

How the demographic makeup of our community influences speech perception

Shiri Lev-Ari^{a)}

Psychology of Language, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands

Sharon Peperkamp

Laboratoire de Sciences Cognitives et Psycholinguistique (Ecole Normale Supérieure, École des hautes études en sciences sociales, Centre national de la recherche scientifique); Département d'Études Cognitives, Ecole Normale Supérieure, PSL Research University, 29 rue d'Ulm, 75005 Paris, France

(Received 29 September 2015; revised 29 April 2016; accepted 4 May 2016; published online 6 June 2016)

Speech perception is known to be influenced by listeners' expectations of the speaker. This paper tests whether the demographic makeup of individuals' communities can influence their perception of foreign sounds by influencing their expectations of the language. Using online experiments with participants from all across the U.S. and matched census data on the proportion of Spanish and other foreign language speakers in participants' communities, this paper shows that the demographic makeup of individuals' communities influences their expectations of foreign languages to have an alveolar trill versus a tap (Experiment 1), as well as their consequent perception of these sounds (Experiment 2). Thus, the paper shows that while individuals' expectations of foreign language to have a trill occasionally lead them to misperceive a tap in a foreign language as a trill, a higher proportion of *non-trill* language speakers in one's community decreases this likelihood. These results show that individuals' environment can influence their perception by shaping their linguistic expectations. © 2016 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4950811>]

[CGC]

Pages: 3076–3087

I. INTRODUCTION

Speech perception is influenced by our experience. Much research has examined how the phonological properties of the input we receive shape our perception of sounds in our native language and in foreign languages we do not speak (e.g., Kuhl *et al.*, 1992; Werker and Tees, 1984). In this paper we show that our experience, and, specifically, the demographic makeup of our community, can also influence our perception of sounds in a foreign language by influencing our expectations about the way that foreign languages sound.

Our perception of phones is shaped by the distribution of cues in the input we received. Infants raised in different language communities end up having different phonological categories. Furthermore, our experience distorts our perception of the perceptual space, honing our attention to phonetic cues relevant in our language and reducing our sensitivity to cues that are irrelevant. Thus, infants raised in an English-speaking environment can distinguish the Hindi dental-retroflex /ʈ-ʈ/ contrast at 6-months of age but they lose this capacity before their first birthday (Werker and Tees, 1984). This is believed to be due to their learning from experience

that the difference between these two sounds is irrelevant for phoneme categorization in their native language. Complementarily, infants' skill at categorizing phonemes in their native language improves during the same time window (Kuhl *et al.*, 2006). Interestingly, perception keeps on being shaped by further input that is received, even during adulthood, as can be evidenced by changes to perception following changes in exposure. For example, intensive exposure to a second language can change the category boundaries and the cues listeners are sensitive to (Flege, 1995).

In a similar manner to the way in which linguistic experience shapes our ability to understand foreign sounds, it also influences our ability to recognize and discriminate foreign languages. In general, people are quite poor at identifying languages that they do not speak, but their linguistic background determines which languages they would be more successful at recognizing (Bradlow *et al.*, 2010). Furthermore, while there are specific properties that lead all people, regardless of their linguistic background, to find certain languages more similar to each other than others, linguistic background can influence which phonetic properties listeners attend to when discriminating between languages (Barkat and Vasilecu, 2001). Moreover, it has even been argued that simply being bilingual can improve one's ability to discriminate between unfamiliar foreign languages by improving one's ability to tune to relevant aspects of the stimuli (Marks *et al.*, 2003).

Perception, however, is also sensitive to the social context. In particular, the expectations we have of the speaker

^{a)}Also at: Laboratoire de Sciences Cognitives et Psycholinguistique (Ecole Normale Supérieure, École des hautes études en sciences sociales, Centre national de la recherche scientifique); Département d'Études Cognitives, Ecole Normale Supérieure, PSL Research University, 29 rue d'Ulm, 75005 Paris, France. Electronic mail: shiri.lev-ari@mpi.nl

can influence how we interpret their speech. In fact, use of information we have about the speaker is an integral aspect of speech processing. There is no one-to-one mapping between sounds and their targets, as different speakers might produce the same phoneme (e.g., /a/) very differently, and, in fact, one English speaker's /a/ might be another speaker's /ɔ/ (Peterson and Barney, 1952). At the same time, information about the speaker can help us disambiguate the input, as production varies systematically along certain demographic variables. For example, males and females differ in the size of their vocal tract, and consequently, in their formant frequencies. The boundary between different vowels they produce therefore lies at different formant frequencies. Listeners are sensitive to this relationship between speakers' sex and their productions, and therefore interpret the same token differently depending on whether they believe they are listening to a female or a male speaker (Johnson *et al.*, 1999). Similarly, information about a speaker's age can be useful in interpreting their speech, and listeners indeed make use of it (Drager, 2011; Koops *et al.*, 2008). When a language change is in progress, people of different ages often speak differently. Accordingly, listeners sometimes interpret the same sound differently depending on the age they believe the speaker to be. For instance, listeners residing in Houston, TX, an area going through a process of unmerging of the PIN and PEN vowels might show phonological competition between *rinse* and *rent* when they believe the speaker is an older adult (for whom the two vowels should be merged), but not when the speaker to be is middle-aged (Koops *et al.*, 2008, also see Drager, 2011). Similarly, listeners use the relation between ethnicity and final /t/ deletion to interpret word meaning, and thus are more likely to initially interpret /mæs/ as *mast* if they believe the speaker is African American but as *mass* if they believe the speaker to be Caucasian (Staum Casasanto, 2008).

The accuracy of listeners' expectations can consequently influence success in comprehending the speech. Thus, listeners are better at understanding Chinese-accented English embedded in noise if they expect the speaker to be a native speaker of Chinese than if they expect the speaker to be a native speaker of English (McGowan, 2015), and conversely, they are better at understanding native speech when they do not expect a foreign accent (Babel and Russell, 2015).

Our knowledge of patterns of usage can also lead us to misperceive speech. For example, expecting a speaker to be non-native because of ethnicity cues can lead us to hear a foreign accent even when there is none (Rubin, 1992). Furthermore, even misconceptions we have about the way people with different characteristics speak can influence the way we perceive speech. Residents of Detroit, for example, believe that while Canadians produce a raised diphthong for the canonical American English diphthong /au/, they themselves produce the canonical one (in fact, they raise their diphthong similarly to Canadians). These expectations influence their perception, leading them to hear the same sound as a canonical American English /au/ when they believe the speaker resides in Michigan but as a raised diphthong when they believe the speaker is Canadian (Niedzielski, 1999).

Research so far, then, shows that expectations can influence speech perception. Furthermore, previous evidence shows that expectations do not even have to be grounded in accurate linguistic knowledge, but that both accurate and inaccurate expectations can influence speech perception, and that they can therefore improve or impair perception. Here we show that an environmental factor—the demographic makeup of one's community—can lead to both correct and false linguistic expectations, and that these expectations, in turn, influence speech perception. Furthermore, we focus here on a different type of expectation than has been investigated beforehand, expectations regarding what a foreign language is likely to sound like. This focus extends previous research to a type of expectation that might be particularly relevant during second language learning.

II. EXPERIMENT 1

The goal of Experiments 1a and 1b was to test whether the demographic makeup of one's community can shape linguistic expectations. Here we focus on one aspect of demographic makeup—the linguistic background of people in one's geographic environment. The properties of the languages spoken in an individual's ambient environment can influence that individual's expectations and stereotypes, and consequently, perception. We tested this by examining how the proportion of speakers of foreign languages with and without an alveolar trill (from here on, trill) influences listeners' expectations regarding whether foreign languages in general are likely to have such a trill. The results of this study can thus allow us, at a second stage, to test whether the demographic makeup of one's community can influence trill perception in foreign languages.

