Journal of Phonetics (1997) **25,** 421–436

Gestural drift in a bilingual speaker of Brazilian Portuguese and English

Michele L. Sancier* and Carol A. Fowler

Haskins Laboratories and University of Connecticut, U.S.A.

Received 13th November 1996, and in revised form 15th April 1997

We report three experiments exploring the occurrence of perceptually-guided changes in speech production by a speaker well past the critical period for language acquisition. A first experiment shows that listeners sharing our speaker's native language (Brazilian Portuguese) can distinguish her productions in that language as having been produced either after recent experience in Brazil or after recent experience producing and listening to English in the United States. In contrast, native English speaking listeners cannot distinguish our speaker's English productions by recent experience. Acoustic measurements of our speaker's voiceless stops produced in both Brazilian Portuguese and English show that, whereas her VOTs are always shorter for productions in Brazilian Portuguese than in English, VOTs of stops produced in both languages are shorter after a several month stay in Brazil than after a several month stay in the United States. We offer a theoretical account of the findings.

1. Introduction

This investigation examines perceptually-guided changes in speech-production (here "gestural drift") in a bilingual speaker and listeners' perception of these changes in the speaker's productions over time. Anecdotal reports of speakers, rather than published research, initially motivated our research. In these reports, geographically-displaced speakers well past the ostensible critical period for language acquisition (e.g., Lenneberg, 1967), were found to have acquired an accent whose origin could be found in the speech of the ambient language community. For example, a colleague of ours, a native speaker of British English, who has been living in the United States for many years, reports that his relatives in England tell him that he speaks with a "[gɑstli] American accent," pronouncing especially the first vowel in "ghastly" with what is still a most decidedly British accent to American ears. A second, now cross-linguistic example, is that of a young woman who is a graduate student in the United States and who has been in the routine of returning home to her native Brazil twice a year during summer and winter vacations. She translated for us her father's words to her on her arrival home from a prolonged stay in the United States, that her speech was "so explosive". Voiceless stops in Brazilian Portuguese are unaspirated, whereas in American English they tend to be aspirated.

*Author to whom all correspondence should be addressed at: Haskins Laboratories, 270 Crown St, New Haven, CT 06511, U.S.A.

0095-4470/97/040421 + 16 $25.00/0/jp970051

Accordingly, the "explosiveness" of her Brazilian-Portuguese stops may reflect an influence of the American-English speech that surrounds the speaker when she is in the United States. The latter cross-language case is the focus of our study. These are just two of many anecdotal reports that imply that pronunication learning toward the phones of an ambient dialect or language can occur well after the "critical period" for language acquisition.

In our view, if this pronunciation learning does occur, and listeners perceive that it has occurred, it is interesting, particularly with respect to considerations of its origins. When it occurs among speakers of different dialects of a common language, the drift might be ascribed to the speaker's attempts at social affiliation; that is, imitation of ambient speech may occur by choice. However, such an explanation cannot account for a speaker's Brazilian-Portuguese speech drifting toward American English in an ambient American-English language environment. There is no social motivation for it to do so. Why else might drift occur?

One possible source of gestural drift may be an underlying disposition of listeners/speakers to imitate the speech they hear; that is, gestural drift may indicate a tendency to imitate the ambient language, a tendency that prelinguistic infants have been shown to exhibit (Boysson-Bardies, Hallé, Sagart & Durand, 1989; Boysson-Bardies & Vihman, 1991; Kent & Murray, 1982; Kuhl & Meltzoff, 1996; Vihman & Boysson-Bardies, 1994; Whalen, Levitt & Wang, 1991). Put another way, perception of speech may foster imitation.

In addition to being a possible indicator of the tendency to imitate speech, gestural drift may also be of interest because, when it gives rise to second language (L2) influences on first language (L1) phones, it may expose the kinds of crosslinguistic-category correspondences proposed by Flege (1987). In one account of these correspondences, Flege (1987) explores the notion of "similar" phones. A similar phone is a production in L2 that closely resembles an L1 production such that the L2 phone is identified with the L1 phone. For example, in the case of our Brazilian-Portuguese speaker, the more aspirated [t^h]s of English and the less aspirated [t]s of Brazilian Portuguese may constitute similar phones for her.[1]

Flege (1987, 1995) suggests that under certain conditions an individual forms a category in L1, say, American-English /t/, that ultimately prevents the establishment of a distinct L2 category for a similar phonological segment, say, French /t/. This prevents the production of the L2 segment from being "authentic" (unaccented). This influence of an L1 production category on the authenticity of a similar L2 phone production is called "interference." Accordingly, these L1 and L2 productions are tied to one another, and, because they belong to the same L1 /t/, they are in correspondence. With this correspondence in place, the L2 production also influences the L1 production; that is, there is a "restructuring" of the L1 category.

Flege (1987) focused on the prevention of authentic production of an L2 phone due to influence by an established L1 category, as observed across monolingual and bilingual groups of speakers, rather than, as we do here, on within-speaker gestural drift; however, his data provide an indication that our anecdotal evidence of gestural drift may be

[1] We use brackets to indicate a specific phone as uttered by a speaker or as it can be uttered by a speaker. Slashes generally refer to a speaker's production category (a gesturally based phonological category) which, at first, is the L1 production. However, in descriptions of other research reports, we also use slashes if the authors had used them, whether or not that usage conforms to our own.

meaningful. Among other groups, Flege investigated native speakers of American English living in Paris and native speakers of French living in Chicago. He also investigated groups of monolingual speakers to estimate phonetic norms in each language. Flege measured VOT in productions of initial [t] and [t^h], respectively, in French and English words. We are interested in the seven Americans living in Paris, who had massive exposure to the French language in France for the 12 years prior to the study; they were married to native speakers of French and had French-speaking children. We are also interested in the age-matched group of French subjects who had been living in Chicago for the 12 years prior to the study; four of them were married to native English speakers at the time of the study. English [t^h] productions by the Americans living in Paris yielded a mean VOT of 56 ms, compared with 77 ms for English monolinguals. Analogously, French [t] productions by the native French speakers living in Chicago yielded a mean VOT of 51 ms, compared with 33 ms for French monolinguals. These data suggest that the VOTs of American-English speakers' American-English [t^h] decreased in the French environment, drifting toward French [t], and that the VOTs of French speakers' French [t] increased in the American-English environment. Compatible evidence has been reported by Major (1992) on changes in American-English speakers' VOTs in their English and Brazilian-Portuguese stops after 12–35 years living in Brazil.

