

Pronunciation in Foreign Language: How to Train?

Effects of different kinds of training in perception and production of the Italian consonant /ʎ/ by adult German learners

Manuela Macedonia^{1,2}

¹Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

²Johannes Kepler University, Linz, Austria

Correspondence: Manuela Macedonia, Max Planck Institute for Human Cognitive and Brain Sciences, Research Group Neural Mechanisms of Human Communication, Leipzig, Germany.

Received: August 17, 2013 Accepted: September 4, 2013 Online Published: October 21, 2013

doi:10.11114/jets.v2i1.205

URL: <http://dx.doi.org/10.11114/jets.v2i1.205>

Abstract

This study investigates the role of perception and sensory motor learning on speech production in L2. Compared to natural language learning, acoustic input in formal adult instruction is deprived of multiple sensory motor cues and lacks the imitation component. Consequently, it is possible that inaccurate pronunciation results from training. Inaccuracy manifests itself in the use of suppletive sounds. For the Italian phoneme /ʎ/ [gl] like in "paglia" (It. straw), native Germans often produce the suppletive phoneme /l/. The Motor Theory of Speech Perception provides theoretical underpinning for the interdependency between perception and production: Thereafter, speech is perceived by reference to the articulator gestures necessary to produce it. Furthermore, imitation is a mechanism driving learning, particularly language acquisition. Accordingly, we hypothesized that training with sensory motor cues together with imitation induces the development of articulatory motor programs. They enable learners to accurately discriminate and pronounce the Italian phoneme /ʎ/. In a between subjects experiment, we trained 49 native Germans to perceive and produce minimal pairs of syllables containing /ʎ/ and /l/ embedded in vocalic contexts. Participants were randomly divided into three subgroups according to the following training conditions: 1) acoustic/imitation (AI), 2) audiovisual/motor task (AVM), and 3) audiovisual/imitation (AVI). The stimuli, consisting of audio files and video clips, were presented in two training blocks totalling 408 stimuli and responses per participant. Responses in stimulus discrimination and reproduction were recorded. The results show that participants discriminated both sounds /ʎ/ and /l/, pre- and post-training equally well. Sound discrimination reached ceiling, independently of the training participants had received. However, training did not improve production accuracy which persisted in being inaccurate until the end of the experiment. We attribute the results in production to insufficient training, and we discuss the findings in terms of age-related resiliency in L2 learning.

Keywords: pronunciation, perception, foreign language, learning, Motor Theory of Speech Perception, imitation, sensorimotor learning

1. Introduction

Learning how to speak a new language as an adult requires, among other skills, the formation of recognition but also the formation of reproduction patterns for the speech sounds specific to that language. It is well known that adult learners are proficient at recognizing sounds of the foreign language but they are not usually able to reproduce them to the same degree of accuracy. Frequently, learners' production of sounds is distorted by interfering patterns from their mother tongue such as category assimilation and perceptual interference (Iverson et al., 2003). Learners accommodate L2 sounds to L1 language phonemic categories (Navarra & Soto-Faraco, 2007). By contrast, in native performance there is no gap between accuracy in sound detection and accuracy in sound production in adult age.

1.1 Is Listening to Sounds Enough to Accurately Reproduce Them?

The difference between native and non-native language in speech production may reflect differences in perception. More specifically, it might be related to the kind of stimuli with which adult learners are provided (Iverson et al, 2005) and additionally to the way in which those stimuli are trained. During L1 acquisition, a

child is flooded with multimodal, i.e., sensorimotor stimuli (Kuhl, 2010). In formal language instruction, learners are trained by means of listening and comprehension activities or silent reading activities. L2 pedagogy holds that overall L2 development is driven by the development of listening skills (Berne, 1998). This view has not changed in practice over the last few years. Thus, foreign language learners, compared with L1 acquirers, often receive an impoverished input lacking visual cues on how sounds are articulated. This must have an impact on natural perception and learning mechanisms. It is well known that visual input plays an important role in speech perception. McGurk & MacDonald (1976) provided evidence for the effects of visual information on sound perception. Visual information can reinforce auditory perception but can also distort it if it does not match with acoustic information (McGurk-effect).

