Sound Production as a Teacher vs Learner

How do the goos and gas of infancy turn into communicative language? One of the first steps in a child's speech development is learning to produce the sounds of his or her native language. Acquiring this skill depends on a lengthy trial-and-error process that requires a great deal of repetition. The focus of this project is to create a view of how children learn sounds by exposing them to new words in different social contexts, and giving them two chances to repeat and refine their pronunciations of these new sounds. **I'm interested in taking a step into looking at how children can train to produce novel sounds in words they haven't heard before.** How do humans learn to produce new sounds and pronounce new words? And further, does the context in which one learns these new sounds affect the rate of learning?

Repetition in Early Language Learning

Speech production relies on three components, and a deficit in any one of them can lead to difficulty communicating. The first is the development of the motor system; we need to have the muscles and neural connections in place to make the gestures that produce sound.

The second component is imitation; we need to be able to hear and copy the sounds around us. Zebra finches are a great example of this imitation. They learn their song by copying their father's tune through a repetition-based, trial-and-error process. As the finch seeks to improve his song, his basal ganglia produces random signals that slightly alter the tune. These small alterations in each repetition inch the young finch's call slowly toward its stable mature form, and, once it is there, the basal ganglia lends to modulation of song variability dependent on social context (Tumer and Brainard 2007, Kao and Brainard 2006).

The third component is active exploration and production. Can we hear whether our utterances are incorrect and make educated guesses toward correcting them? Zebra finches are able to detect when their tune is incorrect and can alter its pitch accordingly (Andalman & Fee, 2009). As it turns out, human infants learning to speak are not so different from zebra finches learning their song; and infants' babbling is an important bridge towards learning to produce native sounds (Fagan, 2015). In the early stages of babbling (4-6 months old), babies around the world make the same noises in repetitive strings, called canonical babbling (Fagan, 2015). Babies' first babbles are fairly consistent, regardless of their linguistic environment (Robb and Bleile, 1994). However, with experience, infants progress from this undifferentiated babbling to producing sounds that resemble their native tongue. During this process, babbling becomes less repetitive, and infants begin to add new sounds to their phonetic repertoire (Stefanuto and Vallée, 1999).

Speech Perception and Categorization

Historically, some prominent theories of speech perception suggest listeners lean on their knowledge of speech production as a model to parse speech. Motor Theory and Analysis by Synthesis, first touched on by Liberman and Cooper in the 1950s, postulate that speech perception is tied to production through a two-way path: an individual turns sounds into meaning through the intermediary step of identifying and decoding which gestures produce those sounds, rather than by parsing the sounds themselves. In other words, Motor Theory suggests that the motor system has both the role of producing and of detecting and understanding sounds (Liberman et al., 1967).

Support for this theory comes from human categorization of sounds (Liberman et al., 1967). Human infants are initially able to detect and differentiate all phonemes and their relevant gestures. As infants grow older, they become more attuned to their native language, and their perception for non-native speech declines. This coincides with the native-language specialization of babbling (Kuhl et al., 2006). We begin to group acoustically different sounds into a single category, and thus perceive multiple different utterances to be a single sound. For example, when asked to label sounds on the continuum for voiced stops, listeners strictly labeled them as *b*, *d*, or *g*, in well-defined categories, with abrupt shifts between them (Liberman et al., 1957). Looking specifically at /d/, though, we can see that, when the sound is followed different vowel sounds in the English repertoire, it results in utterances contain different pronunciations of the /d/ itself. Nevertheless listeners perceive them all as one category of sound (Liberman et al., 1967). As we become more adept at speaking our own language, we begin to categorize sounds, but as a consequence, we can less easily differentiate and learn new languages and foreign sounds (Kuhl et al., 2006, Kuhl et al., 2014). Even further, most adults better distinguish speech sounds of different phonetic categories than two of the same type because of this tendency toward categorization (Liberman et al., 1967).