Following the above-mentioned anecdotal reports and suggestive evidence of Flege and Major, we designed three experiments, two to test the perceptibility of gestural drift and one to measure the drift in the VOTs of a speaker exposed recently to an L1 or L2 language environment. Our speaker was the native speaker of Brazilian Portuguese described earlier.

Our first experiment was designed to determine whether Brazilian Portuguese listeners detected a change in our speaker's accent depending on her recent exposure either to English (her L2) or to Brazilian Portuguese (her L1). We hypothesize that listeners who are native to a speaker's language or dialect can hear an accent due to the speaker's recent experience in a different language environment; specifically, for our speaker, we predict that Brazilian-Portuguese listeners in Brazil will hear an accent, due to recent experience in the US, in her Brazilian-Portuguese productions.

We are not aware of anecdotal evidence of a complementary sort—that is, that listeners hear foreign-accented speech as less foreign accented after recent experience in the listeners' native language community. However, because Flege (1987) and Major (1992) do provide evidence for long-term production changes, we investigate in a second experiment, analogous to the first, whether American-English listeners are able to hear an increase in accent, due to recent experience in Brazil, in our speaker's English speech.

Finally, having obtained evidence in Experiment 1 that our speaker's Portuguese speech sounds more accented when she has been recently exposed to English but, in Experiment 2, that English listeners do not hear an accent change in our speaker's English, we look for changes in her productions. Following our anecdotal report and related studies of Flege and Major, we focus here on changes in this speaker's VOTs as a function of the speech of the ambient language community. We predict that a stay in the US of several months will lead to increases in VOTs of our speaker's native (Brazilian Portuguese) unaspirated stops. Correspondingly, we expect that a stay in Brazil will reduce our speaker's L1 VOTs. Because any shift in the speaker's Brazilian-Portuguese VOTs in the US environment appears to require that our speaker identify English and Brazilian-Portuguese voiceless stops as "similar" in Flege's sense, we expect our speaker's English voiceless stops to shift in parallel with her Portuguese stops. If the last prediction

is borne out, we will have to explain why the changes in our speaker's Brazilian-Portuguese speech are audible to Brazilian Portuguese listeners but the changes in her English are not audible to English listeners.

2. The speech corpus for Experiments 1–3

Our speaker was a 27-year-old female native speaker of Brazilian Portuguese who had attained an advanced level of proficiency in American English at the time of the study. She has normal speech and hearing. The speaker began learning English at age 15. Although no outside access to English was available to her while she lived in Brazil, her English teachers were all native speakers of American or British English. However, her initial exposure consisted in once-a-week classes and three-times-a-week classes in high school. She had no classes in college. Intensive exposure to English began in 1991 (about 4 years before our data collection took place) at an institute in Brazil, where she studied for 6 months just before coming to the US. On arrival in the US, she took a 3 month intensive course in Boston where she used language-laboratory tapes. In addition, she lived with an American family and typically spoke only English.

Since then, as a graduate student she has gone back and forth between Brazil and the US twice a year, staying in each place for months at a time. She speaks mainly English while in the US and mainly Portuguese while in Brazil.[2]

In our study, on each of 6 days, the speaker translated 12 sentences from English into Portuguese and 12 sentences from Portuguese into English in five randomized blocks each, yielding 60 sentences spoken in each language per day. (Please refer to the Appendix for the sentences used and the phones measured for VOT.) We generated sentences so that we could look for changes in a variety of phones; so far, we have looked only at VOT. The same material was used on all 6 days; that is, with regard to wording and content, the 12 sentences were always the same. These sentences were compiled with the aid of a native speaker of Brazilian Portuguese, not our subject, to try to ensure that prompts would elicit an intended translation. Use of translation rather than reading was meant to focus our speaker's attention on what she was saying, not on how she was saying it. For the English sentences, she was prompted by a native speaker of American English, and for the Portuguese sentences, she was prompted either by a native speaker of Brazilian Portuguese (in the first session) or by a native speaker of Castilian Spanish (in the remaining sessions), who understands and speaks Portuguese. For example, the native speaker of English said to the subject "Tell JR, 'They don't have bread'," consistently eliciting from the speaker "Eles não têm pão.", which is the Portuguese translation. During a given day, then, this sentence was uttered five times, in five different randomizations.

[2] On many practical grounds, our speaker was almost uniquely ideal for our purposes, and that is why we ran her in our study. She has the requisite fluency in two languages and spends months of each year in the two language communities. She was willing to make herself available for testing immediately after her trip to Brazil. And she is being trained as an experimental psychologist and so could collect listening data for us while she was in Brazil. In two other respects, she is less than ideal. First, although she is not a phonetician, she has some training in phonetics. Second, although we did not discuss the purpose of our study with her, she was likely aware of its general purpose, because she attended research discussion meetings in which, among other topics, research ideas on this topic were discussed. We believe that our results render any influence of these factors very unlikely in that our effects average only 5 ms, a magnitude of change that would be very difficult to effect intentionally.

We recorded the speaker after a 4.5-month stay in the US right before leaving for Brazil, upon return from Brazil after a 2.5-month stay, and once again, right before leaving for Brazil after 4 months in the US. Each "session" involved recordings on 2 successive days. English-translation and Portuguese-translation portions of the sessions were conducted one after the other in every session.