A high proportion of the population of the United States is bilingual and speaks at home a language other than English. According to the United States Census Bureau (2013), in 2011 20.5% of the population reported speaking a language other than English at home. By far, the most common foreign language spoken at home in the U.S. is Spanish, accounting for 62% of the cases. One of the most noticeable and stereotypical phonetic differences between Spanish and English is the existence of a trill (a “rolled r”) in Spanish. While Spanish also contains a tap (a “non-rolled” r, similar to the way *t* in *butter* is produced in American English), this consonant is less acoustically salient, and is therefore not as stereotypically associated with Spanish as a trill. Trill production varies across varieties of Spanish as well as across individual speakers of the same variety. For example, it has been argued that its most common variants are assimilated in some regions in Paraguay and northern Argentina (Hualde, 2005), and in Dominican Spanish it has been variably described as either pre-aspirated trill (Lipski, 1994) or a pre-breathy voice followed by one or two taps (Willis, 2007). Furthermore, variability in trill production has also been linked to sociolinguistic factors (Adams, 2002; Diaz Campos, 2008). Most types of trill, however, share a perceptually salient property that make them different from all English consonants, i.e., the occurrence of more than one period of vibration, and American English speakers can distinguish a trill from a tap, regardless of whether they are able to produce it (Johnson, 2008).

The fact that English contains a tap but not a trill might enhance even further the relative salience of trills over taps for English listeners. A preliminary explicit online survey we conducted confirmed this observation. In that survey, 20 native English speakers residing in the U.S. listened to 15 syllables consisting of a consonant followed by the vowel /a/ produced by a native Spanish speaker from Argentina. For each syllable, participants were asked to indicate whether they thought that it existed in Spanish. They had the option of indicating that they do not know. Participants were also asked what their native language was (they all responded that it was English) and where they resided. All 20 participants (100%) responded that the trill-initial syllable exists in Spanish. No other syllable was responded to affirmatively by all participants, even though 9 out of the 15 initial consonants were Spanish. Only 60% of the participants thought that the tap-initial syllable exists in Spanish.¹ Trill, then, is the most stereotypical and prominent sound in Spanish from the perspective of native speakers of American English.

Due to the prevalence of Spanish over all other foreign languages in the U.S., American English speakers might come to expect foreign languages in general to have a trill, more so the more they hear Spanish around them, and especially if there are only few other foreign language speakers in their community. That is, just as people who are exposed more often to apples and oranges than to pawpaws and kumquats are more likely to think about the former when fruits are mentioned, so might individuals who are more often exposed to Spanish be more likely to activate Spanish sounds in the context of a foreign language, and this greater activation can influence expectations. In contrast, being exposed to a greater proportion of non-trill foreign languages should reduce such likelihood. Indeed, people tend to perceive and evaluate events in relation to available exemplars that they have for these events, a phenomenon also known as the availability heuristic (Tversky and Kahneman, 1973). Importantly, available exemplars are not necessarily the most common or representative ones (e.g., Tversky and Kahneman, 1973). Experiments 1a and 1b test this hypothesis that listeners' expectations of foreign languages to have a trill depends on the proportion of Spanish and non-trill foreign languages in their environment. Specifically, we used census data to predict participants' expectations regarding whether words containing a trill are more likely to be a real word in an unspecified language than words containing a tap. We tested this by presenting participants with pairs of recorded "words," and asked them to guess which of the two is a real word in either an unspecified European language (one condition) or Spanish (another condition). The Spanish condition served as a type of control as the stereotype of Spanish as containing a trill is quite strong and prevalent across the U.S. Therefore, unlike the case of the unspecified European language, the demographic makeup of individuals' communities should not influence their expectations of Spanish words to include a trill, as these expectations should already be at ceiling. Target pairs contrasted a trill with a tap. We tested whether the proportion of Spanish speakers and other foreign

language speakers in participants' community predicted the likelihood of selecting a trill over a tap.

A. Method

1. Participants

Forty-eight participants participated in Experiment 1a and 40 participated in Experiment 1b. All participants were recruited via Mechanical-Turk (www.mturk.com).² In both Experiments 1a and 1b, participants were invited to participate in a language guessing game. Fourteen of these participants were excluded because they were not native English speakers (Experiment 1a: $N=2$; Experiment 1b: $N=3$), because they reported previously living in a European country whose dominant language includes a trill (Experiment 1b: $N=2$), because they answered the catch trials incorrectly (Experiment 1b: $N=2$), because they did not answer the majority of the items (Experiment 1a: $N=1$), because they provided the same response on all trials (Experiment 1b: $N=1$), or because they had previously participated in a related experiment (Experiment 1a: $N=1$; Experiment 1b: $N=2$). Analyses were therefore conducted over 74 participants (Experiment 1a: Spanish condition: $N=21$, unspecified European language condition: $N=22$; Experiment 1b: Spanish condition: $N=16$, unspecified European language condition: $N=15$).

2. Stimuli

Thirty-two Spanish words with intervocalic tap (e.g., *acabaron* "they finished," *giro* "I turn"; note that in Spanish orthography, tap is always represented as a single r) were selected. In intervocalic position, taps contrast with trills in Spanish, but none of the selected words forms a minimal pair with a word with a trill. Thirty-two additional words in Spanish that contain neither a trill nor a tap were selected as fillers. A native speaker of Argentinian Spanish recorded all target words in two versions: the correct pronunciation with a tap, and an incorrect pronunciation with a trill (e.g., as if they were *acabarron*, *girro*; note that in Spanish orthography, an intervocalic trill is always represented as rr).³ One feature known to distinguish trills from taps is their duration (Henriksen, 2015). The taps in our stimuli were indeed shorter than the trills (33 ms vs 63 ms; $t(31)=10.23$, $p<0.001$). Only one version of each filler word was recorded.

In Experiment 1a, each target word was paired with another target word, such that one word in each pair contained a trill and the other one contained a tap, counterbalanced across the two versions of the experiment (e.g., *acabarron - giro* and *acabaron - girro*). Eighteen of the filler words were selected and randomly grouped in pairs. Overall, there were thus 25 pairs.

In Experiment 1b, each of the 32 target words (which all contain a tap) was paired with its trill version (e.g., *acabaron - acabarron*), such that in half of the pairs the trill appeared first, and in the other half the tap appeared first. The 32 filler words were modified by replacing a phoneme, or by deleting or inserting phonemes. Each filler word was paired with its modified version (e.g., *ocupo - ocupa*, *luchando - chando*).

There were 64 pairs in total. Both experiments included a Spanish and an unspecified European language condition.

3. Procedure

Participants in both Experiment 1a and 1b first answered a few background questions including reporting what their native language is, which languages they speak, where they currently live, and all previous locations in which they had lived.⁴

Experiments 1a and 1b had identical instructions. The instructions in the Spanish condition informed participants that only one of the words in each pair was a Spanish word, and asked them to guess which of the two it was. The instructions in the unspecified European language condition were similar but said that only one word of each pair was a “word in a European language (e.g., Swedish, Czech, French).” Participants were asked to guess which of the two words is a word in such a European language. We provided Swedish, Czech, and French as exemplars of European languages, since participants are unlikely to have strong stereotypes regarding Swedish and Czech, especially regarding whether they have trills or taps, and their stereotype of French is unlikely to include a trill. Thus, even though Swedish and Czech in fact have a trill, we expected participants to respond differently than in the Spanish condition, since their expectations about the languages differ. Condition (Spanish, unspecified European language) was manipulated between participants. Word pairs were presented in one fixed random order. Participants were not timed and responded to all items at their own pace. Each word was presented in a separate audio file that participants played by clicking. Participants therefore controlled the inter-stimulus interval themselves and were free to listen to an audio file more than once if they wished to. Experiment 1b also included three catch trials in which the audio file asked participants to select a specific response. The goal of these catch trials was to ensure that participants are listening to the audio files before selecting their response.

