

Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship^{a)}

Erika S. Levy^{b)}

Program in Speech and Language Pathology, Department of Biobehavioral Sciences, Teachers College, Columbia University, 525 W 120th Street, Box 180, New York, New York 10027

Franzo F. Law II

Program in Speech-Language-Hearing Sciences, Graduate School and University Center, The City University of New York, 365 Fifth Avenue, New York, New York 10016-4309

(Received 15 October 2009; revised 25 May 2010; accepted 30 June 2010)

Second-language (L2) speech perception studies have demonstrated effects of language background and consonantal context on categorization and discrimination of vowels. The present study examined the effects of language experience and consonantal context on the production of Parisian French (PF) vowels by American English (AE) learners of French. Native AE speakers repeated PF vowels /i-y-u-œ-a/ in bilabial /bVp/ and alveolar /dVt/ contexts embedded in the phrase /raCVCa/. Three AE groups participated: speakers without French experience (NoExp), speakers with formal French experience (ModExp), and speakers with formal-plus-immersion experience (HiExp). Production accuracy was assessed by native PF listeners' judgments and by acoustic analysis. PF listeners identified L2 learners' productions more accurately when the learners had extensive language experience, although /y-u-œ/ by even HiExp speakers were frequently misidentified. A consonantal context effect was evident, including /u/ produced by ModExp misidentified more often in alveolar context than in bilabial, and /y/ misidentified more often in bilabial than in alveolar context, suggesting cross-language transfer of coarticulatory rules. Overall, groups distinguished front rounded /y/ from /u/ in production, but often in a non-native manner, e.g., producing /y/ as /^hu/. Examination of perceptual data from the same individuals revealed a modest, yet complex, perception-production link for L2 vowels.

© 2010 Acoustical Society of America. [DOI: 10.1121/1.3466879]

PACS number(s): 43.70.Bk, 43.70.Kv, 43.70.Mn, 43.71.Hw [AL]

Pages: 1290–1305

I. INTRODUCTION

Foreign accents are a testament to the difficulty adults face when learning speech sounds in a non-native language. Among the non-native speech sounds that have been reported as particularly difficult for native speakers of American English (AE) to produce accurately are the Parisian French (PF) front rounded vowels /y/ (in *eu* /y/ “had”) and /œ/ (in *eux* /œ/ “they/them”) (Battye *et al.*, 2000), as well as PF back rounded /u/ (in *où* /u/ “where”) (Flege, 1987; Flege and Hillenbrand, 1984).

Most studies investigating *production* of PF front rounded vowels by native speakers of English have examined the vowels produced primarily in alveolar context (Flege, 1987; Flege and Hillenbrand, 1984; Oh, 2008). Yet AE listeners' *perception* of PF vowels varies as a function of the consonants surrounding the vowels in patterns reflective of coarticulation in their native language (e.g., Gottfried, 1984; Levy, 2009a, 2009b; Levy and Strange, 2008; Strange *et al.*, 2009; see also Bohn and Steinlen, 2003), raising the question of whether production may similarly be affected by context. Nonetheless, conclusions regarding second language

(L2) category formation have been drawn, in part, on the basis of studies of PF /y/ produced in alveolar context (e.g., Flege, 1987, 1995; Flege and Hillenbrand, 1984).

In order to reexamine previously-drawn conclusions of L2-category formation in light of recent information regarding phonetic variations influenced by context, Section II of this article describes an investigation of the effects of consonantal context on the production of PF front rounded vowels by AE L2 learners of French. Language experience effects on production were also explored. An integral aspect of this experimental design was a comparison of production accuracy with perceptual patterns in two consonantal contexts for the same L2 learners. Thus, in Section III, questions regarding the perception-production relationship were addressed through a comparison of the production data with Levy's (2009a, 2009b) perceptual findings.

II. NATIVE AE SPEAKERS' PRODUCTION OF PF VOWELS

A. Introduction

1. Perceiving front rounded vowels

An early categorical perception study using synthetic vowels in isolation suggested that native speakers of English had no trouble discriminating Swedish front rounded vowels from each other (Stevens *et al.*, 1969). Similar ease was ob-

^{a)} Portions of this work were presented at the Acoustical Society of America meetings held in Paris, in June, 2008, and in Portland in May, 2009.

^{b)} Author to whom correspondence should be addressed. Electronic mail: levy@tc.columbia.edu

served for L2 contrasts involving French and German front rounded vowels in studies using natural speech (e.g., Best *et al.*, 1996; Flege and Hillenbrand, 1984; Polka, 1995; Polka and Bohn, 1996). However, more recent studies using more complex stimulus materials with experimental tasks requiring higher memory load indicate that such patterns are not robust for non-native listeners, whether they are naïve to the target language or even highly proficient (Gottfried, 1984; Levy, 2009b; Levy and Strange, 2008).

Levy (2009b) used a categorical discrimination task, in which each of the three sentences in a trial was produced by a different native PF speaker. Experienced listeners (HiExp) did not discriminate the PF /y-u/ contrast significantly more accurately than those who had studied French only in school (ModExp) or who had never studied French (NoExp). Similarly, Levy and Strange (2008) revealed no significant difference between naïve listeners and those with extensive French formal and immersion experience on the contrast.

Unlike perception of the PF /y-u/ contrast, which appeared relatively unaffected by experience, Levy (2009b) found that PF /œ/ was difficult for naïve listeners and listeners with merely classroom French experience to discriminate from other vowels, such as PF /u/. However, with extensive French classroom and immersion experience, the PF /œ-u/ contrast was nearly mastered. (It should be noted that mid front rounded PF /œ/ and /ø/ are rarely contrastive. This article focuses on PF closed syllables; the /œ/ symbol is used to represent this vowel, as opposed to /ø/, which is usually used when discussing open syllables.) Implementing a “cross-language assimilation overlap method,” Levy (2009b) found significant correlations between perceptual assimilation patterns and discrimination accuracy for PF vowels by native AE listeners with all levels of language experience. Thus, the more often listeners assimilated two PF vowels to a single native category, the more discrimination errors they incurred on the contrast, supporting the Perceptual Assimilation Model’s (PAM’s) assertion that assimilation patterns predict discrimination accuracy in naïve listeners (Best, 1995) and L2 learners (Best and Tyler, 2007).

In a perceptual assimilation study, Levy (2009a) found that PF /y/ was assimilated primarily to AE /^hu/, independent of language experience, with consonantal contexts combined. PF /œ/ was assimilated primarily to AE /u/ and /ʊ/ by naïve listeners, with experienced listeners assimilating PF /œ/ primarily to AE /ʊ/ and the more central /ɜ/. PF /u/ assimilated overwhelmingly to AE /u/. Thus, the confusions in Levy (2009b) for PF /y-u/ across groups and for PF /u-œ/ for only the less experienced listeners were predicted by these assimilation patterns.

2. Consonantal context affects L2 vowel perception

Consonantal context has been shown to affect AE listeners’ assimilation (Levy, 2009a) and discrimination (Gottfried, 1984; Levy, 2009b; Levy and Strange, 2008) of PF vowels. For example, Levy (2009a) found that naïve AE listeners and AE learners of French perceptually assimilated the PF high front vowel /y/ to AE /u/ more often in alveolar context than in bilabial context. In bilabial context they assimilated PF /y/ more often to AE /^hu/ than they did in alveolar context. Simi-

larly, the mid front rounded PF vowel /œ/ assimilated more often to AE /u/ in alveolar context than in bilabial. The same listeners discriminated PF /y-u/ contrasts more accurately in bilabial context (PF /y-u/ = 5% errors) than in alveolar context (/y-u/ = 27%) (Levy, 2009b). Consonantal context effects were also found for PF the /œ-o/ contrast. Trends in the same direction were evident for PF /u-œ/ and /y-œ/ contrasts, although they were not statistically significant. Clearly, explanations of perceptual difficulties encountered by naïve listeners and L2 learners cannot rely merely on abstract descriptions of vowel inventories, but should rather consider the articulatory-acoustic properties of vowels in various consonantal contexts.

3. Does consonantal context affect L2 vowel production?

A handful of studies have examined the effects of consonantal context on native speakers’ vowel production and the relationship between phonological inventories and coarticulatory variation (Manuel, 1999). For example, Hillenbrand *et al.* (2001) and Strange *et al.* (2007) demonstrated that native AE speakers’ /u/ and /ʊ/ productions are more “fronted,” i.e., produced farther forward in the oral cavity, resulting in a higher second formant (F2), when uttered in alveolar context than when produced in other contexts (e.g., in isolation or in bilabial context). Strange and her colleagues indicated that fronting is more extensive in English than in French (and German). Levy (2009a) and Levy and Strange (2008) suggested that it is precisely this fronting in AE that may lead AE listeners to perceive back rounded vowels as similar to PF front rounded vowels when they are produced in alveolar context, as this is the context in which PF vowels will assimilate to AE listeners’ (fronted) back-rounded vowel category.

Of interest, then, is whether L2 learners’ productions of front rounded vowels are affected by native coarticulation patterns in ways parallel to their perception of these vowels. This question remains to be answered, as studies of PF vowel production by AE listeners have focused almost exclusively on PF high front rounded /y/ produced in one context only—typically alveolar context (e.g., Flege, 1987; Flege and Hillenbrand, 1984). For example, Flege (1987) examined variations in the acoustic characteristics of AE speakers’ productions of PF rounded vowels /u/ and /y/ as a function of their L2 experience. Forty-two subjects participated, including four groups of native AE speakers with language histories ranging from minimal French experience to having lived in Paris for an average of 11.7 years. The stimuli were natural tokens of /tu/ in English (e.g., “Two little boys”), /ti/ in English (e.g., “TV programs”), /tu/ in French (e.g., “Tous les prêtres”) and /ty/ in French (e.g., “Tu les montres”), read by all participants in phrases and also in sentences created by the participants. PF /u/ was not produced with native-like accuracy (as measured by acoustical analysis) by any group of AE speakers. Those with little French experience produced a very high F2 (higher than for English) for /u/, perhaps targeting PF /y/ instead of /u/. All groups of AE speakers, even those living in France, produced French /u/ with a higher F2 than did native French speakers. Increased French

experience correlated with a lower F2 (approaching native speakers' second formant frequencies), and thus with higher but not native-like accuracy. Unlike the back rounded vowel /u/, the front rounded French /y/ was produced with almost, but not quite, native-like F2 values by the AE groups.

According to [Flege's \(1995\) Speech Learning Model \(SLM\)](#), listeners initially categorize L2 segments by means of a process of "equivalence classification" ([Flege, 1987](#)), assigning the segments to first language (L1) categories, even if they might perceive phonetic differences between L1 and L2 segments. Phones in a second language may be classified as "identical," "similar" or "new" when compared to native language categories. An "identical" phone has nearly the same phonetic properties as a native language phone (e.g., /i/ in AE and Northern German [[Strange et al., 2004](#)]) and is perceived as almost identical to the native category. A "similar" L2 phone, on the other hand, has a clearly identifiable counterpart in a learner's first language, but the L2 phone is discriminably different (at least under some conditions) from its L1 counterpart and is rated as less similar to the native category.

[Flege \(1987\)](#) describes the PF /u/ as a "similar" vowel for native AE speakers, as AE /u/ may be the closest phone to PF /u/, although AE /u/ is characterized by a higher and more variable F2. Flege posits that, at the beginning of French instruction, AE listeners perceptually confuse French /u/ and /y/. Although they are able to perceive auditorily the differences between L1 vs. L2 phones, their representations of the phones are not established. Based on the AE participants' PF /u/ (in alveolar context), which was not produced authentically even by experienced speakers, Flege suggests that with exposure to a language, their representations are modified, but this modification may be limited. Similar L2 phones will continue to be assimilated to L1 categories, resulting in misperceptions and accented productions.