We recorded acoustic and electroglottographic signals. (EGG gives an articulatory estimate of VOT, in that it measures the start of vocal-fold contact for phonation in relation to the release of the stop closure.) The signals were synchronized and recorded using the Haskins Real-Time Physiological Signal Analysis Systems. The audio signal sampling rate was 20 kHz, and the EGG signal sampling rate, 10 kHz, with a 12-bit resolution. Contact electrodes were placed on the laminae of the speaker's larynx; a microphone was used to record the audio signal.

Both the EGG and acoustic signals were monitored for clipping on an oscilloscope before the experiment. The EGG signal was unfiltered, and the audio signal was filtered at 10 kHz. In the first session, pre-emphasis was applied to the audio signal, because high frequency spectral information was originally to be analyzed. For the subsequent two sessions, however, no pre-emphasis was used because we decided that only temporal information would be examined.

3. Experiment 1

In our first experiment, we asked whether Brazilian-Portuguese listeners could tell which of a pair of our speaker's Brazilian-Portuguese sentences, one spoken after a stay in Brazil and one after a stay in the US, sounded more foreign accented to them. The listeners also attempted to note what, if anything, sounded changed.

3.1. *Methods*

3.1.1. *Subjects*

Subjects were 13 male and female native speakers of Brazilian Portuguese in Sao Paulo, Brazil, ranging in age from 18 to 35.[3]

3.1.2. *Materials*

We used the following subsets from our corpus of Brazilian-Portuguese sentences: the first four tokens (taken from the first four randomizations, or blocks, of the five blocks for that day) of each of the 12 sentences spoken on the first day of the second session, right after our speaker's 2.5-month stay in Brazil; and the first four tokens (again, from the first four blocks of the five) of each of the 12 sentences spoken on the first day of the third session, after a 4-month stay in the US. In two cases, we used a token from the fifth block when an utterance from the first four blocks was unusable. We edited silence from the beginning and end of the each sentence. Next, we created four different randomized blocks of the 12 Portuguese utterances such that in each block there were randomized pairings of each of the 12 utterance pairs, one member of each pair drawn from the session right after the stay in Brazil and the other member of the pair, from a session after

[3] We thank Ana Luisa G. P. Navas for locating subjects and for data collection in Brazil.

the stay in the US. Members of a pair were always translations of the same sentence; on half of the trials the sentence spoken after the Brazil experience was first in the pair. In the first three blocks, there was a 1 s pause between the members of a pair, a 4 s pause after the pair, and a 10 s pause between blocks. In the fourth block, everything was the same except that instead of a 4 s pause after each sentence pair, there was a 10 s pause to give time for comments.

3.1.3. *Procedure*

Subjects listened to the audiotape and chose which of two successively presented utterances sounded more American accented. We emphasized in the instructions that listeners should concentrate on how the speaker pronounced her words. On the fourth block, we asked listeners, after attending to the first task, to try to explain how they decided which of the two utterances was more foreign accented. In particular, we requested that they specify in what respects any words or sounds sounded different or strange to them in the sentence they chose as the more accented.

3.2. *Results and discussion*

On average, listeners selected the Brazilian-Portuguese sentences uttered just after the American experience as more accented than those uttered just after the Brazilian experience on 66% of trials. Paired t-tests showed that performance was significantly greater than chance (50%) [by subjects: $t(12) = 4.39$; $p < 0.0009$; by items: $t(11) = 9.22$; $p < 0.0001$]. Every subject but one performed above chance on the task, and every sentence pair was associated with above-chance performance. This study provides strong evidence that our speaker's Portuguese was detectably more accented after several months of exposure to American English. Although overall comments from our listeners as to how they made their judgments were sparse, some of them were salient. Generally, listeners' comments suggested the presence of hyperarticulation, changes in intonation, and changes in nasality after the speaker's stay in the US.

4. Experiment 2

In a second experiment, we asked whether American-English listeners could tell which of a pair of our speaker's English sentences, one spoken after a stay in Brazil and one after a stay in the US, sounded more foreign accented to them. This same group also attempted to note what, if anything, sounded different.

4.1. *Methods*

4.1.1. *Subjects*

Thirty-three[4] male and female native speakers of American English, all of them students at the University of Connecticut-Storrs, participated as listeners and received course credit for their participation.

[4] We ran 20 more American than Brazilian listeners, only because American listeners were more available to us than Brazilian listeners.

4.1.2. *Materials*

The subset of our speaker's English productions that we selected for our listening test was analogous to the subset of Brazilian Portuguese sentences selected for Experiment 1. We chose the first four tokens (taken from the first four blocks of the five for that day) of each of the 12 sentences drawn from the first day of the second session, right after our speaker's 2.5-month stay in Brazil, and the first four tokens (from the first four blocks of the five) of each of the 12 sentences spoken on the first day of the third session, after a 4-month stay in the US. In three cases when one of the first tokens was unusable, we used tokens from the fifth block. We edited the sentences to eliminate leading and trailing silence. Then we created our different randomizations of the American-English utterances, one member of each pair drawn from the session right after the stay in Brazil and the other member of the pair from a session after the stay in the US. Within a pair, members were always translations of the same sentence. Across trials, we counterbalanced which sentence (produced after a stay in the US, or after a stay in Brazil) occurred first. The timing of sentences was as in Experiment 1.

4.1.3. *Procedure*

Subjects listened to the audiotape and chose which of two successively presented utterances sounded more foreign accented. As in Experiment 1, we emphasized in the instructions that listeners should concentrate on how the speaker pronounced her words. On the fourth block, we asked them to provide the bases for their choice of which of each pair of utterances was more foreign accented by specifying in what respect any words or sounds sounded different or strange to them in the sentence they selected as the more accented.

