Language Studies Dossier: Article

DOI: http://dx.doi.org/10.5007/2175-8026.2024.e98089

THE EFFECT OF PERCEPTION TRAINING WITH SYNTHETIC AND NATURAL STIMULI ON THE IDENTIFICATION OF ENGLISH VOWELS BY BRAZILIANS

Elisabeth Ann Bunch Oliveira da Rosa[1*]

Rosane Silveira[1**]

[1]Universidade Federal de Santa Catarina, Florianópolis, Brasil

Abstract

The present study investigates the effectiveness of synthetic versus natural stimuli for perception training on the ability of Brazilians to identify the English vowels /æ-ɛ/. Fifty-six native Brazilians participated in this study, and they were randomly divided into three groups: natural stimuli, synthetic stimuli, and control groups (experimental groups). The participants assigned to the experimental groups underwent perception training sessions, with the natural group listening to recordings containing vowels that did not have their duration altered, while the synthetic group listened to recordings containing vowels that had their duration altered to 350ms. Pre and post training perception tests were administered to measure their progress. Results indicate that perception training with both synthetic and natural stimuli is effective for the identification of /æ-ɛ/.

Keywords: vowel perception; high variability training; synthetic stimuli; English.

[*] MA in Applied Linguistics from the English Graduate Program at Universidade Federal de Santa Catarina. Graduate degree in Linguistics and Spanish from Western Washington University. E-mail: elisabeth.bunch@ymail.com. ORCID: https://orcid.org/0009-0001-1388-4031.

[**] Associate professor at Universidade Federal de Santa Catarina. Faculty member of the English Graduate Program and the Department of Foreign Language and Literature. CNPq research fellow. Coordinator of the Núcleo de Fonética e Fonologia Aplicadas à Lingua Estrangeira research group. E- mail: rosanesilveira@hotmail.com. ORCID: https://orcid.org/0000-0003-0329-0376.

Introduction

Adult learners' perception of second or non-native language (L2) speech is typically patterned by the phonemic inventory of their first or native language (L1). The Speech Learning Model, in its original and revised version, predicts that an L2 sound that is perceived as being similar to an L1 category will be more difficult to acquire than a sound that is perceived as being more different (Flege, 1995; Flege & Bohn, 2021). In the case of native Brazilian Portuguese (BP) speakers learning English, this prediction could be illustrated with vowel pairs such as /æ-ɛ/, as the vowel /ɛ/ exists in the BP vowel inventory but the vowel /æ/, while perceived as being similar to /ɛ/ (Nobre-Oliveira, 2007), does not. BP learners of English may therefore perceive these pairs as the same vowel (/ɛ/). Research has shown that the vowel pair /æ-ɛ/ is in fact one of the most challenging for BP learners of English to perceive in comparison to other vowel pairs, typically assimilating both /æ-ɛ/ vowels into the BP category /ɛ/ (Lima Jr., 2017; Rauber, 2006).

Perception training with synthetic stimuli is one way to assist L2 learners in the formation of new vowel categories. Synthetic stimuli can be created by manipulating a speech sound. Ladefoged (2007) identifies a number of components for a speech sound, namely: the frequencies of the first three formants (F1, F2, F3), the amplitudes of the first three formants, the frequency and amplitude of the voiceless components, and the fundamental frequency of voiced sounds (Ladefoged, 2007, p. 68). Synthetic stimuli refer to the manipulation of one or more of these components, or even their duration. In this study, the duration of the vowel sounds within carrier words will be manipulated to create synthetic stimuli.

Research has shown that synthetic stimuli can be just as, or possibly even more effective than natural stimuli, as with synthetic stimuli, the "subtle and crucial cues of the [acoustic] signal are enhanced, drawing learners' attention to them (and the less important features, attenuated)" (Nobre-Oliveira, 2007, p. 148). Cheng et al.'s (2019) study on temporal exaggeration also supports the use of acoustic exaggeration in the high variability phonetic training method (2019). The present study aims to examine the use of natural versus synthetic stimuli on the perception of the English vowel pair /æ-ɛ/ by Brazilians. The synthetic stimuli maintain the formant frequency cues unaltered while controlling for vowel duration (all vowels were manipulated to have the same duration in the synthetic stimuli). By holding the secondary cue (temporal information) steady, listeners are predicted to pay more attention to the primary cues of formants. The results of this study may be beneficial to L2 teachers and L2 programs, as it will indicate the effectiveness of perception training with natural and synthetic stimuli for L2 learners.

We (a) compare the effect of natural vs. synthetic stimuli for perception training with the vowel pair /æ-ɛ/; (b) examine whether the participants can generalize the gains from their perception training with natural vs. synthetic stimuli to natural stimuli with new voices and carrier words; and (c) examine learning retention with natural vs. synthetic stimuli to natural stimuli.

In order to explore these objectives, this study will answer three research questions:

RQ1: What is the effect of natural versus synthetic stimuli for perception training with the vowel pair /æ-ɛ/?
RQ2: How successfully are BP learners of English able to generalize their perception training with natural versus synthetic stimuli to new voices and carrier words?
RQ3: How successfully are BP learners of English able to maintain the effects of perception training with natural versus synthetic stimuli?

Based on previous studies on the use of synthetic and natural stimuli for perception training (Cheng et al., 2019; Nobre-Oliveira, 2007; Ylinen et al., 2009; Wang et. al, 1999) and on the potential benefits of training with phonetic variability for category formation (Logan et al., 1991; Lively et al., 1993) the following hypotheses (H) were developed.

H1: BP L2 English students that complete perception training with synthetic stimuli will show more improvement in their ability to identify the English vowels /æ-ɛ/ than those that complete perception training with natural stimuli, or those that are in the control group.

H2: BP learners of English that complete perception training with synthetic stimuli will be better at generalizing their learning to new voices and carrier words than those who complete perception training with natural stimuli, or those that are in the control group.

H3: BP learners of English that complete perception training with synthetic stimuli will be able to maintain their learning at a higher level than those that complete perception training with natural stimuli, or those that are in the control group.

The next section presents the review of literature, which provides a framework for the present study and summarizes other studies that contributed to the design. The method, detailing the participants, instruments for data collection, procedures for data collection and analysis will follow, as well as the presentation and discussion of the results. We conclude with some final remarks, a discussion of the limitations of the study, and recommendations for further research.

Review of Literature

In this section we briefly review the perception learning model that will serve as the theoretical background and highlight some acoustic differences between the BP and General American English (GAE) vowels investigated in the present study. Then we turn to studies on phonetic training to address methods for perception training. Studies exploring the effectiveness of perception training and the effectiveness of natural versus synthetic stimuli for perception training will then be reviewed.

Speech learning model

The Speech Learning Model (Flege, 1995) predicts the learning of new L2 categories. In its updated version, the Revised Speech Learning Model (SLM-R), Flege and Bohn (2021) explain that the primary aim of the SLM-R is to "account for how phonetic systems reorganize over the lifespan in response to the phonetic input during naturalistic L2 learning" (Flege & Bohn, 2021, p. 23). The main assumptions of this model are that

1. "the phonetic categories which are used in word recognition and to define the targets of speech production are based on statistical input distributions";

2. "L2 learners of any age make use of the same mechanisms and processes to learn L2 speech that children exploit when learning their L1", and

3. "native versus nonnative differences in L2 production and perception are ubiquitous not because humans lose the capacity to learn speech at a certain stage of typical neuro-cognitive development but because applying the mechanisms and processes that functioned 'perfectly' in L1 acquisition to the sounds of an L2 do not yield the same results" (Flege & Bohn, 2021, p. 23).

The SLM-R provides an explanation for the differences in L1 and L2 learning outcomes, noting that in the early stages of learning, L1 sounds "substitute" L2 sounds, and these existing L1 phonetic categories can block new category formation for L2 sounds (Flege & Bohn, 2021). Flege (2003) explains that new category formation can be predicted based on the perceived phonetic dissimilarity of the L2 speech sounds from the closest L1 sounds, category formation being more likely for those that are more dissimilar. Flege also notes that "if instances of an L2 speech sound category persist in being identified as instances of an L1 speech sound, category formation for the L2 speech sound will be blocked" (p. 10).