A "new" L2 phone, on the other hand, has no phonological counterpart in the native language. [Flege \(1987\)](#) refers to PF /y/ as an example of a "new" L2 phone to American learners of French because the English phonological inventory contains no phonologically distinctive front rounded vowels. However, Flege claims that AE /u/ has phonetic qualities similar to those of /y/ in some phonetic contexts. Flege posits that AE learners are only able to produce /y/ authentically once they have established a new (non L1) phonetic category for it. Eventually, they will develop a new category for PF /y/.

The SLM ([Flege, 1995](#)) predicts that "new" phones, which have no L1 phonological counterpart, will eventually be produced in a more native-like manner than "similar" phones. The greater the "dissimilarity" between L1 and L2 phones, the more likely listeners will learn to discern the differences and establish new categories for the L2 phones. In later versions of the SLM, however, [Flege \(1995\)](#) suggested that even when new L2 categories are created, the categories may rely on different acoustic characteristics from native listeners' categories. Hence, L2 learners may never perceive even "new" L2 phones in a native-like manner.

In a production study that relied on perceptual judgments of native listeners, [Flege and Hillenbrand \(1984\)](#)

found that French /tu/ (in phrases) was produced more accurately (17% errors) than /ty/ (28% errors) by a group of advanced L2 learners of French, as measured through native French listeners' forced choice identification of these vowels (in alveolar context). These findings run contrary to [Flege's \(1987\)](#) claim that "new" segments are ultimately produced more accurately than "similar" segments by experienced speakers. However, the same study by Flege and Hillenbrand also reported acoustic analyses inconsistent with the native listeners' judgments and consistent with Flege's predictions, with /ty/ produced more "authentically" than French /tu/ by native AE L2 learners of French. Studies that demonstrate even advanced L2 learners' inaccurate *perception* of front rounded vowels (e.g., [Gottfried, 1984](#); [Levy and Strange, 2008](#)) are inconsistent with Flege's notion that a new category is established for PF /y/.

If perception of French vowels differs as a function of language experience ([Levy, 2009a, 2009b](#); [Levy and Strange, 2008](#)), as may /tV/ coarticulation ([Oh, 2008](#)), it follows that PF vowels produced by native speakers of English may vary as a function of language experience, as well as consonantal context. Thus, hypotheses drawn on the basis of (conflicting) speech production data found in alveolar context need to be revisited. To determine whether a vowel may be "new" in one context and "similar" in another, for example, it is necessary to collect L2 learners' production data in more than one context.

4. The present study

The aim of the present study was to explore production of PF front rounded /y/ and /œ/ by native AE speakers and to compare performance in bilabial vs. alveolar contexts. This study also included PF /u/ for comparison purposes, as well as PF /i/ and PF /a/ to serve as "control vowels," which would be expected to fare better than the front and back rounded vowels. Native AE speakers who were late learners of French repeated vowels produced by a native PF speaker. Accuracy was assessed through native French speakers' judgments and acoustical analysis. The following questions were asked regarding the AE listeners with no French experience, those with just formal French experience, and those with formal plus immersion experience:

a. *What are the effects of language experience on the production accuracy of French vowels?* Previous literature suggests that increased experience with a second language, especially immersion experience, results in more accurate production of speech sounds in that language. When global pronunciation is examined, a minority of native Dutch ([Bongaerts, 1999](#)) and Anglophone ([Birdsong, 2007](#)) late learners of French achieve native-like pronunciation in French. At the segmental level, some late learners are capable of producing French vowels with native-like duration ([Birdsong, 2007](#)). On the other hand, pronunciation of certain speech sounds is difficult to learn past early childhood and thus may never be native-like, even with extensive experience (e.g., [Flege, 1987](#); [Flege et al., 1999](#)). In the present study, it was expected that vowel production accuracy (e.g., for PF /œ/, which is perceived more accurately with greater experience [[Levy, 2009b](#)]) would generally be higher for individuals

with more language experience, especially those with extensive immersion experience. However, it was predicted that the PF /y-u/ contrast would remain difficult for all groups of listeners to maintain in production. Nevertheless, the reality that the highly-experienced speakers must communicate in their L2 suggests that they may have found a way to produce the sounds more accurately than they perceive them, a phenomenon that has been demonstrated in Japanese speakers' accurate production of /r-l/ despite their inaccurate perception (Sheldon and Strange, 1982; see also Bohn and Flege, 1997, and Gottfried and Beddor, 1988, re: vowel perception/production relationship).

b. What are the effects of consonantal context on the production accuracy of French vowels? Given the results of the perceptual studies reviewed above, suggesting the variation in L2 assimilation and discrimination patterns as a function of consonantal context (Levy, 2009a; Levy and Strange, 2008), it was expected that accurate production of PF front rounded vowels would be more easily achieved in alveolar context than in bilabial context, but with front rounded vowel productions mistaken for back vowels more often in alveolar context than in bilabial.

The possibility of a decrease in consonantal context effects with extensive experience was considered. That is, it was thought that learning an L2 might include learning to produce vowels in a more native-like manner in all consonantal contexts, less affected by native coarticulation patterns. Evidence for a decrease in consonantal context effects on L2 vowel perception and production with greater language experience has been inconclusive. In the perceptual realm, evidence for a decrease in the consonantal context effect with greater language experience was found in Levy and Strange's (2008) discrimination experiment, in perceptual assimilation of PF /y/ (Levy, 2009a), and for some of Levy's (2009b) discrimination pairs (e.g., for PF /y-u/), but not for others (e.g., PF /u-œ/). A decrease in consonantal context effect in assimilation of PF /œ/ with more French experience did not reach statistical significance (Levy, 2009a). Oh's (2008) comparative acoustic study of native French and native AE speakers' productions of /tu/ suggests that language-specific coarticulation is learned along with L2 vowel targets. Thus, in the present study, a decrease in the consonantal context effects with greater experience was cautiously anticipated, as listeners may arguably have learned to disregard their L1 coarticulation while learning to produce French vowels. Theoretical implications were expected to point to consonantal context as a factor to consider when determining whether an L2 speech sound is "similar" or "new," suggesting a context-specific level of analysis in L2 vowel studies. Questions regarding the perception-production link are discussed in Section III.

B. Method

1. Talkers

Three groups of native AE talkers differing in their language experience were recorded. These were a subset of the 39 participants (13 NoExp, 13 ModExp, and 13 HiExp) in Levy's (2009a, 2009b) perceptual studies. Of those 39 par-

ticipants, data from 12 were discarded because of (1) noise or other distraction in the recordings, (2) talkers' inability to imitate the utterances when the consonants changed from bVOWELp to dVOWELt, perseverating on the previous context, (3) inputting errors in the judgment task, resulting in too few stimuli being presented. Thus, data from 27 AE participants were used for the present study. All participants were raised in English-speaking households in the United States. None had had more than a year of instruction in any language with front rounded vowels, aside from French. All passed a bilateral hearing screening at 20 dB at 500, 1000, 2000, and 4000 Hz.

AE speakers were grouped according to language experience; ten were native AE speakers, ages 20–37 years, 7 females and 3 males, with minimal French experience (NoExp Group), i.e., they had had no French instruction and had not interacted with French speakers with any regularity. The second group, AE speakers with moderate French experience (ModExp Group), consisted of 9 native speakers of AE, 6 females, 3 males, ages 22–26 years, who had had formal French training (i.e., French classes), but minimal French immersion. They had begun learning French no earlier than age 13 years (mean age of beginning learning = 16 years, range = 13–21 years) and had received a mean of 3 years of instruction in French (range = 2–4 years), typically in high school and college, and had spent no more than 5 months in a French-speaking country (mean = 1 month, range = 0–5 months). The third group of AE participants, AE speakers with extensive French experience (HiExp Group), consisted of 8 native speakers of AE living in New York City, 5 females and 3 males, ages 20–61 years, who had had extensive formal training and immersion experience in French and were using French regularly in their professions (e.g., instructors, translators) or in their homes (e.g., married to a native French speaker) at the time of testing. They had a mean of 8 years of instruction in French (range = 5–11 years), which began no earlier than 12 years of age (mean = 14 years, range = 13–16 years). All had spent at least a year in a French-speaking country as adults (mean = 3.5 years, range = 1–16 years).

A control group of two native speakers of Parisian French who had been in the United States for less than a month were also recorded. They were ages 26 and 35 years, both males, raised in a monolingual French environment and had minimal L2-exposure, except for English classes in high school. They reported that they did not speak English.

2. Repetition task

A repetition task was implemented because it was one that all three AE groups would be able to perform. That is, for the NoExp group, reading aloud or conversing in French would not have been possible. However, they were able to repeat the PF speakers' utterances, thus information about naïve AE speakers' production could be collected. The stimuli used for the repetition task were produced by a native PF speaker. The PF speaker was a monolingual female who was born and raised speaking only French in the Paris regional area and had been in the United States for less than a year. She produced the phrases "J'ai dit neuf rabVOWELp à

des amis” and “J’ai dit neuf radVOWELt à des amis” (I said nine [nonsense word] to some friends) with the PF vowels /i, y, u, œ, a/. The beginning of the phrases were removed, such that only the portions “rabVOWELp à des amis” and “radVOWELt à des amis” remained. Please see [Levy \(2009a\)](#) for further details regarding the full protocol and methodology. For stimulus verification, 3 native PF listeners in the United States for less than 4 weeks identified each stimulus. These PF listeners incurred 0 errors and rated the stimuli as native French sounding, suggesting that the tokens were valid representations of the vowels.

The AE talkers (and the two native PF talkers) listened over headphones and were asked to repeat just the “rabVOWELp à” or “radVOWELt à” portion of the phrases produced by the native PF speaker. Two repetitions of each token were recorded. These talkers’ French productions were recorded in a sound-attenuated chamber on a Shure SM48 dynamic microphone, with the signals fed through an Earthworks LAB101 Preamplifier to a sound card (Soundblaster Live wave) installed in a Dell Dimension XPS B800 computer. SOUNDFORGE™ software (version 4.5) was used for recording and editing the files.

3. Native speaker judgments

a. Judges Nine adult monolingual native speakers of PF who had been in the United States for fewer than three months were recruited from Columbia University’s Department of French and Romance Philology, la maison française, the International House, the internet, and via word of mouth. These participants were raised primarily in the Paris area, speaking and hearing virtually only French and had not had extensive immersion experience in any other language. All passed a bilateral hearing screening (with the same criteria as for the talkers).

b. Stimuli In preparation for the native speaker judgments, the digital files containing the target utterances (“rabVOWELp à” or “radVOWELt à”) were edited so that only the target /dVOWELt/ or /bVOWELp/ remained for presentation to listeners (e.g., /bip/ and /dit/) with the five PF target vowels. Segmentation was based on changes in the waveform shape and spectrogram along with perceptual appraisal. In order to create stimuli that would contain acceptable speech sounds in either French or English, the “ra” was removed, but the voice bar remained for the voiced stops (/b/ or /d/) and the stimuli were cut at the beginning of closure for /p/ or /t/. Thus, the stimuli began with a bilabial or alveolar stop and ended with a bilabial or alveolar stop with voice onset times appropriate for either language. The files were transferred to a Dell Pentium 4 computer with a Turtle Beach Riviera Sound Card. The native French judges listened to the stimuli via Sennheiser HD 280 pro headphones. Sessions were held in the Speech Production and Perception Laboratory at Teachers College, Columbia University.