4.2. *Results and discussion*

On average subjects identified the American-English sentences uttered after the Brazilian experience as more accented than those uttered right after the American experience in 48% of trials. Performance was not significantly different from chance (50%) [by subjects: $t(32) = -1.20$; $p = 0.24$; by items: $t(11) = -0.92$; $p = 0.38$]. We looked at performance separately on each block but found no improvement in performance over blocks. Finally, the two experimenters, who are well-acquainted with the speaker, also took the listening test, and both performed at chance.

We next compared performance (percent correct minus the 50%-chance level) of American-English and Brazilian-Portuguese listeners and found that the two groups differed significantly, with performance considerably better for the Brazilian listeners [$t(44) = 4.68$; $p < 0.0001$].

Either of two reasons might underlie this difference in performance between American and Brazilian listeners. One possibility is that our speaker's Brazilian Portuguese undergoes gestural drift whereas her English does not, contrary to our prediction in the introduction. Alternatively, the perceptual judgment may be easier for Brazilian listeners, who may be distinguishing accented from unaccented speech, than for American listeners, who must differentiate different degrees of Portuguese-accentedness. Experiment 3 distinguishes these possibilities.

5. Experiment 3

In this final experiment, we asked whether our speaker's VOTs of Brazilian-Portuguese [p] and American-English [p^h], and those of her Brazilian-Portuguese [t] and American English [t^h] would shift in parallel, by correspondence, as a function of the ambient language, or alternatively, whether only her Brazilian-Portuguese [p] and [t] VOTs would shift as a function of the ambient language. We chose to look at the VOTs of these voiceless stops for several reasons. They occurred in sufficient numbers in our corpus; Flege (1987) has shown sensitivity of VOT to English influence on French and vice versa, French being a language, like Portuguese, with unaspirated stops (and, we subsequently discovered, Major (1992) found a similar effect on English VOTs of long-term exposure to Brazilian Portuguese); our speaker's father judged her American-accented Portuguese to be "explosive," finally, according to some of the comments of the Brazilian-Portuguese listeners, the speaker's speech, after a stay in the US, sounded excessively "well-articulated." Both the father's comments and our subjects' might reflect an increase in our speaker's VOTs. We do not suppose, of course, that any changes we may find in our speaker's VOTs exhaust the changes that occur in her speech as a function of recent language experience or exhaust the reasons for Brazilian listeners' above-chance performance in Experiment 1.

5.1. *Methods*

5.1.1. *Subjects*

The speaker was the native speaker of Brazilian Portuguese described earlier.

5.1.2. *Materials*

For analyses performed on each consonant in each language, from the entire corpus of sentences spanning both experiences in the US and that in Brazil, we measured 30 tokens each of six Portuguese words containing [p], 30 tokens each of six Portuguese-[t] words, 30 tokens each of four English-[p^h] words and 30 tokens each of three English-[t^h] words. In most instances the stops were word-initial; however, we measured the [t] in Portuguese capital, the [t^h] in English thermometer (the /t/ as pronounced [t^h] by our speaker) and the [p^h] in English compare. Generally, we only measured /p/ and /t/ VOTs when in a position of stress.

In a few cases, our speaker did not translate consistently (please see the Appendix for alternative translations in parentheses). In these instances, we did not measure the /p/ and /t/ VOTs of the intended translation because they were too few in number. For example, some translations were supposed to yield "talk" or "talking." Our speaker did use "talk" ("talking") a couple of times only; but on all other occasions, she used "speak" ("speaking"). We also avoided measuring /p/ VOT in the proper name "Paolo" when our speaker used it, instead of "Paul," in a sentence translated into English.

5.1.3. *Procedure*

Measurement criteria were established using waveform and spectrographic analyses of the audio signal and the EGG signal. We adhered to the following criteria. In cases in which the waveform signal appeared to precede the EGG signal, VOT was measured

from the beginning of the release burst to vowel onset in the waveform. In cases in which the EGG signal preceded the waveform, VOT was measured from the beginning of the burst to the first glottal pulse of the EGG signal. In most instances, the recorded EGG signal preceded the acoustic waveform in time; however, in the less common instances in which the acoustic waveform appeared to precede the EGG signal, it is evident that voicing had already begun but did not show up in the EGG because of a weakened signal. Again, in such a case, the VOT measurement was made solely from the acoustic signal. Spectrographic analysis was also consistently used to help guide the decision, especially in the few cases of microphone blast that occurred. Although spectrographic analysis lags slightly in time, even behind the acoustic waveform, it offered, nevertheless, a third source of information.

5.2. *Results and discussion*

Outliers, defined as VOTs more than two standard deviations from the mean of their condition, were replaced by condition means. This occurred for 3.8% of Portuguese tokens and 1.9% of English tokens.

In Fig. 1, bar graphs represent the effect of recent language experience on VOT, one graph each for Portuguese [p] and English [p^h], and for Portuguese [t] and English [t^h]. The figure shows that, as expected, VOTs are generally longer in the US sessions than in the Brazil session.

An ANOVA on VOTs of our speaker's /p/, with factors language (Brazilian Portuguese, American English) and recent experience (sessions after several months in the US (US1, US2) and a session after several months in Brazil (BR)), reveals a significant effect both of language [$F(1, 294) = 242.72$; $p < 0.0001$] and of recent experience [$F(2, 294) = 7.03$; $p < 0.001$]. The interaction did not reach significance. The effect of language is significant because the mean VOT for English [p^h] is 22 ms longer than that for Portuguese [p]. As for recent experience, a planned comparison of the two US sessions with the Brazil session [$F(1, 294) = 14.00$; $p = 0.0002$] shows that VOTs following experience in the US are significantly longer (by 6 ms on average) than VOTs following several months in Brazil. The lack of an interaction between the factors indicates that the changes in VOT pattern in the same way in the speaker's Portuguese and English speech.

