Imitation as a Basis for Phonetic Learning After the Critical Period
Author(s): Carol A. Fowler

Proceedings of the 25th Annual Meeting of the Berkeley Linguistics Society (2000), pp. 134-145

Imitation as a basis for phonetic learning after the critical period

CAROL A. FOWLER[1]
Haskins Laboratories
University of Connecticut

1. Overview

Imitation by humans of one another is pervasive. I suggest that its occurrence serves two closely related functions. It fosters interpersonal coordination and, hence, cooperation. When coordination occurs between more and less competent participants in a culture, it facilitates cultural learning by the less competent participant (typically an infant or child). More central to present purposes, however, is the observation that, for humans, imitation is easy to do—so easy, in fact, that infants imitate from birth. I suggest that exploration of the reasons why imitation is easy can be illuminating about the nature of the perceptual systems that are engaged by the imitative disposition. Specifically, exploration of the bases for imitation of speech, in the context of an analysis of the bases for imitation generally, tends to confirm claims that the speech perception system, like every other perceptual system, functions to expose the environment to the perceiver. In the case of speech perception, the exposed aspect of the environment includes the phonological gestures of the vocal tract being produced by a talker. Perception of gestures makes reproduction of those gestures easier than it would be were acoustic speech signals themselves objects of speech perception. Finally, I suggest that the imitative disposition, grounded in the perception of gestures, underlies the subtle phonetic learning that occurs among linguistically displaced individuals even when they are well beyond the critical period for language acquisition.

2. Imitation as an index of cooperativity

In his book, *Using Language* (1996), Herbert Clark suggests that public use of language occurs characteristically in the context of "joint activities." Prototypically, these are cooperative activities involving two or more individuals. Research of three general kinds has shown that individuals engaged in cooperative activities in which talk occurs exhibit a variety of forms of entrainment, many of which are imitative.

Condon (e.g., 1976; 1982) coded video tapes of speakers and listeners for the body movements that interlocutors showed while talk occurred. He suggested that speakers move in time with the rhythms of their own speech (exhibiting "self-synchrony") and that listeners, including infants (Condon and Sanders, 1974), move in time with the speech they hear (exhibiting "interactional synchrony"). (See also Bernieri, Reznick, and Rosenthal, 1988.)

In describing his discovery of interactional synchrony, Condon (1982) observed that he had had "an erroneous view of the universe that communication takes place *between* people" (p. 55). Within-person coordination (say, between the jaw and lips in production of /b/; e.g., Kelso, Tuller, Vatikiotis-Bateson and Fowler, 1984; or among the limbs in locomotion) eventuates in logically independent articulators functioning as a single entity ("a special purpose device" according to Fowler and Turvey, 1978). Condon is commenting, likewise, that between-person coordination during talk forges a unitary system (here, a communicative system) out of the interactions of logically independent individuals (see also, Schmidt and O'Brien, 1997).

LaFrance (1982) reported a different form of entrainment. She found "posture mirroring" on the part of individuals listening to a lecturer under conditions in which the listeners later judged themselves to have been engaged by the speaker. Under those conditions, listeners were inclined to match aspects of the speaker's posture. (For example, if the speaker had placed a hand on his neck, members of his audience tended to adopt the same posture.) Interestingly, LaFrance ran cross-lag correlations between measures of posture mirroring at an earlier time and later ratings of engagement, and between measures of engagement at an earlier time and later ratings of posture mirroring. These measures are sometimes used as a way of drawing inferences about causation from correlational data. She found larger correlations from posture mirroring to engagement judgments ($r = .77$) than in the opposite direction ($r = .58$). If the difference is meaningful in terms of causation, it would imply that dispositional posture mirroring may bring about impressions of rapport more than impressions of rapport foster posture mirroring. That is, the imitative tendency may be primary in relation to the between-person coordination that accompanies talk.

Finally, researchers have found evidence of various kinds of "accommodations" (Giles, Coupland and Coupland, 1991) of cooperating interlocutors' speech. These include convergences in dialect (Giles, 1973), in speaking rate (Street, 1983), vocal intensity (Natale, 1975), and rate and average duration of pauses (Jaffe and Feldstein, 1971).