For the judgment task, the stimuli were entered into the PARADIGM ID program ([Perception Research Systems, 2010](#)) designed to execute perceptual tasks. Stimuli were blocked by consonantal context in order for comparisons to be more appropriate between this task and the perception tasks in [Levy \(2009a, 2009b\)](#), in which stimuli were blocked by con-

text. Half of the listeners were presented with stimuli in bilabial context first and the other half were presented with stimuli in alveolar context first. Stimuli were randomized within each block.

c. Procedure The native speaker judgment task began with instructions presented on the computer screen. Judges were asked to listen to each phrase, paying attention to the vowel (e.g., /dVOWELt/) in the nonsense utterance. After the stimulus was presented, judges were asked to choose a keyword response from the word set “vie, vue, vous, cette, veut, veau, va” (corresponding to the PF vowels /i, y, u, ε, œ, o, a/) to indicate which PF vowel was heard in the target word. They heard the stimulus again and assigned a “goodness” rating of the production on a 9-point Likert scale. A rating of “1” indicated “most foreign-sounding,” “9” indicated “most French-sounding,” and judges were encouraged to use the whole spectrum of the scale. The native judges were presented two categorical tokens of PF /y, u, œ/ and one token of PF /i, a/ per context (i.e., bilabial or alveolar) spoken by each of the 27 AE speakers and 2 native PF controls. Thus, each judge assessed 232 stimuli in each consonantal context.

Prior to the experimental trials, task familiarization took place, in which judges heard PF /i, y, u, ε, œ, o, a/ uttered in /Vb/ context spoken by two female native PF speakers and were presented with the same key word responses as for the experimental trials. Instructions on the computer screen asked them to click on the word that contained the vowel corresponding to the vowel in the word they heard. (Judges then also responded to a goodness rating scale, as described above.) Five of the 9 judges made no identification errors in the familiarization task. Each of the other four made 1 error (out of 14). These judges were given a chance to listen again and to identify all the stimuli. On their second attempt, they made no errors on the 14 stimuli.

C. Results

PF judgments and the acoustic structure of the native AE speakers’ repetitions were analyzed. For all analyses below, although these vowels are discussed as production data, it is recognized that these data reflect both the perception and the production skills required of the repetition task. For example, if a native AE speakers repeated PF /y/ as PF /u/ (as judged by PF listeners), it is not evident whether the speaker 1) misperceived the speech sound as PF /u/ (or as another vowel) and produced it as /u/ (as determined by listener judgments and acoustic analysis) or 2) perceived the speech sound as PF /y/ and produced it incorrectly as /u/ or 3) both perceived and produced the PF /y/ inaccurately. Comparisons of individuals’ perception and production performance are discussed in Section III.

Due to the small number of speakers in each group and the use of ordinal judgment ratings as a dependent measure, nonparametric statistical tests were used. All *p* values reported are based on two-tailed hypotheses and are corrected for ties whenever possible. In addition, all significant *p*-values for pairwise comparisons and post-hoc tests are presented with a Bonferroni correction. Analyses of vowel cat-

egory, language experience effects, consonantal contexts, and subsequent interactions are discussed, first with production accuracy scores, followed by goodness judgment rating scores.

1. Judgment task

a. Data analysis Frequencies of selecting a particular response category were tallied for each French vowel across all AE speakers in each language experience group. Frequencies of the modal response, as well as the other chosen responses for each French vowel stimulus, were determined and converted to percentages of total trials. Thus, the production accuracy percentages indicated how often each particular vowel was identified by judges as the same vowel produced by the PF speaker. The overall median goodness rating for each AE response category was also computed.

b. Vowels As anticipated, all groups' productions of PF /i/ were identified accurately (NoExp=97%; ModExp=99%; HiExp=95%; PF=99%) and received relatively high goodness ratings (median=5.5), indicating that the participants understood the vowel judgment task. PF /a/ was also identified accurately (NoExp=97%; ModExp=93%; HiExp=92%). These vowels were included solely to ensure that the judges were on task. Thus, performance on PF /i/ and /a/ are not discussed further.

In order to determine relative difficulty in producing the vowels of particular interest, a Friedman test was performed. This is a non-parametric test that allows for a comparison of three or more related measures, similar to a 2-way analysis of variance (ANOVA) for repeated measures. Thus, the relative difficulty in production of PF /y/, /u/, and /œ/ was calculated, matching accuracy scores of each vowel for each subject. Results revealed a significant difference in production accuracy for the PF vowels /y/, /u/, and /œ/ [$N=27$, $\chi^2=12.019$, $df=2$, $p=0.002$]. As there was no *a priori* assumption regarding the relative difficulty of these vowels, a Wilcoxon Signed-Rank test, the non-parametric equivalent of a t-test for related measures, was used as a post-hoc test to determine any significant differences. Results of the test indicated that PF /u/ was more accurately produced than both PF /y/ and /œ/ [/u/ vs /y/: $N=27$, $T+=19$, $z=-2.542$, $p=0.033$; /u/ vs /œ/: $N=27$, $T+=21$, $z=-3.149$, $p=0.006$], but no significant difference was found between PF /y/ and /œ/ [$N=27$, $T+=17$, $z=-1.36$, $p=0.174$].

c. Language experience Figure 1 depicts production accuracy of the AE speakers as determined by native PF judges' identification in bilabial context (above) and alveolar context (below). Along the X-axis are the target vowels that the AE speakers produced. The Y-axis represents the percent of productions that were correctly identified by PF judges. Within each vowel category presented is percent accuracy for NoExp, ModExp, HiExp and PF speakers.

The percentage of correctly-labeled tokens by the native listeners for each speaker was tabulated and collapsed across consonantal contexts. A Kruskal-Wallis test, the non-parametric equivalent of a one-way ANOVA, was calculated for PF /y-u-œ/ on the percent-correct of each speaker, grouped by language experience, pooled across context. An overall language effect on performance [$\chi^2=10.741$, df

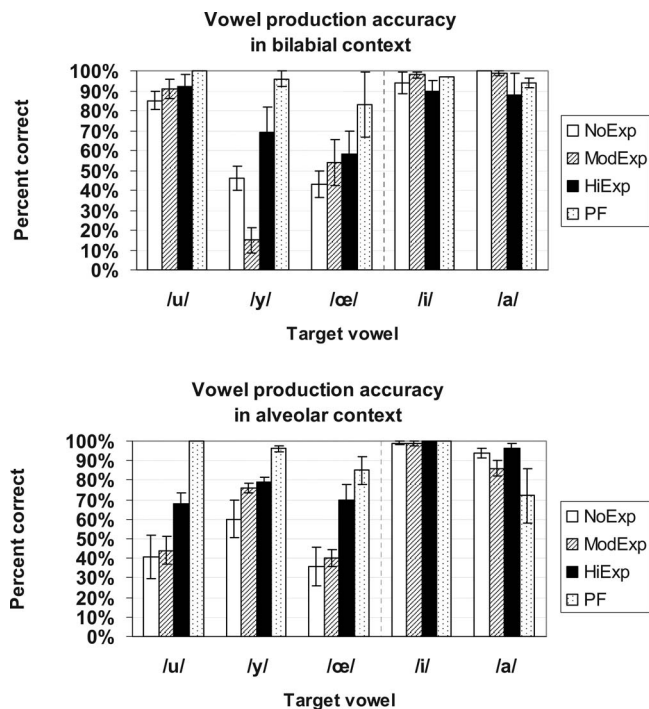


FIG. 1. Percent Parisian French vowels accurately produced by native speakers of American English. Accuracy of production of Parisian French vowels in bilabial /rabVpa/ context (above) and alveolar /radVta/ context (below) by native (female and male) speakers of American English with no (NoExp), formal (ModExp), and formal-plus-immersion (HiExp) French language experience, as determined by native PF listener judgments

=2, $p < 0.01$] was evident. The vowels produced by HiExp were more accurately identified (81% correct) than those produced by ModExp (70%), and NoExp groups (70%). A Mann-Whitney U test for pairwise comparisons, the non-parametric equivalent of a t-test for independent measures, confirmed that HiExp speakers' vowels were more accurately identified than vowels produced by the NoExp [$U=9.0$, $N(HiExp)=8$, $N(NoExp)=10$, $z=-2.757$, $p=0.018$] and ModExp groups [$U=6.0$, $N(HiExp)=8$, $N(ModExp)=9$, $z=-2.889$, $p=0.012$]. (Native PF speakers' vowels were correctly identified on 92% of trials.) The difference between NoExp and ModExp productions was not statistically significant [$U=35.5$, $N(NoExp)=10$, $N(ModExp)=9$, $z=-0.776$, $p=0.438$]. Although the native PF speakers' score was higher than those of ModExp and NoExp groups, that the scores did not approach 100% suggests that identifying vowels in CVC syllables segmented from continuous speech was a difficult task (Miller *et al.*, 1951). In addition, the fact that the PF scores did not reveal ceiling performance reflects primarily the pronunciation of PF /a/ and /œ/ by one of the PF speakers. In alveolar context, this speaker's PF /a/ was identified correctly on only 58% of trials. In bilabial context, his PF /a/ was identified with 92% accuracy. His PF /œ/ was identified accurately on 67% of trials in bilabial context and 78% accuracy in alveolar context. The other speaker's PF /a/ was identified with 97% accuracy in bilabial context and 86% accuracy in alveolar. His PF /œ/ was identified accurately on 100% of trials in bilabial context and 92% accuracy in alveolar context. Identification accuracy for the other vowels (/y, u, œ/) ranged from 92 to 100%.

Regarding productions of the individual PF vowels, it was hypothesized that overall productions of PF /œ/ would be more accurate with experience, whereas PF /y/ and /u/ would not differ. A Kruskal–Wallis test was calculated for each vowel category to compare accuracy of production by each language group. Results revealed no significant difference between groups in overall production of PF /u/ or /œ/ [/u/: $\chi^2=3.895$, $df=2$, $p=0.143$, /œ/: $\chi^2=4.523$, $df=2$, $p=0.104$], but there was a significant difference in /y/ production [/y/: $\chi^2=8.801$, $df=2$, $p=0.012$]. A post-hoc pairwise comparison showed that the ModExp group had less accurate PF /y/ productions than the HiExp group [$U=7.5$, $N(HiExp)=8$, $N(ModExp)=9$, $z=-2.749$, $p=0.018$]. No significant differences were found between NoExp and ModExp groups [$U=31.5$, $N(NoExp)=10$, $N(ModExp)=9$, $z=-1.107$, $p=0.268$] or NoExp and HiExp groups [$U=16.0$, $N(NoExp)=10$, $N(HiExp)=8$, $z=-2.138$, $p=0.099$ (Bonferroni corrected)].

The matrix in Table I lists the target vowel, the responses selected on more than 10% of trials, as well as the median goodness rating for each PF target vowel produced by the three AE groups. The modal response to each target vowel was most often the correct vowel, except for PF /œ/ in alveolar context by NoExp (identified most often as PF /u/), and PF /y/ in bilabial context and PF /u/ in alveolar context by ModExp (identified most often as PF /u/ and /y/, respectively).

d. Context effect on percent correct across vowel category It was predicted that overall performance would be more accurate in bilabial than in alveolar context. In order to test this hypothesis, a Wilcoxon Signed Ranks test was calculated, collapsed across accuracy scores for /y-u-œ/, and paired by context for each participant. Overall, no significant effect of consonantal context [$N=27$, $T+=17$, $z=-1.195$, $p=0.232$] was revealed.

e. Context effect on percent correct within vowel category It was hypothesized that the front rounded vowels (i.e., PF /y/ and /œ/) would be less accurately identified in bilabial context than in alveolar context. A Wilcoxon Signed-Ranks test was calculated for each vowel category, pairing each participant's production accuracy for a given vowel category in bilabial vs. alveolar context as a pairwise comparison to determine whether there was an interaction. No significant difference was found for PF /œ/ [$N=27$, $T+=13$, $z=-0.186$, $p=0.852$], but significant differences were found for PF /y/ [$N=27$, $T+=20$, $z=-2.834$, $p=0.015$] and back vowel /u/ [$N=27$, $T-=24$, $z=-4.056$, $p<0.001$]; /y/ was more accurately produced in alveolar context and /u/ was more accurately produced in bilabial context.