1.2 Hearing and Seeing Articulation: How the Brain Processes Sounds

In the comparison between L1 and L2 learning, another issue is fundamental. In their innate communicative interplay (motherese), caregivers encourage the child to (re) produce sounds, words, and later sentences (Kuhl, 2007). Little by little, L1 learners perceive and articulate sounds of their native language. L2 learners on the other hand play a passive role when perceiving: Learners listen but they are “usually” not required to reproduce what they hear. (Re) Production happens later, when learners attempt to speak in role plays. Thus, there is a prolonged time gap between perception and reproduction and this could influence performance altogether. Interestingly, production also impacts speech perception. In their Motor Theory of Speech Perception, Liberman and Mattingly (1985) claimed that perceiving speech is to perceive articulation, i.e., motor commands and not only acoustic information. Accordingly, if a learner has trained sounds by means of articulation, he/she also perceives them (more) accurately. In a behavioral study by Hazan et al. (2005), Japanese learners of English improved both the perception and the production of consonants through audiovisual perceptual training after being presented with sufficiently salient visual cues. In the last decade, the view that perceptive cues and production of sounds are tightly linked has been shared and also demonstrated in neuroscientific studies (Scott & Johnsrude, 2003). For example, simply viewing lips elicits activity in motor areas connected with sound production in the brain (Nishitani & Hari, 2002). Accordingly, if listeners perceive audiovisual speech, “a motor plan” for the articulation of that sound becomes active in their brains (Skipper, van Wassenhove, Nusbaum, & Small, 2007). In L2 research, Navarra & Soto-Faraco (2007) investigated how viewing lip movement enhances the sensitivity to a particular phonemic contrast in Spanish-Catalan bilinguals differing in their relative dominance of either Spanish or Catalan. They concluded that “visual speech gestures enhance second language perception by way of multisensory integration”. Similar results were achieved in a study by Wang and colleagues (2003) who trained American learners audiovisually to perceive Mandarin tones. In a magnetic resonance imaging experiment, Wilson et al. (2004) made subjects listen passively to monosyllables whilst lying in the scanner. Upon presentation, areas of the brain processing acoustic information became active, but, most interestingly, also motor areas became involved in articulation. Hence, neuroscience has unveiled that speech perception is a sensorimotor process involving auditory and motor areas. Pulvermüller et al. (2006) found evidence for the mapping of single sounds in the motor cortex. They presented participants with syllables including the plosives [p] and [t] that involve lips and tongue in their articulation, respectively. Stimulus presentation activated not only auditory regions in participants’ brains. In fact, portions of the motor cortex specifically involved in actions with lips or tongue also responded to stimuli. In another study by Wilson & Iacoboni (2006), participants listened to native and non-native phonemes. Among the latter, some phonemes were more easily reproducible for English-speakers than others. They elicited activity in auditory brain areas that co-varied with their producibility. On the basis of these results, Wilson & Iacoboni argue that the brain distinguishes between “perceivability” of unfamiliar phonemes and their “producibility”. Accordingly, a learner does not only discriminate between native and non-native phonemes but can also distinguish between perceivable and producible ones. This implies that learners cannot reproduce them all to the same extent, even if they hear them. In fact, immigrants who have spent many years being fully immersed in a foreign country understand the country’s language very well. However, often they cannot reproduce the sounds in a native-like way (Flege, Frieda, & Nozawa, 1997). It seems thus that extensive observation and listening to natives is not enough to create motor programs required for producing the sounds correctly. It is likely—as Buccino et al. (2004) sustain—that actions belonging to the motor repertoire of the observer are mapped onto their motor system, whereas actions not performed by the observer are recognized only on the basis of their visual properties. In an fMRI-study, Buccino and colleagues presented participants with a series of mouth actions performed by humans, monkeys, and dogs, like biting and oral communication (speaking and barking). When participants perceived biting—regardless of the species—they activated motor areas in their brain. This demonstrates that biting was present in subjects’ action repertoire. By contrast, barking did not show the same effect in language regions of the human brain, although it was heard. For these reasons, there must be a difference in perception (performance) and production (performance). It follows, that if a learner does not learn how to produce a

phoneme “correctly”, the phoneme will not enter his or her motor repertoire. The phoneme will only be encoded visually. Hence the sound will be perceivable but not producible. Producibility develops out of motor programmes of articulation that have been learnt. However, if observation of speech is not enough to create functioning motor programs, how can an adult learner acquire good pronunciation in formal instruction?

1.3 Imitation and Phoneme Learning

By observing first language acquisition more closely and comparing it with foreign language learning, a further striking difference emerges: In motherese interactions, babies are not only encouraged by caretakers to reproduce their language, they are also willing to do it and to imitate the sounds and words. In development, innate imitation processes could be the component creating the link between perceivability and producibility (Wilson & Iacoboni, 2006). These processes could initiate articulation programs, making sound reproduction possible. Articulation programs are shaped and reshaped daily during L1 acquisition by means of Hebbian learning (Garagnani, Wennekers, & Pulvermüller, 2009) until the language sound produced matches with the language sound perceived. In fact, when children acquire their first language, they undergo different stages of accurateness in reproduction. Additionally, imitation might release synergies within the language regions. Broca’s area is not only involved in language production. It is one of the core regions of the mirror neuron system, the neurobiological device orchestrating imitation behavior (see for reviews Iacoboni, 2009; Molenberghs, Cunnington, & Mattingley, 2009, and Casile, Caggiano, & Ferrari, 2011). Mirror neurons, originally found in macaque monkeys in the mid-90s, are a special class of neurons that discharge when a person performs an action but also when a person sees this action being performed by somebody else (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). Thus, these neurons mirror action and enable us to understand what another person is doing and/or intends to do (Calvo-Merino, Grezes, Glaser, Passingham, & Haggard, 2006). By “transforming” viewing into action, mirror neurons serve imitation from a neurobiological perspective. According to a recent paper by Glenberg and Gallese (2012), this particular species of brain cells constitute the base of language comprehension, production, and also learning. In this light, the immediate active imitation of what is heard and seen by the learner might represent the bridge towards the creation of the necessary motor programs required for the producibility of the sounds. Thus, imitation could bridge the gap between perception and reproduction.