Research supports the idea that Brazilians face difficulty to form two distinct categories for the English vowel pair /æ-ɛ/ (Nobre-Oliveira, 2007; Rauber, 2006). The phoneme /ɛ/ exists in the Brazilian Portuguese phonetic inventory; however, /æ/ does not. Nobre-Oliveira (2007) found that /æ/ was misidentified as /ɛ/ most of the time. Rauber (2006) reported that the vowel pair /æ-ɛ/ was the most poorly perceived pair (compared to /i-ɪ/ and /u-ʊ/) for BP learners of English. These results suggest that BP learners perceive /æ/ as being similar to /ɛ/ and treat both phonemes as a single category in English, as predicted by the SLM-R.

Table 1 presents baseline acoustic values produced by monolingual speakers for the vowel contrast being examined in the present study. The BP values were provided by Rauber (2006), and the GAE values, by Hillenbrand et al. (1995).

Table 1 - Acoustic baseline values for /ɛ/ and /æ/

	BP		**GAE**			
	/ɛ/		/ɛ/		/æ/	
	Fem.	**Male**	**Fem.**	**Male**	**Fem.**	**Male**
Duration	127	130	254	189	332	278
F1	606	494	731	580	669	588
F2	2282	1908	2058	1799	2349	1952
F3	2912	2614	2979	2605	2972	2601

Source: The authors, based on Hillenbrand et al. (1995) and Rauber (2006).

The BP /ɛ/ average F2 values fall right between the average F2 values of the English /æ/ and /ɛ/, likely causing difficulty for BP learners of English in distinguishing these two vowels, as confirmed by previous research (Lima Jr., 2017; Rauber, 2006). As for duration, Table 1 shows that GAE speakers produce both vowels with a longer duration than the BP /ɛ/. Furthermore, the duration of /æ/ is longer than /ɛ/ for both male and female GAE speakers. This difference in duration for the GAE vowels is salient for Brazilian learners and most likely, as hypothesized by Rauber et al. (2005), Brazilian learners of English rely primarily on duration to distinguish between vowel contrasts such as /ɛ/-/æ/ and /i/-/ɪ/.

The SLM-R therefore would predict difficulty for BP learners of English in acquiring the ability to perceive the /æ/ phoneme, as it is similar to /ɛ/. One way to help learners overcome this difficulty is to enhance the acoustic features of the target vowels by providing phonetic training on synthesized tokens.

Phonetic training

Research has shown that perception training that utilizes repeated exposure to the key sounds is effective in improving learners' L2 perception (Cheng et al., 2019; Wang & Munro, 1999; Carlet et al., 2014; Ylinen et al., 2009; Rato, 2014). In high variability phonetic training (HVPT), listeners are specifically exposed to multiple voices speaking target sounds or words, and the listeners must identify or discriminate between the sounds that are presented. This method was originally developed by Logan et al. (1991), who found that this variability in voices helped listeners to disregard speech differences that were irrelevant to the perception of an L2 contrast and instead focus on the acoustic differences (formant frequencies and/or duration cues) that were in fact important for identifying the L2 sounds. A generalization task, in which participants needed to identify new words produced by familiar and unfamiliar voices was added to the method in a subsequent study by the same authors (Lively et al., 1993). Thus, studies using HVPT have found that input variability is an important factor in the category formation of L2 sounds.

Cheng et al. (2021) explain that vowel discrimination using categorical tests involves both a categorical mode and an auditory mode. In the categorical mode, "listeners map the input onto abstract phonemic categories and using those labels to inform their response", while in the auditory mode, "listeners attend and respond to continuous phonetic detail" (Cheng et al., p. 2). The authors also point out that identification tests draw on the categorical mode, given that listeners have to classify "a stimulus into a phonemic category" (Cheng et al., p. 2).

When performing a vowel perception task, listeners may rely on spectral information (formant frequency) and/or temporal information (duration). Previous studies have shown that the type of acoustic information used by listeners vary across languages, and that reliance on the less important cues may cause L2 learners to have difficulties to perceive and produce vowel contrasts (Colantoni et al., 2015). In a landmark study that illustrates how listeners rely on different spectral differences, depending on their L1, Flege et al. (1997) showed that native speaker of English relied mostly on spectral information to distinguish between /æ/ and /ɛ/, while German and especially Korean and Mandarin learners relied mostly on temporal information to distinguish this vowel contrast.

Two types of perception training that have been researched for L2 perception are forced-choice identification training and discrimination training. According to Pollack and Pisoni (1971), identification training refers to training in which listeners are presented with one stimulus at a time and need to label the sounds into a set of defined categories. On the other hand, discrimination tests involve listening to several stimuli at a time, and the authors explain that there are several types of discrimination training, including ABX. In this type of discrimination test, listeners are presented with three stimuli, and they need to choose whether the third sound is more similar to the first or second sound.

A study with forty-five Spanish/Catalan bilinguals in their first year of a Spanish university English program introduced six thirty-minute sessions of perception training for English vowels to participants (Cebrian et al., 2019). The intervention was implemented with HVPT and utilized both identification and discrimination training. The results of this study showed that training was in fact effective in improving the participants' identification and discrimination of L2 vowels. The four-month delayed post-test scores also revealed long-term effects of the perception training.

Aliaga-García (2009) also conducted a study with advanced Spanish/ Catalan bilingual learners of English, investigating the effects of phonetic training sessions, specifically including perception training with forced-choice identification tasks and discrimination tasks. Participants in this study completed six two-hour training sessions. The results showed that perception training did appear to significantly improve learners' discrimination ability for some of the vowels included in the study, namely, /i ɪ æ ʌ/.

Carlet and Cebrian (2015) looked at the effects of two types of training for one-hundred bilingual Spanish/Catalan learners of English and their perception of English vowels and initial and final stops. The vowels that were trained and tested in this study were /i ɪ æ ʌ ɜ/. This study revealed that for L2 stops, both training methods (ID – identification; DIS – discrimination) were equally effective; however, for vowels, the "ID trainees improved and generalized learning to greater extent than the DIS trainees" (Carlet & Cebrian, 2015, p. 944). The authors provided a possible explanation for their finding concerning the superiority of the ID training method, noting that the superiority of the ID training "might be connected to the presence of labels in the ID task, which provided learners with focus on phonetic form (i.e., phonetic symbols and/or orthography), which is said to impact speech perception" (Carlet & Cebrian, 2015, p. 946).

Considering the results of the studies mentioned in this section and the indications that perhaps there is not a significant benefit to using both training methods and that between the two methods, identification training may be more effective, the present research will utilize identification perception training.

Natural versus synthetic stimuli for perception training

Perception training with synthetic stimuli "has been found to be effective in changing L2 cue weighting of non-native vowel perception" (Cheng et al., 2019, p. 4). An acoustic cue is defined as "any acoustic characteristic of a segment which aids in the recognition of that segment in speech" (Trask, 2006, p. 6), and cue weighting refers to the importance listeners give to certain acoustic cues to distinguish and recognize specific sounds. Cue weighting will be especially important for the present study, as the synthetic stimuli will aim to assist learners in focusing on the formant frequencies as cues. Native English speakers utilize both durational and formant frequencies as cues for identifying vowel sounds, although they rely primarily on the latter (Escudero et al., 2009). These formant frequency cues may be more easily generalized to new contexts.

Cheng et al. (2019) found that synthetic stimuli were in fact useful in assisting learners to focus on the most important cues and essentially ignore the less important features. Their study focused specifically on the role of temporal acoustic exaggeration in perception training for Chinese adults learning the English /i- ɪ/ vowel contrast, with vowel duration being adjusted to 170 ms. Both groups in this study showed significant improvement in their ability to distinguish these two vowels, although the group that listened to acoustically modified stimuli was more successful in generalizing their learning to new speakers and words.