The same analysis of our speaker's VOTs in /t/ reveals a significant effect of language [$F(1, 264) = 1234.94$; $p < 0.0001$]; this effect is significant because VOTs in English are on average 33 ms longer than in Portuguese. The effect of recent experience is also significant [$F(2, 264) = 13.84$; $p < 0.0001$], because VOTs are longer (by 5 ms on average) after the experience in the US than in Brazil. There is also a significant recent experience by language interaction [$F(2, 264) = 22.86$; $p < 0.0001$]. ANOVAs performed on each language separately indicate a significant effect of recent experience in Portuguese [$F(2, 177) = 19.53$; $p < 0.0001$], and in English [$F(2, 87) = 10.53$; $p < 0.0001$]. Planned comparisons of the two US sessions with the Brazil session, performed separately on Portuguese [t] and English [t^h], both yield a significant outcome (Portuguese: [$F(1, 177) = 23.32$, $p = 0.0001$; English: [$F(1, 87) = 4.99$, $p = 0.028$). However, as Fig. 1 shows, the outcome in English was not wholly consistent. Mean VOT for the second US session was the same as that for the Brazil session; both were 10 ms shorter than VOTs collected in the first US session.

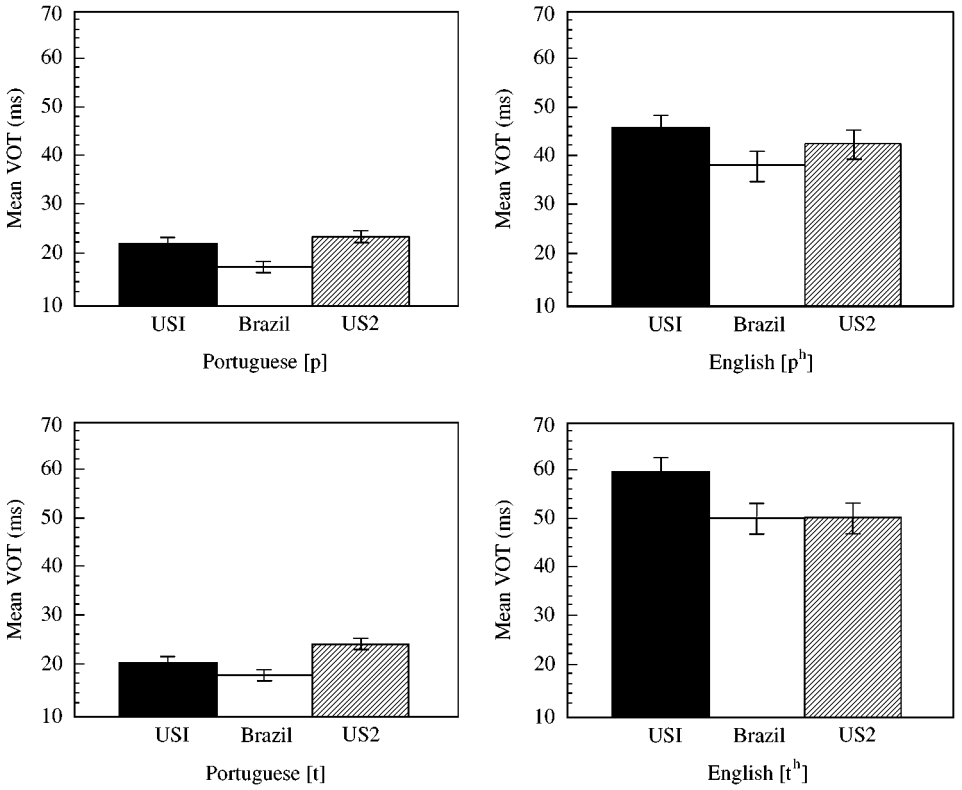

Figure 1. Mean VOTs of Portuguese [p], English [p^h], Portuguese [t] and English [t^h] measured from our native Brazilian-Portuguese speaker after recent experience in the US (US1, US2) and Brazil. Standard errors are marked.

Because we had different numbers of words and, of course, different words in the two languages, we did not use word as a factor in the foregoing analyses. In separate analyses, with the factors recent experience and word, performed on each consonant in each language, we found no interaction of recent experience with word for English [t^h] or Portuguese [t] but significant interactions for both English [p^h] and Portuguese [p]. For English [p^h] three of the four words uttered during both US sessions had longer VOTs than those of these same words uttered during the Brazil session. For the fourth word, the average VOT in the second US session was shorter than that in the Brazil session. For Portuguese [p], every word except one has longer VOTs in both US sessions than in the Brazil session.

6. General discussion

With just a few months of exposure at a time in the US and in Brazil, our speaker manifested gestural drift, that is, a change in VOT reflecting a change in the relative phasing of a laryngeal devoicing gesture and an oral constriction gesture. This difference in VOT was of approximately the same magnitude in the two languages, but it was

audible only to the Brazilian listeners. Tentatively, we ascribe the difference in audibility to the somewhat different tasks confronting the two groups of listeners. Whereas it is likely that Brazilian listeners distinguished our speaker's accented speech, produced after a stay in the US, from unaccented speech, our American listeners had to discriminate speech all of which was accented, but (as our measurements show) differed in degree of accent.

We now attempt to explain why our speaker showed gestural drift. In particular, why did her VOTs in both languages drift toward those of her ambient language community? Most interestingly, why did her Portuguese VOTs drift toward those of American English when she was in the US, and why did the VOTs of her English stops drift toward those of Portuguese stops when she spent time in Brazil? Three ideas may be required to explain the drifting we have observed.

First, as mentioned earlier, we believe that listeners/speakers are disposed to imitate. This may account for why our speaker's American-English voiceless consonant VOTs gravitate toward those of American English when she is in Connecticut, and, of course, why her own Brazilian-Portuguese voiceless consonant VOTs are what they are in Brazil.

