Auditory Perception Influences Speech Motor Learning

Silvia C. Lipski[1], Martine Grice[2], Ingo G. Meister[3]

[1]*Max Planck Institute for Neurological Research Cologne, Lipski@nf.mpg.de*

[2]*IfL-Phonetik, University of Cologne, Martine.Grice@uni-koeln.de*

[3]*Department of Neurology,University Hospital Cologne Ingo.Meister@uk.koeln.de*

Abstract

The present study investigated the effect of auditory training on the acquisition of a new articulatory skill. German speakers were trained to imitate a three-way voicing contrast in intervocalic bilabial plosives. Production training was either accompanied by perceptual training with all three distinctions (full training) or only the native two-way contrast (partial training). We found that full perceptual training improved articulatory performance, and that the improvement in performance was retained during testing three days later. Speakers with only partial training showed no improvement during the main training session; however their articulation improved during the retention task, although the advantage of the group with full auditory training persisted.

The present results indicate that auditory perceptual training plays an important role during early speech motor learning.

1 Introduction

The acquisition of new articulatory gestures has been suggested to be closely related to auditory acuity. According to the model of cortical processing in speech motor learning by [1] auditory targets guide articulatory learning during early stages. The speech motor system receives feedback from a comparison of the auditory target representation with the produced speech sound. The important role of auditory representations for the initial phases of speech motor learning is supported by findings that infants learn to perceive the differences between native phonemes [2] and estimate the relevance of phonetic variation [3] before being able to speak. The essential role of perceptual learning has also been emphasized by research on second language acquisition [4,5].

Our goal was to test the effect of auditory training on the acquisition of a new articulatory skill. Therefore, we systematically varied the amount of perceptual experience with the target speech sound during articulatory training. According to the notion that feedback from auditory representations plays an important role especially in the primary learning stage [1] we predicted that perceptual training enhances the speed of learning and improves the ability to imitate a new speech sound.

Furthermore, we aimed to investigate whether newly acquired articulatory skills are consolidated after a period of rest and sleep. Memory consolidation resulting from sleep has been shown for non-speech motor skills [6] as well as for speech perception [7]. On the basis of these findings we expected that newly acquired articulatory skills could be retained and that speakers could benefit from sleep.

2 Materials and Methods

2.1 Subjects

Thirty female German speakers (mean age = 23.2, range = 20 – 30) were assigned to two groups of equal size for full (Group FT) or partial perceptual training (Group PT) balanced for initial imitation of the new sound, production of native bilabial plosives, perceptual mapping of the new sound to native categories, and age. Subjects were monolingual, and had not begun learning a foreign language before the age of 10 years. Subjects had no or only marginal experience with a language or dialect in which an intervocalic voiceless unaspirated plosive is realized without aspiration.

2.2 Stimuli

Stimuli consisted of disyllabic pseudowords with the target consonant, [b], [p] or [p^{h}] in intervocalic position, and with stress on the first syllable (e.g. ['aba], see Figure 1). Five exemplars for each of the three pseudowords were recorded by a trained male German phonetician.

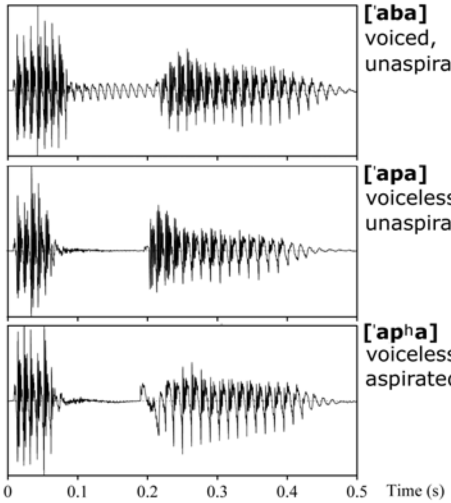

Figure 1: *Examples of stimuli.*

The [aba] stimuli were unaspirated and had a mean of 90.3% of the period of closure containing a periodic signal (range: 74 – 100%). The [apa] stimuli had almost no voicing during closure (mean voicing in closure: 2.34 %, range = 0 – 11.7) and a mean VOT of 6.5 ms, (range: 6.5 – 9.5). The [ap^{h}a] stimuli were completely voiceless with a mean VOT of 26.2 ms (range: 25.1 – 30.8).