Why should interlocutors exhibit these kinds of convergences? In this literature, investigators write that listeners and speakers entrain to one another (Condon, 1976; McGrath and Kelly, 1986) or that they get on one another's wavelength (LaFrance, 1982). These characterizations suggest an idea that imitations or entrainments reflect an effort on the part of individuals to coordinate as indeed they must do in the course of joint activities. Evidence in favor of this interpretation derives from findings of divergences, for example in dialect, in noncooperative interactions. (See Giles, *et al*, 1991 for a review.)

Despite these clear indications that imitation is associated with interpersonal coordination during talk, evidence shows that imitation is not tied exclusively to social, cooperative settings. It occurs dispositionally in humans even in nonsocial settings as shown by research that I will summarize below. I propose, nonetheless, that the "reason why" humans are disposed to imitate reflects the disposition to cooperate, including dispositions for competent members of cultural communities to initiate less competent members into the culture, and dispositions for less competent members to learn from more competent members. Cooperation, in turn requires coordination, and imitation is, perhaps, the most fundamental or primitive way to coordinate oneself with someone else.

Examination of imitation outside of a social context reveals that imitation is very easy for humans to do. Examining why imitation is easy uncovers a basis for imitation in the universal character of perceptual function.

3. The basis for imitation: perception of distal events

Newborn infants imitate facial gestures. For example, Meltzoff and Moore (1977) showed that, confronted with an adult protruding his tongue, 12 to 21 day old infants were more likely to protrude their own tongue than in the context of an adult opening his mouth. Meltzoff and Moore (1983) extended the findings to infants tested on average 32 hours after birth. The reason that infants imitate may have to do with the coordinative tendencies of humans as just suggested and it may have more specifically to do with one way that infants learn from more competent members of their society (see, e.g., Meltzoff and Moore, 1997). These are reasons why the infant imitates, but the question here is how they pull off an imitation. Their performance in the research by Meltzoff and Moore (see Meltzoff and Moore, 1997 for a review) is quite remarkable. The infants can see the adult's tongue but not their own tongue. The infants can proprioceptively feel their own tongue, but not the adult's tongue. In order to imitate, and protrude their tongue (not, say, their lips, which they do not do when they are confronted with a protruding tongue (Meltzoff and Moore, 1977)), infants have to establish what Meltzoff and Moore (1997) call "organ identification." That is, they have to somehow identify their own tongue with the adult model's tongue. Moreover, they have to establish "gesture identification." They have to know what potential action of their own will match or approximate the action of the adult model. Infants must establish these identifications on the basis of information provided by different sensory modalities about the two tongues and their actions or possible actions.

Meltzoff and Moore (1997) suggest that infants achieve organ identification by establishing a supramodal representation of body parts and their actions. In their view, positing a mediational representation is required on two grounds. Infants can exhibit imitative behavior that lags a model's behavior by up to 24 hours (Meltzoff and Moore, 1994), and imitation is not compulsory. If, lacking a mediational representation, sensory inputs were mapped directly onto motor outputs (as in the findings of Rizzolati and colleagues on "mirror neurons," e.g., Rizzolati and Arbib, 1998), Meltzoff and Moore (1997) argue, delayed imitation would not occur and imitation would be compulsory. (However, in fact, despite the existence of mirror neurons that respond both when a monkey grasps something and when it sees something being grasped, imitation is not elicited by observation.) Because the mediating representation is supramodal, the infants' representations of the two tongues can be determined to correspond; that is, organ identification can occur.

Another way to think about the infants' accomplishment does not invoke the idea of a mediational representation. Rather, it invokes the universal character of perceptual systems. The function of perceptual systems universally (see Gibson, 1966), is to allow perceiver/actors to know their environment. Perceptual systems allow that achievement in one general way. They intercept structure in a medium, light for seeing, air for hearing, etc., and they use the structure, not as something to be perceived itself, but as information for the cause of the structure in the world.

When an infant looks at an adult model protruding his tongue, s/he intercepts light that has been causally structured by reflecting off of the adult's tongue. Generally, different visible objects or events (e.g., protruding tongues or lips) structure light distinctively so that light structure tends to specify its cause. Given a patterning in the light as it changes over time (e.g., as the adult produces the action of tongue protrusion), a perceiver can know from the light structure what event occurred in the world.

