INDIVIDUAL DIFFERENCES IN NON-NATIVE PHONETIC PERCEPTION: POTENTIAL LINKS TO NATIVE PERCEPTION, ATTENTION, & MEMORY

Claire Honda[1,2], Meghan Clayards[1,2], Shari Baum[1,2]

[1]McGill University, Montreal, Canada; [2]Centre for Research on Brain, Language and Music, Montreal, Canada.
claire.honda@mail.mcgill.ca, meghan.clayards@mcgill.ca, shari.baum@mcgill.ca

ABSTRACT

Adults differ considerably in their ability to distinguish non-native sounds. The predictors of these individual differences remain poorly understood. We investigated whether differences in the discrimination of unfamiliar non-native phonemes were predicted by differences in performance on two native phonetic perception tasks (2AFC and VAS) and on two non-linguistic cognitive tasks. While preregistered analyses showed no significant predictors of non-native discrimination, exploratory analyses revealed a potential role of consistent native perception and cognitive factors in promoting successful non-native perception. These findings contribute to our understanding of individual differences in non-native phonetic perception, and have implications for encouraging successful acquisition of new languages in adulthood.

Keywords: phonetic perception, gradiency, response consistency, sustained attention, working memory

1. INTRODUCTION

Adults show great variability in their ability to successfully discriminate non-native phonemic contrasts (e.g., [1]). Some of this variability can be attributed to factors including native language background [2], musical ability [3], and auditory acuity [4]. Nonetheless, these factors explain only a portion of the observed variability and may depend on the specific non-native sounds being discriminated, so the predictors of successful non-native perception remain underspecified.

Adults also show individual differences in their perception of native speech sounds. A common way of measuring native phonetic perception is using 2-alternative forced choice (2AFC) tasks, in which listeners hear continua of sounds and classify each sound into one of two categories. A more recent alternative is visual analog scaling (VAS) tasks, in which listeners indicate where each sound falls along a continuous line between two categories. When a participant's responses on these tasks are plotted against the changes in stimulus properties along the continuum, the resulting identification slope can be steep (reflecting a sudden change in perception from one sound to the other) or shallow (reflecting gradual changes in the percept). Listeners differ in the steepness of their identification slopes and also in how consistently they respond to stimuli (e.g., [5]).

It is thought that steep slopes on 2AFC tasks reflect an ability to make fine-tuned and accurate judgments about which category a sound belongs to, which is important for making sense of the variable speech signal [6]. Conversely, shallow slopes on VAS tasks appear to reflect gradient perception, i.e. an ability to distinguish fine-tuned differences between sounds [5]. Gradient perception may grant listeners greater perceptual flexibility by enabling them to activate and manage competing phonemic representations, which facilitates reinterpretation of ambiguous acoustic input as new information becomes available [7]. Such flexibility in perception could be useful when learning a new language given that non-native sounds often sound like ambiguous instances of native sounds [8] (though see [9] for a study where native perception on a 7-point VAS task did not predict non-native perception).

There do appear to be links between native and non-native perception under some circumstances. In participants who underwent perceptual training to identify non-native vowels, better native vowel discrimination predicted better non-native vowel perception [10]. In bilinguals who were divided into poor vs. good perceivers of a non-native contrast, the poor perceivers were less sensitive when processing both native and non-native vowel contrasts [11]. Furthermore, L1 phonetic deficits have been linked to difficulty in L2 phonetic mastery [12]; and successful learning of an L2 contrast recruits the same neural regions involved in processing L1 contrasts [13].

Despite some evidence in support of a relationship between native and non-native phonetic perception, other studies have failed to find such a relationship. Native phonological processing has been shown not to predict learning of a non-native contrast [14], and perception of a continuum of native sounds also did not predict identification of non-native sounds during the same session [9]. As such, the predictors of individual differences in non-native perception and the links between native and non-native perception remain to be clarified.

We aimed to elucidate the potential predictors of successful non-native perception—

including categorization slopes on 2AFC and VAS tasks of native perception—and in so doing, to also clarify the nature of any relationship that might exist between native and non-native perception.

We hypothesized that the ability to distinguish fine-tuned differences in native speech sounds would relate to the ability to accurately discriminate non-native speech sounds. If this were the case, we would find that better non-native perception relates to steeper 2AFC slopes (reflecting an ability to accurately categorize sounds) and shallower VAS slopes (perhaps reflecting gradient perception that is fine-tuned and flexible).