There is evidence that individuals are disposed to imitate what they perceive. We know that both infants (Meltzoff & Moore, 1983) and adults (McHugo, Lanzetta, Sullivan, Masters & Englis, 1985) imitate facial expressions, adults unintentionally. As for imitations of speech, as we indicated in the introduction, a number of investigators have reported imitation by infants (Kent & Murray, 1982; Boysson-Bardies, Hallé, Sagart & Durand, 1989; Boysson-Bardies & Vihman, 1991; Whalen, Levitt & Wang, 1991; Vihman & Boysson-Bardies, 1994; Kuhl & Meltzoff, 1996). As for adults, we do not know of evidence that they are disposed to imitate speech, but evidence does show that adults can imitate speech especially fast. This is shown in investigations by Porter & Castellanos (1980) and Porter & Lubker (1980) following earlier work (e.g., Kozhevnikov & Chistovich, 1965; cited in Porter & Castellanos, 1980). Whereas prototypical choice reaction times are slower than simple reaction times, by 100 to 150 ms according to Luce (1986), Porter & Lubker found choice reaction times that exceeded corresponding simple reaction times by as little as 11 ms when choice responses were imitations of vowels. In fact, imitation latencies were so fast that they approached the lower latency limit for motor gestures of this type according to Porter & Lubker (1980).

We point out that a disposition to imitate speech (and an ability to imitate very rapidly) is more readily understandable from the perspective of a theory in which vocal-tract gestures are perceived [i.e., from the perspective of the direct-realist theory (e.g., Fowler, 1986, 1996) or the motor theory (e.g., Liberman & Mattingly, 1985)] than from a theory in which acoustic signals are mapped onto abstract phonological categories. In particular, a theory that listeners perceive gestures allows us to explain phonetic imitations in the same way that imitations of facial expressions are explained. Infants can imitate facial expressions even when they cannot see their own face, presumably because they see what the experimenter's face is *doing*, and, as it were, they instruct their own face to do the same thing. That is, perceiving the experimenter's face is, effectively, receiving instructions for an imitative response. Research by McHugo *et al.* (1985) cited above, showing that adults unintentionally imitate facial expressions that they see on a videotape, may suggest that these perceived instructions serve as goads for imitation. If listeners perceive the linguistically-significant gestures of the speaker, then perceiving speech is, effectively, receiving instructions for its imitation. As for our speaker, she may

hear a particular phase relation between a laryngeal devoicing gesture and an oral constriction gesture for a voiceless stop and, among other things, this percept may serve as a goad for imitation.

We require a second idea, something like Flege's idea of correspondence between phones in a first and second language, to explain why there is parallel drift in VOTs of stops in L1 and L2, specifically, why our speaker's Portuguese voiceless stop VOTs drift toward those of American English voiceless stops when she is in the US, and why her American-English voiceless stop VOTs drift toward those of Brazilian Portuguese when she is in Brazil.

In this idea, the parallel changes to our speaker's Portuguese and English VOTs may signify that she detects a correspondence between L1 [p] and L2 [p^h] and between L1 [t] and L2 [t^h].[5] Parallel gestural drifting may occur in stops in both languages because the speaker detects these correspondences. Gesturally, voiceless stops in the two languages are very similar. Each has an oral-constriction gesture and a laryngeal devoicing gesture. The stops in the two languages differ largely in the phasing of the two gestures, with the devoicing gesture occurring earlier in the Portuguese stops relative to the oral constriction gesture. In terms of Flege's idea of similar phones, we suppose that our speaker has formed a Brazilian-Portuguese /t/ L1 category that permits an authentic L1 production of Portuguese [t] and an accented production of American English [t^h]; however, because the latter is a "similar" phone, it is constrained by or tied to L1 /t/. In this way, her L1 and L2 realizations may both change when the ambient language changes.

If perceived gestures serve as goads for imitation, we have finally to explain the paired observations that the imitations are measurable and reliably present, but that they are markedly different from the speech they imitate. That is, our speaker's VOTs only change by a few milliseconds. Her Portuguese VOTs remain very short during her stays in the US, and her English VOTs remain long when she is in Brazil. Our speculative proposal here draws on research findings suggesting that recent past experience in memory, if it is very recent, exerts a disproportionately stronger impact on current perception and behavior than more distant past experiences (e.g., Bjork & Bjork, 1992). In application to our data, therefore, after our speaker has been in the US for a few months, both her American-English phones and her Brazilian-Portuguese phones are pulled toward the characteristics of the ambient language owing to recency. Even so, she has quantitatively considerably more distant past experience with native productions of Portuguese stops than she has recent experience with English stops. If these past experiences perceiving and producing a voiceless stop also affect production, but individual recent experiences are relatively more potent than individual distant past experiences, we can explain both why recent experience has a measurable effect at all and why the effect is so small.

The foregoing account of our findings of gestural drift is somewhat superficial in nature in calling on a disparate set of findings to explain our data. Can we find deeper principles than a disposition to imitate, phonological correspondence, and preeminence of recency that might underlie them and explain their conjoint relevance to gestural drift? We suspect that deeper principles might be found in the dynamical systems perspective that in recent years has been imported into the psychological domain to explain a variety of findings in the literature on perception and action (e.g., Kugler & Turvey, 1987;

[5] In fact, we cannot, at present, distinguish this possibility from her detecting the more abstract correspondence between voiceless stops in the two languages.

Turvey, 1990; Thelen & Smith, 1994; Port & van Gelder, 1995) including speech (Tuller, Case & Kelso, 1994; Case, Tuller, Ding & Kelso, 1995). Our suggestions along these lines are tentative, but reflect the direction that we believe our theoretical development of gestural drift will take.

