significant, $\beta = 0.97$, $z = 6.92$, $p < .001$. At training session 1, seniors performed at 83.1% accuracy. Their performance at session 1 was not statistically significantly different from that of juniors', $\beta = 0.02$, $z = 0.076$, $p > .05$. At training session 6, seniors performed at 90.5% accuracy. The interaction term between *Age Group: Seniors* and *Training Session* does not indicate a statistically significant difference between juniors and seniors in the rate of improvement from training sessions 16, $\beta = 0.32$, $z = 1.65$, $p > .05$.

Due to the numeric coding scheme for the variable *Training Session* in the RQ2 model, the average rate of improvement for junior and senior trainees can be described as a linear function for each of the groups (outcome in log-odds):

- $\text{Rate}_{Juniors}$ = 1.57 + 0.19 Œ Training Session_i
- $\text{Rate}_{Seniors}$ = 1.59 + 0.13 Œ Training Session_i

These functions are what Figure 3 shows, albeit on the % correct scale and with added 95% confidence intervals. The confidence intervals indicate that, given hypothetical repeat sampling, and provided that the intervals are estimated using the same procedure for all samples, 95% of samples should contain the true population mean within their confidence interval. There are no assurances that the intervals presently shown include the true population mean.

2.6 Retainment of phonetic learning

For RQ3, we analyzed the subset who completed a delayed post-test in addition to the pre- and post-test ($n = 21$). Accuracy was modeled as a function of *Time* (Pre-test, Post-test, Delayed post-test*) and *Age Group*, with all interactions. The pre-test was included in the model to test whether accuracy at the delayed post-test had dropped to pre-training levels.

Figure 4 shows the estimated marginal means derived from the RQ3 model.

Figure 4: Estimated marginal means derived from the RQ3 model transposed onto the % correct scale. Note that the model compared Delayed post-test (Del. test) on the one hand with Pre-test and Post-test on the other. This is not reflected in the lines on the graph

Recall that this model only represents the subset of participants who completed all three tests. Thus, pre- and post-test results may differ from those of the RQ1 model. At the delayed post-test, junior trainees who completed all three tests performed at 92.8%, down from 97.8% at the post-test. The difference is statistically significant, $\beta = 1.25$, $z = 3.74$, $p < .001$. However, the junior trainees retained much of their learning, as accuracy at the delayed post-test was still higher than at the pre-test, where accuracy was at 73.9%, $\beta = 1.52$, $z = 6.84$, $p < .001$. As for senior trainees who completed all three tests, they performed at 87.1% accuracy at the delayed post-test, down from 91.9% at the post-test. The model provides no evidence that seniors performed worse than juniors at the delayed post-test, $\beta = 0.65$, $z = 1.86$, $p > .05$. Furthermore, the interaction term between *Time: Post-test* and *Age Group: Seniors* does not indicate a difference between juniors and seniors in the rate of attrition from post-test to delayed post-test, $\beta = 0.72$, $z = 1.76$, $p > .05$. The interaction term between *Time: Pre-test* and *Age Group: Seniors* does in fact suggest a smaller total improvement from pre-test to delayed post-test for seniors compared to juniors, $\beta = 0.74$, $z = 2.66$, $p < .01$. However, a post-hoc test conducted on the model's log-odds estimated marginal means indicated that seniors still performed better at the delayed post-test compared to the pre-test, *estimate* = 0.52, $z = 2.15$, $p < .05$.

2.7 Generalizing to another (untrained, syllable-final) phonetic environment

For RQ4, we fitted two separate models to estimate individual trainee gains (in log-odds) on syllable-initial and syllable-final /s, z/. Both models estimated accuracy as a function of *Time*(Pre-test*, Post-test) with random intercepts for participant and fricative, and by-participant Time slopes. To derive individual gain scores for the fricative positions, we extracted participant-specific random effects for Time: Post-test and summed them with the fixed Time: Post-test coefficient for the two models. The resulting scores were then analyzed in a third simple linear model which estimated syllable-final gains as a function syllable-initial gains, interacting with *Age Group*.

Figure 5 shows individual gain scores on syllable-final /s, z/ (the untrained stimulus set) predicted by gain scores on syllable-initial /s, z/ (the trained stimulus set).

