THE RELATION BETWEEN
PERCEPTION AND PRODUCTION
IN L2 PHONOLOGICAL
PROCESSING: THE ROLE OF THE
PHONOLOGICAL LOOP

master's thesis by
Gerda Ana MELNIK

Under the supervision of
Dr Sharon PEPERKAMP

Submitted in June 2016

Submitted in partial fulfillment of the academic requirements for the
Master's degree in Cognitive Science
co-organized by
EHESS – ENS –Université Paris Descartes
Abstract

In this study we looked at the relationship between perception and production in L2 phonological processing. We focused on the perception and production of the French contrast /u/-/y/ by proficient American English learners of French and by a group of native French speakers. An ABX perception task and two production tasks, namely, reading and picture naming, were implemented in this study. We tested the hypothesis that the phonological processing is dependent upon the possibility to use the “phonological loop” during perception and to subvocally rehearse the stimuli.

We therefore manipulated the ISI in the perception task, including a condition with an ISI of 150 ms (short ISI) and an ISI of 1000 ms (long ISI). We expected to observe a significant correlation between perception and production only in the long ISI condition.

As predicted, the results showed that in perception, bilinguals performed worse than native French speakers. Bilinguals perceived the test contrast /u/-/y/ with less accuracy than the control contrast /u/-/a/, reflecting the difficulty of the former for the perception of American English speakers of French. French participants performed well on both contrasts. In production, in the naming task native French speakers produced significantly more distinct categories than bilingual speakers. In reading, the difference between the groups did not reach significance. We propose that this lack of difference could be explained by the faster reading rate of French speakers that resulted in vowel reduction. Contrary to our predictions, performance on the perception task correlated with the pronunciation accuracy of /u/ and /y/ in both ISI conditions. This finding was discussed in light of previous studies on the relationship between perception and production in L2 phonological processing.
Declaration of Originality

Previous studies looked at the perception or production of L2 sounds and only few studies concentrated on the relationship between those two modalities. In our study we will further investigate this relationship focusing on American English speakers of French. This is the first study looking at the role of the phonological loop in the relationship between perception and production of L2 phonological processing.

Acknowledgments

This work was made possible thanks to the help and insights of many people, to whom I wish to express my most heartfelt thanks. First and foremost, I am extremely grateful to my supervisor Dr Sharon Peperkamp. Thanks to her I discovered the precise topics in psycholinguistics that I would like to continue working on in the future. Her guidance throughout the year helped me to deal with methodological problems occurring while preparing the experiments.

I would also like to thank Yue Sun who helped me a lot during my internship in M1, and taught me how to use the recording equipment. He also gave me my first lessons in computer programming. For my internship this year I used some of the scripts in R that Yue left me last year.

I am grateful to Alexandre, Charlotte and Auréliane who kindly agreed to be recorded for my stimuli. I would also like to thank all the members of LSCP in general, who provided such an amazing environment. Being still new to the field, I am indebted to many of them for their kind and professional advice.

Lastly, I thank my family and my dear friends for their never ending support.
# Table of Contents

Abstract .......................................................................................................................... 2
Acknowledgments ........................................................................................................ 3
1. Introduction ............................................................................................................. 6
   1.1. Perception and production of L2 sounds ......................................................... 6
   1.2. Studying the relationship between perception and production of L2 sounds .......................................................................................................................... 7
   1.3. Research Question ......................................................................................... 13
   1.4. Predictions ...................................................................................................... 15
2. Study Methodology ............................................................................................... 17
   2.1. Participants .................................................................................................... 17
   2.2. Stimuli and procedure for the perception task .............................................. 17
   2.3. Production tasks: .......................................................................................... 18
      2.3.1. Stimuli and procedure for the naming task .............................................. 19
      2.3.2. Stimuli and procedure for the reading task ............................................ 19
3. Results .................................................................................................................. 20
   3.1. Production: Results and discussion ............................................................... 20
   3.2. Perception: Results and discussion ............................................................... 25
   3.3. Correlations between perception and production: results and discussion ...................................................................................................................... 26
4. General Discussion ............................................................................................... 32
   4.1. Production ...................................................................................................... 32
      4.1.1. Reading versus naming .......................................................................... 32
      4.1.2. The production accuracy of /u/ versus /y/ by American English speakers of French ........................................................................................................ 33
   4.2. Perception ...................................................................................................... 39
   4.3. Correlation between perception and production ......................................... 40
5. Appendices ............................................................................................................ 45
6. References ............................................................................................................. 48
1. Introduction

1.1. Perception and production of L2 sounds

It is well known that adult speakers have great difficulty producing and perceiving speech sounds that are not used in one's native language (for reviews, see Piske et al. 2001; Sebastián-Gallés 2005). It has been argued that the capacity to tune one's speech processing system to foreign sounds is limited to a critical period that ends at puberty (Lenneberg 1967). Nevertheless, there is evidence that the ability to modify speech perception and production patterns is - to a certain extent - retained well into adulthood (e.g., Flege et al. 1993). The best known models of non-native speech production and perception – the Perceptual Assimilation Model (Best 1995) and the Speech Learning Model (Flege 1995) assume that the success of learning to produce and perceive non-native sounds depends on the phonetic similarities and dissimilarities between L1 and L2 segments. Phonetic similarity or dissimilarity is defined in terms of the articulatory and acoustic characteristics of the linguistically relevant speech sounds (Rojczyk, 2010).

Numerous studies have shown that speech perception becomes attuned to the contrastive sound units of a particular language very early in life. For instance, Kuhl et al. (2006) tested 6 to 12 months old infants. They found a significant increase in performance for the native-language contrast in the first year and a decline in the perception of non-native contrasts over the same time period. Following this attunement to one's native language, the perception of a foreign language that differs from the L1 in some phonological characteristics, can be distorted. These distortions that arise from the mismatch between the properties of the maternal language and the foreign one can manifest themselves in various ways (Sebastián-Gallés 2005; Davidson & Shaw 2011). Different types of “phonological illusions”, i.e. misperceptions of segments in order to ‘repair’ illegal sequences of segments or unknown segments in terms of native categories, are attested: epenthesis (inserting an illusory (epenthetic) phoneme, e.g. Japanese speakers perceiving an illusory vowel /u/ in illegal consonant clusters (Dupoux et al. 1999)); mutations (changing one sound into another, e.g. French speakers perceptually transforming the
onset clusters /tl/ and /dl/, into onset clusters /kl/ and /gl/ (Hallé et al. 1998)); deletions (removing a phoneme to correct illegal clusters, e.g. vowel-adjacent liquids delete Thai (Yun 2012)).

In L2 production, the problems caused by the attunement to one’s native language are even more evident than in perception. Namely, foreign accentedness is one of the most salient features that accompanies foreign language learning in adulthood. The degree of perceived foreign accent depends on multiple factors, such as age of L2 learning, length of residence in an L2-speaking country, gender, formal instruction, motivation, language learning aptitude, amount of native language (L1) use and communicative pressure (Piske et al. 2001). Studies that have investigated the perception and production of non-native phonemes have typically focused on the influence of native language (L1) sound patterns either on the production (Listerri, 1995) or perception (Best, 2001) of the second language (L2). There are, however, relatively few studies on the relation between perception and production in L2 phonological processing.

1.2. Studying the relationship between perception and production of L2 sounds

In one of the earliest studies, on the topic Polivanov (1931/1974) hypothesized that the learners’ deviant productions are due to “the subjective nature” of the perception of sounds. Later on Trubetzkoi (1969), invented the term “phonological sieve”, which became a widely used metaphor for the idea that our L1 phonology filters out those properties of the L2 speech signal which are not relevant to the phonological system of our L1. Thus, according to those theories in the process of learning an L2 system, perception usually precedes production, i.e. learners will more easily accurately identify and discriminate L2 phones than accurately produce these L2 phones, the underlying idea that learners cannot produce sounds accurately without first perceiving differences between them. The Speech Learning Model proposed by Flege (1987; 1995) predicted that the production of L2 contrasts depends on the perception of L2 and on whether the learner has established or not a new phonological category for the non-native phoneme to be learned.