Nobre-Oliveira (2007) researched the effect of synthetic stimuli on intermediate-level native BP speakers of L2 English. She manipulated and enhanced F1 and F2 values of the vowel pairs /i-ɪ/, /æ-ɛ/, and /u-ʊ/ in synthetic stimuli in order to create a larger vowel space, providing an easier distinction for learners to categorize. In order to further enhance the stimuli, this study used isolated vowels (versus vowels within words), which allowed for the control of vowel duration, which was 500ms for all vowels used in the perception training. Her results revealed that synthetic stimuli are at least equally, if not more, effective as natural stimuli in perception training, although the differences were not statistically significant.

Wang et al. (1999) found that fourteen adult native speakers of Mandarin were able to effectively shift their attention from temporal properties to spectral properties in English vowel perception tasks following a short period of training with synthetic stimuli. The words used for the perception training were /hid/, /hɪd/, /hud/, and /hʊd/. Participants underwent perception training sessions with immediate feedback. This study revealed Mandarin listeners' tendency to rely heavily on the temporal cues for the contrast between the English vowels /i-ɪ/, and a lack of a sufficiently strong response to spectral cues in both /i-ɪ/ and /u-ʊ/. The results showed that even a short period of training with feedback may be very effective in assisting L2 learners to shift their attention from temporal to spectral properties in vowel perception tasks.

Having discussed studies investigating how different types of phonetic training can aid L2 speech perception development, we turn now to the method of the present study.

Method

We begin this section by describing the participants, the instruments used, the preparation of stimuli for the pre-test, the perception training, and post-tests, and the procedures for data collection and analysis. The data collection was carried out online. The participants were recruited and asked to read and sign a consent form. This study underwent ethical review and was approved by the *Comitê de* Ética *em Pesquisa com Seres Humanos*[1].

Participants

Fifty-six native Brazilian-Portuguese speakers and learners of English participated as volunteers in this study. These participants were invited to participate through Instagram. Their ages ranged from nineteen to forty-nine (mean = 30.4), 36 females and 20 males. They did not have any hearing or vision impairment at the time of the study. The participants were divided randomly into three groups: 1) those that received perception training with natural stimuli (N=22), 2) those that received perception training with synthetic stimuli (N=15), and 3) those that did not receive any perception training and served as a control group (N=19). Their proficiency level ranged from beginner to proficient users of English as an L2, according to an online vocabulary test (Meara & Miralpeix, 2014) administered to all participants.

Instruments and materials for data collection

We collected data with a background questionnaire, vocabulary test, a perception identification pre-test, post-test, and delayed post-test. The intervention was conducted with perception identification training activities. The background questionnaire was filled out by participants using Google Forms (www.google.com/forms). The perception tests and the perception training were presented to the participants with the help of Gorilla Experiment Builder (Anwyl-Irvine et al., 2018)

The natural and synthetic stimuli were reviewed by native English speakers and Brazilian proficient users of English as an L2 and underwent pilot testing with a limited number of participants prior to data collection. The following sections will describe in more detail the questionnaire as well as the perception pre-test, perception training, and the perception post-tests.

Background questionnaire

An online background questionnaire, set up on Google Forms, was used to introduce the consent form and to collect background information about participants' L2 learning experience, current L2 self-reported level, and knowledge of other languages. The questionnaire also asked participants to report any hearing or vision problems. The full questionnaire can be found in Oliveira da Rosa (2023).

Stimuli preparation

Traditionally, research in the area of perception training has used native speakers' voices for stimuli creation; however, the present study utilized a mixture of both native (American) speakers and non-native (Brazilian) proficient speakers. English as a Lingua Franca (ELF) is now widely used by speakers of various L1s, and a common context for communication is between two non-native speakers (Jenkins, 2009).

A total of twelve speakers (four native American English speakers and eight native BP speakers of English) were recruited and asked to record the stimuli. The recordings were then presented to three listeners (one native American English speaker and two native BP speakers of English) for review. These listeners provided detailed feedback on each individual audio, noting which carrier word they identified, the ease or lack thereof that they had to identify it, and the quality of the audio file. Based on the feedback from these three listeners, we selected recordings from seven speakers (four native American English speakers and three native BP speakers of English). To maintain consistency and to facilitate stimuli collection, we chose recordings from four native American English speakers all from the same state, Washington.

After the final preparation of the stimuli and perception identification tests and training, the stimuli from these seven selected speakers were presented a second time to native American English and native BP speakers of English for review as part of a pilot study. In this pilot, the stimuli were correctly identified by two native American English speakers 100% of the time, and by two native BP speakers of English 88.1% of the time. Further information about the speakers' and listeners' background can be found in Oliveira da Rosa (2022). We also verified the formant and the duration values for the vowels used in the stimuli. Measurements of F1, F2 and F3 values were obtained in PRAAT and indicated that the vowels from both the native American English speakers and the native BP speakers of English varied slightly in vowel duration and formant frequencies. However, all stimuli maintained a coefficient of variation of 0.26 or less for vowel duration, F1, F2, and F3 values, thus indicating that recordings from both the native American English speakers and from the native BP speakers of English were appropriate for use as stimuli in this study, as they displayed low levels of dispersion around the mean values (Banik et al., 2012) and contained the vowel contrast being examined.

The stimuli for this study are composed of twenty-one single-syllable carrier words, divided into seven triads. Following Nobre-Oliveira (2007), we included a distractor within each triad. Each of these triads included one word with the target-vowel /æ/, one word with the target vowel /ɛ/, and one word to be used as a distractor with the vowel /i/. The vowels appeared in each token following a single onset consonant and followed by a single stop consonant. Each of the seven English speakers recorded one triad, a total of three carrier words. These carrier words produced by each speaker can be seen in Table 2.

Table 2 – Carrier words recorded by seven adult speakers

Type of stimuli	**Vowels**	**Carrier words**						
		S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**
Target	**/æ/**	bad	had	mat	bat	vat	sat	sad
	/ɛ/	bed	head	met	bet	vet	set	said
Distractor	**/i/**	bead	heed	meet	beat	veet	seat	seed

Source: Oliveira da Rosa (2022)

To control for the phonemic environments of the target vowels, the stimuli contained some low-frequency vocabulary as carrier words (notably, "heed", "vat", and "Veet"). Because this study involved L2 English students of various proficiency levels, we acknowledge that these low-frequency words, as well as differences in orthography (for example, words like "said" and "head"), could interfere with participants' ability to correctly identify words. As an attempt to reduce task difficulty, the carrier words within each triad were always presented to participants in the same order on screen: the carrier word with the target vowel /æ/ on the left, the carrier word with the target vowel /ɛ/ in the middle, and the carrier word with the distractor vowel /i/ on the right (e.g., "Bad" / "Bed" / "Bead").

The speakers recorded the stimuli on their cellphones, each word in a separate audio file, repeating the same word three times. They sent the audio file via WhatsApp and these files were then transferred to the computer and saved as .wav files. The files were then edited using Audacity to remove background noise and to normalize the amplitudes, the peak amplitudes of all audio files being between 0.900 and 1.001 dB. Each file was previewed to check for audio quality and finally exported as a .wav file. The files were also edited using PRAAT to cut the audio files to include only one repetition of each carrier word and then to adjust the target vowel duration of the synthetic stimuli to 350ms, using a script. This procedure is described in detail in Oliveira da Rosa (2022).

The vowel duration of the synthetic stimuli was manipulated to be 350 ms, purposefully longer than the average vowel production in natural speech with the intention of providing more time for the learners to note formant frequency distinctions, following Nobre-Oliveira (2007). The present study selected 350ms, as it is still longer than the average vowel production of the Brazilian Portuguese /ɛ/, General American English /ɛ/, and General American English /æ/ but allows for higher audio quality as it involves less alteration to the original sound than the values used by Nobre-Oliveria (2007), which was 500 ms.

Perception identification pre-test

The stimuli used in the pre-test included recordings from Speakers 1, 2, 3, and 4 (Table 3). The stimuli for the pre-test included a total of twelve tokens, which were organized into triads. We attempted to evenly distribute the speakers between those whose recordings were used for the pre-test, training, and post-tests, and those whose recordings were used as unfamiliar stimuli only in the post-tests.