Learning New Sounds

Nevertheless, people can, and do, learn to produce new sounds, be it through learning a second language or developmental delays in learning a first language.

Existing literature examines the process by which adults learn a second language. In one study, native Japanese adult speakers were taught to differentiate /r/ and /l/, which are phonetically differentiated in English, but categorically the same in Japanese. Through training, they were able to better perceive the differences between the sounds, but not at a native level.

Learning By Teaching

Separately, studies show that teaching can greatly increase the speed at which students are learning material. The *tutor learning effect*, being that student teachers can learn from their teaching experiences, has been observed when tutors teach others of the same age, of different ages, or reciprocally. When we try to learn information, we do so with certain expectations for how we aim to utilize that information. When students were told they would be expected to teach given material rather than be tested on it, they were better able to freely recall the content they were learning when given a test than students who actually prepared for that exam. But existing studies show the benefit of teaching material knowledge, such as reading, math, and history, not skill (*see review:* Nestojko et al., 2014). Do the beneficial effects of tutoring on teachers extend to skill-based material, such as speech production?

Teaching Speech

Past literature shows that children do, in fact, adjust their speech patterns when they are teaching, depending on the audience. In three related studies, four year olds were asked to explain the workings of a toy to either a two-year-old or an adult. The results showed that, when talking to a two-year-old, the four-year-olds gave shorter utterances, were less inclined to give complex sentences (containing conjunctions, etc.), and more inclined to utter words to grab attention (Shatz & Gelman, 1973). However, this study looks at the level of complete utterances, not phonemes.

Further Questions

In sum, previous studies have found that infants initially learn speech sounds through babbling and repetition, and an infants' perception of non-native speech sounds narrows during the first year of life. It has also, separately, been seen that teaching others leads to better grasp of material than simply learning. However, it is an open question whether teaching also has an effect on skills like non-native sound production.

The current study asks if children learn to produce sounds through repetition at a different rate when teaching versus learning. It looks at whether children achieve better final pronunciations when they're tasked with teaching novel words to someone else, compared to when they're simply repeating the words. And further, it examines if children better able to detect and rectify errors in novel words with English phonology, compared to novel words with novel sounds.

Methods

Participants

30 4- and 5-year-old children (mean age: 4.93, range 4.01-5.85) were recruited from a local children's museum and a university-affiliated nursery school. Ultimately, 7 participants were dropped because they were outside of the target age range or did

not complete the majority of the task. The remaining subjects were randomly assigned to two conditions: Teacher (n = 12) and Learner (n = 11). All participants provided clear assent to participate in the experiment, and their parents provided informed consent in accordance with the requirements of Stanford IRB.

Procedure

All children were told that they would hear and repeat new words in a language called Jabberwocky. Children in the Teacher condition were asked to repeat each word back to a puppet, Sally, who was having trouble learning the language. By contrast, children in the Learner condition were told that they would play a "Say-After-Me" game—that is, they would simply hear each word and repeat it, with the puppet in the room. The environment of the room was matched between conditions.

All children heard eight novel words: four novel, English-sound words and four German words. German words were selected from a sample of common nouns produced by 30-month-old native German-speaking children (Frank et al.). We tried to match word difficulty between the conditions by syllable matching; each condition had three two-syllable words and one one-syllable word. Before the task, all children were given a common English word ("raccoon") as a practice word. They then ran through four words at a time with a break in between. The child repeated each word for two trials, and was then given the option of trying a third time or moving onto the next word. So few children took the option of repeating the word an extra time (only 11 trials total led to an extra repetition), that the data was not included in analysis.

Stimuli

In both conditions, a puppet was present, either as an observer or as a student. The puppet was sitting on the table in front of the child and was introduced as Sally.