As summarized above, PF /u/ was more accurately produced in bilabial context and PF /y/ was more accurately produced in alveolar context. Further, PF /u/ productions were often labeled as PF /y/ in alveolar context and /y/ labeled as /u/ in bilabial context. It was of interest to determine whether PF /u/ was produced *more* accurately than PF /y/ in bilabial context and vice versa in a post-hoc test, in order to add to an understanding of the context-dependent production patterns of /y/ and /u/. The Wilcoxon Signed-Ranks test com-

TABLE I. Native Parisian French (PF) speakers' identification and goodness ratings of PF /i, y, u, œ, a/ in bilabial and alveolar contexts produced by AE talkers with no French experience (NoExp), classroom experience (ModExp) and formal-plus-immersion (HiExp). Boxes indicate target vowels that were perceived primarily as a non-target vowel.

| Target | Bilabial | | | | Alveolar | | | |
|---------|-----------|------------|------------|--------|-----------|------------|------------|--|
| | Perceived | Mdn rating | % Response | Target | Perceived | Mdn Rating | % Response | |
| No Exp | | | | | | | | |
| /u/ | /u/ | 5 | 85 | /u/ | /u/ | 4 | 41 | |
| | /o/ | 4 | 13 | /y/ | /y/ | 3 | 34 | |
| /y/ | /y/ | 2 | 46 | /œ/ | /œ/ | 2 | 23 | |
| | /u/ | 2 | 37 | /u/ | /u/ | 3 | 16 | |
| | /i/ | 3 | 12 | /œ/ | /œ/ | 2 | 16 | |
| /œ/ | /œ/ | 4 | 43 | /œ/ | /u/ | 4 | 42 | |
| | /o/ | 3 | 33 | /œ/ | 4 | 36 | | |
| | /u/ | 3 | 11 | /y/ | 2 | 16 | | |
| /i/ | /i/ | 4 | 94 | /i/ | /i/ | 5 | 99 | |
| /a/ | /a/ | 5 | 100 | /a/ | /a/ | 5 | 94 | |
| Mod Exp | | | | | | | | |
| /u/ | /u/ | 6 | 91 | /u/ | /y/ | 3 | 45 | |
| | /y/ | | | /u/ | /u/ | 4 | 44 | |
| | /œ/ | | | /œ/ | 3.5 | 10 | | |
| /y/ | /u/ | 4 | 75 | /y/ | /y/ | 3 | 76 | |
| | /y/ | 2 | 15 | /u/ | /u/ | 2 | 17 | |
| /œ/ | /œ/ | 4 | 54 | /œ/ | /œ/ | 3 | 40 | |
| | /o/ | 3 | 28 | /y/ | 4 | 28 | | |
| | /u/ | 4 | 10 | /u/ | 3.5 | 23 | | |
| /i/ | /i/ | 5 | 98 | /i/ | /i/ | 6 | 99 | |
| /a/ | /a/ | 5 | 99 | /a/ | /a/ | 6 | 86 | |
| Hi Exp | | | | | | | | |
| /u/ | /u/ | 6 | 92 | /u/ | /u/ | 6 | 68 | |
| | /y/ | | | /y/ | /y/ | 3 | 22 | |
| /y/ | /y/ | 5 | 69 | /y/ | /y/ | 5.5 | 79 | |
| | /u/ | 4 | 16 | /i/ | 2 | 18 | | |
| | /œ/ | | | /œ/ | /œ/ | 4 | 70 | |
| /œ/ | /o/ | 3 | 19 | /u/ | 3 | 15 | | |
| | /u/ | 4 | 17 | | | | | |
| | /i/ | 6 | 90 | /i/ | /i/ | 6 | 100 | |
| /a/ | /y/ | 3 | 10 | | | | | |
| | /a/ | 6 | 88 | /a/ | /a/ | 5 | 96 | |

paring the accuracy of /u/ to /y/ productions in bilabial context confirmed that PF /u/ was more accurately produced [$N=27$, $T+=25$, $z=-4.375$, $p<0.001$] than /y/. In fact, productions of PF /u/ were more accurately identified than PF /y/ in bilabial context for all except two speakers. (Those two HiExp speakers had identical identification percentages for PF /y/ and /u/.) Conversely, PF /y/ was more accurately produced than PF /u/ in alveolar context [$N=27$, $T+=17$, $z=-2.251$, $p=0.048$].

f. Interaction of language effect and context effect

Whether there was a difference in production accuracy between language groups that was dependent on context was examined. A Kruskal–Wallis test comparing productions by language group was marginally significant for bilabial context [$\chi^2=5.744$, $df=2$, $p=0.057$] and significant for alveolar context [$\chi^2=8.468$, $df=2$, $p=0.014$]. Pairwise com-

parisons of the language groups confirmed that both NoExp and ModExp groups performed less accurately than the HiExp group in alveolar context. A Mann-Whitney test was used for post-hoc analyses: [NoExp vs HiExp: $U=11.5$, $N(HiExp)=8$, $N(NoExp)=10$, $z=-2.538$, $p=0.033$; ModExp vs HiExp: $U=11.5$, $N(HiExp)=8$, $N(ModExp)=9$, $z=-2.363$, $p=0.054$]. No significant difference was found between NoExp and ModExp groups [$U=33.0$, $N(ModExp)=9$, $N(NoExp)=10$, $z=-0.981$, $p=0.327$]. Given the finding of no overall significant effect of context, post-hoc analysis of alveolar to bilabial context showed that the NoExp group (8 out of 10) did perform more accurately in bilabial context than in alveolar [Wilcoxon Signed-Ranks test: $N=10$ $T+=8$, $z=-2.492$, $p=0.039$]. There was no significant effect of context for the ModExp or HiExp groups [ModExp: $N=9$ $T+=4$, $z=0$, $p=1$; HiExp: $N=8$ $T+=5$, $z=-0.281$, $p=0.779$].

As mentioned above, the overall accuracy rates of PF /u/ were higher for bilabial than for alveolar context. The Wilcoxon Signed-Ranks test for pairwise comparisons indicated that NoExp and ModExp had significantly more accurate PF /u/ productions in bilabial context than in alveolar [NoExp: $N=10$, $T+=9$, $z=-2.553$, $p=0.033$; ModExp: $N=9$, $T+=9$, $z=-2.67$, $p=0.024$]. It should be noted that all nine ModExp speakers had more accurate productions of /u/ in bilabial context. In fact, for NoExp and ModExp, productions in bilabial context (85 and 91% accuracy, respectively) were identified with more than twice the accuracy of those in alveolar (41 and 44% accuracy). The trend in the same direction (92% accuracy for bilabial, 68% for alveolar) did not reach significance for HiExp, although six out of eight speakers performed more accurately in bilabial context [$N=8$, $T+=6$, $z=-1.873$, $p=0.183$ (Bonferroni corrected)].

Regarding categorization of PF /u/ productions, the NoExp group produced PF /u/ in bilabial context identifiably on 85% of trials, and as back rounded PF /o/ (13%) more often than as front rounded PF /œ/ (2%) or PF /y/ (0%), as determined by native PF judgments. Compared to bilabial context, PF /u/ in alveolar context was identified less often as PF /u/ (41%) and PF /o/ (1%), and more often as PF /y/ (34%), and PF /œ/ (23%). The ModExp group's productions of PF /u/ in alveolar context were nearly evenly divided between identification as PF /y/ (45%) and as PF /u/ (44%) when they were not identified as PF /œ/ (10%). When HiExp speakers' PF /u/ productions were misidentified in alveolar context, they were most often identified as PF /y/ (22%). In bilabial context, only 8% errors were made in identification of PF /u/ uttered by HiExp, and none of the responses was PF /y/.

For PF /y/, all 9 ModExp revealed more accurate productions in alveolar context than in bilabial [$N=9$, $T+=9$, $z=-2.668$, $p=0.024$], a reversal of patterns from PF /u/, which was more accurately produced in bilabial context than in alveolar. In fact, only 15% of ModExp group's /y/ productions were correctly identified in bilabial context, whereas 76% were correctly identified in alveolar. No significant context effect was found for /y/ productions by NoExp [$N=10$, $T+=7$, $z=-1.432$, $p=0.152$] or by HiExp [$N=8$, $T+=4$, $z=-0.676$, $p=0.499$].

On production of PF /y/, the NoExp group performed relatively poorly in bilabial context (46% accuracy). With only formal experience, accuracy was very poor (ModExp = 15%). In fact, the ModExp group's PF /y/ productions were identified most often as PF /u/ (75%) in this context. HiExp speakers produced PF /y/ with 69% accuracy. When PF /y/ was misidentified, it was most often identified as a PF /u/, except for HiExp speakers in alveolar context, whose tokens were most often misidentified as PF /i/ (18%).

g. Analysis of judgment ratings Only the judgment ratings of accurately-identified productions are discussed here; all statistical analyses were computed using the median judgment ratings of correct productions. If all of the productions were incorrectly identified for a particular speaker, the vowel was given a rating score of 0; this occurred 8 times. In general, the range was small for ratings of how French-sounding (9) vs non-native sounding (1) each vowel was, providing less information than production accuracy data.

Overall, analysis of the ratings corroborated with the results described above for percent correct. Generally, not all vowels were rated equally [Friedman test: $N=27$, $\chi^2=23.29$, $df=2$, $p<0.001$]. PF /u/ was judged more French-sounding than PF /œ/ and /y/ [Wilcoxon Signed Rank test: /u/ vs. /œ/: $N=27$, $T+=21$, $z=-4.025$, $p<0.001$; /u/ vs. /y/: $N=27$, $T+=20$, $z=-3.892$, $p<0.001$], but there was no significant difference in ratings between PF /œ/ and /y/. The ratings suggested that AE's productions of front rounded vowels were less French-sounding than those of the other vowels (median rating /y/=2.5, /œ/=4, /u/=5.5, collapsed across groups).

h. Language experience effect on judgment ratings A main effect of language experience was found when PF /y-u-œ/ were collapsed [Kruskal-Wallis Test: $\chi^2=10.477$, $df=2$, $p=0.005$]; productions by the HiExp group were rated higher than those by the NoExp group [Mann-Whitney U: $U=7$, $N(HiExp)=8$, $N(NoExp)=10$, $z=-3.026$, $p=0.006$], but no difference was revealed between NoExp and ModExp, nor between ModExp and HiExp. No overall significant differences were found in ratings across groups for PF /œ/, but a difference was evident for PF /y/ [$\chi^2=9.609$, $df=2$, $p=0.008$] and PF /u/ [Kruskal-Wallis test: $\chi^2=8.373$, $df=2$, $p=0.015$]. Ratings of PF /y/ were more native-like with extensive language experience, rising from 2 for NoExp and ModExp to 5 in HiExp in bilabial context, and following a similar pattern in alveolar context. PF /u/ productions received slightly higher ratings with extensive experience (NoExp bilabial=5, alveolar=4; HiExp bilabial=6, alveolar=6). PF /œ/, in contrast, ranged only from 3–4.5, without showing much, if any, association of native-like production with experience.

i. Context effect on judgment ratings A difference was found in overall rating by context; vowels in bilabial context were rated significantly more French-sounding than those in alveolar context [Wilcoxon Signed Rank test: $N=27$, $T+=18$, $z=-3.539$, $p<0.001$]. Post hoc tests confirmed that PF /u/ was rated as more native-like in bilabial context than in alveolar [Wilcoxon Signed Rank test: $N=27$, $T+=22$, $z=-3.953$, $p<0.001$]. However, the overall native-like ratings

for /u/ differed only by 1 point (i.e., 5 for bilabial, 4 in alveolar context). PF /œ/ and /y/ did not show any context effect on their ratings.