Haptic perception works in the same way. As we explore an object with a hand, the object deforms the skin of the hand. Different objects deform the skin differently, and so patternings of skin deformations, like patternings of light structure, can serve as information for their cause. We feel the rigidity of a pen that we hold, or the softness and flexibility of a blanket that we explore. We feel those things because the skin deformations are intercepted by receptors below the skin, and the haptic perceptual system uses the sensed deformations, not as something to be perceived themselves, but as information for their causal source in the world.

Proprioceptive perception is like haptic perception. Proprioceptors in the oral cavity provide information about its structure and composition (e.g., Bosma, 1967; 1970). Infants can use proprioceptive information to know about their own tongue and its possibilities for action.

Accordingly, rather than suggesting that infants develop a supramodal representation of body parts and the actions in which they participate, we can say that infants perceive the adult model's tongue protruding, and they perceive their own tongue. Organ identification occurs because both sensory modalities yield perception of the causal source of stimulation that they intercept. Organ identification allows, but does not elicit imitation. Accordingly, as Meltzoff and Moore (1997) observe, imitation can be delayed or it can fail to occur at all.

Infants are not the only humans who are disposed to imitate facial expressions. Adults do too. McHugo, Lanzetta, Sullivan, Masters and Englis (1985) presented video clips of Ronald Reagan either on the presidential campaign trail or in televised news conferences after his election to participants who differed in their views of Reagan and his political ideology. In one condition of the experiment, the video clips were presented without sound, and they showed Reagan expressing happiness/reassurance, fear/evasion, or anger/threat. Of the viewers, some favored a Reagan presidency and shared his political views. Others opposed a Reagan presidency. Surface electrodes were placed on the corrugator supercilii muscle of the subjects' brow that is active when people frown and on the zygomaticus major muscle of the cheek that is active when people smile. Both groups of subjects exhibited imitation. When Reagan expressed either fear or anger, the corrugator muscle of viewers was more active than when he expressed happiness. When Reagan expressed happiness, the zygomaticus muscle of viewers was more active than when he exhibited fear or anger. Interestingly, these effects did not interact with viewer group. Whether or not viewers had, by self report, approved of Reagan, they exhibited imitation.

This imitation, like that of the infants, involves cross modal matching of information. Viewers saw Reagan's face, but not their own; they felt their own face proprioceptively, but not Reagan's. The reason for their imitations may be the fundamentally social and cooperative nature of humans; the basis for imitation is perception of "distal" events (that is, events in the environment), not "proximal" stimulation (at the sense organ).

Infants also dispositionally imitate speech. There is no reason to suppose that the reason why they can imitate speech or that the reason why they do is different from the reasons why they can and do imitate facial gestures.

Kuhl and Meltzoff (1996) showed infants a videotape of an adult speaker producing an isolated vowel, for different infants, /i/, /a/ or /u/. Infants were 12, 16 or 20 weeks of age. Any vocalizations by the infants were recorded, and cooing vocalizations that did not overlap with the model's speech were analyzed both acoustically and by transcription. At all three ages, infants' vocalizations were influenced by the vowel to which they were exposed. That is, infants exposed to a model producing /i/ themselves produced vowels judged more /i/-like than infants exposed to /a/ or /u/, and likewise for infants exposed to the other two vowels.

In this research by Kuhl and Meltzoff, infants had to establish organ identification. That is, they had to determine that their own vocal-tract articulators corresponded to the bodily causes of the acoustic signals produced by the adult model. And they had to establish action correspondence; they had to determine which actions of their articulators would match or approximate those of the adult model. At least before they produced any sound, the basis for organ identification was cross modal information: auditory information about the model's organs and actions, proprioceptive information about their own. Either, following Meltzoff and Moore, infants established supramodal representations of their organs and those of the model, or, following Gibson, they perceived the actions of the model's articulators and the possible actions ("affordances" in the terminology of Gibson, 1979) of their own corresponding ones.