Many experimental studies have yielded evidence for the perception-before-production hypothesis. For instance,
Rochet (1995) showed that Canadian English and Brazilian Portuguese learners produced the French vowel /y/ inaccurately because of inaccuracy in their perception of this L2 sound. Likewise, Detey & Racine, (2015) tested the ability of Japanese beginner learners to distinguish three French nasal vowels /ɑ̃ɔ̃ɛ̃/ an AXB discrimination task and compared the perception results with the results of a corpus-based analysis of their production of the same vowels. Their findings support the hypothesis that perception precedes production in the acquisition of a foreign phonological contrast. Similarly, perceptual training is often used to improve both L2 perception and production. For instance, Wong (2015) studied the effects of High Variability Phonetic Training in training both the perception and production of English high-front vowels and high-back vowels by Cantonese ESL learners. Results showed that the training was generally effective in improving the participants' perception of the two vowel pairs and perceptual learning could be transferred to the production domain.

However, some other studies demonstrated the opposite effect, namely that production can precede perception. For example Sheldon & Strange (1982) showed that Japanese learners of English can accurately produce English /r/ and /l/ despite poor perception. Similarly, in a two-year-long study Evans & Iverson (2007) investigated changes in vowel production and perception among university students from the north of England, as individuals adapt their accent from regional to educated norms. They showed that participants adapt their spoken accent after attending university. However, no perceptual changes occurred over time. Moreover, some studies showed that production might precede perception for some sounds but follow perception for others. For example, Zampini (1998) found that English speakers' production of Spanish /p/ was relatively native-like, yet their perception displayed large variation. In contrast, their production of Spanish /b/ deviated more substantially from native speakers than their perception of the same sound. The author proposed that L2 perception precedes production probably only when the target sound is a novel phonetic category, such as the Spanish /b/. When the L2 sound is similar to an L1 category, such as the Spanish /p/, the learners' production may lead their perception. Similar results were obtained by Hao & de Jong (2016) who demonstrated that Korean speakers learning English stops showed a production-leading-perception pattern while English speakers learning Mandarin tones showed an opposite pattern. Hence the
authors suggest that the precedence relationship between L2 perception and production may not be constant, and is likely to be affected by the similarity of the L2 sounds to the L1.

The study of such a causal relationship between the performance in one modality and another also encountered various problems related to the methodological difficulty of assessing and comparing results from production and perception experiments. First, results might differ even within one modality depending on the differences in tasks. For instance, Mack (1989) compared the perception and production accuracy of English-French bilinguals and English monolinguals on the /d-t/ and /i-/ contrasts. She found that bilinguals performed differently from the monolinguals in identification but not in discrimination task. The author proposes that this difference may be caused by the characteristics of these tests. Namely, she argues that compared to discrimination, the identification tests use a language-based response strategy, which may have rendered the identification tests more sensitive to indices of linguistic experience and hence to differences between the bilinguals and the monolinguals. This proposal parallels with the finding that L2 listeners' performance in perception of non-native sounds crucially depends on the nature of the task. The performance gap between native and non-native listeners has been shown to increase as the tasks have greater lexical involvement (Diaz et al. 2012).

Similar issues with task characteristics were reported in studies testing production. For example, Salaberry & Lopez-Ortega (1998) looked at the accuracy of L2 Spanish production across 3 different tasks (narrative task, multiple-choice cloze test, and fill-in-the-blanks cloze test) on 3 discourse-determined grammatical items (subject pronouns, articles, and past tense aspect) among 74 native English speakers. They found different rates of accuracy in L2 production according to task type and grammatical item. Hao & de Jong (2015) looked at the performance of second language learners in an Imitation task compared to that in a Read-Aloud task and an Identification task. They raised the question of whether imitation reflects more production or perception skills. Although at first sight imitation seems to be a production task based on auditory instead of orthographic prompts, it can also be viewed as a perception task using a verbal rather than a written response. They found that the accuracy in Imitation was not always constrained by that in the Identification and Read-Aloud tasks. Therefore even
tapping a single modality can cause complex methodological issues.

This difficulty to assess the performance within on modality is even more salient when comparing performances between modalities. Tsukada et al. (2005) noted several problems inherent to such comparisons. First, it is difficult to directly compare the extent of native-like attainment in L2 production and perception data. This is because the tests used to assess production and perception have no common standard of measurement. For example, tasks used to test perception and production might not be appropriate to evaluate productive and perceptual abilities with the same degree of thoroughness. Second, the tasks used to assess perception and production might be not of equivalent difficulty for L2 speakers. For instance, the task in one modality might be cognitively more demanding than the task in the other. Moreover, previous research has suggested that production and perception might differ as the social consequences of errors in both modalities can be different (e.g. Mack, 1989; Bohn & Flege, 1997). For example, non-native speakers might feel pressure to demonstrate “language loyalty” by speaking with a foreign accent. A similar social pressure might not exist for L2 vowel perception, however.

Fortunately, the relationship between perception and production can also be examined by looking at whether there is a correlation between both modalities. While this does not allow one to examine causation, it does allow one to look at whether there is a relation between the two. For example, a correlation can show whether production and perception accuracy decrease or increase together, independently of which of the modality precedes the other. Results from several experiments have shown moderate correlation between production and perception in L2. Flege (1993) carried out four experiments, which examined the production (in an imitation task) and perception (in a forced choice with adjustment task) of vowel duration cues to the word-final English /t/-/d/ distinction. Both late Chinese English bilinguals and early Chinese English bilinguals produced vowels significantly longer before /d/ than /t/. However, the duration differences produced by the late bilinguals were smaller than those produced by the early bilinguals or by native speakers of English. In perception participants selected the best instances of 'beat' and 'bead' from a natural-edited beat-to-bead continuum, and the best instances of 'bat' and 'bad' from a bat-to-bad continuum. The late Chinese English bilinguals chose stimuli with significantly longer vowels as the best
instances of /d/-final than of /t/-final English words. However, their perceptual differences were smaller than those of the early bilinguals or native English speakers. The size of vowel duration differences observed for the Chinese participants in production and perception showed a modest positive correlation. Another example concerns studies that showed a correlation between auditory acuity and vowel production (Perkell et al. 2004; Perkell et al. 2008). In these studies, participants carried out two tasks: a discrimination task on a vowel continuum and a reading task. The results showed that participants that were better at the discrimination task produced vowels more consistently, with less within-phoneme variability, but spaced further apart in vowel space (larger between-phoneme acoustic distance).

Other studies, however, revealed more puzzling results. Strange (1995), observed that although the segmental production errors of inexperienced L2 learners can often be predicted from their perceptual errors, perceptual difficulties may persist even after segmental production has been mastered. Strange concluded that L2 segmental production and perception by experienced speakers of an L2 may be uncorrelated. Golestani et al. (2006) looked at the anatomical correlates of perception and production and showed that the anatomical differences that predict behavioral measures of phonetic perception and production partially dissociate and that individuals who are faster at learning to perceive foreign speech sounds are not necessarily the ones who are good at correctly pronouncing foreign speech sounds, and vice versa.