4. Community makeup data

Information about the makeup of participants’ current location and each previous location was taken from the latest [United States Census Bureau \(2012\)](#) with the 5-year estimates. Information in the census details what percentage of the residents in a community speaks a language other than English at home, and provides a further break-up into Spanish, other Indo-European languages, Asian and Pacific Islander languages, and other languages. As there is no detailed information about the percentage of speakers of specific languages other than Spanish, we collapsed all other foreign languages together. In general, according to a [United States Census Bureau \(2013\)](#), the most commonly spoken languages in the U.S. after English and Spanish are Chinese (4.8%), Tagalog (2.6%), Vietnamese (2.3%), French (2.1%), and Korean (1.9%). Importantly, none of these languages contains a trill. We acknowledge that our measure might over-estimate the proportion of foreign non-trill language speakers in the community as it might include some speakers

of trill languages that are not Spanish, such as Russian, Swedish, or Malayalam. This over-estimation, though, is likely to be negligible considering the infrequency of such languages in the U.S. We do not know whether the degree to which the measure exaggerates the proportion of non-trill speakers in the community is higher in communities with more non-native speakers, but if that is the case, then the over-estimation of the measure works against our hypothesis, potentially making it more difficult for us to find an effect of this measure.

For each participant, we gathered information about the proportion of Spanish speakers and Other Foreign language speakers for participant’s current place of residence as well as each previous place of residence. We then entered for each participant the highest proportion of Spanish and Other Foreign language speakers across all places they lived in. Two participants had reported living abroad in a country where a non-trill language is spoken. The proportion of Other Foreign Language speakers in that community would therefore be close to 1. To not skew the data too much by adding extreme outliers, however, we entered for these participants the highest proportion of Other Foreign Language Speakers that we found in our sample rather than a value of 1 (but see footnote 5). In general, the proportion of Spanish speakers ranged from 0 (Bethpage, TN) to 0.52 (Ontario, CA; $M = 0.17$), and the proportion of Other Foreign language speakers ranged from 0 (Bethpage, TN) to 0.336 (San Francisco, CA; $M = 0.13$). The two factors modestly correlated ($r = 0.24$; $p < 0.05$).

B. Results

Experiments 1a and 1b had identical goals and only differed in how direct (and therefore, explicit) the comparison of trill and tap words was (*acabaron - girro vs acabaron - acabarron*). The results of the two experiments patterned in identical ways. Indeed, an analysis over the two data sets that included all factors of interests and their interaction with Experiment revealed that Experiment did not interact with any other factor relating to the demographic makeup of the community (all z 's < 1), so the data were collapsed and all analyses are reported over the combined dataset. In general, participants selected the trill 61% of the time. To test whether the proportion of Spanish and Other Foreign language speakers in the communities in which participants have lived influence their expectation that foreign languages have a trill, we ran a mixed-model analysis with Participants and Items as random variables, and Condition (Spanish, Unspecified European Language), Proportion of Spanish Speakers, Proportion of Other Foreign Language Speakers, and the interactions of Condition with Proportion of Spanish Speakers, and Proportion of Other Foreign Language Speakers as fixed variables. The Condition variable was treatment-coded, with Unspecified European Language as the baseline level. The model included intercepts for the random variables, as well as slopes for Condition, Proportion of Spanish Speakers, and Proportion of Other Foreign Language Speakers for the Items variable.

Results revealed a significant interaction between Condition and Proportion of Spanish speakers in a community ($\beta = -4.33$, $SE = 1.93$, $Z = -2.25$, $p < 0.03$; see Appendix A for the full results). Model comparison revealed that this interaction indeed significantly improves the model ($\chi^2 = 4.77$, $p < 0.03$). As can be seen in Fig. 1, the interaction was driven by the fact that while the relationship between the Proportion of Spanish Speakers and trill selection was numerically positive in the Unspecified European Language condition, it was negative in the Spanish condition ($\beta = -2.74$, $SE = 1.61$, $Z = -1.70$, $p < 0.09$). In other words, as predicted, a higher proportion of Spanish speakers in individuals' community increase their expectations that foreign languages will have trills. It does not increase individuals' expectations that Spanish will have a trill, probably because of a ceiling effect, as Americans in general already have this expectation, as also seen in the preliminary explicit survey mentioned in Sec. I. In fact, it seems that a higher proportion of Spanish speakers might even lead to a decrease in such expectation ($\beta = -2.73$). This could be either due to the fact that those who live in an area with more Spanish speakers might be more aware of the existence of taps in Spanish or because they might know the words. After all, in our stimuli, the tap version was always the real word.

Relatedly, some of our participants ($N = 11$; 5 in the Spanish condition) reported speaking some Spanish. To examine whether familiarity with the Spanish words used in the experiment is responsible for the negative association between Proportion of Spanish Speakers and trill selection in the Spanish condition, or the interaction with Condition, we reanalyzed the data while excluding these participants. This analysis showed the same interaction of Condition with Proportion of Spanish speakers ($\beta = -4.26$, $SE = 2.03$, $z = -2.1$, $p < 0.04$), and the magnitude of the negative association between proportion of Spanish speakers and trill selection in the Spanish condition did not decrease, but rather, numerically increased ($\beta = -3.4$). It seems then that learning some Spanish does not override the expectation of

Spanish to be more likely to have a trill than a tap, and might even strengthen it. Indeed, 6 out of the 20 participants in our preliminary explicit survey reported that they know some Spanish; all of them responded that trill exists in Spanish, while only half of them ($N = 3$) thought that a tap existed in Spanish as well. It therefore does not seem to be the case that a higher Proportion of Spanish Speakers in the community leads to a lower trill selection because it leads individuals to learn these words.

In contrast, with regards to the second possibility—that having more Spanish speakers in the community increases the awareness that Spanish has a tap—the distribution of responses in our preliminary survey supports it. If we do a median split over our participants according to the proportion of Spanish speakers in their community, we see that only 30% of the participants who only lived in communities with less than 15% Spanish speakers thought that Spanish has a tap, whereas 90% of those who lived in communities with over 15% Spanish speakers responded that Spanish includes a tap. Despite its small sample size, the results of the preliminary study are revealing and suggest that community makeup shapes up expectations about a language more than learning it does.

Additionally, the analysis revealed a marginal interaction between Condition and Proportion of Other Foreign Language Speakers ($\beta = 3.59$, $SE = 2.13$, $Z = 1.69$, $p < 0.1$). Comparing a model with this interaction to a model without it similarly shows that the interaction marginally contributes to the model ($\chi^2 = 2.74$, $p < 0.1$). For what it is worth, it should be noted that the direction of this interaction was, as predicted, the opposite of the one between Condition and Proportion of Spanish Speakers: a higher proportion of Other Foreign Language Speakers numerically decreases trill selection in the Unspecified European Language condition, but numerically increases it in the Spanish condition ($\beta = 2.81$, $SE = 2.04$, $Z = 1.38$, $p = 0.17$).⁵

As mentioned earlier, we collapsed over Experiment, since Experiment did not interact with any demographic factor (all z 's < 1). While not the focus of our study, it should be

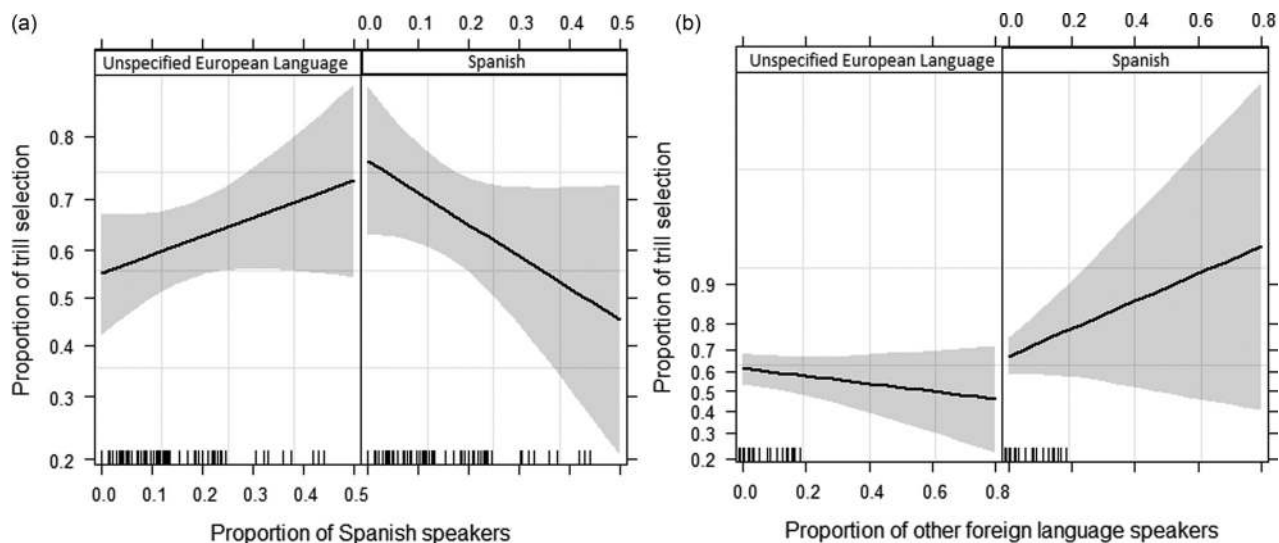


FIG. 1. Probability of trill selection in experiment 1 as a function of proportion of Spanish speakers in the community (a) or proportion of Other Foreign Languages in the community (b), broken down by experimental condition. Bands indicate 95% confidence intervals.