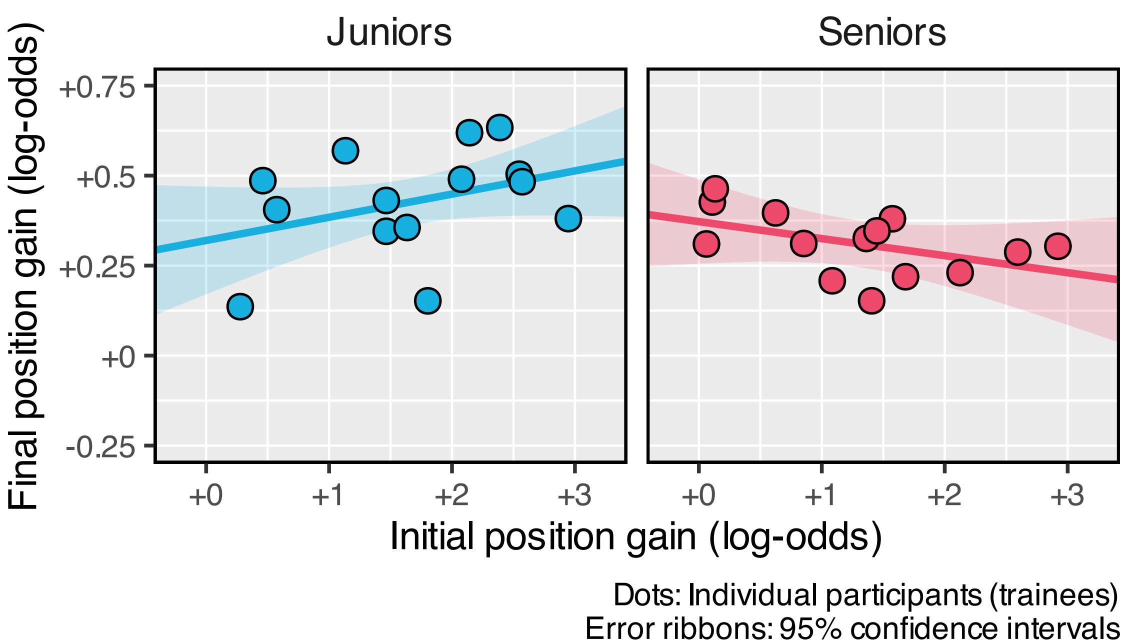

Figure 5: Individual participant gain scores (in log-odds) on syllable-initial and syllable-final fricatives. Trendlines and error ribbons are from the model

The model's intercept is statistically significantly different from 0 log-odds, suggesting that, among junior trainees, zero gains in the trained syllable position would still predict gains on the untrained syllable position, $\beta = 0.32$, $t = 4.41$, $p < .001$. However, further gains in the trained position do not predict further gains in the untrained position, $\beta = 0.06$, $t = 1.65$, $p > .05$. The intercept for seniors is not statistically significantly different from that of juniors, $\beta = 0.05$, $t = 0.57$, $p > .05$. However, the effect of gains in the trained position for seniors is statistically significantly different from juniors, $\beta = 0.11$, $t = 2.09$, $p < .05$. This suggests that, to whatever extent gains in the trained position truly do translate to gains in the untrained position for junior trainees (a proposition not supported by the model), the rate of transfer may be slightly smaller for seniors than for juniors.

2.8 Discussion

Models of nonnative speech learning claim that the capacity for reorganization of phonetic categories remains intact over the whole life span. However, few studies have tested phonetic learning in seniors and how their perceptual performance compares to that of younger listeners. Addressing this issue, the present study examined the perceptual malleability of 14 seniors (aged 60–76) who practiced auditory identification of the English syllable-initial sibilants /s/ and /z/ – a contrast that is proven difficult for many L2 speakers of English.

After ca. 3 weeks of high variability phonetic training, the seniors demonstrated significantly better identification scores on the same 2AFC task (Task III) with a gain score of ca. 12 % judging from the descriptive results. Post-test results for the seniors' age-matched control group did not reveal the same statistically significant improvement in accurate sibilant identification, and we interpret this difference as testament of successful learning after training. Trainees who were available for testing two months after the training had ended also completed the delayed post-test. From this test we collected identification data from 12 seniors who had also completed the same task twice before. While the participants' accuracy level dropped from post — to the delayed post-test, the group was on average more accurate at this last time of testing than they were at pre-test, which indicates at least some retainment of learning. The description of the development in syllable initial sibilant identification for the seniors also applies to the groups of juniors. In fact, the statistical models we ran on the perception data from the syllable-initial (trained) context (RQ1–3) all point to highly similar patterns of nonnative phone identification over time between the two trained age groups. Impressionistically however, we do observe steeper slopes for junior trainees than for the seniors, which is in line with previous findings by Akahane-Yamada and colleagues, who noted that older trainees could improve in nonnative perception yet not as much as younger trainees (Akahane-Yamada, Takada, and Kubo 2002). In the current study on identification of initial /s/ and /z/ we observe greater variance in the seniors' perception scores after training than in those for juniors, and we can speculate if standardized measures of hearing loss or mental acuity could have helped us interpret these differences in learning gains.