German has two voicing categories for plosives. In intervocalic position German voiced [b] is usually realized with partial or complete voicing during closure, voiceless [p^{h}] is produced with no voicing during the closure and a period of aspiration after the release, with a VOT of 30 to 60 ms. Although [p] occurs after sibilants in German, it does not occur intervocalically, and thus constitutes a new sound for German speakers in this context.

2.3 Procedures

Three experimental sessions were conducted. During initial testing, subjects first read ten German words that contained voiced and voiceless intervocalic bilabial plosives. Secondly, subjects categorized [aba], [apa], and [ap^{h}a] according to native voicing categories in a forced choice test with 20 repetitions of each category, presented in random order. Subjects responded by pressing buttons labelled “ABA” or “APA”. Finally, in the imitation task subjects repeated the three stimulus categories, each presented 12 times in random order, one by one, three seconds apart.

Between one and four weeks after initial testing, subjects underwent the training session. This included seven blocks of a perceptual discrimination task and an seven blocks of an imitation task. Perceptual discrimination consisted of an XAB paradigm. Each trial was a sequence of three stimuli (60 sequences per block). The order of sequences and stimuli varied randomly. Subjects decided whether the first item (X) resembled the second or the third one. Group PT discriminated only sequences including [aba] and [ap^{h}a]. Group FT discriminated two types of sequences either including [aba] and [apa] or [apa] and [ap^{h}a]. Thus, during perceptual training Group PT attended to the already familiar contrast. In contrast, subjects in Group FT had to focus on the acoustic differences between the new plosive [p] and native intervocalic [b] or [p^{h}]. The articulatory imitation task was the same as for the initial test. The imitation and perception tasks alternated during the training session.

Three days after the training session, subjects returned for a test of retention. Here, they performed an imitation test equal to the initial test and the production training blocks in the prior session. No perceptual training was administered during retention testing. Subjects were unaware of the number of stimulus types throughout the whole experiment.

2.4 Recording and analysis

Recordings took place in a sound-attenuated room at the Max-Planck Institute for Neurological Research with an AKG C520 headset condenser microphone and M-Audio Fast Track sound card connected to a Dell PC (SR = 44.1 Hz, 16-bit resolution). Stimuli were presented over Sennheiser HD 215 headphones. Praat 5.0.30 [8] was used for Analysis. VOT and voicing during closure were measured from three blocks of testing each including 12 items per stimulus category per subject: initial screening, last block of training, and retention three days after training. VOT was labelled starting at the onset of the oral release until the zero crossing after the first glottal pulse of the following vowel. Closure was labelled from the zero crossing after the final glottal pulse of the first vowel to the onset of the burst. Voicing during closure was measured as the fraction of locally unvoiced pitch frames (length = 10 ms) based on the periodicity detection on the basis of an accurate autocorrelation algorithm [9] implemented in Praat. A relation between voicing during closure and VOT for [p] was tested by using Pearson's product moment correlation coefficient (Spearman's rho). Tests were carried out for each block (initial, last block, retention) and showed that values were not correlated (all p > 0.07). Therefore, VOT and voicing during closure were analysed in separate repeated-measures ANOVA with group (FT, PT) and testing block (initial, last block of training, retention) as independent variables. Bonferroni's corrected pairwise comparisons were performed to test for differences in the individual means.

3 Results

Intervocalic bilabial voiceless plosives in ten native words were produced with a mean of 44.2 ms ±9.2 VOT by Group PT and 45.8 ms ±11.5 for Group FT (T-test: p > 0.1). Intervocalic [b] in the ten German words were produced with a mean of 29.6 % ±19.2 voicing during closure for Group PT[1] and 19.2 % ±13.2 for Group FT, (T-test: p > 0.05). Perceptually, [apa] was categorized as the native voiced plosive 77.9 % ±22 of the times by Group PT and 72.5 % ±20 by Group FT.

Figure 2 displays VOT and voicing values for the new sound [apa] that both groups produced initially, during training, and during retention three days after training. Articulatory and perceptual training both affected the production of the new sound, intervocalic [p]: For VOT, a main effect of testing block was found (F = 8.42; p < 0.001), as well as a significant interaction between block and group (F = 3.34; p < 0.05). Pairwise comparisons revealed that Group FT improved during training: VOTs for [apa] were closer to the target values in the last block of training as compared to initial performance (p < 0.0001). Group FT showed no change in performance during the retention task (p = 0.6).