noted that a model with only the double interactions showed an interaction between Condition and Experiment ($\beta = 0.9$, $SE = 0.4$, $z = -2.23$, $p < 0.03$), reflecting the fact that participants in Experiment 1b (direct comparison) selected more trills in the Spanish condition than in the Unspecified European condition ($\beta = 0.72$, $SE = 0.3$, $z = 2.43$, $p < 0.002$) while this was not the case in Experiment 1a (indirect comparison).

Experiment 1, then, shows that the demographic makeup of the communities people live in can influence their expectations of what languages might sound like. Specifically, it shows that the proportion of Spanish, and potentially also the proportion of Other Foreign Language speakers, can influence people's expectations regarding whether languages are likely to have a trill sound. Experiment 2 tests whether this expectation can consequently influence people's perception of sounds in foreign languages.

III. EXPERIMENT 2

Experiment 2 builds on the results of Experiment 1 that show that the demographic makeup of the communities people have lived in can influence their linguistic expectations by testing whether this demographic makeup can also consequently influence their perception of foreign languages. Specifically, it tests whether a higher proportion of Spanish speakers increases people's misperception of taps as trills when listening to an unspecified foreign language, whereas a higher proportion of Other Foreign language speakers reduces this tendency.

We test this by presenting participants with the tap version of the words from Experiments 1a and 1b, followed by two imitations of the word by two non-native "learners." For all target words, one imitation included a tap and the other a trill. Participants needed to indicate which imitation is closest to the original. Selecting a trill in this task indicates misperception of the tap as a trill. As in Experiments 1a and 1b, we tested whether the proportion of Spanish and Other Foreign language speakers in participants' communities predict their tendency to misperceive taps as trills.⁶

A. Method

1. Participants

Seventy-four participants were recruited via M-Turk. All were residing in the U.S. Twelve participants were excluded because they were not native English speakers ($N = 4$), because they answered the catch trials incorrectly ($N = 3$), or because they had already participated in the other condition in this study ($N = 5$), leaving 62 participants, 37 in the Unspecified European Language condition and 25 in the Spanish condition.

2. Stimuli

The recordings of the 32 tap target words and 32 filler words from Experiment 1 were used. Additionally, two native English speakers capable of pronouncing taps and trills read all target words, once with a tap, and once by substituting the

tap with a trill, as well as all filler words. Figure 2 illustrates the different speakers' productions of taps and trills for the item *cara/carra*. Figure 3 shows boxplots of the durations of taps and trills for each speaker. Like for the native Spanish speaker (see Experiment 1), the mean durations of the taps and trills differed for the two native English speakers [speaker 1: 35 ms vs 78 ms, $t(31) = 12.64$, $p < 0.001$; speaker 2: 31 ms vs 85 ms, $t(31) = 18.23$, $p < 0.001$]. The two native English speakers marginally differed in the duration of their taps [$t(31) = 1.8$, $p < 0.09$]. Neither of them, however, significantly differed from the native Spanish speaker ($p > 0.1$). The two native English speakers also differed a bit from each other in the duration of their trills [$t(31) = 2.01$, $p < 0.05$]. The trills of both native English speakers, however, were significantly longer than those of the native Spanish speakers [$t(31) = 3.86$, $p < 0.001$ and $t(31) = 5.17$, $p < 0.001$]. As the target word in this experiment always contained a tap, the imitators' trill productions were less similar to the native speaker's taps than her own trill productions. Therefore, our stimuli make misperception of taps as trills even less common than in the real world, leading to a particularly conservative test.

We created triplets consisting of a target word followed by an imitation containing a trill by one speaker, and an imitation containing a tap by the other speaker. The inter-stimulus interval between the target word and the first imitation was 1 s. The two imitations were separated by 500 ms. Two versions were created for each target word, such that each speaker contributed a trill imitation for one version, and a tap imitation for the other version. Two lists were created, each containing only one version of each item. For example, in one version of the experiment, the recording of *acabaron* by the native Spanish speaker was followed by an imitation containing a tap from the English speaker 1 and an imitation containing a trill from the English speaker 2, while in the other list in the experiment, the original recording of *acabaron* was followed by an imitation containing a trill from English speaker 1 and an imitation containing a tap from English speaker 2. Having the same speaker provide both the trill and the tap version of each item across the two lists enabled us to control for all other potential differences between the speakers. Within each list, tap imitations preceded trill imitations in half of the items, and trill imitations preceded tap imitations in the other half. Each English speaker was equally likely to produce a trill or a tap and appear either first or second. Filler triplets were also created by coupling the recording of the Spanish speaker with one imitation by each English speaker.

3. Procedure

Participants first answered the same background questions as in Experiment 1. Then they were told that they would evaluate language learners. As in Experiment 1, there were two conditions. In the Unspecified European Language condition, participants were told that "the words are of a made up language based on the average of several European languages, including Swedish, Czech and French." The instructions in the Spanish condition were modeled to be as similar as possible, and therefore stated that "the words are Spanish words that are