The words were compiled into a slide deck. Each slide was visually quite similar, containing a simple blue and green abstract light image, with variations present for the experimenter to keep track of the trial. A bilingual speaker produced the recordings of the words themselves, so the voice for both the novel English and German words matched. Halfway through the task, the children were given a break, as their attention began to wane, and were presented with a short (~11s) video of confetti with cheering noises. At the end of the study the participants were rewarded with a video of Elmo dancing.

In both conditions, the child listened to these words through headphones. This was done in the Teacher condition to explain why Sally could not hear the words for herself, and in the Learner condition to match the Teacher condition.

Coding

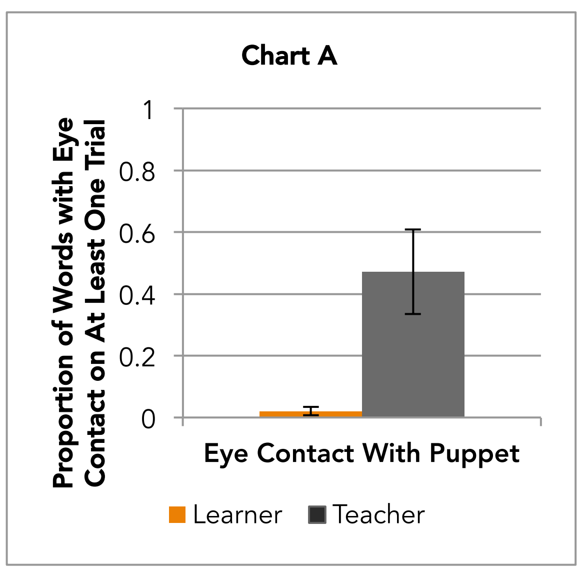
Each trial was scored on a scale from zero to two. Here two meant that, to the coder's ear, the word was pronounced correctly, one meant that the pronunciation had one error, and zero meant the word pronunciation had multiple errors.

Each trial also received a binary eye contact score. If the child was making eye contact with the puppet while saying at least part of the word, the eye contact score was a one; otherwise the score was a zero.

From trial to trial we also measured change and improvement in a binary manner. A one on improvement mean the score increased, a zero meant it decreased or stayed the same, and a negative one meant the score decreased. For change, we included occasions when pronunciation changed, but did not necessarily cause a change in score. If the subject altered their pronunciation (either phonemically or in enunciation), the trial received a one for change, and if the word stayed the same, it received a zero.

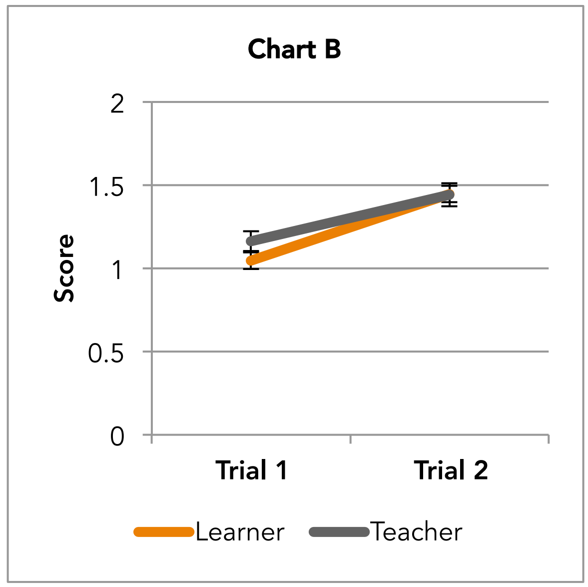
These scores were compiled into change-when-needed and improvement-when-needed, both being binary scales. These look only at instances when the individual got a less than perfect score on trial one. If this was the case and the user changed their utterance, they got a one for change-when-needed, and if they did not they received a zero. Similarly, if the score increased in trial two from trial one when the trial one score was zero or one, improve-when-needed received a score of one and otherwise received a score of zero.