This latter account especially relates speech perception to the universal function of perceptual systems to which I alluded earlier. Dispositional imitation of speech implies extraction of information about articulation from acoustic speech signals as the motor theory (e.g., Liberman and Mattingly, 1985) and direct realist theory (e.g., Fowler, 1986; 1996) of speech perception have long claimed. If perceptual systems universally extract information from stimulation about the causal sources of stimulation in the environment, then speech perception is not special in perceiving gestures, in contrast to the proposal of motor theorists (e.g., Liberman & Mattingly, 1985). It is, in respect to its function, consistent with that of perceptual systems generally.

Infants have to extract information about articulation from acoustic signals in order to learn to speak. Perhaps, however, having learned to speak, they no longer extract articulatory information from acoustic signals because it has become extraneous. Two findings jointly suggest that extraction of articulatory information continues to occur. Imitation of speech by adults occurs dispositionally (e.g., Goldinger, 1998), and it can occur remarkably fast (Porter and Castellanos, 1980; Porter and Lubker, 1980).

Goldinger (1998) collected tokens of spoken words produced by speakers who read them aloud from their printed forms on a computer screen. The speakers then also produced the same words as shadowing responses to words spoken by someone else. (I will call the shadowed speaker the "model".) The former productions, of course, could not be imitations of the model; subjects had not heard the model's speech yet when they performed the reading task. However, shadowing productions could be imitations if speakers are disposed to imitate. To determine whether imitation occurred in productions of shadowed words, Goldinger constructed an AXB discrimination task. X productions were the model's productions that subjects had shadowed. As (or Bs) were the shadowed productions, and Bs (or As) were words produced by the same speaker who produced the A (B) tokens, but these were reading responses to printed words. Listeners were asked to decide which of A or B was more like X. Listeners reliably chose the shadowing production as more like X showing that, in the shadowing task, subjects had imitated. Imitations of low frequency words were more detectable than those of of high frequency words, and repeated exposure to the to-be-shadowed productions increased the detectability (and so, presumably, the strength) of the imitation tendency. However, imitations occurred even to high frequency words and to word tokens that listeners heard for the first time in the shadowing task.

Evidence that speakers can imitate speech remarkably fast comes from research by Porter and colleagues. In Porter and Lubker's research (1980; see also Porter and Castellanos, 1980), speeded vocal responses were collected from subjects who participated in both simple and choice response time tasks. In a generic simple response time task, subjects make a single detection response indicating that any one of a variety of possible stimuli had occurred. For example, subjects might press a response button having detected a light flash of any color. In a choice reaction time task, subjects make different responses to different stimuli.

They might, for example, press a button with the right hand when a red light flashes and with the left hand when a blue light flashes. Characteristically choice response times are slower by 100 ms or more than are simple response times (Luce, 1986). Following earlier work by Chistovich and colleagues (1962), Porter and Lubker showed that this difference can be eliminated or nearly so when the tasks are speech tasks and the choice task invites imitation.

In their simple response time task, Porter and Lubker presented the speech of a model who began by producing the vowel /a/ for an unpredictable period of time between 2 and 5 seconds. The model then shifted unpredictably to one of three vowels, /o/, /i/ or /æ/. The subjects' task was to produce the /a/ vowel along with the model and to shift to /o/ when the model shifted to /o/, /i/ or /æ/. This is a simple response time task because the subjects' shift to /o/ marked only their detection of a shift in the vowel. The choice response task was identical to the simple task except for the subjects' responses. Now, when the model shifted to /o/ the subject shifted to /o/. When the model shifted to /i/ so did the subject, and likewise for /æ/. That is, the choice task was a shadowing task, possibly invoking imitation. In contrast to prototypical comparisons of simple and choice response latencies, Porter and Lubker found that, on comparable trial types (in which both the model and the subject shifted to /o/), latencies in the simple and choice tasks differed by a nonsignificant 11 ms on average. Moreover both response latencies (169 ms and 180 ms respectively) were in the vicinity of prototypical simple, not choice, response latencies.

What enables choice responses to be so fast? My answer, which is similar to that of Porter and Lubker, is that the choice task invites imitation. If listeners to speech recover vocal tract gestures from acoustic speech signals, then (and only then) to hear the model's productions is to receive articulatory instructions for an imitative response.