As for the final research question we were interested in exploring the level of learning. Specifically, we asked if perceptual improvement of the initial /s/ vs. /z/ contrast would extend to improvement in perception of the same contrast in syllable final position, and an identification test (Task II) probed participants' perceptual performance in this context at pre — and posttest as well. Overall, the trainees did improve from pre-test to post-test in how accurately they identified /s/ and /z/ in the syllable final VC stimuli. This improvement was significant for the junior trainees and marginally significant for the senior trainees. How then did this improvement relate to improvement of sibilant perception in the trained position (if at all)? Modelling the gains in Task II (syllable-final /s/ and /z/ identification) as a function of gains in Task III we did not find a clear association between gains in the two tasks. While there was a positive predictive trend for the junior trainees, it was not statistically significant. Interestingly, the trend was slightly negative in the modeled results for the senior trainees, but we are cautious in how we interpret the outcome of the statistical models to RQ4 since there are some obvious problems concerning gain scores on the two different tasks. Recall that the senior trainees happened to more accurately identify syllable-final /s/ and /z/ at Pre-Test than did the junior trainees. With a fixed number of trials on the same test the room for improvement is thus simply smaller when the individual trainee has a relatively higher score at pre-test which necessarily results in a lower gain score, all things being equal.

Just as we are cautious in how we interpret the data from the present study, we also offer thoughts about some of its limitations. Based on the literature (Bohn and Ellegaard 2019), we expected the syllable-initial contrast /s/-/z/to pose substantial challenges to Danish listeners, but > 25% of the recruited participants were excluded after pre-test based on highly accurate (> 90 % correct) sibilant identification in Task III. Consequently, our study is based on quite a small sample which should ideally be increased in future work. Secondly, such a ceiling effect could suggest that the task involving this particular L2 target contrast is perhaps not sufficiently challenging to the larger population of L1 Danish speakers, who are indeed becoming gradually more familiar with English (Lønsmann, Mortensen, and Thøgersen 2022). Additionally, since the status of English has changed over the past 30 years (Lønsmann, Mortensen, and Thøgersen 2022), there may be obvious differences in how members of our demographic will have previously engaged with the language. For example, we can reasonably assume differences in L2 English proficiency and exposure across our two age groups, but without an objective measure we cannot learn the extent to which our perception data is dependent on these relevant variables. Future research may wish to explore the perceptual learnability of even more challenging nonnative speech contrasts, just as it is advisable to control carefully for familiarity with the phonetic inventory of the target language across age groups in as far as it is possible to do so.

3 Conclusion

The present study applied well-known methods from the speech training literature (i. e. HVPT, Logan and Pruitt 1995; Bradlow et al. 1997 to groups of seniors and juniors who were tested on their ability to correctly identify English /s/ and /z/ in CV syllables. The aim was to compare the perceptual malleability of seniors above the age of 60 (whose nonnative speech perception has rarely received scholarly attention), to that of juniors (< 30 years of age), who represent the typical learner group in perceptual training studies. Our results showed that both age groupswere significantly more accurate in identifying the initial sibilants after just ten training sessions over three weeks. Also, both age groups of trainees showed robust training effects in that they maintained higher accuracy on the same test two months after training had ended, with delayed post-test scores differing statistically from pre-test scores for both groups. We controlled for simple test-retest effect by including two age-matched control groups who did not complete the perceptual training, and while we found numeric improvements in these groups between pre — and posttest, improvements in neither of the control groups were significant. Additionally, we found little difference in results on the identification tests of the trained sibilants between the groups of seniors and juniors in the experimental groups. While our sample can in no way be considered representative of the population, the positive learning gains after few training sessions bodes well for the potential of nonnative phonetic learning/strengthening even after the age of 60. Extending studies of L2 speech learning to seniors is an important endeavor in so forth as we wish to test general claims about nonnative speech perception and if these indeed hold true for a group of participants who has been largely overlooked in the literature. But with an aging and increasingly connected population there is an indubitableneed for and interest in foreign language learning, and there are plenty of avenues that the research community will need to address in order to map out how senior learners are different from (or indeed similar to) the younger learners that have mostly informed our understanding of L2 learning.