Results

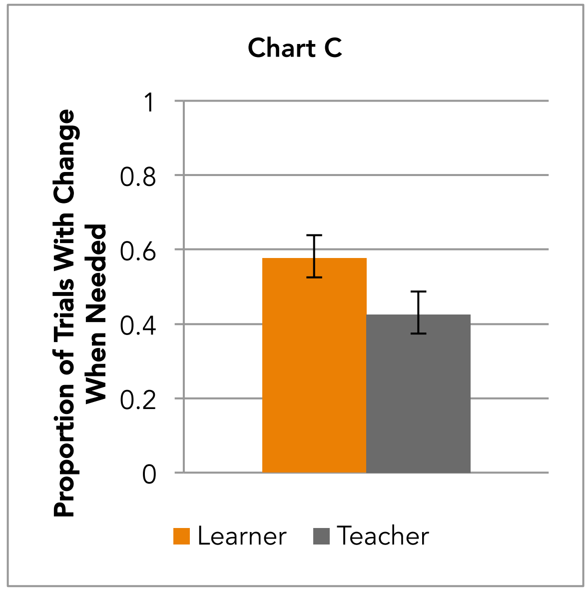


We measured levels of subject engagement in the given task by looking at the amount of eye contact the subject made with the puppet. Chart A shows the proportion of times a child made eye contact with the puppet for at least one of the two trials for any given word, and is sub-divided by condition. The scores show that children in the teacher group (M=0.47, SE=0.14) showed significantly higher levels of eye contact with the puppet, and thus, higher engagement levels by this measure, than those in the learner group (M=0.02, SE=0.01), t(21)=3.4215, p=0.0026.

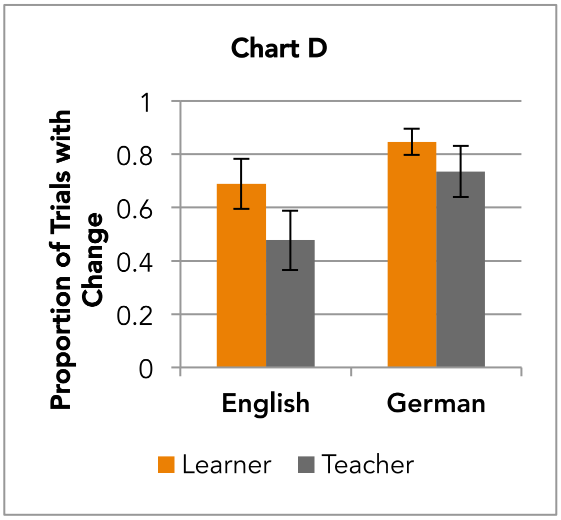
We next looked at the scores on trial one and the scores on trial two for each of the conditions (averaged first by subject and then across all subjects in each condition). Trial one scores for both the Teacher (M=1.05, SE=0.08) and Learner (M=1.45, SE=0.05) conditions were lower than the trial two scores for the Teacher (M=1.16, SE=0.06) and Learner (M=1.44, SE=0.07) conditions. Chart B marks each of these sores in each condition and shows the upward trend in score from trial one to trial two. We see significant change from trial to trial, F(1,42)=28.14, p<0.0001, but no significant differences between conditions F(1,42)=0.83, p=0.3675.