These findings parallel with the observed behavioral dissociations between phonological production and perception deficits in patients with lesions. Praamstra et al. (1991) report a case of auditory verbal agnosia (word deafness). Word deafness is a form of auditory agnosia in which the ability to comprehend auditorily presented speech is profoundly impaired, although patients are not deaf, and may have normal pure tone thresholds (Hickok & Poeppel, 2000). The patient described in Praamstra et al.'s study had deficits in language comprehension and impaired general auditory processing. However, he had fluent speech. The opposite case, involving deficits in production but intact speech production, has been reported by Dronkers (1996) who studied patients suffering of apraxia and thus impaired in their ability to coordinate speech movements but having an intact ability to perceive speech sounds. Another example could be conduction aphasia. People with such aphasia have fluent speech yet produce relatively frequent and predominantly
phonemic speech errors, although their perception and auditory comprehension at the word and conversational level are well preserved (Hickok 2014). Moreover, Stasenko et al. (2015) suggested that the identification (i.e. labeling) of nonword speech sounds may involve the speech motor system, but that the perception of speech sounds (i.e., discrimination) does not require the motor system. They studied this hypothesis by evaluating which aspects of auditory speech processing were affected in a stroke patient with dysfunction of the speech motor system.

The existence of a correlation between perception and production might also depend on some other factors. For instance, Levy & Law (2010) investigated whether a perception-production relationship exists independently of language experience. They studied the production and perception (Levy 2009) of French contrasts /y-u/, /u-œ/ and /y-œ/ by American English learners of French, and found a modest, yet complex, perception-production link for L2 vowels, though only the contrast /y-œ/ revealed compelling evidence for a perception-production relationship for all groups, independently on their proficiency level. Finally, Peperkamp & Bouchon (2011) carried out a study with French-English bilinguals and found that the perception of an English vowel contrast did not correlate with the pronunciation accuracy of the same contrast.

They propose that this result could be explained by the differences in experimental design between their study and previous ones, notably, the differences of ISI (interstimulus interval) in perception tasks. According to the authors, phonological processing is dependent upon the possibility to use the “phonological loop” during perception and to subvocally rehearse the stimuli. This assumption relies on the concept of the phonological loop from the multicomponent model of working memory coined by Baddeley & Hitch (1974). According to this model, the working memory contains a crucial component for language called “the phonological loop”. It “is specialized for the retention of verbal information over short periods of time; it comprises both a phonological store, which holds information in phonological form, and a rehearsal process, which serves to maintain decaying representations in the phonological store” (Baddeley, 1998). Baddeley pointed out to the role of the phonological loop in
learning the novel phonological forms of new words. Jacquemot & Scott (2006) adapted this theory to speech processing, arguing that the phonological loop arises from the cycling of information between the two buffers (phonological store and subvocal rehearsal mechanism), thus mediating between the perception and production (Figure 1).

These suggestions are paralleled with the observation that a correlation between perception and production has been previously found only in studies which used a large ISI (at least 1.2 ms). When, however, the ISI used in the perception task is small (500ms in their task), participants do not have the time to use the phonological loop.

1.3. Research Question

In our study we further investigated the hypothesis proposed by Peperkamp & Bouchon (2011) that the correlation between perception and production in L2 phonological processing is dependent upon the possibility to use the phonological loop during perception and to pronounce the stimuli subvocally. We tested American English-French bilinguals and French monolinguals as control participants. We focused on the French /y-u/ contrast (as in “pull” – “poule”) that does not exist in English and which has been reported as being one of the most difficult for American English speakers to produce (Levy & Law, 2010) and perceive (Levy & Strange, 2008) accurately.
In most languages of the world back vowels that are not low are rounded, while front and central vowels are not. In a minority of world's language (according, to The World Atlas of Language Structures Online only 6.6%) front vowels are rounded. French has three such vowels: /y/ (as in “pull”), /ø/ as in “peu” and /œ/ (as in “boeuf”). Lip rounding involves drawing the corners of the lips together and protruding the lips forward from their normal rest position. Lip rounding is greatest when the vowel is high as the jaw is open as the amplitude of lip adjustment becomes more limited since the lips are being stretched vertically (Dryer & Haspelmath, 2013).

Acoustically, the second formant frequencies (F2) of front rounded vowels /y/ are lower than for front unrounded vowels /i/ because lip rounding increases the length of the oral cavity, lowering the frequencies of the acoustic resonances of the vocal tract. Back rounded vowels (/u/) have even lower F2 frequencies than front rounded vowels /y/ as the backing of the body of the tongue adds to lip protrusion to increase even more the oral cavity. Although acoustically front rounded /y/ is more similar to PF /i/, American English listeners tend to perceive /y/ as more similar to /u/ than to /i/ (Flege & Hillenbrand, 1984). An explanation for this phenomenon was proposed by Levy (2009). According to her, as American English listeners rely on the phonological system of their L1 and in American English round vowels are back vowels, they might associate the feature of roundedness with back vowels, i.e. /u/. It is also important to note, that American English /u/ and French /u/ differ acoustically. American English /u/ has higher F2 frequency than French /u/. Therefore, when native American English speakers attempt to produce French-like /u/, they have to lower the F2 frequency of their L1 /u/.

In order to study the relationship between perception and production we implemented an ABX perception task and two production tasks, namely reading and picture naming. In the perception task, we tested participant’s perception of the French contrast /y-u/ and compared it to their perception of the contrast /u-a/. The latter was chosen as a baseline for performance in this task and hence was used as covariate for the correlation between perception and production.

In order to test the hypothesis on the role of the phonological loop in L2 perception we manipulated the ISI. A condition with an ISI of 150 ms and one with an ISI of 1000 ms was included in the ABX task. As the stimuli were trisyllabic and hence relatively long (mean length 686 ms), the 150 ms ISI was considered to be short enough to block the phonological
loop, while the long one was expected to be sufficiently long to assure that the mechanism of subvocal rehearsal is used. An intra-subject design was chosen so that both ISI conditions were tested for each participant.

As mentioned above, we used a reading and a naming task to assess participants' production of the vowels /u/ and /y/. In naming, which was chosen to prevent a possible interference of orthography in the performance (Escudero et al. 2008; Showalter & Hayes-Harb 2013), participants were recorded while naming objects presented on a screen. The names of objects contained the vowel /u/, /y/ or other vowels (these words were used as fillers). Similarly, in reading participants were recorded while reading a small text in French containing words with /u/ and /y/. The production tasks allowed to calculate the Euclidian distance (distance between two points in Euclidean space) between the French vowels /u/ and /y/ produced by different participants. A larger distance between the vowels would mean that the participant makes a better distinction between these categories.

1.4. Predictions

1. In perception, bilinguals are expected to perform worse than native French speakers. Bilinguals will perceive the test contrast /u/-/y/ with less accuracy than the control contrast /u/-/a/, reflecting the difficulty of the former for the perception of American English speakers of French. French participants are expected to perceive both of these contrasts well, as both of them are native in their L1. However, French participants might also perform slightly worse on the /u/-/y/ contrast as acoustically the distance between /u/ and /y/ is smaller than between /u/ and /a/ and thus it could be easier to perceive a larger acoustic difference than a smaller one. Nevertheless, this difference should not be very salient.

2. In production, native French speakers should possess more distinct categories compared to bilingual speakers. Thus, the Euclidian distance between /u/ and /y/ produced by bilingual speakers was predicted to be significantly lower to

---

1 The distance between two points is the length of the path connecting them. According to the Euclidean distance formula, the distance between two points in the plane with coordinates (x, y) and (a, b) is given by

\[
dist((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2}
\]
the Euclidian distance between those vowels in the production of native French participants.

Regarding the results of reading versus naming, we predicted two possible outcomes. The task that will be cognitively more demanding is expected to be harder and result in less accurate production of bilinguals. On one hand, naming might be cognitively more demanding for bilinguals, as they have to retrieve the lexical item in question. Following this scenario, the difference between bilinguals' and French participants' production accuracy would be larger in naming than in reading as bilinguals would perform worse in naming.