Funding

Research supported by a grant DFF2 0132-00008B from the Independent Research Fund Denmark to OSB.

Ethics statement

OSB (PI) obtained an official approval from Aarhus Universitys Research Ethics Committee on 28 January 2021. Approval Number: 2021-02 Contact: legal@au.dk

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgements

The authors are grateful to Dr. Christoph Draxler and to assistants Mette Petersen, Anders Røn, Ida Krarup, and Geertje Graehn.

References

Akahane-Yamada, Reiko, Tomoko Takada, and Rieko Kubo. 2002. “Effect of Age on Perceptual Learning of Second-Language Phonetic Contrasts.” *The Journal of the Acoustical Society of America* 112 (5-Supplement): 2386–2386. https://doi.org/10.1121/1.4779710.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Best, Catherine T., and Michael D. Tyler. 2007. "Nonnative and Second-Language Speech Perception: Commonalities and Complementarities." In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege,* edited by Murray J. Munro and Ocke-Schwen Bohn, 13–34. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.17.07bes.

Bohn, Ocke-Schwen. 2018. "Cross-Language and Second Language Speech Perception." In *The Handbook of Psycholinguistics,* edited by Eva M. Fernández and Helen Smith Cairns, 213–239. Hoboken, New Jersey: Wiley Blackwell. https://doi.org/10.1002/9781118829516.ch10.

Bohn, Ocke-Schwen, and Anne Aarhøj Ellegaard. 2019. "Perceptual Assimilation and Graded Discrimination as Predictors of Identification Accuracy for Learners Differing in L2 Experience: The Case of Danish Learners' Perception of English Initial Fricatives." In *Proceedings of the 19th International Congress of Phonetic Sciences,* edited by Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, 2071–2074. Canberra and London: Australasian Speech Science / Technology Association / International Phonetic Association.

Bradlow, Ann R., David B. Pisoni, Reiko Akahane-Yamada, and Yoh'ichi Tohkura. 1997. "Training Japanese Listeners to Identify English /r/ and /l/: IV. Some Effects of Perceptual Learning on Speech Production." *The Journal of the Acoustical Society of America* 101 (4): 2299–2310. https://doi.org/10.1121/1.418276.

Chang, Charles B. 2012. "Rapid and Multifaceted Effects of Second-Language Learning on First-Language Speech Production." *Journal of Phonetics* 40 (2): 249–268. https://doi.org/10.1016/j.wocn.2011.10.007.

Christensen, Kaare, Mikael Thinggaard, Anna Oksuzyan, Troels Steenstrup, Karen Andersen-Ranberg, Bernard Jeune, Matt McGue, and James W. Vaupel. 2013. "Physical and Cognitive Functioning of People Older than 90 Years: A Comparison of Two Danish Cohorts Born 10 Years Apart." *The Lancet* 382 (9903): 1507–1513. https://doi.org/10.1016/s0140-6736(13)60777-1.

Deary, Ian J., Janie Corley, Alan J. Gow, Sarah E. Harris, Lorna M. Houlihan, Riccardo E. Marioni, Lars Penke, Snorri B. Rafnsson, and John M. Starr. 2009. "Age-Associated Cognitive Decline." *British Medical Bulletin* 92 (1): 135–152. https://doi.org/10.1093/bmb/ldp033.

Draxler, Christoph. 2011. "Percy — An HTML5 Framework for Media Rich Web Experiments on Mobile Devices." *INTERSPEECH 2011,* 3339–3340.

———. 2014. "Online Experiments with the Percy Software Framework – Experiences and Some Early Results." In *Proceedings of the Ninth International Conference on Language Resources and Evaluation,* 235–340. Reykjavik: European Language Resources Association.

Felker, Emily, Esther Janse, Mirjam Ernestus, and Mirjam Broersma. 2023. "How Explicit Instruction Improves Phonological Awareness and Perception of L2 Sound Contrasts in Younger and Older Adults." *Linguistic Approaches to Bilingualism* 13 (3): 372–408. https://doi.org/10.1075/lab.20113.fel.