The interaction of language experience and context on judgment ratings was significant for both alveolar and bilabial contexts; ModExp received higher ratings than NoExp in bilabial context [Mann-Whitney U: $U=14$, $N(\text{ModExp})=9$, $N(\text{NoExp})=10$, $z=-2.704$, $p=0.021$]. HiExp revealed higher ratings than NoExp in both bilabial and alveolar contexts [bilabial: $U=8.5$, $N(\text{HiExp})=8$, $N(\text{NoExp})=10$, $z=-2.964$, $p=0.009$; alveolar: $U=12.5$, $N(\text{HiExp})=8$, $N(\text{NoExp})=10$, $z=-2.477$, $p=0.039$]. No overall difference between ModExp and HiExp groups was found. Findings of significant context effects on ratings, but not on percent correct across vowel categories, suggest that productions were often recognizable, but non-native like.

Regarding differences in judgment ratings between PF /u/ and /y/, /u/ was rated as more native-like in bilabial context [Wilcoxon Signed-Ranks: $N=27$, $T+=23$, $z=-4.236$, $p<0.001$], but differences in alveolar context were not significant [$N=27$, $T+=12$, $z=-1.056$, $p=0.291$].

2. Acoustic analysis

a. Procedure Acoustic analysis of the French vowel stimuli was performed by means of customized software in “CVCZ,” a specialized MATLAB™ program by Valeriy Shafiro. Onsets and offsets of syllable were selected from a wideband spectrogram and waveform generated by the program. Onsets were defined as the release burst for /b/ or /d/, whereas offset was defined as the beginning of closure for the /p/ or /t/, determined by a decrease in amplitude on the waveform and the offset of energy in the upper formants. The program determined the first three formant frequencies (i.e., F1, F2, F3) at the 25, 50, and 75% points of the syllable at 23-ms. windows, using Linear Predictive Coding (LPC) and Fast Fourier Transform (FFT) analyses. When the computer-generated formants were clearly incorrect (e.g., merged, missing or spurious formants), the experimenters hand corrected them based on their judgments from comparing FFT spectra and (LPC) tracking of formants on the spectrogram. Hand-corrected formants were analyzed by two experimenters. If discrepancies were revealed they were resolved by reanalyzing and discussing the stimuli together. A discrepancy was defined as an interjudge difference in manual measurement of more than 0.28 Barks, which could potentially be a perceptible difference (i.e., just noticeable difference) for vowel perception in sentence context (Kewley-Port and Zheng, 1999). [All transformations from frequency to Bark in this article are based on the formula proposed in Zwicker and Terhardt (1980).] Approximately 1% of F1, 13% of F2 and 15% of F3 values were measured by hand. Formant frequencies at the 50% point were plotted in each context for each vowel to check analysis errors.

b. Results Fig. 2 presents plots of the mid-syllable formant (F1-F2) frequencies of PF /i-y-u-œ-a/ produced by the PF, NoExp, ModExp, and HiExp speakers. The plots represent vowels in bilabial (bVp) context (a) and alveolar (dVt) context (b). Within each context and language experience group, the plots are divided into productions by females vs.

males. Ellipses encompass productions from all subjects in each group, with the symbol in each ellipse representing the mean F1-F2 value.

Productions of /i/ and /a/ were relatively consistent for speakers in all groups. Compared to the PF speakers (both males), the AE speakers displayed generally more spectral overlap. Vowels overlapped more in alveolar context than in bilabial for the NoExp and ModExp speakers. The midpoints of female ModExp /y, u, œ/ in alveolar context were within almost 1 Bark of each other. PF /u/ ranged greatly for females at all levels of experience, with even HiExp productions spanning a range of more than 6 Barks. PF /y/ in bilabial context by females showed particularly large ranges. Most of this variability was between-speaker and is discussed below.

For the PF speakers, there was no F2 overlap for the vowels PF /y/, /u/, and /œ/ in either context. As expected, high vowels /y/ and /u/ had a lower F1 than mid vowel /œ/. Also, the front vowel /y/ had a relatively high F2, the back vowel /u/ had a relatively low F2, and the mid vowel /œ/ fell in between, although still within the range of a front vowel (replicating production of vowels reported in Strange *et al.*, 2009). All three non-native groups maintained the appropriate height distinction in bilabial context; the F1 values were lower for /y/ and /u/ than for /œ/. However, there was a large degree of variability in F2, especially in productions of PF /y/. In alveolar context, NoExp and ModExp groups had very little separation in the mean F1-F2 values for their productions of /y/, /u/, and /œ/. On average, these vowels were produced as high fronted vowels. HiExp did produce PF /œ/ in a more native-like manner, with a relatively higher F1. They also produced PF /u/ with a mean F2 value that was lower than /y/ and /œ/, which is more appropriate for a back vowel, although still not as low as native productions of PF /u/.

Flege and Hillenbrand (1984) reported nearly complete F2 overlap for French /u/ and /y/ productions in alveolar context by their inexperienced group, who had somewhat more exposure to French than the ModExp speakers in the present study in that the inexperienced group in Flege and Hillenbrand’s study had lived in Paris the academic year prior to being tested. With respect to landmarks, it has been theorized that a boundary exists roughly between 1500–1700 Hz. (or roughly 11–12 Barks), distinguishing front from back vowels (Stevens, 2002). If this boundary is used to describe the data in Flege and Hillenbrand (1984), both native English groups on average produced /y/ and /u/ as front vowels in alveolar context. The group equivalent in experience to the ModExp group in the present study had a mean F2 of 12.8 Barks for /u/, 13.1 Barks for /y/. The group equivalent to the HiExp group had an average F2 of 11.6 Barks for /u/, 13.1 Barks for /y/. In contrast, the native French group produced /u/ with a mean F2 of 10.7 Barks, and /y/ with a mean F2 of 13.4 Barks. The authors conclude that /y/ is a “new” category for the L2 learners and that the learners are more accurate at producing a new category than “similar” PF /u/.

The present study partially replicates these results. In alveolar context, both NoExp and ModExp groups produced /u/ as a front vowel on average (12.0 and 12.3 Barks, respec-

Bilabial Context

Alveolar Context

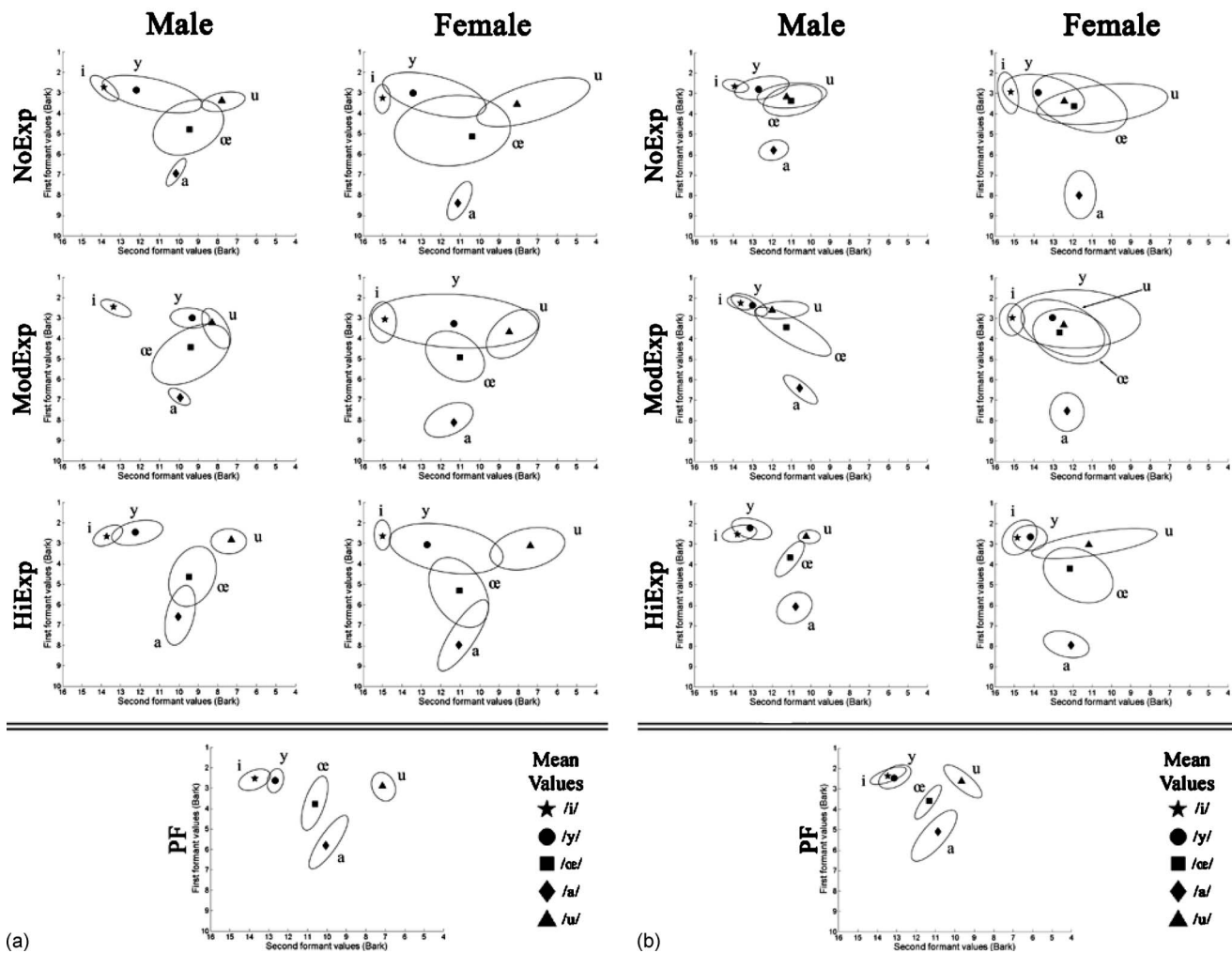


FIG. 2. Acoustic analysis of second-language and native Parisian French vowels. PF /i, y, u, œ, a/ produced by native speakers of American English with no (NoExp), formal (ModExp), and formal-plus-immersion (HiExp) French language experience, and by native Parisian French (PF) speakers, in bilabial /rabVpa/ context (a) and alveolar /radVta/ context (b), summarized over speakers. Symbols represent mean first formant/second formant values by all speakers in each speaker group and are surrounded by ellipses representing all subjects' productions within each group.

tively). Indeed, 8 out of 10 NoExp and all 9 ModExp had an average F2 greater than 11 Barks for /u/. Figure 2 illustrates the considerable F2 overlap for NoExp and HiExp groups' productions of /y/ and /u/. On average, HiExp did produce /u/ as a back vowel (10.8 Barks). Only 2 out of 8 HiExp speakers had an average F2 greater than 11 Barks for /u/. All three groups appropriately produced /y/ as a front vowel (NoExp: 13.4 Barks, ModExp: 13.0 Barks, HiExp: 13.8 Barks).