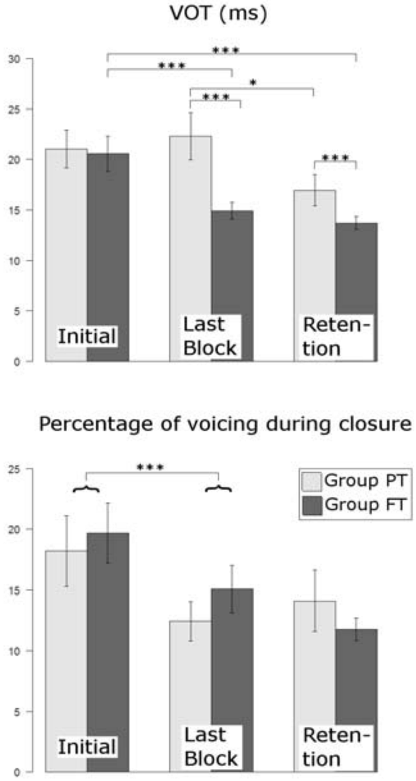

Figure 2: *Mean VOT (top) and percentage of voicing during closure (bottom) with standard error bars for intervocalic [p] (* for p<0.05, *** for p<0.0001).*

[1] One subject in Group PT produced [b] without closure as a glottalized vowel. This subject's data were not included for analysis of native word production

For Group PT no change emerged between the initial and the last block of training (p = 1), indicating that no improvement occurred during training. However, their production improved in the retention task (p < 0.05). VOT values for Group FT (with full perceptual training) were significantly better than those for Group PT (with only partial training) in the last block of the training session (p<0.0001) as well as in the retention test (p<0.001).

Voicing during closure came closer to the target for both groups after training, indicated by a main effect of block (F= 5.14, p < 0.01). No group difference was found. Comparison between individual blocks for both groups pooled showed that speakers improved from the initial test to the end of training (Initial vs. Last block: p < 0.0001, Initial vs. Retention: p < 0.0001). The same level of performance was evident in the retention test (Last block vs. Retention: p = 1).

4 Discussion and Conclusion

The present study shows that the combination of perceptual and articulatory training was beneficial for learning to articulate a new sound. Learners without full perceptual training improved voicing during closure but they did not correct their VOT production by the end of the training session whereas the group with full auditory training improved both parameters considerably. This suggests that perceptual training facilitates articulatory accuracy during initial practice. Our results are in accordance with previous studies on the influence of perception on production, e.g. [5]. By comparing two groups with partial and with full perceptual training we could directly demonstrate the effect of auditory training.

These findings are consistent with the framework proposed by [1] stating that auditory target representations function as a reference for speech motor control during initial stages.

Furthermore, our results suggest that early perceptual training has a lasting effect. Three days after the training session the performance of speakers with partial auditory training had improved. However, it was still inferior to the productions of speakers with complete training who could retain the articulatory improvements they had made during the training session.

The consolidating effect of sleep has been shown for memory, perceptual, and non-speech motor tasks [6,7]. The present results show that new articulatory skills are retained and improve after sleep. Even if learning conditions were suboptimal as in the case of subjects with only partial perceptual training, articulatory performance profited from sleep.

In summary, the results of this study indicate the crucial role of early auditory learning for the establishment of accurate articulatory representations.

5 References

[1] F.H. Guenther. Cortical interactions underlying the production of speech sounds. *J Commun Disord*, 39:350-65, 2006.

[2] J.F. Werker, R.C. Tees. Developmental changes across childhood in the perception of non-native speech sounds. *Can J Psychol*, 37:278-86, 1983.

[3] C. Dietrich, D. Swingley, J.F. Werker. Native language governs interpretation of salient speech sound differences at 18 months. *Proc Natl Acad Sci U S A*, 104:16027-31, 2007.

[4] J.E. Flege. Second language speech learning: Theory, findings, and problems. In: W. Strange, ed. Speech perception and linguistic experience: Issues in cross-language research. Baltimore, MD: York Press, 1995: 233-73.

[5] Y. Wang, A. Jongman, J.A. Sereno. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *J Acoust Soc Am*, 113:1033-43, 2003.

[6] M.P. Walker, T. Brakefield, J. Seidman, et al. Sleep and the time course of motor skill learning. *Learn Mem*, 10:275-84, 2003.

[7] K.M. Fenn, H.C. Nusbaum, D. Margoliash. Consolidation during sleep of perceptual learning of spoken language. *Nature*, 425:614-16, 2003.

[8] P. Boersma, D. Weenink. Praat: doing phonetics by computer. 5.0.30 edn, 2007.

[9] P. Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam*, 17:13, 1993.