In summary, literature reviewed here provides evidence that audiovisual presentation of phonemes and their immediate repetition by means of imitation supports learning. Thus, if L2 learners perceive sound, view articulation, and immediately thereafter imitate the sound, they could create a sensorimotor representation of the sound in a native-like manner. This might enhance sound (re) production performance. Hence, the use of a more “natural”, native-like training must include two steps: 1) native-like audiovisual sound perception, i.e., hearing sound and seeing articulation of the sound; 2) immediate and active (re) production by the learner of the perceived articulatory actions and of the sounds.

Here, taking the above considerations into account, we aim to elucidate the following issues:

- a) Does audiovisual presentation compared to acoustic presentation have an impact on perception and discrimination of L2 phonemes, i.e., / λ / vs. /l/ ?
- b) Do reproduction and imitation of audiovisually perceived phonemes / λ / and /l/ lead to accurate pronunciation of the Italian phoneme / λ / ?
- c) Does mouth motor activity unrelated to pronunciation lead to accurate reproduction of the Italian phoneme / λ / to the same extent as imitation?

We hypothesize that audiovisual cues have an impact on discrimination and that immediate reproduction of the audiovisually perceived Italian phoneme / λ / leads to accurate pronunciation of the sound.

2. Method

2.1 Participants

Forty-nine German-speaking subjects (mean age 26.12, 25 males and 24 females) with no previous knowledge of Romance languages participated in sessions 1 and 2 of the experiment. In session 3 (day 60), only 34 subjects participated. Participants were recruited from the Institute’s database and had no known hearing deficits or neurological disorders. Before the experiment, volunteers gave written informed consent to participate and thereafter they were paid for their participation.

2.2 Materials

The training materials comprised minimal pairs of syllables containing the Italian phonemes / λ / and /l/.

/ λ / is a palatal lateral approximant present in most Romance languages. Its articulation involves the middle portion of the tongue that is raised to the hard palate. By contrast, /l/ is a lateral consonant for which the tongue

blocks the airstream in proximity of the alveoles. Thereby, the airstream proceeds along the sides of the tongue and escapes laterally. When observing the articulation of both sounds, the difference is salient. For / λ / not only the tongue movement is different, which might not be well visible, but also the whole face is involved in articulation by stretching the mouth region (Figure 1a and 1b). Faces articulating /l/ or / λ / cannot be confounded, however, it is difficult for learners to discern how the sound is articulated as the tongue is not easily visible. When learning Italian, German natives tend to substitute the sound with /l/, as the sound / λ / is not present in their phonemic inventory.



Figure 1a/b. The male actor is articulating the syllable /gli/ [li] from the frontal and the side perspective.

In this study, both consonants /l/ and / λ / were embedded in vocalic contexts with the structure consonant + vowel or vowel + consonant + vowel like in /la/ / λ a/ /ala/ /a λ a/, where the vowels were /a/, /e/, /i/, /o/, /u/. Altogether, participants were trained on 102 stimuli: 51 containing /l/ and 51 containing / λ /.

For each syllable to be trained, the training stimulus consisted of a video file and an audio file. Video clips had a duration of 800 to 1600 ms. They showed the mouth region of 2 native Italians (male and female) from a frontal and from a side perspective when producing the syllables (Figures 1 and 2). Audio files were extracted from the videos and were identical in all respects to the video files.

2.3 Training Conditions

Participants were assigned randomly to 3 groups and trained according to the following conditions:

1) Acoustic / Imitation (AI) condition (16 participants)

In this condition, participants were cued to listen to the audio files extracted from the videos, i.e., to the syllables pronounced by the actors (male and female); immediately after presentation, participants were asked to reproduce, i.e., to imitate the syllables. The task was formulated: “listen and repeat aloud what you hear”. AI corresponds roughly to what L2 learners do with listening and comprehension activities.