In bilabial context, a different pattern is found, however. All three groups accurately produced PF /u/ as a back vowel (NoExp: 8.0 Barks, ModExp: 8.4 Barks, HiExp: 7.4 Barks). HiExp and NoExp groups did differentiate PF /y/ from /u/ and produced /y/ as a front vowel (12.5 and 13.0 Barks, respectively). Only 2 out of 8 HiExp and 1 out of 10 NoExp speakers produced /y/ with a mean F2 less than 11 Barks. Note that this is a midpoint measure of F2. As will be described in the Discussion section, many NoExp productions of /y/ were more often a palatalized /y̥/, yielding a high F2 value at midpoint. This high F2 value at midpoint was not

sufficient for native-like production, however, due to the descending F2 trajectory. The ModExp group produced /y/ as a back vowel, with a mean F2 of 10.6 Barks. Specifically, 6 out of 9 ModExp had a mean F2 value less than 11 Barks for /y/. The spectral overlap for /u/ and /y/ in bilabial context is evident for ModExp in Fig. 2. Thus, it is PF /y/ that is not as accurately produced as PF /u/ in bilabial context, calling the status of /y/ as a "similar," more accurately-produced vowel into question. In other words, as found in [Flege and Hillenbrand \(1984\)](#), the ModExp group exhibited a large degree of F2 overlap. However, the direction of assimilation of F2 is the opposite in bilabial context; a *lower* F2 is found for PF /y/ in bilabial context and a *higher* F2 is found for PF /u/ in alveolar context.

D. Discussion

The present study's findings regarding language experience and consonantal context effects in AE speakers' production of PF vowels may be discussed with reference to [Flege's](#)

(1987) acoustical analysis and the acoustical analysis and judgment data in [Flege and Hillenbrand's \(1984\)](#) study. However, caution must be exerted because of the following differences in subject and task variables: First of all, as little is known about naïve AE participants' productions of PF vowels, no comparison can be made with naïve speakers in previous studies. In [Flege's \(1987\)](#) and [Flege and Hillenbrand's \(1984\)](#) previous studies, the talkers with the least experience were more experienced than the ModExp group in the present study. Moreover, in those studies, the PF /tV IV/ stimuli were produced by talkers reading stimuli or generating sentences with phrases such as "Tous les soldats" or "Tu les observes" (p. 711), thus this was strictly a production/reading task, not calling on perception as heavily as the repetition task (required for naïve speakers) involving /radVt a/ or /rabVp a/ (with PF /i-y-u-œ-a/ vowels) in the present study. And finally, in [Flege and Hillenbrand](#), the native French-speaking judges, who were from France and Belgium and had been in the United States for a mean of 12 years, performed an /y-u/ forced-choice task. In the present experiment, an identification task with six PF vowels was administered to native French judges from the Paris area who had been in the United States for fewer than 11 months.

1. PF /u/ and /y/: Language experience and consonantal context

Nevertheless, the context effects revealed in the present study suggest that [Flege's \(1987\)](#) and [Flege and Hillenbrand's \(1984\)](#) conclusions regarding PF /u/ and /y/ may not be applicable to these phones when they are uttered in bilabial context. Moreover, the designation of PF /y/ as "new" and PF /u/ as "similar," even in alveolar context, needs to be reconsidered.

Acoustic analysis in [Flege's \(1987\)](#) and [Flege and Hillenbrand's \(1984\)](#) study found that PF /y/ (in alveolar context) was ultimately produced in a native-like manner by AE speakers, whereas PF /u/ was not. Although the AE speakers in the present study did not produce PF /u/ or /y/ with native-like accuracy in alveolar context (HiExp /u/=68%, /y/=79% accuracy), the present findings are consistent with [Flege's \(1987\)](#) and [Flege and Hillenbrand's \(1984\)](#) results, in that HiExp revealed a trend in the direction of higher accuracy of PF /y/ than /u/ in alveolar context. (In addition, ModExp performed significantly more accurately on PF /y/ than on PF /u/, and when pooled across experience groups, PF /y/ was more accurately produced than /u/.)

In bilabial context, however, the present findings contradict the acoustical analysis in both of [Flege's](#) studies: PF /y/ was produced relatively poorly in bilabial context by all groups (NoExp=46% ModExp=15%, HiExp=69% accuracy). Moreover, in the present study, PF /u/ was produced more accurately than PF /y/ in bilabial context by all three language experience groups (NoExp=85%, ModExp=91%, HiExp=92% accuracy).

[Flege \(1987\)](#) and [Flege and Hillenbrand \(1984\)](#) posited that PF /y/ was a "new" vowel, ultimately produced more accurately by experienced listeners than the "similar" PF /u/ vowel. In the present study, in bilabial context, PF /y/ was relatively poorly produced, not displaying the characteristics

ascribed to a "new" vowel. Moreover, in this context, PF /u/ was not a "similar" vowel, as it displayed near-native performance. Thus, designations as "similar" or "new" vowel based on production in one consonantal context may not apply to the vowel in another context.

Even in alveolar context, although PF /y/ was produced more accurately than PF /u/, it is not evident that PF /y/ was categorized as a "new" vowel. As AE back-rounded vowels are fronted in alveolar contexts, it could be argued that AE participants perceived and produced PF /y/ similarly to their AE (high fronted) /u/ in that context, which, to PF judges, was a relatively authentic PF /y/. That is, a context-dependent level of analysis might suggest that because participants were familiar with a PF /y/-like phone, this "similar" PF /y/ in alveolar context was produced more accurately than the "new" PF /u/ in that context. Generally, PF front-rounded vowels, as well as PF back vowels, fronted by AE speakers in alveolar context, tended to be produced as PF front rounded vowels in the present study.

Why did the ModExp display the least accurate /y/ production in the present study? [Flege and Hillenbrand \(1984\)](#) suggest that /y-u/ confusion may stem from merging PF /y-u/ to a single category for Americans with some formal French experience. This is not the case for naïve speakers in this study. Thus, this may be learned as an (albeit incorrect) strategy through formal instruction. This is further exacerbated by the fact that the PF /y/ is consistently represented by the grapheme "u," which also corresponds to AE /u/. It is also possible that the participants had learned about the potential confusions, had not had enough experience to master them, and remained in a state of learned helplessness ([Hiroto and Seligman, 1975](#)) when confronted with sounds that might be PF /u/ or PF /y/. As individuals who studied a language at school but were not immersed in it are a little-studied, but large, population, further research into perception by such individual may tease out the factors that contributed to their production patterns.

2. PF /œ/: Language experience and consonantal context

Judgments in the present study indicated that PF /œ/ was produced with relatively low accuracy by the HiExp listeners. This mid front rounded vowel was difficult for native PF listeners to categorize, consistent with [Levy's \(2009a\)](#) perceptual assimilation study indicating that PF /œ/ is an "uncategorized vowel," according to the Perceptual Assimilation Model ([Best, 1995](#); [Best and Tyler, 2007](#)), assimilated by AE listeners to several native categories. Thus, the uncategorizable perception of /œ/ may have led to the highly variable non-native productions.

3. Mixed results regarding decrease in context effect

As in the perceptual domain (e.g., [Levy, 2009a](#); [Levy and Strange, 2008](#)), the present study yielded mixed findings regarding a decrease in consonantal context effects on production with increased experience. For PF /u/, the context effect was statistically significant for NoExp and ModExp, not for HiExp, and a trend not reaching significance was

found for the HiExp group. For PF /y/, a context effect was found for ModExp, just a trend in the same direction for HiExp, and, contrary to previous studies, no context effect was found for NoExp. (For PF /œ/, no context effect was found for any group.) On the other hand, the goodness ratings revealed context effects for NoExp and ModExp in some conditions, but no context effect for HiExp in any condition except /u-œ/ in bilabial context, suggesting that when speakers with extensive experience produced French vowels accurately, they generally pronounced them equally well, regardless of context. Thus, mixed support was provided for Oh's (2008) conclusion that learning to produce vowels includes learning their coarticulatory properties.

4. Limitations

The limitations of this study include that the repetition of segmented CVC stimuli may approximate "language-neutral" stimuli, but may not be entirely reflective of "real world" speech. Furthermore, although many stimuli were investigated, a small number of subjects from varied language backgrounds were tested. And finally, although using a repetition task prevented an examination of production without consideration of the perceptual component, this paradigm made it possible to test naïve participants, which was felt to be an important contribution, allowing a better understanding of production skills in the beginning stages of language learning. (In particular, little information exists about PF vowel production by naïve speakers and about PF /œ/ production by L2 learners.) However, examining perceptual performance by the same talkers provides clues to the role of perception. This is to be addressed in the next section.

III. PERCEPTION-PRODUCTION RELATIONSHIP

The few studies that have examined the relationship between speech perception and production of L2 phones have generally found that native language perceptual patterns influence L2 production (see Llisterri, 1995, for an overview). However, the nature of the perception-production relationship is poorly understood and the empirical evidence is sparse and inconclusive, leading Flege *et al.* (1999) to posit a modest correlation between speech perception and production. Understanding the relationship between perception and production is confounded by methodological issues such as different techniques for assessing perception and production (Mack, 1989), different task demands within the perceptual realm (Strange and Shafer, 2008; Werker and Curtin, 2005), group data analyses performed rather than examinations of individual listeners' performance (see Fox, 1982), as well as learning issues such as individuals not updating their production to modified perceptual representations (Flege *et al.*, 1999).

Specifically with respect to the perception-production link for French front rounded vowels, Rochet (1995) reported that native speakers of English identified synthesized front rounded vowels within a high vowel (/i/ to /u/) continuum as more like English /u/ than like /i/. When these same native English speakers repeated a Standard French speaker's /y/ vowel, half of the stimuli were identified as the

French /y/ target by native French listeners. When the vowel was not produced as /y/ by native English speakers, it was produced similarly to a French /u/ or as a vowel somewhere between French /u/ and /y/ according to the native French judges. In contrast, when native Brazilian Portuguese speakers listened to the same French vowel continuum, the Portuguese speakers perceived them more like /i/ than like /u/—and repeated the front rounded stimuli as more /i/-like as judged by the native French listeners, suggesting language-specific relationships between perception and production of French /y/.

Further evidence for the perception-production link for L2 vowels includes Flege *et al.* (1997) finding of a relationship in nonnative listeners' reliance on temporal vs. spectral cues in discrimination and production of AE /i-/ and /ɛ-æ/ pairs. Additionally, Flege *et al.* (1999) found that native Italian speakers' discrimination of English correlated with their intelligibility scores. On the other hand, Gottfried and Beddor (1988) found that Native French speakers who did not use temporal distinctions as a perceptual cue for the French /o-ɔ/ did distinguish the sounds temporally in production. The authors attribute this finding to duration generally not being a reliable perceptual cue in distinguishing French speech sounds. Similarly, Bohn and Flege (1997) found that native speakers of German experienced in English were able to produce the AE /ɛ-æ/ contrast in a native-like manner, but did not discriminate the contrast as did native speakers, relying less on spectral differences than did native AE speakers. The authors suggest that in the initial stages of language learning, perception may lead production. With continued L2 exposure, learners' production abilities may become more native-like, whereas their perceptual skills may lag behind for several years.