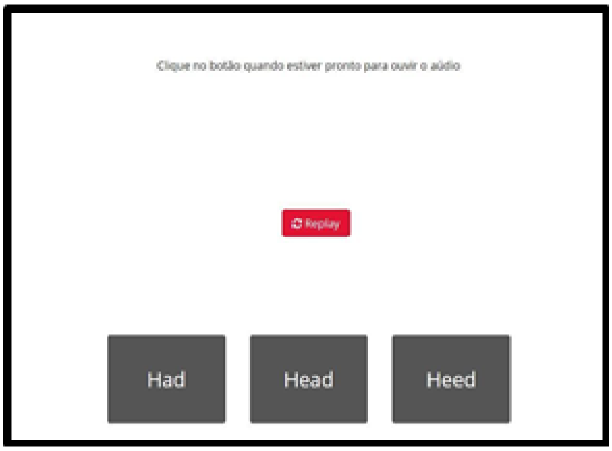
Table 3 – English stimuli for pre-test and training

Type of Stimuli	**Vowels**	**One-syllable token words**			
		S1	S2	S3	S4
Target	/æ/	bad	had	mat	bat
	/ɛ/	bed	head	met	bet
Distractor	/i/	bead	heed	meet	beat

Source: Oliveira da Rosa (2022)

These carrier words were presented to participants using the online research platform Gorilla Experiment Builder. Participants were shown each of the four triads listed in Table 3 a total of six times (twice for each carrier word). Each time that a triad appeared on screen, the participants clicked the "play" button when they were ready. They would then hear the carrier word and need to select which word they heard, selecting a button on the screen accordingly. Each of these carrier words was played twice, and participants were allowed to listen to the same audio up to two times if needed. Figure 1 shows a screenshot of the perception identification pre-test. Note that the button "Replay" appeared after the participant had played the audio the first time, in case the participant would like to repeat the audio.

Figure 1 – Perception pre-test example



Source: Oliveira da Rosa (2022)

The perception identification pre-test and training utilized recordings from one male native English speaker, one female native English speaker, one male Brazilian L2 English speaker, and one female Brazilian L2 English speaker. The stimuli for the pre-test were automatically randomized for the participants via the platform Gorilla Experiment Builder, to prevent presentation order effects. In total, the perception pre-test included twenty-four opportunities for participants to respond (two repetitions of each carrier word within each triad). From these responses, only 16 contained the target contrast /æ-ɛ/.

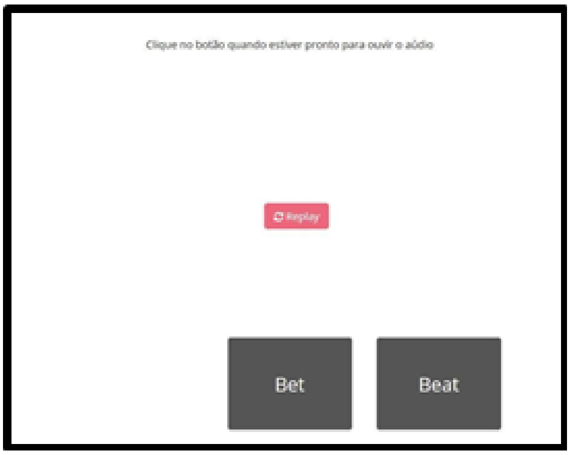
Perception identification training

The participants assigned to the control group did not receive perception training. The participants assigned to the natural stimuli experimental group received three perception training sessions with recordings that did not have their duration altered. The participants assigned to the synthetic stimuli experimental group received three perception training sessions with recordings that had their duration altered to 350ms.

These three training sessions for participants in the two experimental groups utilized the same stimuli from the perception identification pre-test, as seen in Table 3; however, these carrier words were presented to participants either naturally or synthetically, depending on the experimental group to which they were assigned. All the students participating in the perception training listened to a total of twenty-four (either natural or synthetic, depending on their experimental group) stimuli, presented in triads as shown in Figure 1. The training sessions lasted approximately ten minutes.

For each triad, participants heard a stimulus and selected which of the three alternatives displayed on the screen they heard. Each stimulus was presented two times per set, and participants were able to listen to the stimulus up to two times if necessary. Participants received immediate feedback following each response. When they responded correctly, they automatically moved on to the next triad. As shown in Figure 2, when they responded incorrectly, their incorrect answer disappeared from the screen, leaving two options (the remaining two carrier words) for them to choose from for a second time. These three training sessions were conducted over the course of three consecutive days.

Figure 2 – Immediate feedback, incorrect response example



Source: Oliveira da Rosa (2022)

As the duration of both the /æ/ and /ɛ/ (and /i/) vowels were the same for the synthetic stimuli, 350 ms, participants were unable to depend on duration for vowel identification. The purpose of this adjustment to the stimuli in the synthetic group was to assist learners in their cue weighting and depend more consistently on the formant frequency values for vowel identification. Controlling duration was found to successfully change L2 cue weighting for Chinese and Finnish listeners' perception of the English vowels /i-ɪ/ (Cheng et al., 2019; Ylinen et al., 2009, respectively), and native speakers of Mandarin's perception of the English vowels /i-ɪ/ and /ʊ-u/ (Wang et al., 1999). According to Hillenbrand et al.'s (1995) data, the F2 value differences are the most significant (compared to the other formant values) for vowel distinction. These values should be naturally emphasized as the duration is controlled.

Perception identification post-test and delayed perception post-test

As seen in Table 4, the post-test was composed of a combination of familiar tokens (the same recordings from speakers 1, 2, 3, and 4 from the pre-test and training sessions) and new tokens (recordings from speakers 5, 6, and 7 that have not been previously introduced to the participants). The unfamiliar stimuli to be used only in the post-tests were recorded by one male native English speaker, one female native English speaker, and one male Brazilian L2 English speaker. All these stimuli were presented naturally, without alteration to the vowel duration. The perception post-test was conducted in the same format as the perception pre-test, including random presentation of the stimuli, but due to the higher number of stimuli (21 carrier words X 2 repetitions) being presented, took approximately 20 minutes. From the 42 responses, only 28 contained the target vowel contrast.

The delayed perception post-test was the exact same format as the perception post-test and had the same stimuli and duration; however, it was presented to participants six days following the completion of the post-test.

Table 4 – Perception post-test and delayed perception post-test stimuli triads

Type of stimuli	Vowels	Carrier words						
		Familiar speakers and carrier words				**Unfamiliar speakers and carrier words**		
		S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**
Target	/æ/	bad	had	mat	bat	vat	sat	sad
	/ɛ/	bed	head	met	bet	vet	set	said
Distractor	/i/	bead	heed	meet	beat	veet	seat	seed

Source: Oliveira da Rosa (2022)

Procedures for data collection

Participants that replied to the social media invitations and were interested in participating in the study completed the online background questionnaire via Google Forms after reading and signing a consent form. This background questionnaire detailed their experiences learning English, knowledge of other languages, and noted any hearing or vision problems. Participants that were eligible and had access to a computer, equipped with internet access and headphones, scheduled group Zoom video call meetings with one of the researchers. During the first encounter, the participants completed the vocabulary test used to assess English proficiency, as well as the identification perception pre-test. Participants received the web link to begin the vocabulary test through Zoom. Once the participants completed the vocabulary test, they reported their scores (shown on the final report screen on the test platform) to one of the researchers during the video call, and the researcher manually recorded the information.

Then, a second web link was sent to the participants through Zoom, which they used to access the perception pre-test. The participants logged into Gorilla Experiment Builder using a participant ID number pre-assigned by the researcher.

Four or five days later (depending on the time scheduled by each participant), the participants of the experimental groups began the identification perception training sessions. These sessions were completed during a simultaneous Zoom meeting with one of the researchers. The three sessions were completed on three consecutive days.

Following the perception identification training sessions, participants from both the experimental and control groups took a perception identification post-test on Gorilla. Again, the test was completed during a simultaneous Zoom meeting with one of the researchers. The delayed post-test (taken six days after the post-test) was in the same format as the post-test. All scores were automatically saved by Gorilla in spreadsheets available for downloading by the test administrators.