2) Audiovisual / Motor task (AVM) condition (16 participants)

Participants watched the video and listened to the syllables pronounced by the actors (male and female); immediately after presentation participants were asked not to imitate, i.e., not to reproduce the sounds. To ensure that participants were not spontaneously imitating, we controlled for action. Participants were cued to perform a motor task not related to speech, i.e., lip pressing. The task was formulated: “watch the video, listen, and press your lips”. AVM should reconstruct a natural environment in which L2 learning can take place but in which learners do not imitate.

3) Audiovisual / Imitation (AVI) condition (17 participants)

Participants watched the video and listened to the syllables pronounced by the actors (male and female); immediately after presentation, participants were asked to imitate, i.e., to reproduce the sounds, thereby imitating the action necessary to produce it. The task was formulated: “watch the video, listen, and imitate what you hear and what you see”. AVI corresponds to what we suppose is the optimum to learn L2 phonemes at adult age. This procedure is similar to “natural” learning in L1: The child perceives language and action synchronously and is encouraged to imitate and in turn create sensorimotor programs leading to accurate pronunciation.

2.4 Experimental Sessions

On day 1, participants completed a perception test with a non-word repetition task (Barry, Sabisch, Friederici, & Brauer, 2011). Participants were randomly assigned to one of the three training groups following the above training conditions (Group AI, Group AVM, Group AVI).

On Day 2 (after approx. 24 hours), participants performed a perception and a production test.

On Day 3 (after approx. 60 days), after a short ‘refreshing phase’, participants completed a perception and a production test.

2.5 Training Procedures

Participants were first familiarized with the stimuli according to the training condition for 7 minutes. Thereafter, stimuli arranged in minimal pairs like /la// λ a/, /le// λ e/ were presented in 2 training blocks of approximately 20 minutes each, for a total of 408 responses per participant (204 containing /l/ and /204/ containing / λ /).

Stimuli were presented and participants were instructed to press one of two response keys. The left key coded the “familiar” sound /l/, the right key the “new” sound (/ λ /). The experimenter sat in a room next to the lab, monitored the training with a web cam and a head set, and provided feedback online. For a correct response, the experimenter sent a positive smiley symbol, for an incorrect response a negative smiley symbol. Participants could thereby learn to discriminate the two sounds. After discrimination, the stimulus was presented once again and participants were asked to reproduce it depending on the condition in which they were training, i.e., AI, AVM, or AVI. Each speech sound was scored online by the experimenter (an Italian phonetician). Speech production was scored on a scale of 1 to 5, with 1 being the best. Scores were immediately communicated to the participant and sent from the experimenter’s computer to the lab, which appeared on the participant’s screen. This was meant to reinforce good pronunciation and to weaken inaccurate (re) production.

2.6 Perception Testing Procedure

Perception was tested during the first session by collecting the scores generated during the training (log files). Perception during sessions 2 and 3 was assessed by providing the same procedure as in session 1 but omitting the video and the feedback.

2.7 Production Testing Procedure

After the perception task during sessions 2 and 3, participants repeated the syllables they had perceived. Their production was recorded and scored 3 times (2 blind scores) by the same Italian phonetician who trained the participants.

3. Results

3.1 Perception

The mean perception performance was better with audiovisual encoding (AVI and AVM) than with mere acoustic input (AI). Considering the high variability especially in the (AI) condition, we divided the population into low and high proficiency subjects by a median split. Thereafter, we ran a repeated measures ANOVA with training and proficiency as between subject factors. We found a significant main effect of training $.046 F(2,28) = 3.588$. For the combined factor training/proficiency in perception, the result approached significance $F(2,28) = 3.328 p = .051$. This suggests that poor performers were affected by the training they had received and that they had better perception if sensory motor cues were provided.

For the whole population however, training did not impact performance significantly: discrimination by means of sensorimotor input with or without imitation was not better than mere listening to the syllables. Furthermore, decay did not affect results after 60 days. Perception performance stayed constant at ceiling levels.

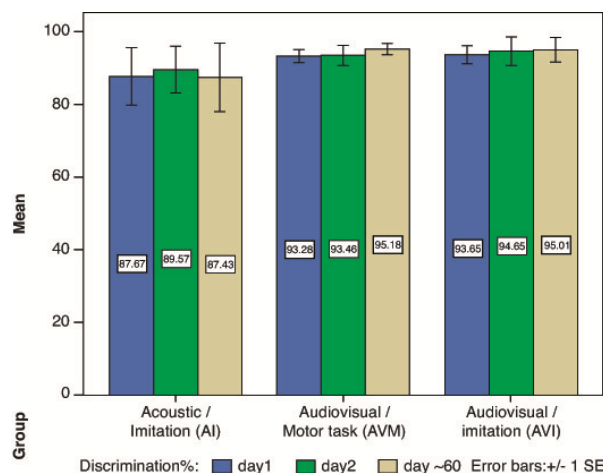


Figure 2. Perception results of all training groups at all test times (day 1, day 2, day ~60).