On the other hand, reading could be expected to be more demanding as it requires to produce many syllables per second. We would thus predict that the difference between bilinguals' and French participants' production accuracy would be larger in reading than in naming due to the poorer performance of bilinguals in reading.

3. Given the hypothesis that the perception-production correlation depends on the possibility to use the phonological loop during perception, we expect to observe a significant correlation between perception and production only in the long ISI condition.


2. Study Methodology

2.1. Participants

Twenty American English speakers (16 females and 4 males) living in Paris were recruited by means of advertisements placed in the International Student Campus and in various student organizations. We focused on speakers of only one dialect of English, as some vowels vary across dialects and we wanted to avoid this variability. They were students staying in France for at least one year (mean length of stay 3.1 years) and they all were proficient speakers of French (aged between 20 and 31). In order to assess speakers' level of bilingualism we used a questionnaire based on the bilingual dominance scale developed by Dunn & Fox Tree (2009). This scale targets three main criteria important in describing dominance: 1. percent of language use for both languages; 2. age of acquisition and age of comfort for both languages; and 3. restructuring of language fluency due to changes in linguistic environments. Further in this study American English speakers of French will be called “bilinguals”. We use this term for simplicity only therefore we will not discuss here the somewhat problematic definition of this term.

An additional 10 native French speakers were recruited as control participants (6 females, 4 males). They were between 21 and 35. They had no hearing or language problems. Data from two bilingual and one French participant had to be discarded because they performed at chance level in the perception task. However, we kept the French participant's data for the production tasks as data from French participants was not used for correlation analysis.

All participants completed the whole experiment (around 45 minutes) and were paid 10 euros for their participation.

2.2. Stimuli and procedure for the perception task

We used an ABX discrimination paradigm. Participants were presented with trisyllabic nonsense words over headphones. They had to determine whether X was the same as A or as B. The non-word in the triplets were produced by different speakers: A and B were produced by two female
native French speakers, X by a male French speaker. The phonetic variability resulting from multiple talkers prevents participants from focusing on low-level acoustic features, thus allowing to tap the higher level (phonological) processing (Barcroft & Sommers, 2005). Recordings of the three talkers were made in a soundproof booth, at 16 bits mono with a sampling rate of 44.1 kHz.

The test phase consisted of 160 trials, eight per triplet, divided over four blocks. In each block, half of the trials concerned the experimental /y/-/u/ contrast, the other half a control contrast (/u/-/a/); the identity of X as well as the correct response (A or B) was counterbalanced and the trials were presented in a random order. The blocks were homogenous for ISI, resulting in two blocks with a short ISI and two blocks with a long ISI. Half of the participants were first presented with a short ISI block, the other half started with a long ISI block. The ISI block types alternated, resulting in a sequence of blocks short ISI → long ISI → short ISI → long ISI for half of the participants and long ISI → short ISI → long ISI → short ISI for the second half.

Twenty triplets of trisyllabic French pseudo-words ending in /ã/, /ɛ̃/, /õ/ were created (e.g. /sanuvɛ̃/-/sanyvɛ̃/-/sanavɛ̃/) (Appendix). Members of each triplet differed only in the vowel of the second syllable, i.e. /u/, /y/, or /a/. We took into account the findings of Levy (2010) who showed that American English made more errors on discriminating pairs involving front vs. back rounded vowels (such as /y/ and /u/) in alveolar as opposed to bilabial contexts. Thus, we controlled for the consonantal context: in half of the stimuli the target vowel was preceded by an alveolar consonant (/t/, /d/, /n/), in the other half it was preceded by a bilabial consonant (/p/, /b/, /m/).

Each block of the experiment started with a warm-up phase of 5 trials, during which participants received feedback as to whether their responses were correct. In the case of an incorrect response or no response within 3500ms the trial was repeated until the correct response was given.

2.3. Production tasks:

Participants performed two production tasks, a picture naming task, and a reading task. The former minimizes the influence of orthography, but has the disadvantage that participants might not know the intended word or produce a
different one. The recordings for both production tasks were made in a soundproof booth.

2.3.1. Stimuli and procedure for the naming task

We selected 120 pictures of imageable nouns, 30 containing /u/, 30 /y/, and 60 fillers with neither of these vowels (Appendix). The nouns were likely to be familiar to all participants. Pictures were presented in a random order one by one on the screen. Participants were asked to name the object they saw and to press a button to proceed to the following picture.

2.3.2. Stimuli and procedure for the reading task

A short story in French containing 10 content words with /u/ and 10 with /y/ was created for the reading task (Appendix). Target words always occurred at the end of a sentence. In order to control for the consonantal context as in the perception task, the consonant preceding the crucial vowel was bilabial for half of the stimuli and alveolar for the other half. Participants were recorded while reading the story; they were instructed to read as naturally as possible.
3. Results

3.1. Production: Results and discussion

For each production task, we measured the first two formants (F1 and F2) of all tokens of /u/ and /y/ occurring in the target words, using the software Praat (version 6.0.17). The waveform and the wideband spectrogram of the production data were visualized and an annotation text was added for each audio file to segment and label the target vowels. To manually segment the vowels, boundaries were set at zero crossings. After segmentation, the mean values of the first two formants (F1 and F2) were automatically measured for each target vowel. The mean F1 and F2 values of each vowel were calculated, as well as the Euclidean distance between them, i.e., the space in Hertz (Hz) between the vowels /u/ and /y/ in the F1/F2 space. In order to identify outliers we used Mahalanobis distance metric measure, which estimates the distance between a point and a distribution. Thus, this measure allows us to capture the natural variability of the target vowel space (Majewski et al. 1979). Outliers beyond 2.5 standard deviations from the center of each category were discarded. Twenty tokens (17 in Reading and 3 in Naming) were also discarded because those vowels were aspirated. Figures 2 – 5 show scatterplots for individual productions of /u/ and /y/ by bilingual and French participants in naming and reading, respectively. Each square represents one participants' performance.

The Euclidian distance between /u/ and /y/ was computed again after discarding outlier tokens. The Euclidean distance between both categories (in the scatter plots, clouds of green versus red points) for each participant was taken as a measure of their production accuracy. This measure was used for statistical analyses.
Figure 2. Mean F1 and F2 frequencies (Hz) for the tokens of /u/ and /y/ produced by bilingual participants in the reading task. Each square represents the productions of one participant. Colours represent the productions of the vowels /u/ and /y/.

Figure 3. Mean F1 and F2 frequencies (Hz) for the tokens of /u/ and /y/ produced by French participants in the reading task. Each square represents the productions of one participant.
Figure 4. Mean F1 and F2 frequencies (Hz) for the tokens of /u/ and /y/ produced by bilingual participants in the naming task. Each square represents the productions of one participant.

Figure 5. Mean F1 and F2 frequencies (Hz) for the tokens of /u/ and /y/ produced by French participants in the naming task. Each square represents the productions of one participant.
All statistical analyses in this study were carried out using the R statistical computing software, version 3.1.1.

A t-test was carried out for each of the production tasks in order to compare the scores of bilinguals and controls. As predicted, native French controls produced /u/ and /y/ more accurately than bilinguals in the naming task (t(22.1)=2.8, p = 0.009), the mean distance between those categories produced by French participants being 703 Hz, while that of bilinguals being 535 Hz. However, in the reading task the difference of production scores for bilinguals and controls did not reach significance (t(25.3)=2, p = 0.057, Bilingual Mean =491 Hz; French Mean = 644 Hz) (Figure 6).

Figure 6. Distance between the categories /u/ and /y/ (Hz) produced by participants in naming and reading. The colours represent the bilinguals' vs. native French speakers' performance.