2. METHODS

The methods, exclusion criteria, and analyses for this study were preregistered based on a pilot study (https://osf.io/ez5qh/?view_only=8e4a1498e04f4ee0946752ee93b9ce71).

2.1. Participants

139 monolingual English speakers (97 females) were recruited through Prolific.co. They were aged 18-35 (mean: 25), right-handed, born and living in the United States or Canada, and had no language, cognitive, or hearing impairments. They completed the study at home on Gorilla.sc, using their own headphones.

2.2. Native phonetic perception tasks

Two native phonetic perception tasks were used. On both tasks, participants listened to two minimal pairs (*bet–bat* and *dear–tear*, publicly available at https://osf.io/369my/ [15]). The minimal pairs were manipulated so that each one varied systematically in two relevant acoustic cues (formant frequency and vowel duration for *bet–bat*, voice onset time and onset F0 for *dear–tear*). Each cue varied in 5 steps, and each version of the first cue was paired with each version of the second cue, resulting in 25 stimuli per pair. Stimuli whose cue values were both at the extremes (i.e., step 1 or step 5) consequently sounded clear and unambiguous, while stimuli with more intermediate cue values sounded more ambiguous. The same stimuli were used in both tasks.

On each trial of the 2AFC task, participants chose between two options on the screen via mouse click to indicate what they heard (e.g., *bet* or *bat*). On each trial of the VAS task, the screen displayed a slider with the two members of the minimal pair on opposite ends, and participants indicated where along the continuous slider they perceived the stimulus to be (from 0 to 100). Each stimulus from each minimal pair was presented 5 times per task. Participants completed the VAS task first to avoid biasing responses based on the categorical demands of the 2AFC task.

2.3. Non-native phonetic perception task

Participants discriminated German vowels and consonants (/øː/–/œ/, /yː/–/ʏ/, /ʃ/–/ç/) that are perceptually challenging sounds for native English speakers [16]. To construct the stimuli, three native German speakers were each recorded producing 14 minimal pairs containing these phonemes. Stimuli were presented in a 3-interval oddity task. On each trial, participants heard three words—one from each speaker—and clicked “1”, “2”, or “3” to indicate which word sounded different or “None” if all three words sounded the same. There were 12 trials per minimal pair: half were switch trials, where one word was the other member of the minimal pair (e.g., “Kirche”, “Kirche”, “Kirsche”), and the other half were catch trials, where all 3 words were the same.

2.4. Cognitive tasks

To determine whether any observed relationships between performance on the speech tasks might be driven by individual differences in general cognitive capacity, participants completed two non-linguistic cognitive tasks. A version of the Continuous Performance Task (CPT) was used as a measure of sustained attention [17], and a backwards digit span task was used as a measure of working memory [18]. On the AX-CPT task, participants saw a string of letters and had to press one key whenever they saw the letter X after the letter A or press a different key for any other letter combination. There were 200 trials: 140 AX trials (A followed by X), 20 AY trials (A followed by a consonant other than X), 20 BX trials (B followed by X), and 20 BY trials (B followed by a consonant other than X). On the backwards digit span task, participants listened to a recorded series of numbers and had to type them out in the reverse order. The number of digits to be recalled increased every three trials, starting with two digits and increasing to a maximum of ten. When a participant incorrectly answered all three trials of a difficulty level, the task was terminated.

2.5. Preparatory data analysis

Participants were excluded if they reported phonetic training, German exposure, or very good listening or speaking ability in a language other than English. They were also excluded for low-effort task performance based on preregistered criteria related to reaction times and accuracy. This left 91 participants with complete data across all tasks.

2.5.1. Native phonetic perception tasks

Two mixed-effects logistic regression models were fit to the 2AFC data. In both models, the outcome was participants' responses (first model: responses to *bet–bat* stimuli; second model: responses to *dear–tear*). Fixed effects were the first and second acoustic cues, and correlated random effects were by-participant random intercepts and by-participant random slopes for each acoustic cue. The random slopes coefficients for both cues were extracted as the variables of interest, because they quantify how each participant's use of a given cue differs from the group average [15]. These random slopes coefficients—4 per participant—were reduced to two dimensions using principal component analysis (PCA) in R [19] using the lme4 package [20] and the prcomp() function.

The rotated logistic function in [5], which models response slope (gradiency), was fit to participants' VAS responses in MATLAB [21]. This generated two slopes and two consistency measures per participant (*bet–bat*, and *dear–tear*). For each trial, the difference between the participant's actual VAS response and the response predicted by the rotated logistic was calculated. The standard deviation of these residuals was averaged per minimal pair to provide an estimate of each participant's response consistency. The four final VAS variables—2 gradiency measures and 2 consistency measures—were reduced to two PCA components in R as for the 2AFC task.