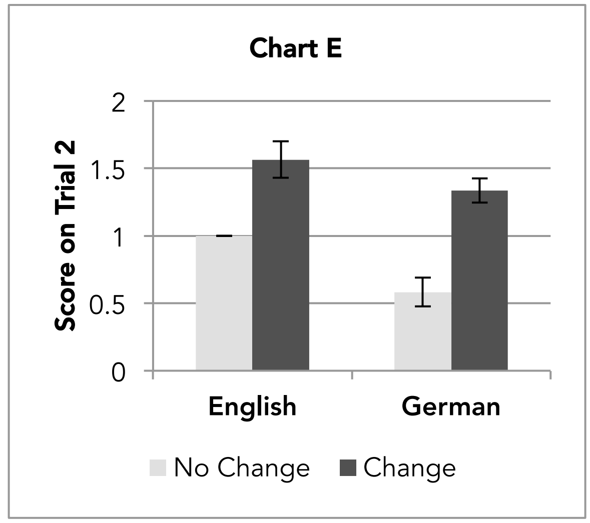


Looking at the subject's ability to self-detect and attempt to correct mistakes, we examined the proportion of times they chose to change their utterance from trial to trial when a change was needed. In other words, if the subject received a less than perfect score on the first trial, did they pronounce the word differently in the second trial. As Chart C illustrates, the subjects in the Learner condition (M=0.57, SE=0.06) changed their pronunciations when needed from trial one to trail two more often than those from Teacher condition (M=0.43, SE=0.05), t(21)=1.8425, p=0.0792.



Finally, to look at the significance of these changes, we looked into the proportion of trials with change when needed (Chart D) and the effect these changes have on scores (Chart E). We see that English and German learners (M=0.69, SE=0.09; M=0.85, SE=0.05) showed higher degrees of change than the English and German teachers (M=0.48, SE=0.11; M=0.73, SE=0.10) respectively, $F(1,41)=3.45$, $p=0.0704$. We also see that overall, subjects attempted more change when it was needed for German words than for English words, $F(1,41)=5.57$, $p=0.0231$. When the subject attempted to change their utterance on trial two in order to correct mistakes from trial one, we see higher trial two scores for both English (M=1.56, SE=0.11) and German (M=1.34, SE=0.09) than we see when no attempt at changes have been made (M=1.00, SE=0.00; M=0.58, SE=0.14). We see that change very strongly indicates an improvement in trial two scores, $F(1,64)=41.72$, $p<0.0001$, and that trial two scores are higher in English words than German words, $F(1,64)=5.84$, $p=0.0185$.





Discussion

The goal of this study was to build on existing research regarding production of native versus non-native speech, and to see if there was any benefit to learning a skill by teaching it over simply learning it. We hypothesized and wanted to confirm that repetition would lead to improvement, but that, given that the conditions had the same number of repetitions and opportunities for trial-and-error, final scores across the conditions would be similar. We believed children would attempt to correct mistakes more often in the teacher condition, as they had greater incentive—doing right by a student—to pronounce a word correctly. Finally, we believed, that children would attempt to correct their mistakes more often for English words than German words, since they would be able to perceive English sounds better, and that corrections would lead toward improvement.

We first verified that the conditions yielded different behavior and showed this by comparing eye contact with the puppet, as a measure of engagement, across the two conditions. We saw that subjects in the teaching condition made eye contact with the puppet nearly half of the time, while learners did so only about 2% of the time. This confirmed differing behavior between conditions.

Then, to examine our hypotheses, we looked at how scores changed from trial to trial and across conditions. We were able to show statistical significance in improvement; from trial one to trial two scores increased on average, showing that, as hypothesized, subjects became better at producing the words with repetition. This matches previous literature regarding babbling as an infant's means of repetition-based trial-and-error in learning to produce speech sounds. However, we did not observe any difference in scores between the two conditions. We observed that teaching versus learning has little impact when it comes to learning a new skill, but we must keep in mind that the subjects were teaching a puppet, not another child. Existing literature focuses on children either actively teaching or preparing to teach other children. It would be interesting to see if our study's results change when we have the subject teach another child.

We then look at correction attempts; literature shows that teachers improve more quickly than learners, so we hypothesized that teachers in our study would choose to attempt to correct their mistakes—by changing their utterance from trial one to trial two—more often than learners (but that they would not necessarily improve at a faster rate given they had the same number of trials). We ended up finding the opposite of this prediction; learners chose to attempt to correct themselves after making a mistake in trial one more often than teachers. One reason for this might be that the subjects in teaching positions had concerns over being perceived as knowledgeable, so they were hesitant to change their utterance from one trial to the next out of fear of diminishing their own authority.