3.2 Production

All trainings had a similar impact on production. Additionally, variability among subjects did not differ

depending on the training. Our hypothesis that audiovisual input followed by imitation would enhance accuracy in reproduction could not be confirmed either in the short (day 2) or in the long time range (day 60). Instead, we recorded consistent performance decay of time compared to perception. The repeated measures ANOVA showed that the factor time affected performance significantly: $F(1, 33) = 11.2$ $p = .002$ whereas the combined factors $\text{time} \times \text{group}$ and the group factor alone lost significance respectively $F(2, 31) = 0.015$ $p = 0.985$ and $F(2, 31) = 0.015$ $p = .712$.

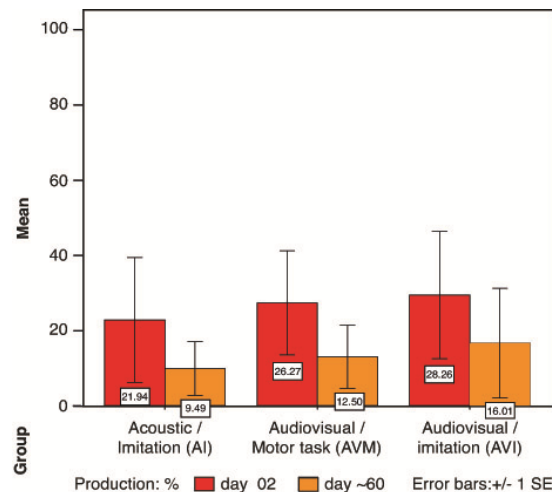


Figure 3. Production results of all training groups tested on day 2 and ~60

3.3 Relationship between Perception and Production

Results comparing perception with production show a huge gap between the two types of performance. Subjects of all groups could discriminate well between /l/ and /ɺ/ with a high degree of accuracy at all times. In contrast, the accuracy in production was very low after training and decayed dramatically after the interval between both tests (60 days). The different kinds of training did not show any effects either for perception or for production.

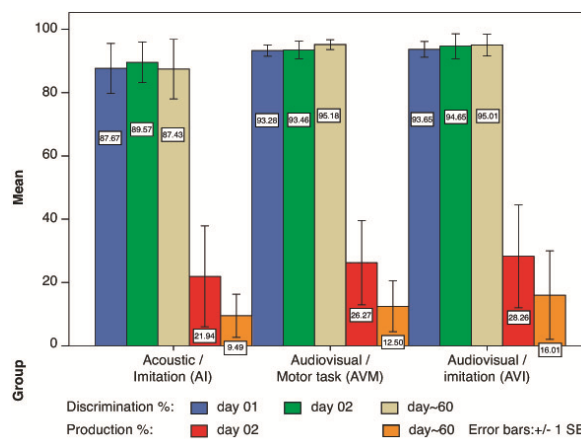


Figure 4. Difference in perception and production performance

3.4 Analysis of Errors on Day 60

On day 2 of the experiment, subjects had a correct production rate of 24.82% (mean). On day 60, it sank to 12.67% (mean). The performance decay was high for all training groups i.e., 87.33% (mean) of the potential production was performed as “error”. Interestingly, although learners were aware of the difference and had correctly discriminated the sounds, they varied error performance and articulated instead:

- the minimal contrast consonant /l/ that was present in perceptual training and perceived with high accuracy. At all time points this substitution represented the main error type;
- the vocal /j:/ or the combination /l/ + /j/ with different lengths of interruption between both sounds. This error is close to /ɺ/ in acoustic perception even to native ears, but natives discriminate both sounds even if the difference is minimal;
- plosives like /k/ and /g/ but also /d/. Plosives combined with aspirations are also present in the error analysis.

This could indicate that the learners do not know how to articulate the sound / λ / and produce their own “creations” of it;

d) the last kind of error is the omission of the consonant [l] and its substitution with silence followed by /j:/. The length of the sound is deprived of its liquid component. This variation of / λ / is also present in central Italian dialects, where /j:/ is an allophone of / λ /.