2.5.2. Non-native phonetic perception task

From participants' responses on the oddity task, the non-parametric sensitivity index *A* was calculated, which takes into account hits (correct selection of the odd item on a switch trial) and false alarms (incorrect selection of an item as being odd on a catch trial).

2.5.3. Cognitive tasks

Bin scores were calculated from participants' AX-CPT responses [22]. These scores account for both reaction time (RT) and accuracy. Lower bin scores indicate better sustained attention as reflected by higher accuracy and/or smaller RT differences between non-switch (AX) and switch (non-AX) trials.

The highest number of digits successfully recalled on the backwards digit span task was taken as a measure of working memory.

2.6. Primary data analysis

A multiple regression model was fit with oddity *A* scores as the outcome. Predictor variables were the first two principal components derived from the PCA of the 2AFC coefficients, the first two components derived from the PCA of the VAS measures, and the two control predictors (AX-CPT bin scores and digit span levels). We predicted that the 2AFC and VAS measures would relate to oddity scores after accounting for the control predictors.

3. RESULTS

There were individual differences in performance on all tasks, as anticipated. Examples of four participants' responses on the VAS task are shown in Figure 1, illustrating differences both in slopes and in the consistency of responses.

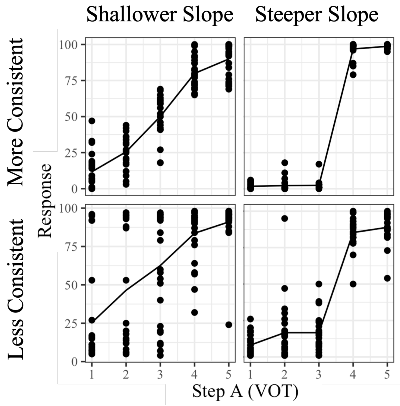

Figure 1: Four different participants' response curves for dear-tear cue A on the VAS task, illustrating individual variability.

The preregistered multiple regression model revealed that, contrary to our hypothesis, none of the predictors was significantly related to non-native perception scores ($p > 0.5$ for all).

Cook's distance was calculated to check for participants who might be particularly influencing the model's fit. This revealed one participant with higher influence than the others ($d = 1.6$), who upon closer examination had responded to the VAS task in a more categorical way than most (primarily using endpoints instead of the whole range). The multiple regression model was refit without this influential participant. The resulting exploratory model is summarized in Table 1. The 2AFC and VAS slope measures were again not significant predictors. Instead, the first principal component derived from the VAS variables was a significant predictor of non-native perception. This component primarily reflected consistency (see Table 2). Figure 2 shows non-native perception plotted against VAS consistency (averaged across both contrasts), revealing how more consistent VAS responders tended to have better non-native discrimination. The attention and memory control measures were also marginally significant ($p = 0.052$ and $p = 0.056$, respectively). We contextualize and elaborate upon these findings below.

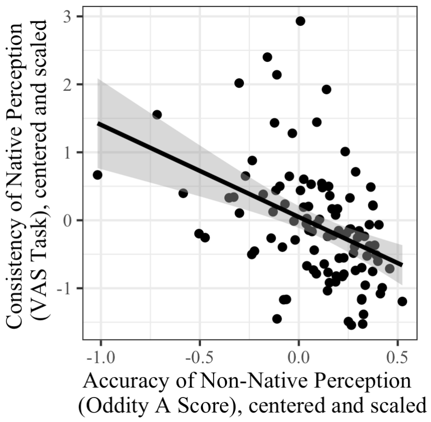

Figure 2: Scatterplot of non-native discrimination and native VAS consistency (higher values indicate less consistency and better non-native perception).

Coefficient	$\hat{\beta}$	$SE(\hat{\beta})$	t	p
(Intercept)	0.078	0.025	3.129	0.002
2AFC principal component (PC) 1	-0.008	0.021	-0.353	0.725
2AFC PC2	<0.001	0.023	0.002	0.999
VAS PC1	0.099	0.029	3.391	**0.001**
VAS PC2	-0.033	0.028	-1.194	0.236
AX-CPT bin score	-0.055	0.028	-1.974	**0.052**
Backwards digit span	-0.056	0.029	-1.936	**0.056**
Multiple R^2 *= 0.217, Adjusted* R^2 *= 0.161, Residual SE = 0.231 (df = 84), n = 91*				

Table 1: Regression table for the multiple regression model predicting oddity A scores. Model equation: Oddity A score ~ 2AFC PC1 + 2AFC PC2 + VAS PC1 + VAS PC2 + AX-CPT bin score + Backwards digit span.