This is contrary to our predictions as we expected French participants to have more distinct categories /u/ and /y/ and this should have been reflected in both tasks. Thus why do French participants produced the target vowels more accurately than bilinguals only in naming but not in reading? We hypothesized that this difference between the tasks could be explained by the different speech rates of bilinguals versus French speakers in reading and naming tasks. Namely, French speakers were likely to read faster as it is known that vowels
tend to be reduced in L1 compared to L2 reading as L1 reading approaches spontaneous speech which is often subject to vowel reduction (Spilková & Dommelen, 2007). In order to test this hypothesis we measured the duration of each target vowel produced by the participants (Figure 7.).

![Figure 7](image_url)

**Figure 7.** Duration of the target tokens produced by bilingual and French participants in naming and reading. Colours represent the performance of bilingual vs. native French speakers.

We performed a linear mixed effects analysis to check for interaction between Group and Task, i.e. if the effect of group was larger in reading than in naming. The R package lme4 (Bates, Maechler & Bolker, 2012) was used to carry out these analyses. Group (bilingual vs. French) and Task (reading vs. naming) were entered into the model as fixed effects. As random effects, we had intercepts for participants. The analysis revealed a main effect of Group ($\beta = 0.1$, $t = 30.2$). $p$-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. We found a significant interaction between Group and Task ($\chi^2(1) = 12.06, p = 0.0005$).

We then performed a linear mixed effects analysis of the relationship between Duration of target tokens and Group (bilingual vs. French) separately for reading and naming. For each of the tasks we entered into the model Group (bilingual vs. French) as fixed effect. As random effects, we had intercepts for participants. Here too, $p$-values were obtained
by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. In naming task Group affected Duration ($\beta = 0.1$, $t = 30$, $\chi^2(1) = 8.08$, $p = 0.004$), lowering it by about 21 ms, (standard error 7 ms). In reading, Group also affected Duration ($\beta = 0.1$, $t = 23$, $\chi^2(1) = 20.07$, $p < 0.001$), lowering it by about 37 ms (standard errors 7 ms).

Thus, French speakers produced significantly shorter tokens than bilinguals, this difference being more salient in reading.

### 3.2. Perception: Results and discussion

For the perception task, the correctness rates were submitted to two ANOVAs, one for each of the ISI conditions, with the factors Group (bilinguals vs. French) and Contrast (/u/-/y/ vs. /u/-/a/) (Figure 8.)

**Figure 8.** Mean percent of correct responses in the perception task by bilingual and French participants in both ISI conditions (short ISI on the left side of the graph, long ISI on the right side). Colours represent the contrasts on which participants were performing, i.e. the control contrast /u-a/ and the test contrast /u-y/.
For the short ISI condition (left side of the graph) the ANOVA revealed main effects of Group ($F(1, 25) = 26.77, p < 0.001$) and Contrast ($F(1, 25)=38.66, p < 0.001$), as well as an interaction between these two factors $F(1, 25) = 10.70, p = 0.003$). The interaction was due to the fact that the bilinguals were more accurate on the /u-/a/ than on the /u-/y/ contrast ($F(1,17)=36.26, p<.0001$; whereas the English showed no difference ($F(1,8) = 3, p=0.1$). Similar results were obtained for long ISI condition (right side of the graph), where we found main effects of Group ($F(1, 25) = 30.57, p < 0.001$) and Contrast ($F(1, 25) = 27.892, p < 0.001$), and an interaction between these two factors ($F(1,25) = 8.871, p = 0.006$). The interaction between Group and Contrast in long ISI condition was due to the fact that the bilinguals were more accurate on the /u-/a/ than on the /u-/y/ contrast ($F(1,17)=25.24, p<.001$); whereas the French showed a very small difference between the perception of both contrasts. As predicted they perceived /u-/a/ slightly better than /u-/y/ ($F(1,8) = 6.1, p = 0.04$).

Thus we observe as predicted that bilinguals perceived the test contrast /u-/y/ with less accuracy than the control contrast /u-/a/. French participants, on the other hand had no difficulty in perceiving either of the contrasts.

### 3.3. Correlations between perception and production: results and discussion

In order to investigate whether perception and production were correlated in bilingual participants, we carried out a linear regression for each type of ISI and each production task between the production score on the one hand and the correctness rate on the /u-/y/ contrast in the perception task on the other hand. The correctness rates on the /u-/a/ contrast in the perception task were entered as a covariate.

The results of the correlation analysis were contrary to our hypothesis. We predicted that the correlation between perception and both production tasks will be observed only in the long ISI condition as this would allow the participants to use the phonological loop and rehearse the stimuli subvocally. Nevertheless, we found significant correlations between perception and production in both ISI conditions and both production tasks:

- Short ISI naming: $r^2 = 0.276, p = 0.035$
- Short ISI reading: $r^2 = 0.379, p = 0.011$
Long ISI naming: $r^2 = 0.254, \ p = 0.043$;

Long ISI reading: $r^2 = 0.41, \ p = 0.008$.

In order to graphically represent the correlations and to take into account the covariates we produced the following plots. The size of the points reflect the influence of the covariate (perception score on the test contrast /u/-/a/).

**Figure 9.** Correlation between accuracy of production in the reading task (Hz) and the percentage of correct response in the perception of the test contrast /u/-/y/ in the long ISI condition. The size of the points represent the influence of the covariate (correct perception of the control contrast /u/-/a/).
Figure 10. Correlation between accuracy of production in the naming task (Hz) and the percentage of correct response in the perception of the test contrast /u/-/y/ in the long ISI condition.

Figure 11. Correlation between accuracy of production in the reading task (Hz) and the percentage of correct response in the perception of the test contrast /u/-/y/ in the short ISI condition.
Figure 12. Correlation between accuracy of production in the naming task (Hz) and the percentage of correct response in the perception of the test contrast /u/-/y/ in the short ISI condition. The size of the points represent the influence of the covariate (correct perception of the control contrast /u/-/a/).
4. General Discussion

4.1. Production

4.1.1. Reading versus naming

The present study yielded mixed findings regarding the production accuracy of bilingual versus French speakers. We expected bilingual participants to produce /u/ and /y/ less accurately than native French speakers (with less distance between the two vowels on the Euclidian space) and hence, to have less distinct categories /u/ and /y/.

In naming we indeed found French participants to be significantly more accurate than bilinguals as they produced tokens /u/ and /y/ that were distributed further apart from each other compared to bilinguals' productions. However, in reading the difference between French and bilingual participants did not reach significance.

We predicted two possible types of outcomes regarding the results of both production tasks. First, naming might have been hard for bilinguals as it involved the retrieval of lexical items in L2. Thus, the difference between bilinguals' and French participants' production accuracy would be higher in naming as bilinguals would perform less accurately in a cognitively more demanding task.

A second possibility was that reading could be a more demanding task as it involves producing multiple syllables per second. Following this scenario, the difference between bilinguals' and French participants' production accuracy would be higher in reading than in naming as bilinguals would be less accurate in reading.

At first sight, the first explanation at least partly corresponds to our results as we observed a significant difference between bilinguals and French speakers in naming but not in reading (although the initial prediction was that there will be less difference in reading but that there will still be some difference. What we found was no significant difference at all). However, an inspection of the data and of the plot (Figure 7.) suggests a third, unpredicted, explanation for the results. Namely, the lack of significance between the two groups in reading could be caused not by a poor performance of bilinguals but by the performance of French
participants. In fact, a post hoc analysis of the duration showed that French participants produced significantly shorter tokens of /u/ and /y/, and this effect was especially salient in reading. As the tokens of /u/ and /y/ produced by French speakers where shorter, the distance between those vowels was reduced. Vowel reduction occurs under many conditions and in various prosodic contexts such as in unstressed position, at lower prosodic boundaries and increased speech rate. This phenomenon is often seen as a durational shortening, as vowel reduction is often accompanied by the shortening of the vowel (Mooshammer & Geng, 2008). Lindblom (1963) described this durational shortening as a “target undershoot”, i.e. a formant undershoot occurring because the articulators do not reach the vowel-specific target due to temporal constraints. Thus, the lack of difference between French and bilingual productions in reading could be explained by the reduced productions of French participants and not by a lower accuracy of bilinguals. This issue with the performance of control subjects should be taken into account in further studies.