Procedures for data analysis

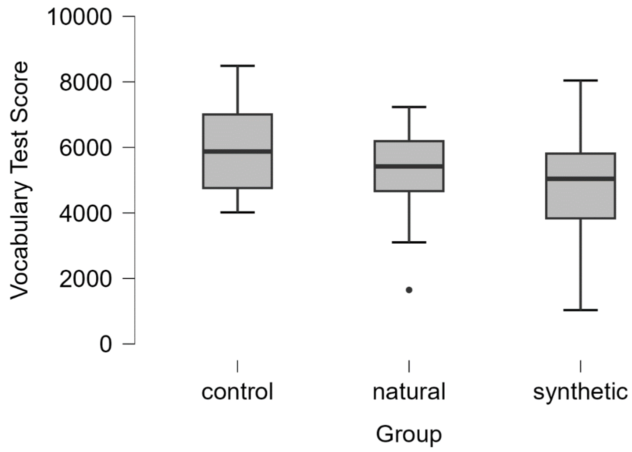
To achieve the objectives of the present study, a quantitative analysis was conducted. The independent variable is group condition (natural stimuli for perception training, synthetic stimuli for perception training, control with no perception training). The dependent variables are (i) identification accuracy at the pre-test; (ii) identification accuracy at post-test for familiar stimuli; (iii) identification accuracy at post-test for unfamiliar stimuli; and (iv) identification accuracy at the delayed post-test. Normality tests (Shapiro-Wilk) were run to verify the distribution of the data set and help to decide on the type of test to be used to compare means between and within groups. Further information about the statistical tests used is provided in the following section, as we report the results for each research question. All statistics were run on JASP 0.17.3.

Results and Discussion

The assessment of the participants' proficiency level was based on their performance on the Lognostic vocabulary test (Meara & Miralpeix, 2014) and the maximum score possible was 10,000. As no attempt was made to separate the participants according to proficiency level, a one-way ANOVA was run to verify whether the three groups were similar in terms of overall proficiency level. As Figure 3 shows, the participants of the control group obtained a slightly higher mean than the other groups (control: $M = 5{,}885$, $SD = 1{,}417$; natural: $M = 5{,}327$; $SD = 1{,}426$; synthetic: $M = 4{,}870$; $SD = 1{,}940$). The ANOVA indicates no significant difference among the three groups, according to proficiency level ($F(2, 53) = 1.77$, $p = .18$).

The control group had participants of a slightly higher proficiency level, and the two experimental groups displayed a greater range of proficiency level, including a few participants who scored very low in the vocabulary test, especially in the synthetic group.

Figure 3 - Proficiency level means for the three groups of Brazilian listeners



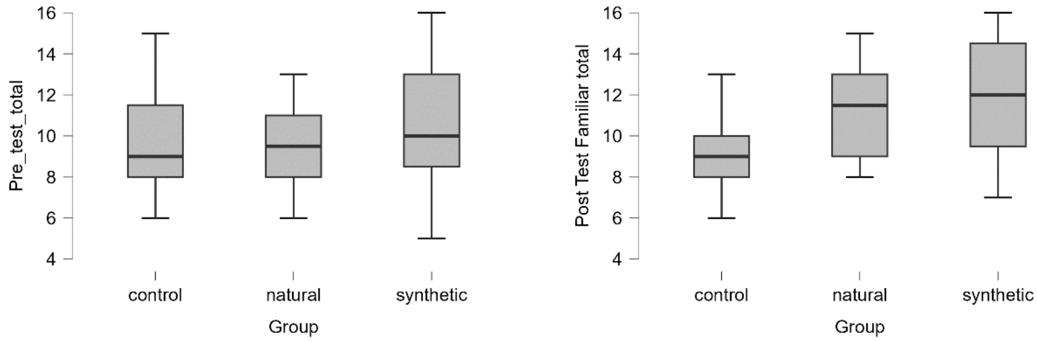
Source: The authors

We organized the results section into four subsections, one for each research question that guided the present study, and a final section discussing the results.

Effects of Perception Training

The first question guiding our study was whether perception training including either natural or synthetic stimuli would affect the perception of the vowel pair /æ-ɛ/ by Brazilian listeners. Our hypothesis was that the participants that completed perception training with synthetic stimuli would show more improvement in their ability to identify /æ-ɛ/ than those that completed perception training with natural stimuli, or those who received no perception training (Cheng et al., 2019; Nobre-Oliveira, 2007). Figure 4 shows the results comparing the participants' performance in the pre-test and post-test, including 16 familiar tokens only (i.e., tokens that were included in both the pre-test and post-test administered to all groups, as well as in the perception training sessions).

Figure 4 – Performance in the pre-test and post-test (familiar tokens)



Source: The authors

In the pre-test, the three groups obtained similar means, with slightly higher means for the synthetic and control groups (control: M = 9.94, SD = 2.67; natural: M = 9.54, SD = 2.08; synthetic: M = 10.26, SD = 1.82). A one-way ANOVA was run to compare the pre-test means between the three groups. The results showed that the groups did not differ significantly in the pre-test (F (2, 53) = .34, p = .71). As for the post-test results, the two experimental groups obtained higher means and higher variability than the control group (control: M = 9.10, SD = 1.82; natural: m = 11.27, SD = 2.35; synthetic: M = 11.60, SD = 3.04). Another one-way ANOVA was run to verify whether the means from the post-test differed between groups. The results showed a significant difference (F (2, 53) = 5.83, p = .005), and the Tukey post-hoc tests, displayed in Table 5, confirmed that the two experimental groups differed from the control group, but no significant difference was observed between the two experimental groups. Effect size for these differences were assessed with Cohen's d, revealing strong effect size for the comparisons involving the experimental groups versus the control groups.

Table 5 - Post-hoc comparisons for post-test results

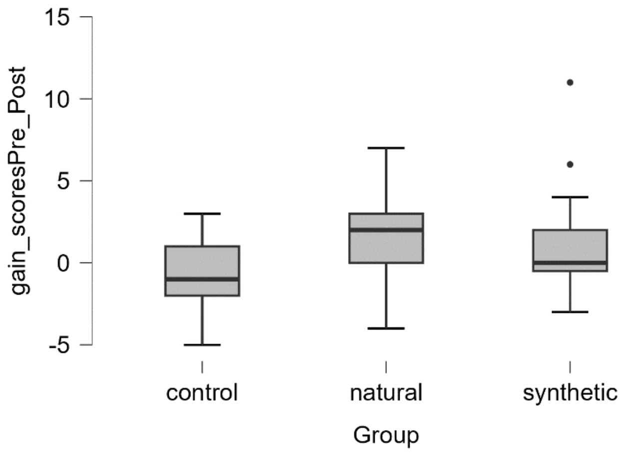
	Mean Difference	Cohen's d	p
control – natural	-2.16	-0.90	.01
control - synthetic	-2.49	-1.03	.01
natural - synthetic	-0.32	-0.13	.91

Note. P-value adjusted for comparing a family of 3

Source: The authors

To check for differences from the pre-test to the post-test, gain scores were calculated (post-test scores minus pre-test scores). As Figure 5 shows, the control group obtained negative gain scores (M = -.84, SD = 2.03), while the two experimental groups showed positive gain scores (natural: M = 1.72, SD = 2.79; synthetic: M = 1.33, SD = 3.49). It is worth pointing out the high degree of variability observed for gain scores, especially for the synthetic group. Another one-way ANOVA was run to verify whether the gain scores from pre to post-test differed between groups. The results showed significant differences in gain scores (F (2, 53) = 4.79, p = .01). Post-hoc tests, displayed in Table 6, confirmed that the two experimental groups differed from the control group, but no significant difference was observed between the two experimental groups. Effect size for these differences were assessed with Cohen's d, revealing a strong effect for the comparisons involving the experimental groups versus the control group.