Tracking a certain observable characteristic of a system such as a phase relation between two moving articulators [e.g., between two hand-held pendulums functioning as virtual limbs (e.g., Sternad, Amazeen & Turvey, 1996), or, in our case, between an oral constriction gesture and a laryngeal devoicing gesture], permits observation of learning in that system. Change in the state of a biological system over time may be understood in terms of the changes that may occur in its stable or preferred phase relations (initially, its "intrinsic coordinative dynamic") as reflected in a layout of attractors, or a so-called potential function. When a new potential function for a newly learned phase relation is added to an already existing one, the original potential function undergoes change. That is, depending on the nature of the added potential function, stabilities (attractors) in the original potential function may shift, or become more shallow (weaker) or deeper (stronger). In addition, new attractors may be formed and old ones lost.

Zanone & Kelso (1992; see also, Schöner, Zanone & Kelso, 1992) offer a starting point for examining the links among perception, learning, and action and, therefore, gestural drift. In one experiment, Zanone & Kelso (1992) showed what happened when participants learned a 90-degree phase relation between the movements of their two index fingers. Learning this new phase relation changed the original potential function away from the intrinsic coordinative dynamic (attractors at 0 and 180 degrees) and toward the goal of 90 degrees. However, learning the 90-degree phase relation took some practice and was not as easy to produce as the 0 and 180 degree phase relations. This may indicate that there is also an influence from the original intrinsic coordinative dynamic, which may play a role in preventing the new phase relation from appearing immediately.

By analogy, then, for our speaker as a beginning L1 producer, a potential function emerged for [p] (or, equivalently at this time, /p/), for example, reflective of the characteristic phase relation between the lip closing gesture and the laryngeal devoicing gesture found in the ambient-language productions of the Brazilian-Portuguese speakers surrounding her. Subsequently, as an adult, when she learned American-English [p^h], its production was dominated by her Portuguese intrinsic coordinate dynamic, the stronger of the two attractors. That is, for our speaker, [p^h] constituted a new potential function reflecting far less experience than her original intrinsic coordinative dynamic, [p], and so was incorporated into (or added to) the original potential function in close proximity to the attractor for [p]. Together, attractors [p^h] and [p] compose a new intrinsic coordinative dynamic, /p/; they may be said to be in correspondence.[6] The Portuguese VOT phase relation, ostensibly the stronger attractor of the two, reflecting the original potential function, is the influence of L1 on L2, ensuring that L2 productions are accented. In turn, the attractor for the L2 oral constriction/devoicing phase relation influences the L1 attractor merely by its coming into existence. It thus influences the L1 category, the intrinsic coordinative dynamic /p/ of the Portuguese gestural phase relation (remember that, initially, the L1 production [p] and the L1 category /p/ are the same). This is in keeping with the potential function undergoing change with a newly learned phase relation.

[6] In this conceptualization, contrary to Flege's, the phones are produced differently not because of realization rules but because they constitute different potential functions.

Continuing in this hypothetical vein, then, a phonological category, put in dynamical terms, is characterized as a "clumping together" of attractors in a potential function that emerges from adding new potential functions to the original intrinsic coordinative dynamic potential function. Presumably, the original potential function manifests the stronger attractor because of the greatest amount of experience producing and perceiving category members. Clumped attractors emerge as preferred VOT phase relations in a potential function through experience with ambient speech. The fact that our speaker's attractors, by correspondence, may drift together, toward the incoming oral-constriction/devoicing phase relation characteristic of the ambient language environment, suggests that the preferred L1 phase relation may actually constitute a flexible L1 phonological category, /p/, to which all our speaker's subsequent productions of /p/ are tied, thus essentially expanding her category. This dynamical correspondence, from attractor proximity, may underlie the notion of a phonological category.

Subsequent trips back and forth between the two language environments give rise to additional learning that affects oral-constriction/devoicing phasing, as reflected in the potential function, although new attractors are not being formed. Trips back and forth between Brazil and the US mean that our speaker's potential function must repeatedly undergo change, indicating that learning never stops. Further evidence that learning never stops even after the critical period for language acquisition is supplied, for example, by Flege's data, previously described (1987) on native speakers of French and English living, respectively, in Chicago and Paris. With 12 years of immersion in L2, there are larger changes in VOTs, implying larger changes in the potential function, than we see in our speaker. We expect our future research exploring this learning after the critical period to expose further the nature of gestural drift and to test the viability of a dynamical account of its occurrence.

This research was supported by NICHD Grant HD01994 to Haskins Laboratories. We thank Bruce Kay and Michael Turvey for their comments on our General Discussion and Anders Löfqvist for help with data processing.

References

Bjork, R. A. & Bjork, E. L. (1992) A new theory of disuse and an old theory of stimulus fluctuation. In *From learning processes to cognitive processes: essays in honor of William K. Estes II* (A. F. Healy, S. M. Kosslyn, R. M. Shiffrin, eds.), pp. 35–68. Hillsdale, NJ: Lawrence Erlbaum Associates

Boysson-Bardies, B. & Vihman, M. M. (1991) Adaptation to language: evidence from babbling and early words in four languages, *Language*, **61,** 297–319

Boysson-Bardies, B., Hallé, P., Sagart, L. & Durand, C. (1989) A crosslinguistic investigation of vowel formants in babbling, *Journal of Child Language*, **16,** 1–17

Case, P., Tuller, B., Ding, M. & Kelso, J. A. S. (1995) Evaluation of a dynamical model of speech perception, *Perception & Psychophysics*, **57,** 977–988

Flege, J. E. (1987) The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification, *Journal of 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* (W. Strange, ed.), pp. 233–277. Baltimore, MD: York Press

Fowler, C. A. (1986) An event approach to the study of speech perception from a direct-realist perspective, *Journal of Phonetics*, **14,** 3–28