In short, Goldinger's (1998) study shows that dispositional imitation occurs among adults even in nonsocial settings; Porter and Lubker's study (1980) shows that imitating (or at least shadowing) is like producing a response in a simple response task, not a choice task. It must be supposed that listeners extract gestural information from acoustic speech signals, and their doing so appears to serve as a goad for an imitative response.

Possibly, the disposition to imitate can help to explain some phonetic learning that takes place even among adults and even when there is no obvious social motivation for the learning.

4. Imitation as a basis for phonetic learning after the critical period for language acquisition.

It is well known that speakers gradually lose the ability to acquire a new language with native fluency. Whether or not they acquire fluency with the syntax of the language, if they learn the language after the "critical period" (e.g., Lenneberg, 1967), they speak with an accent. Even so, it appears that phonetic learning continues to occur over the life span, albeit at a much reduced rate than the rate characteristic of infants and young children. Many of us are aware of examples of people who move from one dialect or language community to another who are judged by their acquaintances "back home" to speak with an accent representative of their new language community. Members of their new community, however, easily detect their original accent. That is, there is an accent migration toward that of the ambient dialect or language, but it is far from complete.

When a shift in language community is across dialect but within language (for example, among English speakers, when a person moves from New England to a southern state of the US, or from the US to England), the reason or reasons for the dialect shift is/are unclear. The shift in accent may reflect an intended effort to affiliate with members of the new language community, or it may reflect an unintended disposition to imitate, regardless of the social benefits or costs of the shift.

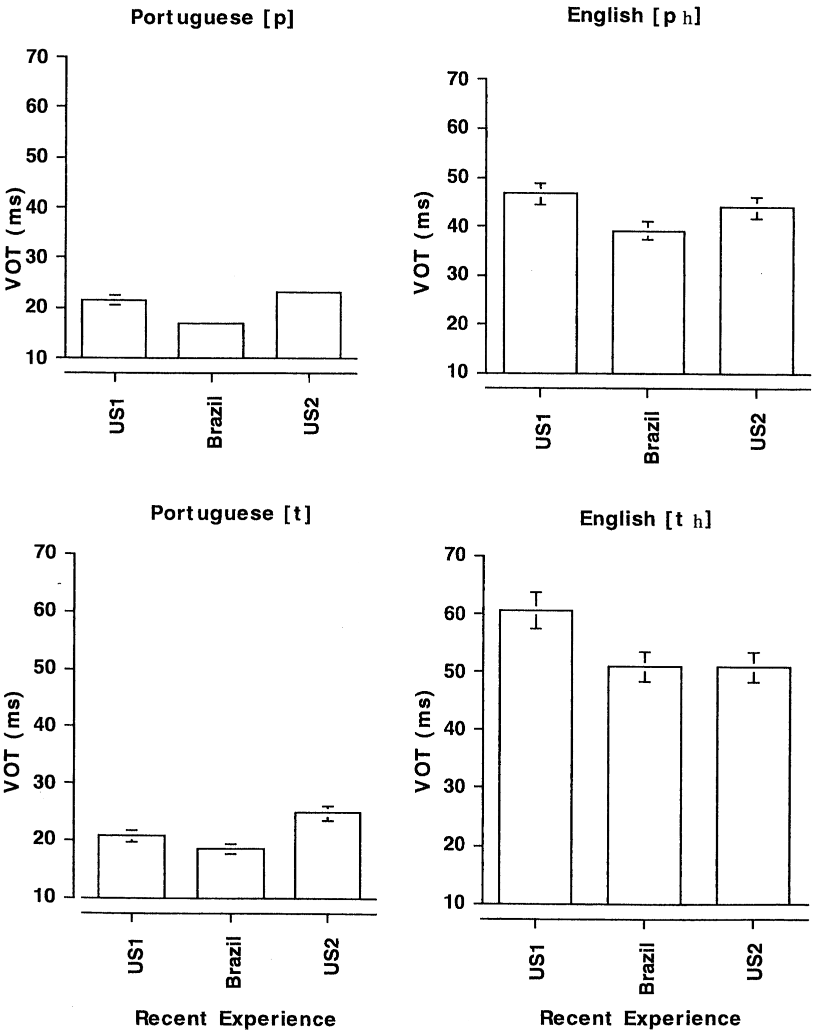
Phonetic learning after the critical period that occurs cross linguistically may be more interpretable. Flege (1987) has shown that native speakers of English who have lived in France for an average of 12 years and who speak French as their principal language, produce English voiceless stops with shorter VOTs than do monolingual speakers of English. Likewise native French speakers who have lived in the Chicago area for an average of 12 years and speak English as their principal language, produce French voiceless stops with VOTs longer than those of monolingual French speakers. The acquisition of French-accented English by native English speakers or of English accented French by native French speakers does not carry any obvious social benefits. Accordingly, it is more likely that the change is an unintended consequence either of the exposure to a different language or of producing a different language or of both sources of effect. A disposition to imitate is a likely source of the phonetic learning.