Figure 5 – Gain scores from pre-test to immediate post-test



Source: The authors

Table 6 - Post-hoc comparisons for gain score results in the pre-test and the post-test

	Mean Difference	Cohen's d	p
control – natural	-2.56	-.92	.01
control - synthetic	-2.17	-.78	.07
natural - synthetic	0.39	.14	.90

Note. P-value adjusted for comparing a family of 3

Source: The authors

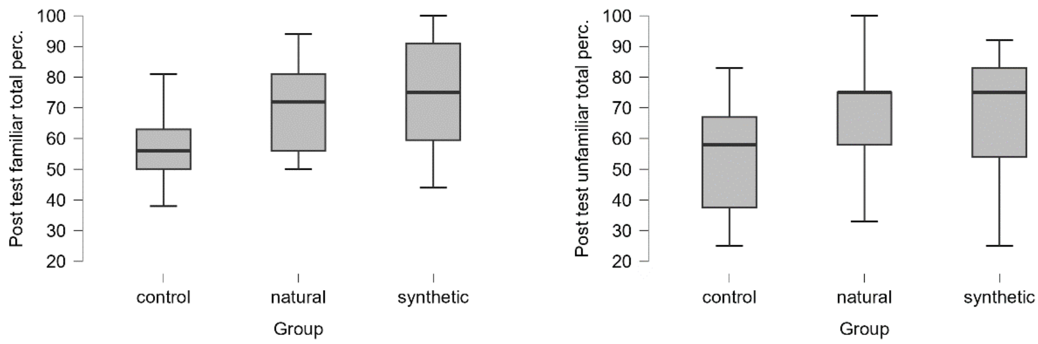
Hypothesis 1 is partially confirmed, given that perception training seemed to have helped both experimental groups outperform the control group, but the synthetic group did not outperform the natural group.

Generalization of perception training

Our second research question inquired about whether the participants would be able to generalize their perception training with natural versus synthetic stimuli to unfamiliar voices and carrier words. We hypothesized that the participants that completed perception training with synthetic stimuli would be better at generalizing their learning than those who completed perception training with natural stimuli (Cheng et al., 2019); no difference in performance was expected for the control group.

In addition to the 16 familiar tokens, the post-test included 12 unfamiliar tokens, which allowed us to assess whether the participants managed to generalize any gains in perception to new tokens. Because the number of familiar and unfamiliar tokens was different, we converted the test scores into percentages to run this analysis. As Figure 6 shows, the experimental groups outperformed the control group with both familiar (natural: M = 70,50, SD = 14,79; synthetic: M = 72,66, SD = 19,01; control: M = 57,05, SD = 11,37) and unfamiliar words (natural: M = 68,54, SD = 16,27; synthetic: M = 67,73, SD = 18,88; control: M = 55,57, SD = 19,24) in the post-test. The mean scores for the unfamiliar words were slightly lower than for the familiar words for the three groups though.

Figure 6 – Comparison between familiar versus unfamiliar words in the post-test



Source: The authors

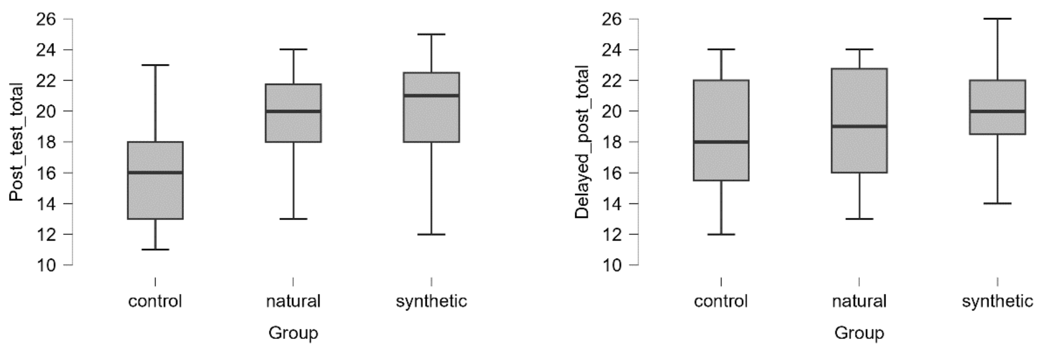
Because the numbers in this analysis represent percentages and the data were not normally distributed, we ran non-parametric Wilcoxon paired-samples tests to verify whether each group performed differently when perceiving familiar versus unfamiliar tokens.

The results showed no significant difference for either group (control: p = .73; natural: p = .67; synthetic: p = .61). In this case, the non-significant results add partial support to our second hypothesis, in the sense that the unfamiliar words were not significantly more difficult than the familiar words for both experimental groups, thus indicating that these groups were able to generalize perception learning to unfamiliar tokens. However, the same can be said for the performance of the control group, and the synthetic group did not outperform the natural group with unfamiliar words as predicted.

Retention of perception training effects

The third research question examined how much the two experimental groups managed to retain perception learning in a delayed post-test. Our hypothesis, based on Nobre-Oliveria (2007), was that participants that completed perception training with synthetic stimuli would be able to maintain their learning at a higher level than those that completed perception training with natural stimuli, while the control group would display no different results. The synthetic stimuli maintain the formant frequency cues unaltered while controlling for vowel duration (all vowels were manipulated to have the same duration). By holding the secondary cue (temporal information) steady, listeners are predicted to pay more attention to the primary cues of formants. For this analysis, we combined both familiar and unfamiliar tokens (number of tokens for each test = 28), given that the participants showed no significantly different performance with these two types of tokens in the immediate post-test. Figure 7 shows different performance patterns across tests for the three groups. While in the immediate post-test there was a clear difference in the means obtained by the two experimental groups (natural: M = 19.50, SD = 3.15; synthetic: M = 19.86, SD = 3.70) when compared to the control group (M = 15.78, SD = 3.39), in the delayed post-test, this difference is not as clear cut, with both control (M = 18.47, SD = 3.73) and natural (M = 18.90, SD = 3.87) groups obtaining similar means, and the synthetic group obtaining a slightly higher mean (M = 19.93, SD = 3.39).

Figure 7 – Performance in the immediate post-test and the delayed post-test

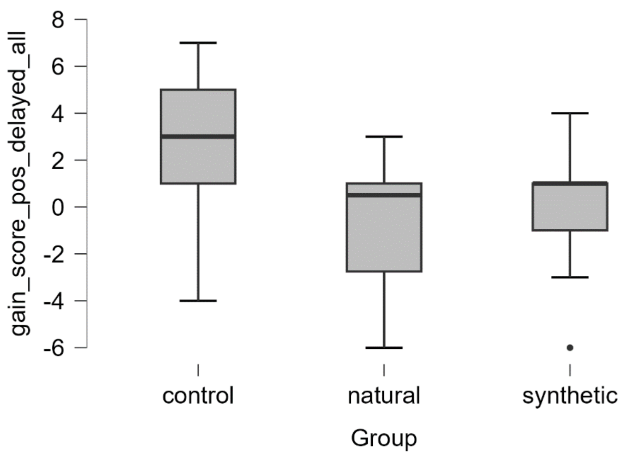


Source: The authors

A one-way ANOVA comparing the results for the delayed post-test shows no significant differences across the three groups (F(2, 53) = .67, p = .51), which is an indication that the initial difference observed in the immediate post-test (see discussion for first research question) is no longer observed when we look at the delayed post-test scores. To verify that the change in scores for the control group was not due to error, we ran paired-samples t-test to compare the means of the two post-tests for each group. The results corroborate the fact that there was no significant change for the experimental groups across the two post-tests (natural: *t*(21) = -.97, p = .34, d = .20; synthetic: *t*(14) = .09, p = .92, d = .02), which confirms they were able to retain their learning in the delayed post-test. But the results also confirm that the control group significantly improved performance in the delayed post-test (*t*(18) = 3.80, p = .001), with a strong effect size (d = .87). The control group improvement might be related to their slightly higher proficiency scores obtained in the Vocabulary Test used in the present study. Alternatively, we can speculate that learners from the control group might have had a chance to focus on the vowel contrast tested here and got motivated to search for information and practice the vowel contrast after taking the perception pre-test and post-test.

We calculated gain scores by subtracting the delayed post-test scores from the immediate post-test scores for each group. Figure 8 displays the gain score results for each group, and it reveals that the control group was the one with higher gains and higher dispersion (M = 2.68; SD = 3.07). The two experimental groups did not show much change in gain scores, with the synthetic group obtaining a slightly higher mean and high dispersion (M = .06, SD = 2.63) and the natural group, a slightly lower (negative gain score) mean and high dispersion too (M = -.59, SD = 2.84).