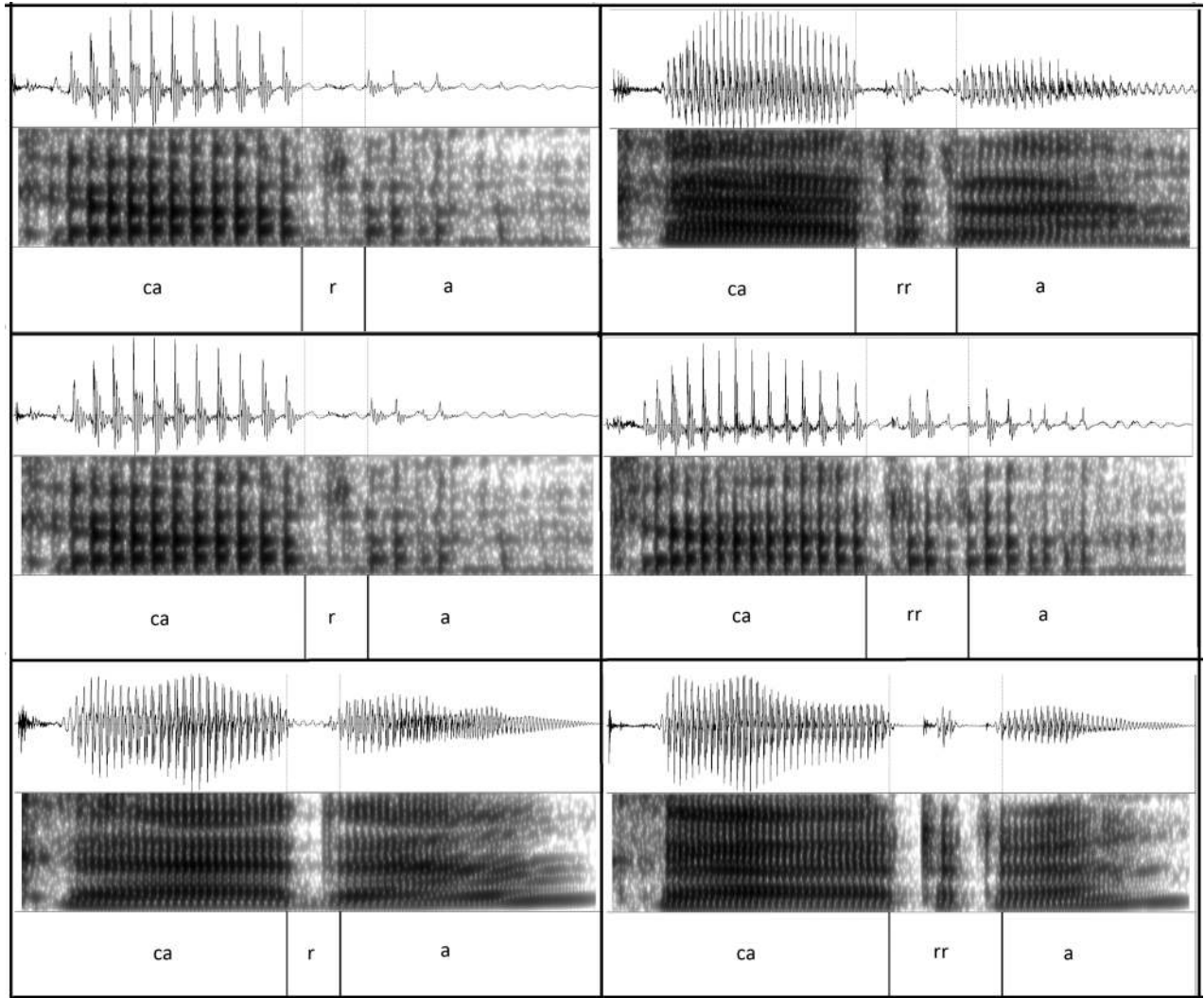


FIG. 2. Spectrograms of the tokens *cara* (i.e., with a tap) and *carra* (i.e., with a trill) from each of the speakers in our experiments. Top panels: native Spanish speaker (Experiments 1 and 2); middle and bottom panels: native English speakers (Experiment 2).

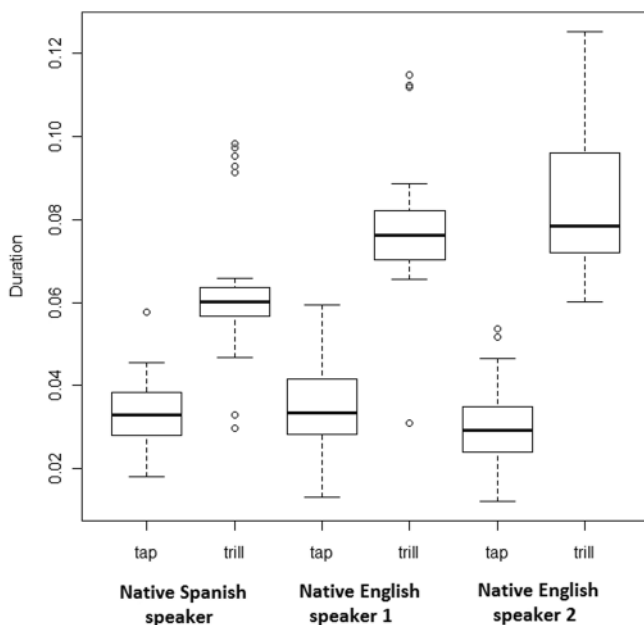


FIG. 3. Trill and tap durations of speakers in Experiments 1 and 2.

based on the average of several Spanish dialects, including European Spanish, Argentinian Spanish, and Mexican Spanish.” The instructions in the two conditions were otherwise identical, and asked participants to indicate which of the two imitations they heard of each word is more similar to the original. The words were presented as originating in several languages rather than one with the aim of targeting more general expectations rather than expectations of a particular language, as well as reducing the chance that participants might not believe the manipulation because they would think that the words do not sound like words in the specific provided language. Participants were not timed and responded at their own pace. They could click on an audio file more than once if they wished to. Items were presented in one fixed random order.

4. Community makeup data

As in Experiment 1, information about the proportion of Spanish speakers and the proportion of other foreign language speakers in each of the locations in which participants had resided was gathered from the latest [United States Census Bureau \(2012\)](#) with the 5-year estimates. The proportion of

Spanish speakers ranged from 0.003 (Orono, ME) to 0.925 (Hiialeah, FL; $M=0.2$). The proportion of Other Foreign Language speakers ranged from 0.009 (Smithfield, NC) to 0.577 (Rosemead, CA; $M=0.16$). The proportion of Spanish speakers did not correlate with the proportion of Other Foreign Language speakers ($r=0.06$, n.s.).

B. Results

An examination of the data revealed that some participants had a strong preference for one of the speakers over the other one. Participants' comments at the end of the experiment suggest that this was mostly due to the intonation that speaker 2 used. To control for the bias that such irrelevant preference might exercise, we coded for each participant their likelihood of selecting each speaker on filler trials. Four of the participants consistently selected one of the speakers more than 90% of the time in the filler trials (as well as in the target trials). They were excluded. For the remaining participants ($N=58$; $N=35$ in the Unspecified European Language condition, and $N=23$ in the Spanish condition), we entered the participant's baseline likelihood of selecting the speaker in the filler trials as a fixed variable according to identity of the trill speaker in that trial. That is, for the trials in which speaker 1 produced the trilled version, we entered participants' likelihood of selecting speaker 1 in the filler trials as a covariate, and for the trials in which speaker 2 produced the trilled version, we entered participants' likelihood of selecting speaker 2 in the filler trials.

On average, participants selected the trilled imitation of the tap 26% of the time. To test whether the demographic makeup of participants' communities influenced the participants' likelihood of selecting the trilled imitation, we conducted a mixed-model analysis with Participants and Items as random variables, and Speaker Bias, Condition (Spanish, Unspecified European Language), Proportion of Spanish speakers, Proportion of Other Foreign Language Speakers, and the interactions of Condition with Proportion of Spanish speakers and the Proportion of Other Foreign Language Speakers as fixed factors. The Condition variable was

treatment-coded with Unspecified European Language as the baseline level. The model included intercepts for the random variables as well as slopes for Speaker Bias, Condition, Proportion of Spanish Speakers, and Proportion of Other Foreign Language Speakers for the Items variable, and a slope for Speaker Bias for the Participants variable.

The analysis revealed an effect of Speaker Bias ($\beta=3.45$, $SE=0.81$, $Z=4.23$, $p<0.001$; see Appendix B for the full results) indicating that participants were more likely to select the imitation of the speaker whose imitations they also preferred on filler trials. Crucially, the analysis also revealed an effect of Proportion of Other Foreign Language speakers at the model's baseline (Unspecified Language condition: $\beta=-5.66$, $SE=1.85$, $Z=-3.06$, $p<0.01$) and an interaction between Condition and Proportion of Other Foreign Language speakers ($\beta=5.22$, $SE=2.21$, $Z=2.36$, $p<0.02$).⁷ Model comparison indicated that this interaction significantly improves the model ($\chi^2=4.16$, $p<0.05$). As can be seen in Fig. 4, the interaction reflects the fact that the higher the proportion of other foreign language speakers in the communities in which participants had lived, the less likely participants are to mishear taps as trills in the Unspecified European Language condition, as expected. Also as expected, a proportion of other foreign language speakers in the community did not influence performance in the Spanish condition ($\beta=-0.45$, $SE=1.23$, $Z=-0.36$, $p=0.72$). In contrast to our expectation, however, the Proportion of Spanish Speakers did not significantly influence performance.