Flege, James Emil, and Ocke-Schwen Bohn. 2021. “The Revised Speech Learning Model (SLM-r).” In *Second Language Speech Learning*, 1st ed., edited by Ratree Wayland, 3–83. Cambridge University Press. https://doi.org/10.1017/9781108886901.002.

Flege, James Emil, Naoyuki Takagi, and Virginia Mann. 1995. “Japanese Adults Can Learn to Produce English /I/ and /l/ Accurately.” *Language and Speech* 38 (1): 25–55. https://doi.org/10.1177/002383099503800102.

Gabrys-Barker, Danuta, ed. 2018. *Third Age Learners of Foreign Languages*. Bristol: Multilingual Matters. https://doi.org/10.21832/GABRYS9405.

Garibaldi, Camilla Linn, and Ocke-Schwen Bohn. 2019. “Production and Perception of Danish Front Rounded /y/: A Comparison of Ultimate Attainment in Native Spanish and Native English Speakers.” In *Romance-Germanic Bilingual Phonology*, edited by Mehmet Yavas, Margaret Kehoe, and Walcir Cardoso, 121–136. Toronto: University of Toronto Press.

Harrington, Jonathan, Sallyanne Palethorpe, and Catherine Watson. 2000. “Monophthongal Vowel Changes in Received Pronunciation: An Acoustic Analysis of the Queen’s Christmas Broadcasts.” *Journal of the International Phonetic Association* 30 (1–2): 63–78. https://doi.org/10.1017/S0025100300006666.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. “The Weirdest People in the World?” *Behavioral and Brain Sciences* 33 (2–3): 61–83. https://doi.org/10.1017/s0140525x0999152x.

Horslund, Camilla Søballe, and Ocke-Schwen Bohn. 2022. “Assimilation Patterns Predict L2 Identification Accuracy of English Initial Consonants.” *Journal of Second Language Pronunciation* 8 (2): 218–247. https://doi.org/10.1075/jslp.21024.hor.

Jähi, Katri, Paavo Alku, and Maija S. Peltola. 2015. “Does Interest in Language Learning Affect the Non-Native Phoneme Production in Elderly Learners?” Paper presented at International Congress of Phonetic Sciences (ICPhS’15).

Kliesch, Maria, Nathalie Giroud, Simone E. Pfenninger, and Martin Meyer. 2018. “Research on Second Language Acquisition in Old Adulthood: What We Have and What We Need.” In *Third Age Learners of Foreign Languages*, edited by Danuta Gabrys-Barker, 48–76. Bristol: Multilingual Matters. https://doi.org/10.21832/9781783099412-006.

Kubo, Rieko, Masato Akagi, and Reiko Akahane-Yamada. 2015. “Dependence on Age of Interference with Phoneme Perception by First- and Second-Language Speech Maskers.” *Acoustical Science and Technology* 36 (5): 397–407. https://doi.org/10.1250/ast.36.397.

Kubo, Rieko, Reiko Akahane-Yamada, and Masato Akagi. 2012. “Effects of Perceptual Training on the Ability of Elderly Adults of Japanese Speakers to Identify American English /r/-/l/ Phonetic Contrasts.” *The Journal of the Acoustical Society of America* 131 (4-Supplement): 3272–3272. https://doi.org/10.1121/1.4708230.

Kubo, Rieko, Reiko Akahane-Yamada, and Tsuneo Yamada. 2000. "Influence of Aging on Perceptual Learning of Second-Language Phonetic Contrasts." *The Journal of the Acoustical Society of America* 108 (5-Supplement): 2653–2653. https://doi.org/10.1121/1.4743899.

Lively, Scott E., John S. Logan, and David B. Pisoni. 1993. "Training Japanese Listeners to Identify English /r/ and /l/. II: The Role of Phonetic Environment and Talker Variability in Learning New Perceptual Categories." *The Journal of the Acoustical Society of America* 94 (3): 1242–1255. https://doi.org/10.1121/1.408177.

Logan, John S., and John S. Pruitt. 1995. "Methodological Issues in Training Listeners to Perceive Non-Native Phonemes." In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research,* edited by Winifred Strange, 351–378. Baltimore: York Press.