Figure 8 – Gain scores between immediate post-test and delayed post-tests



Source: The authors

The one-way ANOVA yielded significant results for gain scores (F (2, 53) = 7.13, p = .002), and post-hoc tests, displayed in Table 7, indicated that the mean gain scores of the control group are significantly higher than the gain scores of both the natural group (p = .002) and the synthetic group (p = .02), with a strong effect size. No significant difference was observed between the two experimental groups (p = .77).

Table 7 – Post-hoc comparisons for gain score results in the immediate and the delayed post-tests

	Mean Difference	**Cohen's d**	**p**
control – natural	3.27	1.14	.00
control - synthetic	2.61	0.91	.02
natural - synthetic	-0.65	-0.22	.77

Note. P-value adjusted for comparing a family of 3

Source: The authors

Once again, our research hypothesis is partially corroborated, given that the results of this study confirm that identification perception training aided both experimental groups to maintain perception learning when performing a delayed post-test, but no experimental condition was superior to the other. Surprisingly though, the control group obtained higher means in the delayed post-test.

Overall discussion

Based on the assumption that adult L2 learners maintain their ability to develop new vowel contrasts (Flege, 1995; Flege & Bohn, 2021), this study examined the perception of the English vowel pair /æ-ɛ/ by three groups of Brazilian Portuguese English learners. The main objective was to investigate whether providing the participants in the experimental groups with synthetic or natural stimuli for perception training with the target vowel contrast would affect the way they perceive this contrast.

Previous research has shown that both synthetic and natural stimuli are effective in perception training (Cheng et al., 2019; Nobre-Oliveira, 2007). However, in Cheng et al.'s (2019) study, it was shown that participants who trained with synthetic stimuli were more successful in generalizing their learning to new speakers and words, given that the modified stimuli helped them focus on primary acoustic cues. In Nobre-Oliveira's (2007) study, the data indicated that synthetic stimuli may be more effective for perception training, although the data were not statistically significant. The present study confirms that both types of stimuli are effective methods of perception training for the contrast /æ-ɛ/, but no support is provided to the superiority of training with synthetic stimuli, neither in the immediate post-test (with both familiar and unfamiliar words) nor the delayed post-test.

Nobre-Oliveira's (2007) study showed that participants who trained with natural stimuli maintained their learning, with the same score from post-test to follow-up test, and that participants who trained with synthetic stimuli showed a slight, non-significant improvement from post-test to follow-up test. Our results confirm that both experimental groups maintained perception learning in a delayed post-test administered six days after the training sessions. We also found slightly higher scores for the group who received synthetic stimuli perception training, but like in Nobre-Oliveira (2007), no statistical significance was found between the two experimental groups in the delayed test. An intriguing finding, though, when looking into the raw scores of the delayed post-test, is that the control group obtained higher means than in the immediate post-test, which is an indication that this group somehow developed their perception of the contrast /æ-ɛ/. One possibility is that by taking the perception test three times and being exposed to controlled input, learning took place for the control group, who was also slightly more proficient than the experimental groups. Furthermore, we cannot rule out the possibility that the control group participants engaged in explicit learning activities or had additional exposure to the vowel contrast during the research duration.

Overall, our study confirms the effectiveness of intervention with perception training, for both natural and synthetic stimuli, and the learning of such training is likely to be generalized to new words and speakers. Furthermore, this learning is mostly maintained after a six-day interval. Partial support was provided to the three research hypotheses, as we found no significant difference between perception training with synthetic and natural stimuli. Furthermore, possible testing effects may have helped the participants in the control group start learning the target vowel contrasty as well, as demonstrated by the delayed post-test results.

Final Considerations

Speech perception is a crucial component in L2 development. The present study is based on the principle that adult learners can access learning mechanisms that will allow them to create new categories and contrasts for L2 speech (Flege & Bohn, 2021). This task can be facilitated if learners have access to rich input that aids them in the identification of phonetic cues that are important to distinguish categories in the L2. In that sense, we tested the effect of two types of perception training (natural stimuli and synthetic stimuli) on Brazilian participants' ability to identify the English vowel contrast /æ-ɛ/. Our results indicated that both types of training can aid Brazilian listeners with perceiving this contrast with familiar and unfamiliar tokens, thus adding support to previous studies that highlighted the relevance of perception training in L2 speech learning. An unexpected result, though, was that even the control group demonstrated the ability to learn the contrast in the delayed post-test, thus indicating that exposure to controlled input in the form of a perception identification test, administered multiple times, can trigger learners' attention to phonetic features that are important to distinguish vowels.

The present study has limitations that might have impacted the results reported here. Notably, we were unable to control for proficiency level within each group, and this might have impacted the participants' performance in the perception identification tests and somehow account for the large variability observed in the data. We also ended up with an uneven number of participants within each group, given that some participants did not conclude the different stages of data collection. Furthermore, the perception training sessions were short, which could account for the lack of significant differences between the two experimental groups and the time lapse between the administration of the immediate post-test and the delayed post-test was short. Future studies should try and minimize these methodological shortcomings. On the bright side, our results show that perception training is an important asset for learning L2 speech contrasts. Simple exposure to abundant examples that highlight the importance of perceiving similar sounds as different categories in the L2 can already trigger learners' attention, and this is an essential step for them to start developing new contrasts that suit the L2 phonemic inventory.

Acknowledgments

The authors are thankful to CAPES and CNPq for the research grants that supported this study.

Note

1. Ethics Board review number: 4.622.381

References

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In Recent research in second language phonetics/phonology: perception and production (pp. 2–32). Newcastle upon Tyne: Cambridge Scholars Publishing.

Anwyl-Irvine, A. L., Massonié, J., Flitton, A., Kirkham, N.Z., & Evershed, J. K. (2018). Gorilla experiment builder. Available at: https://gorilla.sc/. Access on: 01 jun 2021.

Banik, S., Kibria, B. M. G., & Sharma, D. (2012). Testing the Population Coefficient of Variation. *Journal of Modern Applied Statistical Methods*, *11*(2), 325–335. https://doi.org/10.22237/jmasm/1351742640

Carlet, A., & Cebrian, J. (2015). Identification vs. discrimination training: Learning effects for trained and untrained sounds. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, Scotland.

Carlet, A., & Cebrian, J. (2014). Training Catalan speakers to identify L2 consonants and vowels: A short-term high variability training study. *Proceedings of the 7th International Symposium on the Acquisition of Second Language Speech. Concordia University working papers in applied linguistics* (Vol. 5, pp. 85-98).

Cebrian, J., Carlet, A., Gorba, C., & Gavaldà, N. (2019). Perceptual training affects L2 perception but not cross-linguistic similarity. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.*). Proceedings of the 19th International Congress of Phonetic Sciences*, *Melbourne, Australia*. (Vol. 1, pp. 929-933).

Cheng H. S., Niziolek, C. A., Buchwald, A., McAllister T. (2021). Examining the Relationship Between Speech Perception, Production Distinctness, and Production Variability. *Frontiers in Human Neuroscience*, *28(15)*, 660948. https://doi.org/10.3389/fnhum.2021.660948.

Cheng, B., Zhang, X., Fan, S., & Zhang, Y. (2019). The Role of Temporal Acoustic Exaggeration in High Variability Phonetic Training: A Behavioral and ERP Study. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01178

Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. A. H. (2009, September 1). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America*, *126*(3), 1379–1393. https://doi.org/10.1121/1.3180321

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A., Meyer; N., Schiller (Ed.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319–395). Mouton de Gruyter.

Flege, J. E. (1995). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*(4), 425–442. https://doi.org/10.1017/s0142716400066029

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25(4)*, 437-470.

Flege, J. E., & Bohn, O. S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). chapter, Cambridge: Cambridge University Press.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111. https://doi.org/10.1121/1.411872

Ladefoged, P. (2007). Vowels and consonants. 2. ed. Oxford, UK: Blackwell Publishing.

Lima Jr, R. M. (2017). The influence of metalinguistic knowledge of segmental phonology on the production of English vowels by Brazilian undergraduate students. *Ilha Do Desterro a Journal of English Language, Literatures in English and Cultural Studies*, *70*(3), 117–130. https://doi.org/10.5007/2175-8026.2017v70n3p117

Lively, S. E., Logan, J. S., Pisoni, D., B. (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.