4.1 An experiment

Recently Michele Sancier and I (1997) examined a case of phonetic learning after the critical period that occurred over a much shorter time span than the 12 years of Flege's subjects. We studied the speech of a native speaker of Brazilian Portuguese who was fluent in English but did not learn English until she was in her teens. At the time of the study, she was 27 years old. She had told us anecdotally that on arriving home in Brazil for a visit, she was told by her father that her speech was "explosive." We inferred that her voiceless unaspirated stops of Portuguese, like those of Flege's French speakers living in Chicago, had become more aspirated, and that is what our subject's father heard as explosiveness.

To test this idea, we recorded our speaker producing sentences of both English and Portuguese that included a number of /p/ and /t/ consonants. We recorded her on three occasions. On two occasions, she had been in the United States, at University of Connecticut, where she was a graduate student, for at least four months. On another occasion (between the two recording sessions just mentioned), she had just returned to the United States from Brazil where she had spent 2.5 months. We measured the VOTs of her voiceless /p/s and /t/s in both her English and her Portuguese speech. The results are shown in (1). Our speakers' VOTs of Portuguese stops were shorter than those of her English stops. Accordingly, the two sets of stops were categorically distinct for this speaker. However, the Portuguese and English VOTs changed in parallel as a function of recent language experience. In both languages, VOTs were longer when the speaker had spent four months in the United States before being recorded than when she had just returned from Brazil after a 10 week stay. Although these effects are small in absolute terms, averaging about 6 ms, they are statistically significant in both languages.

(1) VOTs (and standard error bars) for a native speaker of Brazilian Portuguese producing /p/s and /t/s in both English and Portuguese on three occasions: after a 2.5 month stay in Brazil (Brazil) or after four month stays in the US (US1, US2). (Standard errors are very small for Portuguese /p/, and two are invisible in the graph.)



Why do our speaker's VOTs change at all; why do they change in both languages when the speaker is exposed to (and speaks) just one of the languages; why are the changes so small?

We conclude that our speaker's VOTs change due to the disposition to imitate speech that Goldinger's research has found for adults. That disposition is fostered by the nature of speech perception, in which information about phonetic gestures is extracted from acoustic speech signals. Our speaker hears a different phasing of the oral constriction gesture for /p/ and /t/ with the devoicing gesture when she is in Brazil than when she is in the US. Hearing a different phasing fosters imitation.

As to why English VOTs changed in parallel with Portuguese VOTs when our speaker was in a Portuguese speaking environment and why Portuguese VOTs changed in the English speaking environment, we turn to Flege's (1987) notion of equivalence classifications. As infants are exposed to a language, quite early on, they learn to ignore such irrelevant acoustic differences among tokens of a phonological category as differences in speaker identity (Kuhl, 1980). More generally, they attune themselves to the acoustic distinctions that are contrastive in their language. This attunement may lead second-language (L2) learning individuals to have difficulty establishing, as an independent category, an L2 category that is sufficiently similar to one in the native language (L1). Flege notes that equivalence classification need not imply that the L2 and L1 categories are wholly equated in production, and, indeed, our subject did distinguish her L1 and L2 voiceless stops in VOT.

As to why the changes in VOT are so small, we suppose that our speaker's VOTs reflect a lifetime of exposure to and production of voiceless stops. In some memory research, very recent experience has been found to have a disproportionate effect on memory relative to the effects of more remote experiences (Bjork and Bjork, 1992), and that is why 2.5 to 4 months of exposure to a language can have a measurable effect at all. However, the effect is small because the cumulative effects of our speaker's production and perception experience prior to those 2.5 to 4 months must be much larger than the effects of recent experience. We note that the VOT changes in Flege's research alluded to earlier (1987) were larger than ours.