Another methodological problem concerning the reading task was related to the position of words containing target vowels. In the reading task the target vowels /u/ and /y/ were embedded in words sentence-finally as this position was expected to be emphasized by sentence stress. However, in several cases this resulted in the devoicing of those vowels by French speakers. Thus, we had to discard 17 such tokens. We can hypothesize that if target words occurred in other positions, we would have had less devoiced tokens.

4.1.2. The production accuracy of /u/ versus /y/ by American English speakers of French

Previous studies on the production of French /u/ and /y/ by American English talkers (Flege 1987; Levy & Law 2010) yielded different findings regarding the production accuracy of those sounds. Although the main goal of our study was to look at the relationship between perception and production of bilinguals, the data we collected also allowed us to inspect the production of /u/ and /y/ of our participants and compare the results to the results of the above mentioned studies.

Figure 13. presents plots of the mean F1 and F2 values of /u/ and /y/ produced by bilinguals and French monolinguals in naming and reading for female participants, while Figure 14.
presents the data for male participants. The data for female and male participants is split as the formant frequencies of vowels produced by male and female speakers tend to differ, i.e. vowels produced by female speakers have on average higher F1 and F2 values than those produced by male speakers. Each vowel on the plot represents the mean F1 and F2 values for each of the vowels /u/ and /y/ for each participant in each speaker group.

**Figure 13.** Mean F1 and F2 values for the vowels /u/ and /y/ produced by female French and bilingual participants. Each letter on the graph represents the mean productions for /u/ and /y/ for one participant.

**Figure 14.** Mean F1 and F2 values for the vowels /u/ and /y/ produced by male French and bilingual participants. Each letter on the graph represents the mean productions for /u/ and /y/ for one participant.

The mean frequency values of F1 and F2 for both tasks and both groups for the /y/ and /u/ vowels are presented in Table 1 and 2.
For both groups, both production tasks and both female and male participants, there was no overlap between /u/ and /y/. The front vowel /y/ had a relatively high F2, the back vowel /u/ had a relatively low F2. The F1 of French and bilingual speakers producing /u/ and /y/ did not differ or differed only slightly (the maximum difference of F1 between both groups was in female reading, were bilinguals produced /u/ 23 Hz lower than French participants). However, there was a clear difference between groups across tasks and genders for the production of F2 in both vowels.

Bilingual female participants produced /u/ with a higher F2 than French participants (92 Hz higher in naming and 98 Hz higher in reading) and the /y/ with a lower F2 than French participants (144 Hz lower in naming and 104 Hz lower in reading).
Male participants showed a similar tendency: bilingual participants produced /u/ with a higher F2 than French participants (108 Hz higher in naming and 57 Hz higher in reading) and the /y/ with a lower F2 than French participants (21 Hz lower in naming and 78 Hz lower in reading).

Overall, these results show the influence of the L1 on the production of L2 of the bilingual participants. Bilingual participants' native language - American English - has no /y/ but has an /u/ that is in between French /u/ and /y/, according the F2 frequency. Thus, they have to produce French high rounded vowels by lowering the F2 for /y/ and raising the F2 for French /u/.

\[
\text{/y/} \leftarrow \text{EN /u/} \rightarrow \text{FR /u/}
\]

Flege (1987) studied the production of /u/ and /y/ by two groups of female French speakers (monolinguals and French learners of English) and three groups of female English speakers of French differing in their level of proficiency. The English subjects in Group C had spent several periods of time in France (average of 1.3 years) while the subjects in Group D were married to native speakers of French and had been living in Paris for an average of 11.7 years at the time of the study. In our study the American English speakers have on average spent in France 3.1 years, thus they are between group C and group D in Flege's study according to their experience of French.

As in our study, in Flege's study all three groups of English subjects produced French /u/ with a mean F2 frequency that was higher, and therefore more English-like, than the French monolinguals. This difference between French and English /u/ was related to the level of proficiency in French. The F2 frequencies of French /u/ were decreasing across the groups as their proficiency was increasing. This shows that American English learners of French gradually improved their production accuracy of French /u/, producing it with much lower F1 frequencies than in English.

However, participants in English groups produced French /y/ with only slightly lower mean F2 frequencies than the group of French monolingual subjects. Moreover, there were no clear differences for the production of /y/ across groups of different proficiency. In our study, however,
bilinguals differed more from French speakers in the production of /y/ than in the production of /u/ (except for naming by male participants, where bilinguals produced almost French /u/ almost like native speakers of French).

In Flege's study all three native English groups produced smaller /u/- /y/ differences than native French speakers. The amount of these differences was related to French-language experience. The subjects in the most proficient group produced a mean 422 Hz difference; the subjects in Group C produced a mean 379 Hz difference; and the subjects in Group B a mere 61 Hz difference. This is consistent with our finding that native speakers of French have more distinct categories /u/ and /y/, and hence produce larger differences between them, compared to bilinguals. Flege's study shows that during the process of learning American English speakers acquire better distinctions between the two categories and thus gradually become more native-like. It is important to note that in his study the distinctions between the two categories increase with language experience due to the gradual improvement of the production of French /u/.

The author explains this effect by referring to “equivalence classification” (Flege & Hillenbrand 1984), “a basic cognitive mechanism which permits humans to perceive constant categories in the face of inherent sensory variability found in the many physical exemplars which may instantiate a category”. According to this theory, L2 phonemes that have an acoustically similar counterpart in the L1 of a learner will be hard to perceive and produce. On the other hand, L2 sounds that have no counterpart in the L1 of the learner, and thus are “new”, will be learned easier. Following this assumption, French /y/ is a “new” phone for native American English speakers as they had no major difficulty in producing it accurately even at the beginning of the learning and only slightly differed from native French speakers. In contrast, Flege suggests that French /u/ should be considered a “similar” sound as it is similar to the English /u/. This is why, beginner learners of French had extreme difficulty in producing it accurately. Even though the production accuracy increased across groups with language proficiency, even highly experienced speakers of French did not produce French /u/ in a native-like manner, while they almost reached nativelikeness in the production of /y/.

As mentioned above, we found a different pattern, as /y/ in our study was on the overall produced by bilinguals with less accuracy than /u/.
Levy & Law (2010) looked at the production of /u/ and /y/ by three groups of American English speaker of French of different levels of proficiency and compared their accuracy to a group of native French controls. A repetition task was used to elicit the productions. The highly experienced group of American English participants in Levy and Law's study best corresponds to our group of American English speakers (they spent at least one year in a French speaking country and had extensive formal training). On average, in alveolar context the non-experienced groups and the moderate-experienced group produced /u/ as a front vowel, while the highly experienced group did produce /u/ as a back vowel in bilabial context. All three groups appropriately produced /y/ as a front vowel in the alveolar context. In bilabial context, all three groups accurately produced PF /u/ as a back vowel. The non-experienced and the highly experienced group produced /y/ accurately as a front rounded vowel. However, the moderate-experienced group produced /y/ as a back vowel, according to the F2 values. Thus, following the equivalence classification theory proposed by Flege (1987), Levy & Law found that in alveolar context French /u/ is a “similar” vowel for non-experienced and moderate experienced group. However, contrary to Flege's prediction, in bilabial context it was the vowel /y/ that was more difficult to produce for moderate-experienced talkers. Therefore, Levy & Law showed that a vowel can be assimilated as a “new” or “similar” sound depending on the consonantal context and the level of L2 experience.