Logan, J. S., Lively, S. E., Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America. 80(2)*, 874-886.

Meara, P. & Miralpeix, I. (2014). *Lognostics*, 2014. Available at: https://www.lognostics.co.uk/tools/.

Nobre-Oliveira, D. (2007). *The effect of perceptual training on the learning of English vowels by Brazilian Portuguese speakers* [Unpublished Doctoral Dissertation]. Universidade Federal de Santa Catarina, Brazil.

Oliveira da Rosa, E. A. B. (2022). *The effect of perception training with synthetic and natural stimuli on BP learners' ability to identify the English vowel /æ-ɛ/***.** [Unpublished Doctoral Dissertation]. Universidade Federal de Santa Catarina, Brazil.

Pollack, I., & Pisoni, D. (1971). On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science*, *24*(6), 299–300. https://doi.org/10.3758/bf03329012

Rato, A. (2014). Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese. *Concordia Working Papers in Applied Linguistics*, 5, 529-546.

Rauber, A. S., Escudero, P., Bion, R. A. H., & Baptista, B. O. (2005, September). The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. In Interspeech (pp. 2913-2916). http://doi:10.21437/Interspeech.2005-769.

Rauber, A. (2006). *Perception and production of English vowels by Brazilian EFL speakers* (dissertation). Pós-Graduação em Inglês, Florianópolis, SC.

Wang, X., & Munro, M. J. (1999, August). The perception of English tense-lax vowel pairs by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. In Proceedings of the 14th international congress of phonetic sciences (Vol. 3, pp. 125-128). Berkeley, CA: University of California.

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2010). Training the brain to weight speech cues differently: A Study of Finnish second-language users of English. *Journal of Cognitive Neuroscience*, *22*(6), 1319–1332. https://doi.org/10.1162/jocn.2009.21272

Recebido em: 05/01/2024
Aceito em: 15/10/2024

APPENDIX A

Stimuli Vowel Duration and Formant Frequency Values

Table 8 – Acoustic information for males /ɛ/ vowel

Context	Vowel duration	F1	F2	F3
Males - Average English /ɛ/ (HILLENBRAND ET. AL, 1995)	*189 ms*	*580 Hz*	*1799 Hz*	*2605 Hz*
Males - Average BP /ɛ/ (ESCUDERO ET. AL, 2009)	*123 ms*	*518 Hz*	*1831 Hz*	*2722 Hz*
/ɛ/ in carrier word "head" (S2)	182 ms	537.54 Hz	1741.97 Hz	2596.68 Hz
/ɛ/ in carrier word "said" (S4)	231 ms	633.80 Hz	1797.51 Hz	2614.52 Hz
/ɛ/ in carrier word "vet" (S6)	189 ms	606.38 Hz	1746.37 Hz	2311.70 Hz
/ɛ/ in carrier word "bet" (S8)	221 ms	489.76 Hz	1678.23 Hz	2468.79 Hz
Mean Score of four recordings	205.75 ms	566.87 Hz	1741.02 Hz	2497.92 Hz
Standard Deviation (SD)	20.70	56.67	42.32	121.32
Coefficient of Variation (CV)	0.10	0.10	0.02	0.05

Table 9 – Acoustic information for females /ɛ/ vowel

Context	Vowel duration	F1	F2	F3
Females - Average English /ɛ/ (HILLENBRAND ET. AL, 1995)	*254 ms*	*731 Hz*	*2058 Hz*	*2979 Hz*
Females - Average BP /ɛ/ (ESCUDERO ET. AL, 2009)	*141 ms*	*646 Hz*	*2271 Hz*	*2897 Hz*
/ɛ/ in carrier word "bed" (S1)	230 ms	635.42 Hz	1842.15 Hz	2267.77 Hz
/ɛ/ in carrier word "set" (S5)	151 ms	833.13 Hz	1898.64 Hz	2479.60 Hz
/ɛ/ in carrier word "met" (S7)	218 ms	766.33 Hz	1831.83 Hz	2474.33 Hz
Mean Score of three recordings	199.67 ms	744.96 Hz	1857.54 Hz	2407.23 Hz
Standard Deviation (SD)	34.76	82.12	29.37	98.64
Coefficient of Variation (CV)	0.17	0.11	0.02	0.04

Table 10 – Acoustic information for males /æ/ vowel

Context	Vowel duration	F1	F2	F3
Males - Average English /æ/ (HILLENBRAND ET. AL, 1995)	*278 ms*	*588 Hz*	*1952 Hz*	*2601 Hz*
/æ/ in carrier word "had" (S2)	318 ms	713.28 Hz	1628.77 Hz	2347.81 Hz
/æ/ in carrier word "sad" (S4)	300 ms	800.79 Hz	1703.31 Hz	2267.36 Hz
/æ/ in carrier word "vat" (S6)	224 ms	737.83 Hz	1687.18 Hz	2367.25 Hz
/æ/ in carrier word "bat" (S8)	281 ms	689.49 Hz	1637.86 Hz	2432.15 Hz
Mean Score of four recordings	280 ms	735.35 Hz	1665.28 Hz	2353.64 Hz
Standard Deviation (SD)	35.28	41.47	31.65	58.79
Coefficient of Variation (CV)	0.13	0.06	0.02	0.25

Table 11 – Acoustic information for females /æ/ vowel

Context	Vowel duration	F1	F2	F3
Females - Average English /æ/ (HILLENBRAND ET. AL, 1995)	*332 ms*	*669 Hz*	*2349 Hz*	*2972 Hz*
/æ/ in carrier word "bad" (S1)	259 ms	824.71 Hz	1895.47 Hz	2222.11 Hz
/æ/ in carrier word "sat" (S5)	180 ms	1039.93 Hz	1683.09 Hz	2465.56 Hz
/æ/ in carrier word "mat" (S7)	345 ms	889.01 Hz	1565.47 Hz	2393.55 Hz
Mean Score of three recordings	261.33 ms	917.88 Hz	1714.68 Hz	2360.41 Hz
Standard Deviation (SD)	67.38	90.20	136.56	102.11
Coefficient of Variation (CV)	0.26	0.10	0.08	0.04

APPENDIX B

Raw Scores for the Post-Test and the Delayed Pos-Test

Table 12 – Control Group Raw Scores (Familiar tokens)

Participant	Pre Test	Post Test	Delayed Post Test
P005	14	12	10
P015	7	8	12
P019	13	10	14
P153	12	10	13
P031	9	8	8
P045	6	7	10
P067	9	11	10
P072	8	8	7
P074	7	7	11
P089	11	6	8
P094	11	9	12
P109	9	10	14
P110	9	8	9
P111	7	10	10
P121	15	13	14
P129	11	8	9
P133	8	9	8
P135	9	8	10
P147	14	11	13

N = 16 for each test (familiar tokens only).

Table 13 – Natural Stimuli Group Raw Scores (Familiar tokens)

Participant	Pre Test	Post Test	Delayed Post Test
P016	11	13	13
P018	12	12	13
P021	13	15	12
P026	10	9	10

P027	6	13	6
P033	12	12	14
P055	13	15	9
P148	11	9	12
P077	8	10	10
P087	7	14	12
P088	9	10	14
P154	9	15	13
P093	10	13	14
P105	9	12	10
P107	10	9	9
P123	12	8	8
P132	10	9	7
P134	8	11	9
P140	6	8	8
P142	8	8	10
P145	8	11	12
P117	8	12	6

N = 16 for each test (familiar tokens only).

Table 14 – Synthetic Stimuli Group Raw Scores (Familiar tokens)

Participant	Pre Test	Post Test	Delayed Post Test
P008	10	11	13
P036	13	13	15
P046	9	10	8
P052	7	7	7
P056	9	7	12
P060	8	8	9
P108	9	15	14
P112	5	16	12
P115	13	15	14
P119	10	12	11
P125	13	12	15
P138	5	9	7
P141	14	14	11
P143	16	15	14
P103	13	10	13

N = 16 for each test (familiar tokens only).