One worry we had was that the proportion of Spanish speakers did not reach significance because of a couple of participants who had resided in quite atypical communities in terms of the proportion of Spanish speakers (92.5% and 83.3%, both more than 2.5 standard deviations away from the mean). These participants might have unduly influenced the results, both because their experience is atypical, and because these values are outliers. Excluding these participants, however, did not lead to any significant changes in the results.⁸

It is unclear why the Proportion of Spanish Speakers played a bigger role in Experiment 1. One possibility relates to the more

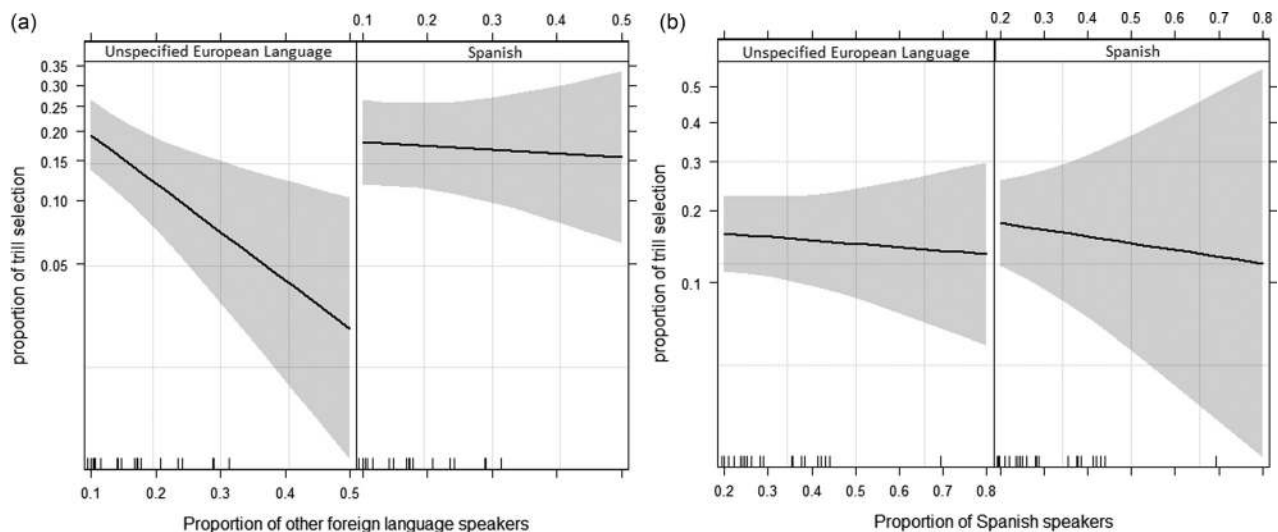


FIG. 4. Probability of trill selection in Experiment 2 as a function of proportion of other foreign language speakers in the community (a) or proportion of Spanish speakers (b), broken down by experimental condition. Bands indicate 95% confidence intervals.

meta-linguistic and explicit nature of the task in Experiment 1. In both Experiment 1a and Experiment 1b participants were explicitly asked about their linguistic expectations, whereas Experiment 2 was presented as a perceptual similarity task. Experiment 1 therefore might have induced participants to rely more on their *explicit* knowledge of stereotypes about Spanish and the relevant aspects of their experience, whereas the similarity judgment task in Experiment 2—where no mention is ever made of the fact that the imitations contrast phonemically—might have led participants to rely more on their *implicit* knowledge and expectations. The proportion of Spanish speakers in one's community can influence both explicit awareness of the sounds in Spanish and implicit associations between Spanish and certain sounds. In general, explicit and implicit knowledge and attitudes can differ in their content and in their predictive power on behavior (Greenwald *et al.*, 2009). While the relative predictive power of explicit versus implicit knowledge and attitudes for different tasks and behaviors is still not fully understood, it seems likely that explicit knowledge and attitudes would influence behavior more when the link to the knowledge is clear. Therefore, if a higher proportion of Spanish speakers in one's community increases explicit knowledge more than it increases implicit associations (because, for example, implicit associations are already common in the population at large), then a proportion of Spanish speakers in the community might influence the performance on an explicit meta-linguistic task more than it would influence tasks that do not directly ask about the existing knowledge. Further research, however, is needed to test whether reliance on different types of expectations is indeed the source of the difference between the two experiments.

Nonetheless, Experiment 2 shows the predicted interaction of proportion of other foreign language speakers and condition, such that in the unspecified language condition, but not in the Spanish condition, trill selection decreases with a higher proportion of other foreign language speakers in one's community. Thus, the study shows that the demographic makeup of an individuals' community can influence the way they perceive sounds in foreign languages.

IV. GENERAL DISCUSSION

An individuals' perception of speech is influenced by a variety of factors. One of the factors that shape perception is the individuals' past experience. Previous research on the role of past experience in shaping perception has focused on the shaping of categories by the distribution of different features in the input, mainly during first and second language acquisition (e.g., Flege, 1995; Werker and Tees, 1984). The experiments reported here show how past experience can also shape our linguistic expectations, and consequently, our perception. This is in line with research that shows that expectations regarding the *speaker* can influence speech perception (e.g., Drager, 2011; Koops *et al.*, 2008; Niedzielski, 1999).

But how do expectations influence perception? There are at least two possible mechanisms that could underlie such effects. One potential mechanism operates at the perceptual level. According to this account, expectations, or even simply aspects of the context (speaker identity, location, topic, etc.), lead certain exemplars (i.e., stored episodic tokens; Johnson, 1997) to be

more activated than others, pulling perception toward them. For example, an expectation to hear trills will enhance the activation of trilled exemplars more than of tap exemplars, and the heard token will be assimilated toward them. While listeners might differ in their past exposure to Spanish trills, or trills in general, it is unlikely that our participants have never heard any trill in their lives. As the preliminary study mentioned in Sec. I shows, all participants without exception, including Americans who have never lived in a community with any Spanish speakers, identified the audio recording of the trill as a Spanish sound. This indicates that they have some representation of the sound. Previous research on the effects of listeners' expectations of speakers on perception often proposed such a perceptual account (e.g., Drager, 2011; Hay and Drager, 2010; Niedzielski, 1999). In many cases, the exemplars that pulled perception were sounds that—like the trills in the present case—are not represented as phonologically distinct categories in participants' phonological grammar, but belong to regional accents, or are merged with another category in participants' own use (e.g., Hay and Drager, 2010; Koops *et al.*, 2008).

Another potential mechanism operates at the memory level. According to this account, the effects do not occur at the stage of initial perception but during the matching stage. This account assumes that the activation of the presented token quickly decays. Such decay might be particularly quick when the sound is foreign or less familiar, as recall is better for familiar than unfamiliar input. For example, input from one's first language is remembered better than input from one's second language, words are remembered better than nonwords, and phonotactically frequent stimuli are remembered better than phonotactically infrequent stimuli (e.g., Gathercole *et al.*, 1999; Thorn and Gathercole, 1999). Thus, during the matching task, when participants must recall the sound, they rely on reconstruction of the sound from memory, and this reconstruction process might be biased by expectations. If this is the case, then participants' responses might be more biased by their expectations the longer the delay between the presentation of the first token and the presentation of the imitations. As our target words differed in length, we measured the delay between the beginning of the tap phoneme in the target token and the beginning of the tap or trill phoneme in the first imitation of the word. This delay ranged from 1278 to 1930 ms ($M = 1610$). Adding delay to the model revealed a main effect of delay, in addition to the reported effects and interaction ($\beta = 1.52$, $SE = 0.73$, $z = 2.07$, $p < 0.04$). The longer the delay was, the more trill selection (i.e., misperceptions) there were. Models that included interaction of delay with other factors did not converge reliably and did not suggest that any interactions exist. The effect of delay suggests that misperception is partially due to the difficulty of holding the percept in memory. At the same time, since delay did not seem to interact with Condition or the proportion of Spanish or Other Foreign Language Speakers, some of the effect might be due to processes taking place at the initial perception stage. It is important to note, however, that this is a *post hoc* analysis, and it is possible that the longer and shorter words that were used in the experiment differ on certain aspects (e.g., the degree to which they sound Spanish), which

might lead to the delay effect or, in contrast, mask an interaction between delay and one or more other factors.