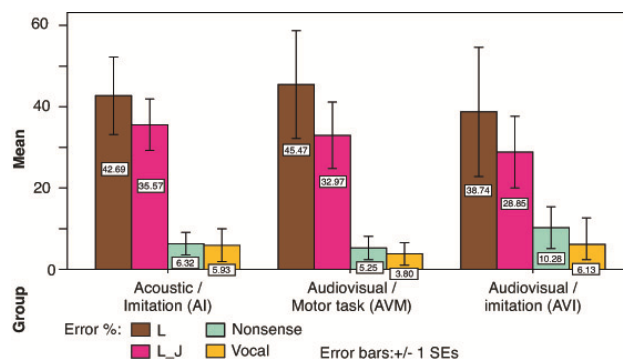


Figure 5. Error analysis on day ~60

Altogether, the results show that perception of the Italian phonemes [λ] and [l] was performed equally well across all training groups at all times with a high degree of accuracy. The sensory motor cues helped low proficiency learners enhance their perception. These results nearly reached significance. The production performance, however, was very poor and no training procedure could lead to significant improvement. This suggests that poor production is not related (only) to poor perception. Training did not reduce the gap between perception and production at any time. The transfer between perception and articulation for the Italian sound / λ / did not occur. In fact, the error analysis makes clear that participants hardly created the new motor program they needed. Instead, they slipped into native language articulation and produced sound substitutions. This provides further support for the motor theory of speech perception. Thereafter the learners would perceive influenced by the motor programs of their first language. Then the learners would reproduce by substituting the target phoneme with similar native sounds. In sum, the data do not confirm our hypothesis, i.e., that audiovisual cues have an impact on discrimination and that immediate reproduction of the audiovisually perceived Italian phoneme / λ / leads to accurate pronunciation of the sound.

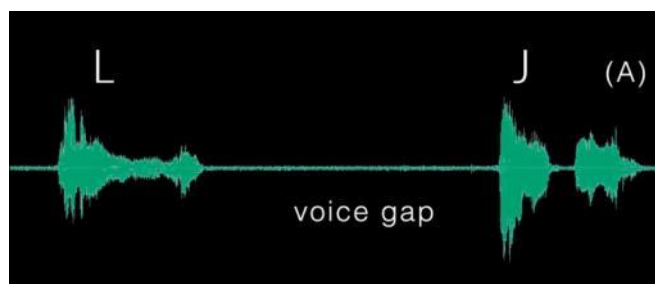


Figure 6. Spectrogram of error [l_ja] articulated instead of [λ a]. The difference is the substitution of [λ] through the phoneme-combination [l] and [j]. A voice gap between [l] and [j] is clearly audible to natives and also visible in the spectrogram.

4. Discussion

This study comparing three different training procedures on adult learners shows that our population was remarkably resilient to visual cues and imitation when learning pronunciation of the phoneme / λ . Several mechanisms affecting L2 pronunciation acquisitions during adult age are well described in literature. In a review article, Piske and colleagues (2001) list extrinsic factors like length of residence in an L2-speaking country and formal instruction that—of course—nurture skill, and intrinsic factors like age of acquisition. The latter seems to be the most important predictor of degree of foreign accent. In a recent study, Archila-Suerte (2012) and colleagues investigate the relationship between age of skill acquisition and proficiency in perception of L2 sounds. The authors attribute L2 pronunciation proficiency to exposure age to non-native sounds. This is not new to professionals dealing with bilingualism (Abrahamsson & Hyltenstam, 2009; Flege et al., 2006; Flege, MacKay, & Meador, 1999; Hopp & Schmid, 2013). From the time when the Italian educator Maria Montessori described "sensitive periods" in human development, it is well known that there are learning phases in which the child

responds particularly well to stimuli provided. In the case of language, Montessori (1953) observed that sensitive periods occur from birth until around six years of age—a view shared by the majority of educators. Most interestingly, neuroscience has been devoted to this topic over the last few years and sensitive periods are also explained in terms of brain development (Dehaene-Lambertz, Hertz-Pannier, & Dubois, 2006). Uylings (2006) reviews several neurobiological factors linked with sensitive periods comprising neurogenesis, cell migration, axonal development, and brain plasticity. The author also considers extrinsic factors that impact brain growth positively in the form of environmental enrichment or negatively through social isolation and neglect. A. Kral (2013) reviews literature on auditory sensitive periods. On this basis, the author describes auditory sensitive periods as developmental phases in which naive cortical networks create patterns of L1 sounds by exposure to L1. Particularly impressive insights also come from studies on brain development investigating white matter maturation. White matter consists of fiber tracts—myelinated axons—embedded in glial cells. In the case of language, these fibers connect Broca's and Wernicke's areas, involved respectively in language production and perception, into a network (Friederici, 2009; Pujol et al., 2006). At birth, these fiber pathways already exist and support certain functions like discrimination between speech and other sounds, and also between different phonemes (Perani et al., 2011). Recently, different pathways have been discovered connecting language core regions (Brauer, Anwander, Perani, & Friederici, 2013): ventral fibers, involved in phonetic and phonological processing are present at birth. Whereas, dorsal fibers, which are associated with syntax functions, mature later. This might explain why newborns (Ramus, Hauser, Miller, Morris, & Mehler, 2000) and infants (Cheour et al., 1998) are already sensitive to L1 sounds and why accurate syntax production occurs later in childhood (Tomasello, 2005). Strikingly, white matter maturation terminates at a certain age of development. Lebel et al. (2008) conducted a longitudinal study with 202 subjects (5-30 yrs) in which they examined the maturation of 10 major fiber tracts in the brain. They found that inferior fronto-occipital fiber pathways connecting language areas reached their peak of maturation at the age of 5. Thereafter, this process slowed down and reached its lowest point around the age of 10. Maturation then proceeded at the same level until the age of 30. This issue could provide evidence for the fact that only very early bilingualism (at approximately age 2) leads to native-like L2 acquisition (Goorhuis-Brouwer & de Bot, 2010) and to accurate phonetic production (Archila-Suerte et al., 2012). In sum, these studies explain sensitive periods in terms of brain development and functions. For our data under discussion, the outcomes from this research provide insights for resilience phenomena in acquisition of native-like L2 pronunciation.