The results of our study are more consistent with Levy and Law's (2010) findings. In our study /y/ was produced with less accuracy than /u/ in naming and reading by female participants and in reading by male participants. Thus, /y/ was not an easier vowel for bilinguals to produce, as was suggested by Flege (1987). However, male participants in naming produced /u/ more accurately than /y/. We should note, that we had only four male subjects in each of the groups, thus it is not clear whether this result is representative. Further studies should look at whether there could be a difference in production of French /u/ versus French /y/ between male and female bilinguals.
4.2. Perception

As predicted, bilingual speakers were less accurate in the perception task than native French speakers. Bilinguals had more difficulty in perceiving the test contrast /u/-/y/ than the contrast /u/-/a/ as the former does not exist in English, while the later does. Gottfried (1984) who studied AE listeners' (and French controls') perception of French vowels in an ABX discrimination paradigm, also found that Native French listeners performed significantly more accurately than AE listeners. This reflects the strong effect of the listeners’ native language on their phonological categorization.

It is worth noting that /u/-/y/ is acoustically less distant than /u/-/a/, therefore independently on the language background it should be slightly harder to perceive. However, this difficulty should be minimal. We can observe this effect in the perception results of French participants' who performed slightly worse on the /u/-/y/ contrast than on /u/-/a/. Nevertheless, they were overall very good at discriminating both contrasts. Thus, the significant difference between French and bilingual participants in the perception of those two contrasts is due to the difficulty of bilinguals to perceive a hard L2 contrast that does not exist in their L1.

Our results are consistent with the findings of Levy (2008) who examined in an ABX discrimination task the perception of French vowels by American English (AE) listeners with and without French language experience. On the overall, both groups of American English listeners in their experiment performed worse than French control participants. For both American groups the /u/-/y/ contrast was the most difficult one (the other test contrasts were /i/-/y/, /u/-/œ/, /y/-/œ/), especially compared to the easy control contrast /i/-/u/ and /i/-/œ/. French controls made very few errors on all contrasts. Levy found a strong effect of consonantal context on the performance of listeners' vowel discrimination. The consonantal context had a significant impact on naive listeners' perception. The inexperienced listeners made more errors on discriminating the /u-y/ contrast in alveolar context than in bilabial context. There was no effect of consonantal context on the accuracy of experienced listener's perception. The author suggests that experienced American-English speakers of French learned coarticulatory rules due to the exposure to the L2 during the
learning process. In our study we controlled for the consonantal context, as half of our stimuli were presented in a bilabial and the other half in alveolar context. Moreover, we tested only experienced listeners, therefore following Levy’s finding, they should already have acquired robust context-independent representations.

### 4.3. Correlation between perception and production

We predicted that the correlation between perception and both production tasks will be observed only in the long ISI condition as this would allow the participants to use the phonological loop and rehearse the stimuli subvocally. Nevertheless, the results of correlation analysis appeared to be contrary to our hypothesis as we found significant correlations between perception and production in both ISI conditions and both production tasks.

One explanation for these results could be that the correlations were due to the covariate (the percentage of correct response on the baseline contrast /u/-/a/) that we added to the analysis. We thus carried out a post hoc analysis of the correlation between perception and production without the covariate. A linear regression was carried out for each type of ISI and each production task between the production score on the one hand and the correctness rate on the /u/-/y/ contrast in the perception task on the other hand. All correlations remained significant:

(Short ISI naming: $r^2 = 0.3$, $p = 0.011$;
Short ISI reading: $r^2 = 0.2$, $p = 0.204$;
Long ISI naming: $r^2 = 0.34$, $p = 0.007$;
Long ISI reading: $r^2 = 0.44$, $p = 0.002$).

Another explanation for why we found correlations between perception and production across tasks and ISI conditions while other studies did not, could be related to task particularities, differences in the L2 proficiency of participants or difference in the choice of evaluating production accuracy. Table (3) presents the comparison of previous studies on the relationship between perception and production in L2 adult learners, that found or did not find a correlation between perception and production scores.
<table>
<thead>
<tr>
<th>Paper</th>
<th>Subjects</th>
<th>Perception task(s)</th>
<th>Production task(s)</th>
<th>Target</th>
<th>Results</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Procedures</th>
<th>Measures</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flege (1993)</td>
<td>Inexperienced Mandarin late learners (LOR 1.1)</td>
<td>1. Identification 2. Adjustment</td>
<td>Vowel durations in English words ending in /t/ and /d/</td>
<td>Perception:</td>
</tr>
<tr>
<td></td>
<td>Inexperienced Taiwanese late learners (LOR 1.2)</td>
<td></td>
<td></td>
<td>The larger the difference in vowel duration between preferred examples of beat and bead, and bat and bad, the less accented the non-natives’ production was judged to be.</td>
</tr>
<tr>
<td></td>
<td>Experienced Taiwanese late learners (LOR 5.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Childhood L2 learners (LOR 12.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peperkamp &amp; Bouchoin (2011)</td>
<td>17 French-English bilingual</td>
<td>ABX</td>
<td>English /i/-/ɪ/ and /ɪ/-/u/</td>
<td>No correlation</td>
</tr>
<tr>
<td>Flege et al. (1997)</td>
<td>20 German, 20 Spanish, 20 Mandarin, and 20 Korean speakers</td>
<td>Identification (synthetic beat-bit and bat-bet continua)</td>
<td>English /i/ /ɪ/ /æ/</td>
<td>Correlation</td>
</tr>
<tr>
<td>Perception: Levy (2009)</td>
<td>13 were native AE speakers: no French instruction and little interaction with French speakers</td>
<td>AXB</td>
<td>French vowels /i-/ɪ-/u-/œ-/a/</td>
<td>Modest, yet complex, perception-production link for L2 vowels. Only the contrast /y-œ/ revealed compelling evidence for a perception-production relationship independent of language experience and context.</td>
</tr>
<tr>
<td>Production: Levy &amp; Law (2010)</td>
<td>11 native Japanese speakers (formal education only, no immersion)</td>
<td>Pretest &amp; posttest = minimal pair identification task training: perceptual</td>
<td>English /r/-/l/</td>
<td>No correlation</td>
</tr>
<tr>
<td>Bradlow et al. (1997)</td>
<td>13 native Japanese speakers (formal education only, no immersion)</td>
<td>Pretest &amp; posttest = repetition → nativelikeness judgment</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.

**Proficiency of participants:** Levels of proficiency in L2 differ across studies. Although some studies define different groups depending on their level in L2, there is considerable variation in the criteria chosen for such classifications. For instance, some studies take into account formal training in L2 (Flege et al., 1997, Levy & Law 2010, Bradlow et al. 1997), while others concentrate only on the immersion time in L2 speaking countries (Flege 1993). Even if the criteria for grouping participants correspond, the final classification into highly experienced and non-experienced groups might differ. For example, in Flege et al.'s study (1997) participants were classified into groups according to their length of residence in the L2 speaking country. Thus, in this study non-experienced learners stayed in the US for 0.7 years, while experienced, for 7.3 years. However, in Levy & Law's study (2010) the highly experienced group lived in the L2 speaking country for a median of only 1.4 years but had extensive formal education in their L2. Thus, according to their length of residence they were closer to the non-experienced group of Flege et al's study (1997) but the amount of formal education they received was much higher. Although it has been shown that immersion is crucial for L2 phonological acquisition, it is unclear how to accurately compare the effect of formal education vs. immersion.