	PC1	PC2
2AFC variables		
bet-bat acoustic cue A	-0.360	0.537
bet-bat acoustic cue B	-0.170	0.766
dear-tear acoustic cue A	-0.647	-0.266
dear-tear acoustic cue B	-0.651	-0.232
VAS variables		
bet-bat slope	0.395	-0.217
dear-tear slope	0.259	0.813
bet-bat consistency	-0.632	-0.262
dear-tear consistency	-0.614	0.472

Table 2: Correlations between the original 2AFC (top) and VAS (bottom) variables and the first two principal components extracted from them. PC1 and PC2 explained 52% and 29% of the variance in 2AFC responses; PC1 and PC2 explained 40% and 29% of the variance in VAS responses respectively.

4. DISCUSSION

Our objective was to investigate possible predictors of successful non-native perception. Contrary to our hypothesis, we did not find a relationship between fine-tuned or gradient native perception (as measured by 2AFC and VAS tasks) and accurate non-native perception. However, we found preliminary evidence that consistent native perception and non-linguistic cognitive factors (attention and memory) may play a role in discriminating unfamiliar language sounds.

The capacity to consistently perceive native sounds could conceivably help with reliably and accurately perceiving non-native sounds as well. The importance of consistent perception is hinted at by work showing that the ability to consistently assimilate a given non-native phoneme to a native category has been linked to having greater non-native perceptual proficiency [23] and a larger non-native vocabulary [24]. Future work should investigate the potential functions and implications of consistency during native and non-native phonetic perception.

Our attention and memory measures were marginally related to non-native discrimination, which is also supported by some previous work. For instance, non-native phonemes that are attended to are learned better than ones not attended to [25], and attention on the CPT has been related to non-native listening test scores [26]. It has also been shown that people with better working memory learn non-native phonemes faster and use more optimal learning strategies [27]. Attention and memory likely contribute to promoting robust processing and representation of non-native sounds.

The lack of strong relationships observed between our variables could be due in part to differences between the tasks used to measure native and non-native perception (e.g., the number of stimuli presented per trial, unmanipulated vs. ambiguous tokens). It could also be that native and non-native perception are simply not strongly linked – at least not initial non-native perception and the native constructs we measured ([9]). Perhaps native and non-native perception involve different mechanisms or strategies to some extent, as suggested by research showing that native speakers integrate phonetic information more gradiently while non-native speakers show a more categorical pattern [28].

If future work replicates our finding that consistent native perception, attention, and memory relate to the success of non-native perception, then this could lead to the development of brief pre-screenings to identify language learners who would benefit from more support during learning. Overall, our findings contribute to the understanding of individual differences in native and non-native phonetic perception, and we encourage further investigation of the factors (such as consistency) that may play a role in the successful perception of non-native phonemes.

5. REFERENCES

[1] K. Hattori and P. Iverson, "English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy," *J Acoust Soc Am*, vol. 125, no. 1, p. 12, 2009.

[2] J. E. Flege, O.-S. Bohn, and S. Jang, "Effects of experience on non-native speakers' production and perception of English vowels," *J. Phon.*, vol. 25, no. 4, pp. 437–470, Oct. 1997.

[3] L. R. Slevc and A. Miyake, "Individual Differences in Second-Language Proficiency: Does Musical Ability Matter?," *Psychol. Sci.*, vol. 17, no. 8, pp. 675–681, Aug. 2006.

[4] M. Kachlicka, K. Saito, and A. Tierney, "Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound," *Brain Lang.*, vol. 192, pp. 15–24, May 2019, doi: 10.1016/j.bandl.2019.02.004.

[5] E. C. Kapnoula, M. B. Winn, E. J. Kong, J. Edwards, and B. McMurray, "Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 43, no. 9, pp. 1594–1611, Sep. 2017, doi: 10.1037/xhp0000410.

[6] A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The discrimination of speech sounds within and across phoneme boundaries.," *J. Exp. Psychol.*, vol. 54, no. 5, pp. 358–368, 1957.

[7] E. C. Kapnoula, J. Edwards, and B. McMurray, "Gradient Activation of Speech Categories Facilitates Listeners' Recovery From Lexical Garden Paths, But Not Perception of Speech-in-Noise," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 47, no. 4, pp. 578–595, 2021.