4.2 A replication

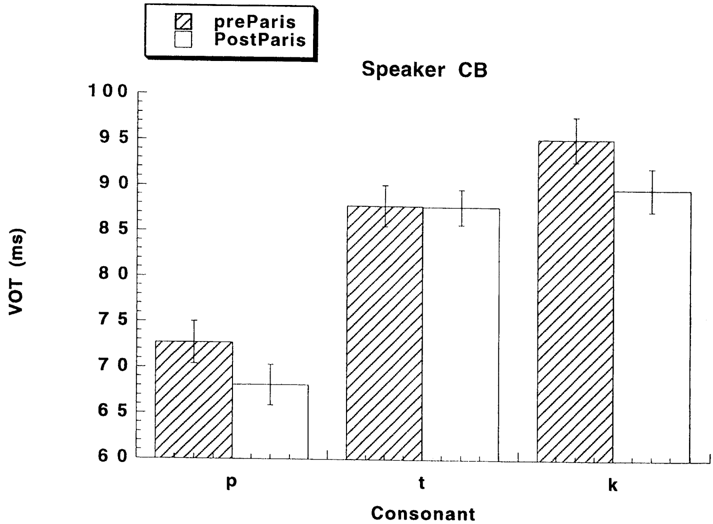
Sancier and I have recently had an opportunity to replicate our earlier experiment. An adult native speaker of English and her children (ages 11 and 14 years) spent several months in Paris. The adult speaker had some fluency in French and largely spoke French while in Paris. Her children had less fluency in French and spent less time speaking French. We recorded all three speakers producing isolated words in both French and English. Words began with /p/, /t/, or /k/. Half were cognates in the two languages; half were phonologically similar, but were not cognates. We recorded the speakers (in fact, they recorded themselves) just after arriving in Paris and shortly before returning.

To date, we have measured the English productions by the adult. (For present purposes, measuring the French productions is of interest mainly to the speaker herself who hoped to improve her French while there. For our purposes, it is less interesting to show that her French VOTs reduced than to show that her English VOTs did so.) We measured this speaker first because she spoke considerably more French and less English than did her children. We are, however, measuring the children's English speech as well. Presumably the children heard considerable

French even if they did not speak much. This may tell us something about effects of passive exposure to new gestural phasings.

The figures in (2) show the overall durations of our speaker's English VOTs for words beginning with /p/, /t/ and /k/. The speaker shows statistically significant reduction of her English VOTs overall (on average by 3 ms); the effects, for unknown reasons, were absent for the consonant /t /.

(2) VOTs (ms) of a native speaker of American English before exposure to a largely French-only language environment and after exposure.



5. Conclusion.

Throughout the lifespan, humans are disposed to imitate one another. Imitation may be a primitive or fundamental kind of entrainment that humans exhibit because of their fundamentally social and cooperative nature. For purposes of understanding speech perception, it is more interesting to focus on the bases for imitation. Imitation requires identification of one's own body parts and their potential actions with corresponding ones of the model to be imitated. Characteristically, the correspondences have to be made cross modally. I have suggested that cross modal identifications that underlie imitation are possible even in infants, because of the universal character of perceptual function. Perceptual systems universally, and including those responsible for speech perception, use information in stimulation at the sense organs as information for their distal source in the world. Perceivers detect the distal source. This makes cross modal "organ identification" easy, because correspondences can be established using a common metric.

Language users exhibit phonetic learning throughout the lifespan, and they exhibit it even when learning is undesirable (as when our native Portuguese speaker developed American accented Portuguese). I suggest that components of the explanation for continuous phonetic learning include the tendency of humans to imitate, coupled with the ease of imitation due to a speech perception that achieves perception of phonetic gestures.

Notes

[1] Preparation of the manuscript was supported by NICHD grant HD-01994 to Haskins Laboratories.

References

Bernieri, F., Reznick, J., Rosenthal, R. (1988). Synchrony, pseudosynchrony and dissynchrony: Measuring the entrainment process in mother-infant interactions. *Journal of Personality and Social Psychology*, **54**, 243-253.