However, language learning is not only a matter of nature. Nurture plays a major role and training is basic in L2 education considering the lack of environmental input if learners are taught L2 in their home countries. Referring back to our results, we assume that the training procedure should be refined. In order to create sensorimotor programs that enable subjects to articulate the target sounds with a maximum of accuracy, we suggest that:

- the frequency of training sessions is increased; in fact, one hour of training is a short time compared to the innumerable hours spent by a child acquiring pronunciation;
- /N/-embedding is extended and varied by adding multisyllabic contexts, thus making training less monotonous;
- perception training is enriched by the variation of the frequent substitution /l_j/ to raise awareness of this error.

Learning accurate pronunciation in adult age might be successful at least in a portion of the learners. However, it must require extensive focussed training and considerable time.

Acknowledgements

I thank Prof. Dr. Angela D. Friederici and Dr. Jonas Obleser for input on different topics related to this study and for helpful discussion on the experiment's results.

This study has been funded by a research grant from the Cogito Foundation Wollerau.

References

- Abrahamsson, N., & Hyltenstam, K. (2009). *Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny*. *Language Learning*, 59(2), 249-306. <http://dx.doi.org/10.1111/j.1467-9922.2009.00507.x>
- Archila-Suerte, P., Zevin, J., Bunta, F., & Hernandez, A. E. (2012). Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Bilingualism*, 15(1), 190-201. <http://dx.doi.org/10.1017/S1366728911000125>
- Barry, J. G., Sabisch, B., Friederici, A. D., & Brauer, J. (2011). Encoding: the keystone to efficient functioning of verbal short-term memory. *Neuropsychologia*, 49(13), 3636-3647. <http://dx.doi.org/10.1016/j.neuropsychologia.2011.09.018>

- Berne, J. E. (1998). Examining the relationship between L2 listening research, pedagogical theory, and practice. *Foreign Language Annals*, 31(2), 169-181.
- Brauer, J., Anwander, A., Perani, D., & Friederici, A. D. (2013). *Dorsal and ventral pathways in language development*. *Brain and Language*. In press. <http://dx.doi.org/10.1016/j.bandl.2013.03.001>
- Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., & Benuzzi, F. (2004). Neural Circuits Involved in the Recognition of Actions Performed by Nonconspicuous: An fMRI Study. *Journal of Cognitive Neuroscience*, 16(1), 114-126. <http://dx.doi.org/10.1162/089892904322755601>
- Calvo-Merino, B., Grezes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Curr Biol*, 16(19), 1905-1910. <http://dx.doi.org/10.1016/j.cub.2006.07.065>
- Casile, A., Caggiano, V., & Ferrari, P. F. (2011). The mirror neuron system: A fresh view. *Neuroscientist*, 17(5), 524-538. <http://dx.doi.org/10.1177/1073858410392239>
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., & Alho, K. (1998). Development of language-specific phoneme representations in the infant brain. *Nat Neurosci*, 1(5), 351-353.
- Dehaene-Lambertz, G., Hertz-Pannier, L., & Dubois, J. (2006). Nature and nurture in language acquisition: anatomical and functional brain-imaging studies in infants. *Trends in Neurosciences*, 29(7), 367-373. <http://dx.doi.org/10.1016/j.tins.2006.05.011>
- Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics*, 34(2), 153-175. <http://dx.doi.org/10.1016/j.wocn.2005.05.001>
- Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, 25(2), 169-186. <http://dx.doi.org/10.1006/jpho.1996.0040>
- Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, 106(5), 2973-2987. <http://dx.doi.org/10.1121/1.428116>
- Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends in Cognitive Sciences*, 13(4), 175-181. <http://dx.doi.org/10.1016/j.tics.2009.01.001>
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(2), 593-609. <http://dx.doi.org/10.1093/brain/119.2.593>
- Garagnani, M., Wennekers, T., & Pulvermüller, F. (2009). Recruitment and consolidation of cell assemblies for words by way of hebbian learning and competition in a multi-layer neural network. *Cognitive Computation*, 1(2), 160-176. <http://dx.doi.org/10.1007/s12559-009-9011-1>
- Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, 48(7), 905-922. <http://dx.doi.org/10.1016/j.cortex.2011.04.010>
- Goorhuis-Brouwer, S., & de Bot, K. (2010). Impact of early English language teaching on L1 and L2 development in children in Dutch schools. *International Journal of Bilingualism*, 14(3), 289-302. <http://dx.doi.org/10.1177/1367006910367846>
- Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47(3), 360-378. <http://dx.doi.org/10.1016/j.specom.2005.04.007>
- Hopp, H., & Schmid, M. S. (2013). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics*, 34(2), 361-394. <http://dx.doi.org/10.1017/S0142716411000737>
- Iacoboni, M. (2009). Imitation, Empathy, and Mirror Neurons. *Annual Review of Psychology*, 60, 653-670. <http://dx.doi.org/10.1146/annurev.psych.60.110707.163604>
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., & Kettermann, A. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1), B47-57. [http://dx.doi.org/10.1016/S0010-0277\(02\)00198-1](http://dx.doi.org/10.1016/S0010-0277(02)00198-1)
- Kral, A. (2013). Auditory critical periods: A review from system's perspective. *Neuroscience*, 247, 117-133.
- Kuhl, P. K. (2007). Cracking the speech code: How infants learn language. *Acoustical Science and Technology*, 28(2), 71-83. <http://dx.doi.org/10.1250/ast.28.71>