The possibility of parallels between perception and production domains can be explored when perception and production data are collected from similar groups of individuals. For example, Levy (2009a) indicated that participants assimilated front rounded vowels primarily to back rounded vowels and Levy (2009b) found that participants confused front rounded vowels more often with back rounded vowels in a discrimination task. Similarly, in the present study, front rounded vowels, when produced inaccurately, were more often produced as back vowels than as front vowels, as judged by native PF listeners. Acoustic analysis in the present study further demonstrated that inaccurate productions were a matter of an F2 (front-back) deviation from native production. As another example, Levy (2009b) found that the PF /y-u/ contrast was discriminated relatively accurately (NoExp = 84%, ModExp = 81%, HiExp = 88% accuracy, contexts combined). Perception appears to precede production for this contrast in that in the present study, production of PF /y/ in bilabial context only reached 69% accuracy, and was most often confused with PF /u/ (16% of responses). PF /u/ in alveolar context only reached 68% accuracy and was frequently confused with /y/ (22% of responses).

In both perception and production, Levy (2009a, 2009b) and the present study found that HiExp groups tended to perform "more differently" from the ModExp than the ModExp did from the NoExp, suggesting a more powerful

influence of extensive formal and immersion experience than only classroom experience. This general finding is consistent with [Flege and Hillenbrand's \(1984\)](#) results; the F2s for /y-u/ productions were more similar to each other for L2 speakers with less experience than for those with more experience. [Levy \(2009a, 2009b\)](#) and the present study suggest that context affects both perception (assimilation and discrimination) and production (repetition), in that front rounded vowels were confused with/assimilated to back vowels in both domains more often in alveolar context than in bilabial context.

Thus, generally, a relationship, albeit a complex one, was found between perception and production. Whether the relationship had predictive value remained to be investigated. That is, would individuals who were able to perceive PF /œ/ relatively accurately, for example, also produce PF /œ/ relatively accurately? Results of this experiment and previous studies have shown that both perception and production patterns vary based on language experience and consonantal context. Of interest is whether a perception-production relationship exists independent of language experience and consonantal context, in which more accurate PF /œ/ perceivers would also be more accurate PF /œ/ producers, regardless of experience and context, pointing to a phonological skill that may not be learned through language experience.

The availability of data on discrimination, assimilation and production on the same participants with three levels of French language experience permitted an examination of the perception-production relationship on an individual level. It would have been beyond the scope of this paper to discuss predictions for all of the possible vowel pairs in the production experiment. Thus, the perception-production relationship for three of the vowel pairs involving front rounded vowels that caused some of the most (and most persistent) perceptual difficulties for native AE participants were examined: PF /y-u/, /u-œ/, and /y-œ/.

The following questions were asked:

- (1) Does French vowel perception (as measured by perceptual assimilation and discrimination tasks) by AE listeners predict their production accuracy?
- (2) Is there a perception-production relationship independent of language experience and consonantal context?

The expectation was that in this study, in which PF speech sounds produced in two consonantal contexts were compared to the same native AE individuals' perception of such speech sounds, a (complex) relationship between perception and production would be revealed. It was predicted, based on anecdotal data and related studies (e.g., [Deutsch et al., 2009](#); [Gottfried, 2007](#); [Norton et al., 2005](#)) that, although language experience and consonantal context would account for most of the perception-production correlation, some evidence of skill (or lack thereof) beyond these factors would be revealed.

A. Data analysis

The talkers in the present production study were a subset of the participants whose perceptual assimilation patterns

were tested in [Levy \(2009a\)](#) and whose discrimination was tested in [Levy \(2009b\)](#). Thus, the perceptual data of only the subset (10 NoExp, 9 ModExp, and 8 HiExp) were used in the analysis of the perception-production relationship. The "cross-language assimilation overlap method" ([Levy, 2009b](#)) was implemented to examine the perception-production relationship. An overlap score was defined as the percentage of trials that an individual (in [Levy's 2009a](#)) perceptually assimilated two PF speech sounds to the same AE vowel, based on the premise that sounds assimilated to the same native category would be difficult to differentiate ([Best, 1995](#)). Each participant's overlap score was computed. Partial correlations were implemented to compare the cross-language assimilation overlap data (e.g., for PF /y-u/, /u-œ/, and /y-œ/) with the same individuals' production accuracy, as judged by PF listeners for each vowel (i.e., PF /y/, /u/, /œ/).

B. Results

The scatter plots in [Fig. 3](#) display the correlation between cross-language assimilation overlap and front rounded vowel production accuracy (as determined by native listener judgments) for PF /y-u/, /u-œ/ and /y-œ/. The hypothesis was that the percentage of perceptual overlap of two vowels would be negatively correlated with production accuracy for one or both of the vowels. Additional measured variables were also associated with more accurate production (e.g., a significant correlation between context and production accuracy for /y/ and /u/ was found). Thus, it was necessary to perform a partial correlation analysis to determine correlations between perception and production accuracy that were independent of any measured variable that was also significantly correlated with accurate production.

In determining whether there was a correlation between perceptual overlap and production accuracy of PF /u/ and /y/, it was necessary first to control for context. More perceptual overlap of /y/ and /u/ was found in alveolar context than in bilabial ($r=-0.457$, $p=0.001$). The partial correlation analysis controlling for context revealed that no association was found between PF /u/ production and perceptual overlap of /y-u/ ($r=-0.153$, $p=0.274$); nor between PF /y/ production and perceptual overlap ($r=-0.166$, $p=0.235$). That is, more frequent assimilation of PF /y/ and /u/ to the same native categories was not associated with less (or more) accurate production of PF /y/ or /u/.

Perceptual overlap of PF /u/ and /œ/, on the other hand, was negatively correlated with PF /œ/ production by native speakers ($r=-0.275$, $p<0.05$), indicating that non-native speakers who had less perceptual overlap between PF /u/ and /œ/ had more identifiable productions of PF /œ/. However, language experience was also negatively correlated with perceptual overlap ($r=-0.556$, $p<0.001$). A subsequent partial correlation analysis controlling for language experience revealed that perceptual overlap of PF /u/ and /œ/ was no longer negatively correlated with production of /œ/ ($r=-0.129$, $p=0.358$), suggesting that language experience had driven the correlation, as shown in the second scatter plot of [Fig. 3](#). Context was not significantly correlated with perceptual overlap of /u/ and /œ/; thus, no partial correlation was

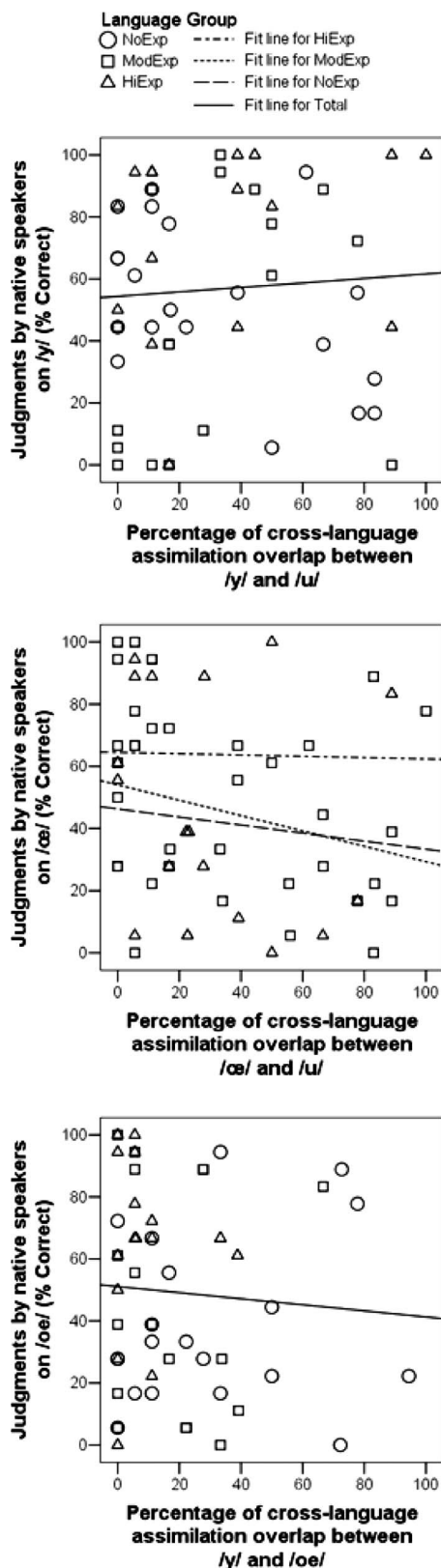


FIG. 3. Perception-production relationship. Correlation between AE listeners' perception (cross-language assimilation overlap) and production (repetition) for PF /y-u/ (top), /œ-u/ (middle), and /y-u/ (bottom). American English with no (NoExp) are represented by circles, formal (ModExp) by squares, and formal-plus-immersion French language experience (HiExp) by triangles

calculated, controlling for context. (Correlations of perceptual overlap of PF /u/ and /œ/ with PF /u/ production were

marginally significant [$r=-0.252$, $p=0.067$], but became non-significant. in a partial correlation for experience [$r=-0.164$, $p=0.24$].)

In analyzing for the correlation between perceptual overlap and production accuracy for PF /y/ and /œ/, it was again necessary to control for context. There was a larger degree of perceptual overlap in alveolar context ($r=-0.317$, $p=0.02$). Results of the partial correlation controlling for context suggest that production of PF /y/ was negatively correlated with perceptual overlap ($r=-0.389$, $p<0.005$); i.e., less perceptual overlap between /y/ and /œ/ was associated with more accurate production of /y/. A partial correlation controlling for both language experience and consonantal context did not reverse any of the previously reported significant findings, suggesting a perception-production correlation for PF /y-œ/ independent of experience or context. (Language experience was negatively associated with overlap: [$r=-0.384$, $p=0.004$]. A correlation of perceptual overlap of PF /y/ and /œ/ with PF /œ/ production was not statistically significant overall [$r=-0.071$, $p=0.61$]. Controlling for context, it was also non-significant [$r=-0.055$, $p=0.693$]. When context and language group were controlled for, the correlation remained non-significant [$r=0.082$, $p=0.563$].)

An additional finding was that, consistent with Levy's (2009a, 2009b) results, when the discrimination and assimilation data of only the subset of participants in the perception tasks were analyzed, a strong negative correlation between cross-language overlap and discrimination was found for all three vowel pairs /y-u/ ($r=-0.590$, $p<0.001$), /u-œ/ ($r=-0.656$, $p<0.001$); /y-œ/ ($r=-0.607$, $p<0.001$), supporting the PAM's (Best, 1995) and PAM-L2's (Best and Tyler, 2007) claim that assimilation patterns predict discrimination skills in these naïve listeners and L2 learners. This further suggests that the present production findings may also be relevant to the talkers' discrimination patterns. Moreover, despite the different strategies and resources being used in perceptual assimilation and AXB discrimination tasks, the relevance of which is inherent in the Automatic Selective Perception (ASP) model (Strange and Shafer, 2008), these two measures of perception appeared to be correlated with the same faculties.

C. Discussion

All L2 language groups performed adequately in discriminating the PF /y-u/ pair; their accuracy was not native-like, but it was well above chance. That PF /y/ was perceived as most similar to AE palatalized /^hu/ may have proven helpful for discrimination of PF /y-u/, but not for production of /y/. Indeed, PF /y/ in bilabial context was frequently produced as /^hu/. Using the terminology of the PAM (Best, 1995), all groups may have had category-goodness, if not 2-category, assimilation of PF /y-u/ (as AE /^hu-u/), yielding relatively good discrimination for all groups. On the other hand, productions of PF /y/ that are /^hu/-like are unlikely to be recognized by a native speaker as instances of PF /y/.