Lønsmann, Dorte, Janus Mortensen, and Jacob Thøgersen. 2022. "Er engelsk stadig et fremmedsprog i Danmark? Et spørgsmål om kollektiv sproglig identitet." *Nydanske Sprogstudier* 61:126–179. https://doi.org/10.7146/nys.v1i61.132242.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Ramscar, Michael, Peter Hendrix, Cyrus Shaoul, Petar Milin, and Harald Baayen. 2014. "The Myth of Cognitive Decline: Non-Linear Dynamics of Lifelong Learning." *Topics in Cognitive Science* 6 (1): 5–42. https://doi.org/10.1111/tops.12078.

Reifegerste, Jana. 2021. "The Effects of Aging on Bilingual Language: What Changes, What Doesn't, and Why." *Bilingualism: Language and Cognition* 24 (1): 1–17. https://doi.org/10.1017/S1366728920000413.

Roediger, Henry L., and Andrew C. Butler. 2011. "The Critical Role of Retrieval Practice in Long-Term Retention." *Trends in Cognitive Sciences* 15 (1): 20–27. https://doi.org/10.1016/j.tics.2010.09.003.

Romano-Hvid, Carmit. 2003. *To See the Signs and Hear the CALL: A Speech Perception Training Study in a Computer Assisted Language Learning Environment.* Frederiksberg, Denmark: Samfundslitteratur.

Saito, Kazuya, Keiko Hanzawa, Katya Petrova, Magdalena Kachlicka, Yui Suzukida, and Adam Tierney. 2022. "Incidental and Multimodal High Variability Phonetic Training: Potential, Limits, and Future Directions." *Language Learning* 72 (4): 1049–1091. https://doi.org/10.1111/lang.12503.

Sereno, Joan A., and Yue Wang. 2007. "Behavioral and Cortical Effects of Learning a Second Language: The Acquisition of Tone." In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege,* edited by Ocke-Schwen Bohn and Murray J. Munro, 239–258. Amsterdam: Benjamins. https://doi.org/10.1075/lllt.17.22ser.

Shannon, Robert V., Angela Jensvold, Monica Padilla, Mark E. Robert, and Xiaosong Wang. 1999. "Consonant Recordings for Speech Testing." *The Journal of the Acoustical Society of America* 106 (6): L71–L74. https://doi.org/10.1121/1.428150.

Tamminen, Henna, Teija Kujala, Risto Näätänen, and Maija S. Peltola. 2021. "Aging and Non-Native Speech Perception: A Phonetic Training Study." *Neuroscience Letters* 740:135430. https://doi.org/10.1016/j.neulet.2020.135430.

Tamminen, Henna, Maija S. Peltola, Teija Kujala, and Risto Näätänen. 2015. "Phonetic Training and Non-Native Speech Perception – New Memory Traces Evolve in Just Three Days as Indexed by the Mismatch Negativity (MMN) and Behavioural Measures." *International Journal of Psychophysiology* 97 (1): 23–29. https://doi.org/10.1016/j.ijpsycho.2015.04.020.

Trapp, N. L., and Ocke-Schwen Bohn. 2002. "Training Danish Listeners to Identify English Word-Final /s/ and /z/: Generalization of Training and Its Effect on Production Accuracy." In *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech,* edited by Allan James and Jonathan Leather, 343–350. Klagenfurt: University of Klagenfurt.

Werker, Janet F., and John S. Logan. 1985. "Cross-Language Evidence for Three Factors in Speech Perception." *Perception and Psychophysics* 37 (1): 35–44. https://doi.org/10.3758/BF03207136.

Werker, Janet F., and Richard C. Tees. 1984. "Phonemic and Phonetic Factors in Adult Cross-Language Speech Perception." *The Journal of the Acoustical Society of America* 75 (6): 1866–1878. https://doi.org/10.1121/1.390988.

Yamada, Reiko A. 1993. "Effect of Extended Training on /r/ and /l/ Identification by Native Speakers of Japanese." *The Journal of the Acoustical Society of America* 93 (4-Supplement): 2391–2391. https://doi.org/10.1121/1.406052.

Zhang, Wei, Yi Liao, and Hoang Trung Truong. 2024. "High Variability Phonetic Training Facilitates Categorical Perception of Mandarin Lexical Tones in L2 Older Adults: A Link to Auditory Processing." *Language Teaching Research,* Advance online publication. https://doi.org/10.1177/13621688241282259.