Hay *et al.* (2006), who tested how expectations regarding the geographic background of a speaker influence perception of that speaker's vowels, similarly examined the role of memory in underlying their effect of expectations. Specifically, they examined whether the delay between presentation of the target vowel and the end of the sentence in which it appeared, the point at which participants were tested on its identification, modulated the influence of expectations on participants' performance in the task. They did not find support for such an effect, even though the delay contrast in that experiment was quite large, as they presented participants with full sentences containing a target word in either the middle or the end of the sentence. It seems, then, that the effect is likely to be, at least partially, perceptual in nature. Further research should examine whether and how the role of each mechanism depends on the task and the context.

Some research has found that ambient exposure might have different effects on immediate processing versus long-term representations. In particular, it has been shown that covert speakers of the New York dialect, that is, people who spent their entire lives in the New York region but do not produce the dialect in their habitual speech, show the same short-term priming of dialect variants as overt speakers of the New York dialect, but unlike them, do not show any long-term semantic priming of the dialect (Sumner and Samuel, 2009). Several differences, however, make it difficult to draw inferences from these results that apply to performance in our studies. First, covert speakers are people who fully understand the dialect, whereas the ambient languages in the case of our participants are languages that they do not speak or understand. Relatedly, Sumner and Samuel (2009) examined processing of words that the listeners have stored in long-term memory. We examined processing of words that the participants might have never encountered, that they do not understand, and of which they therefore do not have long-term representations. Last, our study focuses on the effect of ambient languages on linguistic expectations, and consequently on perception. Sumner and Samuel do not examine expectations, and it is unclear how expectations might influence their pattern of results.

Interestingly, the experiments reported here show that past experience can shape our perception not only by exposing us to relevant distributional information but also by exposing us to potentially irrelevant information. For example, the proportion of Spanish speakers in one's community is not informative regarding the likelihood of other languages to have a trill. Similarly, the proportion of Other Foreign Language speakers is not necessarily informative about the likelihood of European languages to have a trill, especially since the common other foreign languages in the U.S. are mostly languages spoken in East Asia. Future research should investigate which factors and conditions govern the generalizations that individuals make from the input they received over their lifetime.

The results of these experiments extend previous research on the role of expectations in speech perception. Unlike that research, however, the present experiments focus on expectations about *languages* rather than expectations about

speakers. Such linguistic expectations are particularly interesting as they can have implications for language learning. Do expectations of certain sounds lead language learners, similarly to our participants, to misperceive and perhaps even misproduce the sounds of the language they are learning? The results of our preliminary survey mentioned in Sec. I suggest that this is likely to be the case, as half of those who reported knowing some Spanish were still under the impression that Spanish has a trill but not a tap. Anecdotal evidence suggests as well that this is the case. In fact, the native Spanish speaker who recorded the stimuli for these experiments and was oblivious to the goal of the experiment, remarked during the recording of the trill version of the words (which are mispronunciations of tap words) that she feels as if she is imitating Americans who try to speak Spanish. She commented that they tend to over-produce trills in cases where a tap should be produced. This means that the environment we live in might determine our success at acquiring new languages by influencing our expectations regarding them. While we focused on Spanish trills and taps in this study, we expect it to be true for other stereotypical sounds in other languages, such as the front round vowels in French and the pharyngeal sounds in Arabic.

To conclude, this set of studies adds to our understanding of how our environment can influence the way we process speech not only via bottom-up distributional learning but also by shaping our expectations.

ACKNOWLEDGMENT

This work was supported by Fondation Pierre-Gilles de Gennes and the Agence Nationale de la Recherche under Grant Nos. ANR-10-LABX-0087 and ANR-10-IDEX-0001-02. We would like to thank Alex Cristia, Alex Martin, and Maggie Zander for recording the stimuli, and Page Piccinini for analyzing the audio files.

APPENDIX A: DEMOGRAPHIC DETAILS OF PARTICIPANTS IN EXPERIMENT 1

States in which participants currently reside:

Unspecified European Language condition:

AL, AZ ($N=2$), CA ($N=4$), FL ($N=4$), GA, IN, MA, MD, ME, MI, MN, MS ($N=2$), NC ($N=3$), NY ($N=2$), OK, PA ($N=4$), RI, TX ($N=2$), UT, VA, WA, WI.

Spanish condition:

AR, AZ, CA ($N=3$), CO, CT, FL ($N=3$), GA ($N=3$), HI, IL ($N=2$), MA ($N=2$), MI, NY ($N=2$), OH ($N=4$), OR ($N=2$), PA, TN ($N=3$), TX ($N=3$), WA, WI, WY.

Average maximal percentage of Spanish speakers and of other foreign language speakers in participants' community makeup:

	% Spanish speakers	% Other Foreign Language speakers
Spanish condition	16%	12%
Unspecified European Language condition	20%	14%

APPENDIX B: RESULTS OF EXPERIMENT 1

	β	SE	z	p
Intercept	0.47	0.16	2.88	<0.01
Condition (Spanish)	0.26	0.22	1.17	0.24
% Spanish speakers	1.60	1.16	1.375	0.17
% other foreign language speakers	-0.78	0.68	-1.14	0.25
Condition \times % Spanish speakers	-4.33	1.93	-2.25	<0.03
Condition \times % other foreign language speakers	3.59	2.13	1.69	<0.1

APPENDIX C: DEMOGRAPHIC DETAILS OF PARTICIPANTS IN EXPERIMENT 2

States in which participants currently reside:

Unspecified European Language condition:

AL, AR, CA ($N=3$), CT, FL ($N=4$), GA ($N=2$), ID, IL, LA, MA, MD, ME, MI, NC ($N=2$), OH, OK, OR ($N=2$), PA ($N=3$), SC, TX ($N=3$), VA ($N=3$), WV ($N=2$).

Spanish condition:

CA ($N=2$), FL ($N=2$), GA ($N=2$), IA, IL, NC ($N=4$), NE, NJ, NY ($N=3$), OH ($N=2$), OR, TN, TX ($N=2$), WA, WI

Average maximal percentage of Spanish speakers and of other foreign language speakers in participants' community makeup:

	% Spanish speakers	% Other Foreign Language speakers
Spanish condition	18%	11%
Unspecified European Language condition	22%	11%

APPENDIX D: RESULTS OF EXPERIMENT 2

	β	SE	z	p
Intercept	-3.33	0.45	-7.36	<0.001
Trill speaker bias	3.40	0.82	4.17	<0.001
Condition (Spanish)	0.11	0.29	0.39	0.70
% Spanish speakers	-0.40	0.80	-0.50	0.62
% other foreign language speakers	-5.66	1.85	-3.06	<0.01
Condition \times % Spanish speakers	-0.40	1.91	-0.21	0.83
Condition \times % other foreign language speakers	5.22	2.21	2.36	<0.02

¹One potential caveat is that participants might have been less likely to respond affirmatively to the syllable with the tap, because it does not occur word-initially in Spanish. Therefore, hearing it without a preceding sound might have sounded less typical. While we cannot rule out this possibility, it seems unlikely considering another aspect of the findings discussed in the results of Experiment 1—respondents living in areas with more Spanish speakers were more likely to indicate that a tap exists in Spanish. Had the negative responses been due to the lack of fit between the common context and the manner of presentation in this study, more Spanish in one's environment should not have increased the likelihood of an affirmative response. Additionally, considering the fact that the speech stream is continuous, non-Spanish speakers would probably not be able to tell whether taps can appear at word-initial position, as they would not be able to segment the speech stream into separate words.

²Web-based experiments have been growing in popularity over the past decade. Such studies have not only proven to give reasonable results, direct comparisons of performance on web- and lab-based experiments showed that they are quite comparable, and as the population is more representative, they even generalize better to the real world (Buhrmester *et al.*, 2011; Crump *et al.*, 2013; Horton *et al.*, 2011; Mason and Suri, 2012). Furthermore, M-Turk has been used successfully in previous speech perception studies as well, and again, when performance was compared to lab-based experiments, the same effects replicated across platforms (e.g., Chodroff and Wilson, 2014; Kleinschmidt and Jaeger, 2012; Yu and Lee, 2014).