- Kuhl, P. K. (2010). Brain Mechanisms in Early Language Acquisition. *Neuron*, 67(5), 713-727. <http://dx.doi.org/10.1016/j.neuron.2010.08.038>
- Lebel, C., Walker, L., Leemans, A., Phillips, L., & Beaulieu, C. (2008). Microstructural maturation of the human brain from childhood to adulthood. *NeuroImage*, 40(3), 1044-1055. <http://dx.doi.org/10.1016/j.neuroimage.2007.12.053>
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1-36. [http://dx.doi.org/10.1016/0010-0277\(85\)90021-6](http://dx.doi.org/10.1016/0010-0277(85)90021-6)
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746-748. <http://dx.doi.org/10.1038/264746a0>
- Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2009). Is the mirror neuron system involved in imitation? A short review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, 33(7), 975-980. <http://dx.doi.org/10.1016/j.neubiorev.2009.03.010>
- Montessori, M. (1953). [The Absorbent Mind.] *La Mente del bambino* (Seconda edizione.) Milano.
- Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *Psychol Res*, 71(1), 4-12. <http://dx.doi.org/10.1007/s00426-005-0031-5>
- Nishitani, N., & Hari, R. (2002). Viewing lip forms: Cortical dynamics. *Neuron*, 36(6), 1211-1220.
- Perani, D., Saccuman, M. C., Scifo, P., Awander, A., Spada, D., & Baldoli, C. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences of the United States of America*, 108(38), 16056-16061. <http://dx.doi.org/10.1073/pnas.1102991108>
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29(2), 191-215. <http://dx.doi.org/10.1006/jpho.2001.0134>
- Pujol, J., Soriano-Mas, C., Ortiz, H., Sebastián-Gallés, N., Losilla, J. M., & Deus, J. (2006). Myelination of language-related areas in the developing brain. *Neurology*, 66(3), 339-343. <http://dx.doi.org/10.1212/01.wnl.0000201049.66073.8d>
- Pulvermüller, F., Huss, M., Kherif, F., Del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, 103(20), 7865-7870. <http://dx.doi.org/10.1073/pnas.0509989103>
- Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288(5464), 349-351. <http://dx.doi.org/10.1126/science.288.5464.349>
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26(2), 100-107. [http://dx.doi.org/10.1016/S0166-2236\(02\)00037-1](http://dx.doi.org/10.1016/S0166-2236(02)00037-1)
- Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex*, 17(10), 2387-2399. <http://dx.doi.org/10.1093/cercor/bhl147>
- Tomasello, M. (2005). *Constructing a language: a usage-based theory of language acquisition*. Cambridge, Mass. London: Harvard University Press.
- Uylings, H. B. M. (2006). Development of the Human Cortex and the Concept of “Critical” or “Sensitive” Periods. *Language Learning*, 56, 59-90. <http://dx.doi.org/10.1111/j.1467-9922.2006.00355.x>
- Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, 113(2), 1033-1043. <http://dx.doi.org/10.1121/1.1531176>
- Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *NeuroImage*, 33(1), 316-325. <http://dx.doi.org/10.1016/j.neuroimage.2006.05.032>
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat Neurosci*, 7(7), 701-702. <http://dx.doi.org/10.1038/nn1263>