**Tasks:** Different studies used different tasks in order to assess the production and perception accuracy of participants. As discussed in the introduction, results might differ even within one modality depending on the differences in tasks. Thus, the comparison of experiments using different tasks might be problematic. For instance, production has been tested either in a reading task (Flege et al. 1997, Peperkamp & Bouchon 2011), or in a repetition task (Levy & Law 2011, Bradlow et al. 1997), or both (Flege 1993). However, as noted by Levy & Law, it is unclear to what extent repetition reflects perception and production skills as both are involved in a repetition task. For example, a native speaker of American English might repeat French /y/ as French /u/ for three distinct reasons: first, the speaker might have misperceived the sound...
as French /u/ and therefore produced it as /u/; second, the speaker might have perceived the sound accurately as French /y/ but produced it incorrectly as /u/; third, the speaker might have both perceived and produced the French /y/ inaccurately, i.e. the speaker might have first perceived the French /y/ as /u/ or as another vowel and afterwards produced it incorrectly as /u/.

Similarly, in perception, some studies used a discrimination task (Peperkamp & Bouchon 2011, Levy 2009), others an identification task (Flege 1993, Flege et al. 1997, Bradlow et al. 1997), Flege (1993) also used the method of adjustment. As shown by several studies, discrimination and identification tasks might not tap into the same level of L2 phonological processing and thus trigger different results (Mack 1989, Diaz et al. 2012, Stasenko et al. 2015).

Production accuracy measurements: The methods used to evaluate production data might also influence the outcomes of the various studies. Some of them used acoustic analysis in order to measure the production accuracy of participants (Flege 1993, Levy 2009), others chose the method of nativelikeness judgement (Peperkamp & Bouchon 2011, Levy & Law 2010, Bradlow et al. 1997) and identification by native speakers (Peperkamp & Bouchon 2011). As nativelikeness judgement is a discrete measure, while acoustic analysis concerns continuous data, it might be problematic to compare the results of those measures.

Finally, we should note that although task-wise the most similar study to ours is the study by Peperkamp & Bouchon (2011), their results were the most different from what we found, i.e. they did not find any correlation at all between L2 perception and production, while we obtained modest but significant correlations across tasks and conditions. According to the results the most similar study to ours was the study by Flege et al. (1997). They also tested the relationship between the perception and perception of L2 vowels and also found a significant correlation between the performance in both modalities. Moreover, Flege, who probably studied the question the most extensively, found moderate correlations between the two modalities in the majority of his studies. He looked at the relationship between perception and production of non-native consonants (Flege 1993, Flege & Schmidt 1995, Schmidt & Flege 1995) and vowels (Flege et al. 1997, Flege et al. 1999, McAllister et al. 2002), demonstrating that the correlation is well established, although not very strong.
How can our results be explained in light of the theory of the phonological loop? Our finding that perception and production are correlated even in the short ISI condition suggests that the phonological loop did not interfere in the online phonological processing of bilingual participants. This should be seen in view of Baddeley’s (1998) idea that the phonological loop is essentially a mechanism involved in the learning of novel phonological forms of novel words. According to the author, the phonological loop allows to store unfamiliar sound patterns, while their representation is being gradually recorded in long term memory. Thus, the phonological loop might be involved in the learning process of French in our bilingual participants, while its role could not be captured in an experiment testing online processing of foreign sounds. This seems to be contrary to the version of the phonological loop theory proposed by Jacquemot & Scott (2006), who emphasized the role of the phonological loop in phonological processing. However, our study focussed on L2 processing, while Jacquemot & Scott mostly looked at L1 processing. As L1 and L2 phonological processing differ in many aspects, the role of the phonological loop in each of them might be different.

Further research could concentrate on the role of the phonological loop in the phonological processing of L1 versus L2. Moreover, it would be useful to make an extensive comparison of the methods used to study perception and production in order to define the most appropriate and accurate way to study the relationship between the two.
5. Appendices

Text for the reading task:

Ce matin quand le réveil a sonné Pierre ne l'a pas entendu. Il a dû être réveillé par Lise, son amour. Sans perdre une seconde, il a sauté sous la douche. Ensuite il a rapidement avalé une tartine à la confiture. Pour aller plus vite, il a voulu prendre sa voiture. Hélas, il n'a pas trouvé ses clés, et pourtant il les a cherchées partout ! Il a donc décidé de prendre le bus. Des embouteillages l'ont fait perdre encore une vingtaine de minutes. Enfin arrivé à son travail, il s'est fait terriblement mal au genou. En fait, la lampe dans son bureau ne marchait plus et il s'est cogné contre le mur. Il fallait donc bien changer l'ampoule. Pendant qu'il s'en occupait, un coup de fils de son chef l'a interrompu : il lui a demander de terminer un dossier avant midi. Pierre s'est donc lancé dans le travail, mais peu de temps après il a été dérangé par une mouche ! Celle-ci a commencé à lui tourner au-dessus de la tête comme un vautour. Pierre ne s'est pas trop énervé au début. Mais ça n'arrêtait pas et finalement il était à bout ! A un moment donné la mouche s'est posée sur son pouce. Pierre a essayé de l'attraper, mais il a renversé une tasse de café sur son pull. C'était donc une bataille perdue. Enfin bref, il s'est dit, demain ça ira mieux, sans doute.

Words for the naming task:

ampoule
bambou
bouchon
bougie
bouteille
bouton
chou
cou
coude
couronne
coussin
couteau
douche
fourchette
genou
goutte
dournal
kangourou
loupe
mouche
moustache
ours
pouce
poule
poupee
roue
souris
tatouage
trousse
yaourt
allumette
autruche
bulle
bus
caactus
capuche
ceinture
chaussure
confiture
cube
ecureuil
fumee
fusil
jupe
legumes
luge
lunettes
nuage
peluche
perruche
plume
pull
statue
sucre
tortue
turban
voiture
Triples for the ABX discrimination task:

<table>
<thead>
<tr>
<th>Bilabial:</th>
<th>/u/</th>
<th>/y/</th>
<th>/a/</th>
<th>Alveolar:</th>
<th>/u/</th>
<th>/y/</th>
<th>/a/</th>
</tr>
</thead>
<tbody>
<tr>
<td>kabouzin</td>
<td>kabuzin</td>
<td>kabazin</td>
<td>katoudin</td>
<td>katudin</td>
<td>katadin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vépoubin</td>
<td>vépubin</td>
<td>vépabin</td>
<td>tidouzin</td>
<td>tiduzin</td>
<td>tidazin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>namouguin</td>
<td>namuguin</td>
<td>namaguin</td>
<td>sanouvin</td>
<td>sanuvin</td>
<td>sanavin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>liboupan</td>
<td>libupan</td>
<td>libapan</td>
<td>rétouun</td>
<td>rétunan</td>
<td>rétanan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gapouman</td>
<td>gapuman</td>
<td>gapaman</td>
<td>midoukan</td>
<td>midukan</td>
<td>midakan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rémOlán</td>
<td>rémulan</td>
<td>rémalan</td>
<td>lanoufan</td>
<td>lanufan</td>
<td>lanafan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>timoubon</td>
<td>timubon</td>
<td>timabon</td>
<td>fadouton</td>
<td>fadoton</td>
<td>fadaton</td>
<td></td>
<td></td>
</tr>
<tr>
<td>véboumon</td>
<td>vébunon</td>
<td>vébamon</td>
<td>bétougon</td>
<td>bétugon</td>
<td>bétagon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mipounon</td>
<td>mipunon</td>
<td>mipanon</td>
<td>ganousson</td>
<td>ganusson</td>
<td>ganasson</td>
<td></td>
<td></td>
</tr>
<tr>
<td>jabounin</td>
<td>jabunin</td>
<td>jabanin</td>
<td>jidoulan</td>
<td>jidulan</td>
<td>jidalan</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6. References