Previous literature shows that individuals are more attuned to native speech and have difficulty perceiving non-native sounds. So, when we further divided the proportion of times the subject chose to change their utterance from trial one to trial two after an incorrect trial one attempt, we believed we would see English having a greater number of corrections than German, simply because the subject would be better able to perceive their own mistake. However, we saw that German word pronunciations saw greater variance between the two trials. This might be because the children are less able to recall what they have heard or said when they are dealing with unfamiliar sounds and the lack of memory leads to variation.

We have seen that subjects choose to attempt to correct their mistakes with repetition, but this alone does not determine how repetition leads to improvement. To look at that, we have examined how the score on the second trial differs when the subject chooses to change their utterance and when they choose not to. We see that there are significantly higher final scores for both languages when subjects chose to change their utterance when needed. This further supports not only that repetition leads to improvement, but also how trial-and-error plays a role. The subjects were able to attempt to say the word, perceive where they were incorrect, and then attempt to fix those mistakes the next time around. Given many repetitions of this process, we can see how an individual might learn how to say a word or, even, speak a language.

It is notable to mention that, while our study focused on a small age range, early piloting was conducted on adults and older children, and both of these populations showed slightly different behaviors. One further avenue of study would be further exploring how this trial-and-error and active exploration method of language learning persists across ages.

Overall, we saw no great evidence for higher rates of improvement or final performance across conditions. Teaching did not show any great benefit over simple learning when it comes to skills rather than the attainment of knowledge or facts. However, we did see further support for the role of repetition and trial-and-error in language learning. And further, we see active, deliberate exploration. Subjects improved from trial to trial, on average, meaning that the changes weren't introduced by random signaling, but rather were educated changed. We see here that humans learn to produce sounds by making attempts, determining what went wrong, and then attempting to correct those mistakes.

References

Andalman, A.S. & Fee, M.S. (2009). A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc Natl Acad Sci USA, 106,* 12518–12523.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. American Educational Research *Journal, 19,* 237–248.

Fagan, M.K. (2006). Why repetition? Repetitive babbling, auditory feedback, and cochlear implantation. *Journal of Experimental Child Psychology, 137,* 125-136.

Frager, S, & Stern C (1970). "Learning by Teaching." *The Reading Teacher, 23,* 403-17.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (in press). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*

Kao, M. H. & Brainard, M. S. Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *J. Neurophysioi, 96,* 1441–1455 (2006)

Kuhl, P.K., Ramírez, R.R., Bosseler A., Lin J. F., & Imada T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proc. Natl. Acad. Sci. U S A, 111,* 11238–11245.

Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci., 9,* F13–F21.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). "Perception of the speech code". *Psychological Review, 74*, 431–461.

Liberman, A.M., Harris, K.S., Hoffman, H.S., & Griffith, B.C. (1957). "The discrimination of speech sounds within and across phoneme boundaries". *Journal of Experimental Psychology, 54,* 358–368

Nestojko, J. F., Bui, D. C., Kornell, N., & Bjork, E. (2014). Expecting to teach enhances learning and organization of knowledge in free recall of text passages. *Memory & Cognition, 42,* 1038–1048.

Shatz, M., & Gelman, R. (1973). The Development of Communication Skills: Modifications in the Speech of Young Children as a Function of Listener. *Monographs of the Society for Research in Child Development, 38,* 1-38.

Stefanuto, M., & Vallée, N. (1999). Consonant systems: From universal trends to ontogenesis. *Proceedings of the 14th International Congress of Phonetic Sciences*, 3, 1973–1976.

Robb M., Bleile K. (1994) Consonant inventories of young children from 8 to 25 months. *Clinical Linguistics and Phonetics, 8*, 295-320.

Tumer, E.C., & Brainard, M.S. (2007). Performance variability enables adaptive plasticity of “crystallized” adult birdsong. *Nature, 450*, 1240–1244.