The significant correlation for the PF /u-œ/ pair, but non-significant partial correlation for the same pair suggests that this relationship was driven by language experience. As par-

ticipants gained experience, they learned both to perceive and to produce the vowels in a more native-like manner. Only the HiExp group had a modal response to more central /ɜ/ as in “heard.” More centralized perception was related to more native-like production of PF /œ/; when participants’ performance was analyzed individually, if PF /œ/ was not perceived as a back vowel, it was produced more accurately.

The partial correlations between production accuracy of /œ/ and both perception measures for /y-œ/ presented the most compelling evidence for a perception/production relationship independent of language experience and context. For this contrast, regardless of language background or context in which the vowels were presented, individuals who assimilated PF /y/ and PF /œ/ more often to separate categories (than did their peers) also produced these vowels more intelligibly than their peers, suggesting that gaining language experience is not the only route to mastery of front rounded vowel perception and production.

It should be noted that not all correlations were statistically significant, and when they were, they were not large, supporting Flege *et al.*’s (1999) claim of a modest perception-production correlation. Such findings are also consistent with Bradlow *et al.*’s (1997) suggestion that the relationship between Japanese listeners’ perceptual learning of English /r/ and /l/ and their resultant production skills reflected considerable individual variation in the degree of participants’ perceptual vs production learning. Clearly, additional quantitative and qualitative methods are needed to explain better the individual differences in mapping patterns. Such methods may improve the accuracy of predictions of the difficulties individuals may encounter in their L2 speech sound production as a function of their language experience and of their perceptual mastery of speech sounds in various phonetic contexts.

The limitations of using parametric correlation analyses on these data are noted. Assuming that language experience as measured here is an ordinal scale, the smaller-than-desired N may violate assumptions of a Pearson’s partial correlation. However, there is no non-parametric equivalent to a partial correlation. Correlations not accounting for multicollinearity were verified using Spearman’s Rank Order correlation, the non-parametric equivalent.

IV. CONCLUSION

In summary, the present study revealed more accurate production of PF /y/ than PF /u/ in alveolar context when pooled across language experience groups, consistent with acoustic analyses in Flege (1987) and Flege and Hillenbrand (1984) (cf. Flege and Hillenbrand judgment data), although in the present study, PF /y/ uttered by the most experienced speakers in alveolar context (79% accuracy) was still not produced at the 96% accuracy level of native PF speakers. It is suggested that when examined in alveolar context, front rounded PF /y/ is not a “new” category, but is “similar” to AE (fronted rounded) /u/ in alveolar context to AE speakers. Thus, it may have been the L2 vowel’s familiarity, rather

than its novelty, that resulted in its more accurate production than PF /u/, which is produced farther back in the oral cavity than AE /u/ in this particular context.

In extending previous L2 vowel learning studies to other consonantal contexts, it is evident that it is not possible to define /y/ as a “new” or “similar” phoneme for AE learners of French, independent of context. In bilabial context, “similar” PF /u/ (in that context) was produced more accurately than “new” PF /y/ (in that context) and in alveolar context, “similar” /y/ was produced more accurately than “new” PF /u/. Thus, when the vowels are examined context by context, contrary to the SLM’s (Flege, 1995) hypothesis, “similar” vowels may be pronounced more accurately than “new” vowels. To a modest extent (Flege *et al.*, 1999), perceptual performance predicted and probably preceded production. Only for PF /y-œ/ did the perception-production correlation suggest a responsible variable, perhaps talent, independent of language experience and consonantal context.

Findings from the present study indicate that L2 vowel production accuracy may vary as a function of speakers’ language background, the particular vowel targeted, and the consonantal context in which the vowel is produced. The extensive language experience of the HiExp group here suggests that reaching native-like production may not be feasible for certain L2 vowels.

Empirical knowledge gathered on the basis of perceptual and acoustic similarity of L1 and L2 vowels using current theoretical models of speech perception and production makes partial predictions regarding intelligibility of L2 vowels. Future use of quantitative measures, such as the cross-language assimilation overlap metric, are expected to help test theoretical assumptions and shed light on the complex perception-production link in L2 learning. The robust consonantal context effects on production of PF /u/ and /y/ by individuals with a few years of classroom language-learning experience suggest that models of L2 speech perception and production might consider that the ease of non-native phoneme learning is most appropriately assessed at a context-specific level.

ACKNOWLEDGMENTS

This research was supported, in part, by a grant to the first author by the Provost’s Office at Teachers College, Columbia University. The authors also wish to thank Marina Faygenbaum, Victoria Hatzelis, Ana de la Iglesia, Jamy Rodriguez, Valeriy Shafiro, Winifred Strange, the Teachers College Speech Production and Perception Laboratory, and the Speech Acoustics and Perception Laboratory for their thoughtful assistance.

Battye, A., Hintze, M.-A., and Rowlett, P. (2000). *The French Language Today: A Linguistic Introduction* (Routledge, London).

Best, C. T. (1995). “A direct realist view of cross-language speech perception,” in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Best, C. T., Faber, A., and Levitt, A. (1996). “Assimilation of non-native vowel contrasts to the American English vowel system,” *J. Acoust. Soc. Am.* **99**, 2602.

Best, C. T., and Tyler, M. D. (2007). “Nonnative and second-language speech perception: Commonalities and complementarities,” in *Language*

- Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 13–34.
- Birdsong, D. (2007). "Nativelike pronunciation among late learners of French as a second language," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 99–116.
- Bohn, O.-S., and Flege, J. E. (1997). "Perception and production of a new vowel category by adult second language learners," in *Second-Language Speech*, edited by A. James and J. Leather (Mouton de Gruyter, Berlin), pp. 53–73.
- Bohn, O.-S., and Steinlen, A. K. (2003). "Consonantal context affects cross-language perception of vowels," in Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona: Causal Productions, pp. 2289–2292.
- Bongaerts, T. (1999). "Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners," in *Second Language Acquisition and the Critical Period Hypothesis*, edited by D. Birdsong (Erlbaum, Mahwah, NJ), pp. 133–159.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997). "Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production," *J. Acoust. Soc. Am.* **101**, 2299–2310.
- Deutsch, D., Dooley, K., Henthorn, T., and Head, B. (2009). "Absolute pitch among students in an American music conservatory: Association with tone language fluency," *J. Acoust. Soc. Am.* **125**, 2398–2403.
- Flege, J. E. (1987). "The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification," *J. Phonetics* **15**, 47–65.
- Flege, J. E. (1995). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 233–277.
- Flege, J. E., Bohn, O.-S., and Jang, S. (1997). "Effects of experience on non-native speakers' production and perception of English vowels," *J. Phonetics* **25**, 437–470.
- Flege, J. E., and Hillenbrand, J. (1984). "Limits on phonetic accuracy in foreign language speech production," *J. Acoust. Soc. Am.* **76**, 708–721.
- Flege, J. E., MacKay, I. R. A., and Meador, D. (1999). "Native Italian speakers' perception and production of English vowels," *J. Acoust. Soc. Am.* **106**, 2973–2987.
- Fox, R. A. (1982). "Individual variation in the perception of vowels: Implications for a perception-production link," *Phonetica* **39**, 1–22.
- Gottfried, T. L. (1984). "Effects of consonant context on the perception of French vowels," *J. Phonetics* **12**, 91–114.
- Gottfried, T. L., and Beddor, P. S. (1988). "Perception of temporal and spectral information in French vowels," *Lang Speech* **31**, 57–75.
- Gottfried, T. L. J. (2007). "Music and language learning," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O. S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 221–258.
- Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (2001). "Effects of consonant environment on vowel formant patterns," *J. Acoust. Soc. Am.* **109**, 748–763.
- Hiroto, D. S., and Seligman, M. E. P. (1975). "Generality of learned helplessness in man," *J. Pers. Soc. Psychol.* **31**, 311–327.
- Kewley-Port, D., and Zheng, Y. (1999). "Vowel formant discrimination: Towards more ordinary listening conditions," *J. Acoust. Soc. Am.* **106**, 2945–58.
- Levy, E. S. (2009a). "Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French," *J. Acoust. Soc. Am.* **125**, 1138–1152.
- Levy, E. S. (2009b). "On the assimilation-discrimination relationship in American English adults' French vowel learning," *J. Acoust. Soc. Am.* **126**, 2670–2682.
- Levy, E. S., and Strange, W. (2008). "Perception of French vowels by American English adults with and without French language experience," *J. Phonetics* **36**, 141–157.
- Llisterri, J. (1995). "Relationships between speech production and speech perception in a second language," in Proceedings of the 15th International Congress of Phonetic Sciences, edited by K. Elenius and P. Branderud, Stockholm: KTH/Stockholm University, Stockholm, Sweden, pp. 92–99.
- Mack, M. (1989). "Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals," *Percept. Psychophys.* **46**, 187–200.
- Manuel, S. Y. (1999). "Cross-language studies: Relating language-particular patterns to other language-particular facts," in *Coarticulation: Theory, Data and Techniques*, edited by W. J. Hardcastle and N. Hewlett (Cambridge University Press, Cambridge, England), pp. 179–198.
- Miller, G. A., Heise, G. A., and Lichten, W. (1951). "The intelligibility of speech as a function of the context of the test materials," *J. Exp. Psychol.* **41**, 329–335.
- Norton, A., Winner, E., Cronin, K., Overy, K., Lee, D. J., and Gottfried, S. (2005). "Are there pre-existing neural, cognitive, or motoric markers for musical ability?" *Brain Cogn* **59**, 124–134.
- Oh, E. (2008). "Coarticulation in non-native speakers of English and French: An acoustic study," *J. Phonetics* **36**, 361–384.
- Perception Research Systems, www.perceptionresearchsystems.com (Last viewed May, 2010).
- Polka, L. (1995). "Linguistic influences in adult perception of non-native vowel contrasts," *J. Acoust. Soc. Am.* **97**, 1286–1296.
- Polka, L., and Bohn, O. S. (1996). "A cross-language comparison of vowel perception in English-learning and German-learning infants," *J. Acoust. Soc. Am.* **100**, 577–592.
- Rochet, B. L. (1995). "Perception and production of second-language speech sounds by adults," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 379–410.
- Sheldon, A., and Strange, W. (1982). "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception," *Appl. Psycholinguist.* **3**, 243–261.
- Stevens, K. N. (2002). "Toward a model for lexical access based on acoustic landmarks and distinctive features," *J. Acoust. Soc. Am.* **111**, 1872–1891.
- Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., and Öhman, S. (1969). "Cross-language study of vowel perception," *Lang Speech* **12**, 1–23.
- Strange, W., Bohn, O.-S., Trent, S. A., and Nishi, K. (2004). "Acoustic and perceptual similarity of North German and American English vowels," *J. Acoust. Soc. Am.* **115**, 1791–1807.
- Strange, W., Levy, E. S., and Law, F. II (2009). "Cross-language categorization of French and German vowels by naïve American listeners," *J. Acoust. Soc. Am.* **126**, 1461–1476.
- Strange, W., and Shafer, V. L. (2008). "Speech perception in second language learners: The re-education of selective perception," in *Phonology and Second Language Acquisition*, edited by M. Zampini and J. Hansen (Cambridge University Press, Cambridge, England).
- Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., and Nishi, K. (2007). "Acoustic variability within and across German, French and American English vowels: Phonetic context effects," *J. Acoust. Soc. Am.* **122**, 1111–1129.
- Werker, J. F., and Curtin, S. (2005). "PRIMIR: A developmental model of speech processing," *Lang. Learn. Dev.* **1**, 197–234.
- Zwicker, E., and Terhardt, E. (1980). "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," *J. Acoust. Soc. Am.* **68**, 1523–1525.