Fowler, C. A. (1996) Listeners do hear sounds not tongues, *Journal of the Acoustical Society of America*, **99,** 1730–1741

Kent, R. D. & Murray, A. D. (1982) Acoustic features of infant vocalic utterances at 3, 6, and 9 months, *Journal of the Acoustical Society of America*, **72,** 353–365

Kozhevnikov, V. A. & Chistovich, L. A. (1965) *Speech articulation and perception.* Washington, D.C.: Joint Publications Research Service
Kugler, P. N. & Turvey, M. T. (1987) *Information, natural law, and the self-assembly of rhythmic movement*, Hillsdale, NJ: Lawrence Erlbaum Associates
Kuhl, P. & Meltzoff, A. (1996) Infant vocalization in response to speech: vocal imitation and developmental change, *Journal of the Acoustical Society of America*, **100,** 2425–2438
Lenneberg, E. H. (1967) *Biological foundations of language.* New York: John Wiley & Sons
Liberman, A. M. & Mattingly, I. G. (1985) The motor theory of speech perception revised, *Cognition*, **21,** 1–36
Luce, R. D. (1986) *Response times: their role in inferring elementary mental organization.* New York: Oxford University Press
Major, R. C. (1992) Losing English as a first language, *The Modern Language Journal*, **76,** 190–208
McHugo, G. J., Lanzetta, J. T., Sullivan, D. G., Masters, R. D. & Englis, B. G. (1985) Emotional reactions to a political leader's expressive displays, *Journal of Personality and Social Psychology*, **49,** 1513–1529
Meltzoff, A. N. & Moore, M. K. (1983) Newborn infants imitate adult facial gestures, *Child Development*, **54,** 702–709
Port, R. & van Gelder, T. (1995) *Mind as motion.* Cambridge, MA: MIT Press
Porter, R. J. & Castellanos, F. X. (1980) Speech production measures of speech perception: rapid shadowing of VCV syllables, *Journal of the Acoustical Society of America*, **67,** 1349–1356
Porter, R. J. & Lubker, J. F. (1980) Rapid reproduction of vowel-vowel sequences: evidence for a fast and direct acoustic-motor linkage in speech, *Journal of Speech and Hearing Research*, **23,** 593–602
Schöner, G., Zanone, P. G. & Kelso, J. A. S. (1992) Learning as change of coordination dynamics: theory and experiment, *Journal of Motor Behavior*, **24,** 29–48
Sternad, D., Amazeen, E. L. & Turvey, M. T. (1996) Diffusive, synaptic, and synergetic coupling: an evaluation through in-phase and antiphase rhythmic movements, *Journal of Motor Behavior*, **28,** 255–269
Thelen, E. & Smith, L. B. (1994) *A dynamic systems approach to the development of cognition and action.* Cambridge: MIT Press
Tuller, B., Case, P. & Kelso, J. A. S. (1994) The nonlinear dynamics of speech categorization, *Journal of Experimental Psychology: Human Perception and Performance*, **20,** 3–16
Turvey, M. T. (1990) Coordination, *American Psychologist*, **45,** 938–953
Vihman, M. M. & Boysson-Bardies, B. (1994) The nature and origins of ambient language influence on infant vocal production and early words, *Phonetica*, **51,** 159–169
Whalen, D. H., Levitt, A. G. & Wang, Q. (1991) Intonational differences between the reduplicative babbling of French- and English-learning infants, *Journal of Child Language*, **18,** 501–516
Zanone, P. G. & Kelso, J. A. S. (1992) Evolution of behavioral attractors with learning: nonequilibrium phase transitions, *Journal of Experimental Psychology: Human Perception and Performance*, **18,** 403–421

Appendix

Translations into Portuguese (with measured stops underlined)

Tell xx, "They don't have bread."
R: Eles não t̲êm p̲ão.

Tell xx, "It's a great cultural event."
R: É um grande acontecimento cultural.

Tell xx, "I have many books."
R: (Eu) t̲enho muitos livros.

Tell xx, "I did not go to New York."
R: Eu não fui a Nova Yorque.

Tell xx, "Brasília is the capital of Brazil."
R: Brasília é a capital do Brasil.

Tell xx, "I go to the theatre."
R: (Eu) vou ao t̲eatro.

Tell xx, “My mother has a good heart.”
R: Minha mãe tem um bom coração.

Tell xx, “We have plenty of time.”
R: Nós temos muito tempo.

Tell xx, “The banana is green.”
R: A banana está verde.

Tell xx, “Peter asked for bread.”
R: Pedro pediu pão.

Tell xx, “What is the capital of Portugal?”
R: Qual é a capital de Portugal?

Tell xx, ‘My name is Ana Luisa Pinto.”
R: Meu nome é Ana Luisa Pinto.

Translations into English (with measured stops underlined)

Diga a Michele, “Fala com o Paulo.”
R: Talk (Speak) to Paul (Paolo).

Diga a Michele, “São três e meia.”
R: It’s three-thirty.

Diga a Michele, “Olha para o mundo.”
R: Look to the world.

Diga a Michele, “A ópera é um acontecimento cultural.”
R: The opera is a cultural event.

Diga a Michele, “A moça está falando.”
R: The girl is talking (speaking).

Diga a Michele, “Informação cultural.”
R: Cultural information.

Diga a Michele, “Fala com o novo chefe.”
R: Talk (speak) to the new boss (chief).

Diga a Michele, “Isto é una pérola.”
R: This is a pearl.

Diga a Michele, “Há trinta e três pessoas na sala de aula.”
R: There are thirty-three people in the classroom.

Diga a Michele, “Compara estes dois poemas.”
R: Compare these two poems.

Diga a Michele, “A moça vive em Nova Yorque.”
R: The girl lives in New York.

Diga a Michele, “O termómetro está sobre a mesa.”
R: The thermometer is on the table.