[8] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Learning & Language Teaching*, vol. 17, O.-S. Bohn and M. J. Munro, Eds. Amsterdam: John Benjamins Publishing Company, 2007, pp. 13–34.

[9] P. Fuhrmeister, "Neural Correlates of Native-language Speech Perception and Non-native Speech Sound Learning," Doctoral thesis, University of Connecticut, 2020.

[10] A. Lengeris and V. Hazan, "The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek," *J. Acoust. Soc. Am.*, vol. 128, no. 6, Dec. 2010.

[11] B. Díaz, C. Baus, C. Escera, A. Costa, and N. Sebastián-Gallés, "Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language," *Proc. Natl. Acad. Sci.*, vol. 105, no. 42, pp. 16083–16088, Oct. 2008.

[12] R. Sparks and L. Ganschow, "The Impact of Native Language Learning Problems on Foreign Language Learning: Case Study Illustrations of the Linguistic Coding Deficit Hypothesis," *Mod. Lang. J.*, vol. 77, no. 1, pp. 58–74, 1993.

[13] N. Golestani and R. J. Zatorre, "Learning new sounds of speech: reallocation of neural substrates," *NeuroImage*, vol. 21, no. 2, pp. 494–506, Feb. 2004.

[14] F. S. Earle and D. T. Arthur, "Native phonological processing abilities predict post-consolidation nonnative contrast learning in adults," *J. Acoust. Soc. Am.*, vol. 142, no. 6, pp. EL525–EL531, Dec. 2017.

[15] M. Clayards, "Differences in cue weights for speech perception are correlated for individuals within and across contrasts," *J. Acoust. Soc. Am.*, vol. 144, no. 3, pp. EL172–EL177, Sep. 2018.

[16] R. Mayr and P. Escudero, "Explaining individual variation in L2 perception: Rounded vowels in English learners of German," *Biling. Lang. Cogn.*, vol. 13, no. 3, pp. 279–297, Jul. 2010.

[17] C. K. Conners, J. N. Epstein, A. Angold, and J. Klaric, "Continuous Performance Test Performance in a Normative Epidemiological Sample," p. 8, 2003.

[18] D. Weschler, "Wechsler Adult Intelligence Scale, (WAIS-IV)," *San Antionio TX Psychol. Corp.*, 2008.

[19] R Core Team, "R: A Language and Environment for Statistical Computing." Vienna, Austria, 2020.

[20] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using **lme4**," *J. Stat. Softw.*, vol. 67, no. 1, 2015.

[21] The MathWorks Inc., "MATLAB." Natick, Massachusetts, 2015a.

[22] M. M. Hughes, J. A. Linck, A. R. Bowles, J. T. Koeth, and M. F. Bunting, "Alternatives to switch-cost scoring in the task-switching paradigm: Their reliability and increased validity," *Behav. Res. Methods*, vol. 46, no. 3, pp. 702–721, Sep. 2014.

[23] Y. Kang and J. Schertz, "The influence of perceived L2 sound categories in on-line adaptation and implications for loanword phonology," *Nat. Lang. Linguist. Theory*, vol. 39, no. 2, May 2021.

[24] R. L. Bundgaard-Nielsen, C. T. Best, and M. D. Tyler, "Vocabulary Size is Associated with Second-Language Vowel Perception Performance In Adult Learners," *Stud. Second Lang. Acquis.*, vol. 33, no. 3, pp. 433–461, Sep. 2011.

[25] E. Pederson and S. Guion-Anderson, "Orienting attention during phonetic training facilitates learning," *J. Acoust. Soc. Am.*, vol. 127, no. 2, pp. EL54–EL59, Feb. 2010, doi: 10.1121/1.3292286.

[26] S.-P. Lee, "Computer-Detected Attention Affects Foreign Language Listening but Not Reading Performance," *Percept. Mot. Skills*, vol. 123, no. 1, pp. 33–45, Aug. 2016.

[27] J. R. McHaney, R. Tessmer, C. L. Roark, and B. Chandrasekaran, "Working memory relates to individual differences in speech category learning: Insights from computational modeling and pupillometry," *Brain Lang.*, vol. 222, Nov. 2021.

[28] F. Desmeules-Trudel, "Spoken word recognition in native and second language Canadian French: Phonetic detail and representation of vowel nasalization," Doctoral thesis, University of Ottawa, 2018.