Bjork, R., & Bjork, E. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (pp. 35-68). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Bosma, J. F. (Ed.). (1967). *Symposium on oral sensation and perception.* Springfield, Illinois: Charles C. Thomas.

Bosma, J. F. (Ed.). (1970). *Second symposium on oral sensation and perception.* Springfield, Illinois: Charles C. Thomas.

Chistovich, L., Klass, I., & Kuz'man, I. (1962). The process of speech sound discrimination. *Voprosy psikhologii*, **6**, 26-39. (Cited in Porter and Lubker, 1980.)

Clark, H. (1996). *Using language.* Cambridge: Cambridge University Press.

Condon, W. (1976). An analysis of behavioral organization. *Sign Language Studies*, **13**, 285-318.

Condon, W. S. (1982). Cultural microrhythms. In M. Davis (Eds.), *Interaction rhythms: Periodicity in communicative behavior* (pp. 53-67). New York: Human Sciences Press.

Condon, W. S., & Sanders, L. W. (1974). Neonate movement is synchronzied with adult speech: Interactional participation and language acquisition. *Science*, **183,** 99-101.

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.

Fowler, C. (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.

Fowler, C., & Turvey, M. (1978). Skill acquisition: An event approach with special reference to searching for an optimum of a function of several variables. In G. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 2-40). New York: Academic Press.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J. J. (1966). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, **15**, 87-105.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: communication, context and consequence. In H. Giles, J. Coupland, & N.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, **105**, 251-279.

Jaffe, J., & Feldstein, S. (1970). *Rhythms of dialogue*. New York: Academic Press.

Kelso, J. A. S. , Tuller, B. H., Vatikiotis-Bateson, E. & Fowler, C. A. (1984). Functionally-specific articulatory cooperation following jaw perturbation during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 812-832.

Kuhl, P. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian J. F. Kavanagh C. A. Ferguson (Eds.) *Child phonology, Volume 2: Perception.* (pp. 41-66). New York: Academic Press.

Kuhl, P., & Meltzoff, A. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, **100**, 2425-2438.

LaFrance, M. (1982). Posture mirroring and rapport. In M. Davis (Eds.*)*, *Interaction rhythms: Periodicity in communicative behavior* (pp. 279-292). New York: Human Sciences Press.

Lenneberg, E. (1967). *Biological foundations of language.* New York: Wiley.

Liberman, A., & Mattingly, I. (1985). The motor theory revised. *Cognition,* **21**, 1-36.

Luce, R. D. (1986). *Response times.* New York: Oxford University.

McGrath, J., & Kelly, J. (1986). *Time and human interaction.* New York: The Guilford Press.

McHugo, G., Lanzetta, J., Sullivan, D., Masters, R., & Englis, B. (1985). Emotional reactions to a political leader's expressive displays. *Journal of Personality and Social Psychology*, **49**, 1513-1529.

Meltzoff, A., & Moore, M. (1977). Imitation of facial and manual gestures by human neonates. *Science*, **198,** 75-78.

Meltzoff, A., & Moore, M. (1983). Newborn infants imitate adult facial gestures. *Child Development,* **54**, 702-709.

Meltzoff, A., & Moore, M. (1994). Imitation, memory, and the representation of persons. *Infant Behavior and Development*, **17**, 83-99.

Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting*, **6**, 179-192.

Natale, M. (1975). Social desirability as related to convergence of temporal speehc patterns. *Perceptual and Motor Skills*, **32**, 790-804.

Porter, R., & 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., & Lubker, J. (1980). Rapid reproduction of vowel-vowel sequences: Evidence for a fast and direct acoustic-motoric linkage. *Journal of Speech and Hearing Research* , **23**, 593-602.

Rizzolati, G. & Arbib, M. (1998). Language within our grasp. *Trends in Neuroscience*, **21**, 188-194.

Sancier, M. and Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics,* **25,** 421-436.

Schmidt, R. C. & O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology*, **9**, 189-206.

Street, R. L. (1983). Noncontent speech converges in adult-child interactions. In R. N. Bostrom (Eds.). *Communication yearbook* (pp. 369-395). Beverly Hills, CA: Sage.