³We specifically designed the study such that the real words would always have a tap, because we wanted to ensure that the predicted increase in trill selection in the Unspecified European Language condition as a result of a higher proportion of Spanish speakers in the community could not be a consequence of having heard those words previously.

⁴Participants were also asked (1) what are the most common languages that are spoken in their area and (2) what are the most common languages that they hear. Unfortunately, almost all participants responded "Spanish" to both questions, so no further analyses were conducted on responses to these questions.

⁵We also ran this analysis while keeping the value of 1 for the Proportion of Other Foreign Language Speakers for those who lived abroad. The results looked the same. There was a significant interaction between Condition and Proportion of Spanish speakers in a community ($\beta = -4.26$, $SE = 1.88$, $Z = -2.27$, $p < 0.03$), and a marginal interaction between Condition and Proportion of Other Foreign Language Speakers ($\beta = 3.47$, $SE = 2.07$, $Z = 1.67$, $p < 0.1$).

⁶Note that expectations should not influence perception of the imitation, as the imitators were presented as learners whose pronunciation needs to be evaluated, and therefore prone to mispronounce words and not use the native phoneme inventory.

⁷Conducting the same analysis while keeping the values of 1 for the Proportion of Other Foreign Language Speakers for those who had lived abroad leads to the same results: an effect of Speaker Bias ($\beta = 3.42$, $SE = 0.8$, $Z = 4.27$, $p < 0.001$), an effect of Proportion of Other Foreign Language Speakers ($\beta = -3.31$, $SE = 1.26$, $Z = -2.62$, $p < 0.01$), and an interaction of Condition with Proportion of Other Foreign Language Speakers ($\beta = 3.23$, $SE = 1.42$, $Z = 2.285$, $p < 0.03$).

⁸It might be worth noting that this exclusion did lead to the emergence of a pattern that is numerically similar to the one found in Experiment 1: A higher proportion of Spanish speakers was numerically associated with a higher trill selection in the Unspecified European Language condition and with a lower trill selection in the Spanish condition. As said, this pattern was not significant (Unspecified European Language condition: $\beta = 1.23$, $SE = 0.95$, $z = 1.30$, n.s.; Spanish condition: $\beta = -1.73$, $SE = 1.86$, $z = -0.94$, n.s.).

Adams, C. (2002). "Strong assimilation and prestige: A sociolinguistic study in the Central Valley of Costa Rica." Ph.D. dissertation, University of California, pp. 1-195.

Babel, M., and Russell, J. (2015). "Expectations and speech intelligibility," *J. Acoust. Soc. Am.* **137**(5), 2823-2833.

Barkat, M., and Vasilescu, I. (2001). "From perceptual designs to linguistic typology and automatic language identification: Overview and perspectives," in *Proceedings of the Eurospeech*, pp. 1065-1068.

Bradlow, A., Clopper, C., Smiljanic, R., and Walter, M. A. (2010). "A perceptual phonetic similarity space for languages: Evidence from five native language listener groups," *Speech Commun.* **52**(11), 930-942.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2011). "Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?," *Perspect. Psychol. Sci.* **6**(1), 3-5.

Chodroff, E., and Wilson, C. (2014). "Burst spectrum as a cue for the stop voicing contrast in American English," *J. Acoust. Soc. Am.* **136**(5), 2762-2772.

Crump, M. J. C., McDonnell, J. V., and Gureckis, T. M. (2013). "Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research," *PLoS One* **8**(3), e57410.

Díaz-Campos, M. (2008). "Variable production of the trill in spontaneous speech: Sociolinguistic implications," in *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, pp. 47-58.

Drager, K. (2011). "Speaker age and vowel perception," *Lang. Speech* **54**(1), 99-121.

- Flege, J. (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 Press, Timonium, MD), pp. 229–273.
- Gathercole, S. E., Frankish, C. R., Pickering, S. J., and Peaker, S. (1999). "Phonotactic influences on short-term memory," *J. Exp. Psychol.: Learn., Memory, Cognit.* **25**(1), 84–95.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). "Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity," *J. Person. Soc. Psychol.* **97**(1), 17–41.
- Hay, J., and Drager, K. (2010). "Stuffed toys and speech perception," *Linguistics* **48**(4), 865–892.
- Hay, J., Nolan, A., and Drager, K. (2006). "From fush to feesh: Exemplar priming in speech perception," *Linguistic Rev.* **23**(3), 351–379.
- Henriksen, N. C. (2015). "Acoustic analysis of the rhotic contrast in Chicagoland Spanish: An intergenerational study," *Linguistic Approach. Bilingual.* **5**(3), 285–321.
- Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). "The online laboratory: Conducting experiments in a real labor market," *Exp. Economics* **14**(3), 399–425.
- Hualde, J. I. (2005). *The Sounds of Spanish* (Cambridge University Press, Cambridge).
- Johnson, K. (1997). "Speech perception without speaker normalization: An exemplar model," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic Press, San Diego, CA), pp. 145–165.
- Johnson, K., Strand, E. A., and D'Imperio, M. (1999). "Auditory-visual integration of talker gender in vowel perception," *J. Phonetics* **27**(4), 359–384.
- Johnson, K. E. (2008). "Second language acquisition of the Spanish multiple vibrant consonant," Doctoral dissertation, ProQuest, pp. 1–237.
- Kleinschmidt, D., and Jaeger, T. F. (2012). "A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation," in *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, pp. 605–610.
- Koops, C., Gentry, E., and Pantos, A. (2008). "The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas," Univ. PA Work. Papers Linguistics **14**(2), 93–101.
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (2006). "Infants show a facilitation effect for native language phonetic perception between 6 and 12 months," *Develop. Sci.* **9**(2), F13–F21.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). "Linguistic experience alters phonetic perception in infants by 6 months of age," *Science* **255**, 606–608.
- Lipski, J. (1994). *Latin American Spanish* (Longman, New York), pp. 1–392.
- Marks, E. A., Bond, Z. S., and Stockmal, V. (2003). "Language experience and the representation of phonology in an unknown language," *RESLA* **16**, 23–31.
- Mason, W., and Suri, S. (2012). "Conducting behavioral research on Amazon's Mechanical Turk," *Behav. Res. Methods* **44**(1), 1–23.
- McGowan, K. B. (2015). "Social expectation improves speech perception in noise," *Lang. Speech* **58**(4), 502–521.
- Niedzielski, N. (1999). "The effect of social information on the perception of sociolinguistic variables," *J. Lang. Soc. Psychol.* **18**(1), 62–85.
- Peterson, G. E., and Barney, I. I. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- Rubin, D. L. (1992). "Nonlanguage factors affecting undergraduates' judgments of non-native English speaking teaching assistants," *Res. Higher Ed.* **33**, 511–531.
- Staum Casasanto, L. (2008). "Does social information influence sentence processing?," in *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, Washington, DC, pp. 1–6.
- Sumner, M., and Samuel, A. G. (2009). "The role of experience in the processing of cross-dialectal variation," *J. Memory Lang.* **60**, 487–501.
- Thorn, A. S., and Gathercole, S. E. (1999). "Language-specific knowledge and short-term memory in bilingual and non-bilingual children," *Quart. J. Exp. Psychol. A* **52**(2), 303–324.
- Tversky, A., and Kahneman, D. (1973). "Availability: A heuristic for judging frequency and probability," *Cognit. Psychol.* **5**(2), 207–232.
- United States Census Bureau. (2012). "Selected social characteristics in the United States 2008–2012," American Community Survey (5-year estimates).
- United States Census Bureau. (2013). "Language use in the United States: 2011" (<http://www.census.gov/prod/2013pubs/acs-22.pdf>) (Last viewed September 28, 2015).
- Werker, J. F., and Tees, R. C. (1984). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," *Infant Behav. Develop.* **7**, 49–63.
- Willis, E. W. (2007). "An acoustic study of the 'pre-aspirated trill' in narrative Cibaëño Dominican Spanish," *J. Int. Phon. Assoc.* **37**, 33–49.
- Yu, A. C., and Lee, H. (2014). "The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study," *J. Acoust. Soc. Am.* **136**(1), 382–388.