



Perception and Production of Vowel Contrasts in German Learners of English

Helena Levy

GRK „Frequency effects in language“, University of Freiburg

helena.levy@frequenz.uni-freiburg.de

Abstract

This paper discusses the relationship between the perception and the production of vowel contrasts in German learners of English. The underlying study investigates whether German subjects' ability to distinguish between the English front vowels /e/ and /æ/ on the level of production improved after training only perception. The main hypothesis was that perception training would improve production. One perception test and one production task were followed by a session of perception training during which the subjects did not speak. After the training, the two modalities were tested again. Results supported the hypothesis, maintaining that the production of the intended contrast improved after training. Perception, however, remained almost the same. Overall results suggest that training perception in an EFL (English as a foreign language) context may have positive effects on pronunciation.

Index Terms: speech perception, perception-production link, second language acquisition (SLA)

1. Introduction

The link between perception and production has been of great interest in various areas of linguistic research and there is an ongoing debate on how the two might be related. When we produce sounds we automatically listen to them, which is why a connection seems obvious. One would assume that perceiving two sounds as different is a necessary precondition to actually produce them. However, such a link still has not been proven without doubt [1, p. 273]. We do not know if being able to accurately perceive a contrast between two vowels really helps in producing this contrast, or whether training perception can improve a learner's ability in pronouncing sounds of the target language. In the area of child language research, studies have reported "a positive correlation between speech sound discrimination and articulation ability – especially between poor discrimination and misarticulation [...]" [2, p. 1]. Furthermore, experiments on sensorimotor integration showed that lexical effects in perception resulting from altered auditory feedback lead to a change in production [3]. Similarly, in SLA research, there are theories that claim that perception is a prerequisite for the production of the sounds of a second language [4]. If such a sequence could be proven, it would be beneficial both for language teachers and for the theoretical knowledge about language learning [5].

2. Theory and related work

Studies concerned with the mutual influence of auditory perception and speech production have been able to show that speakers who produce a contrast more accurately also perceive

it more distinctively [6] and that sounds are recognized better if they have been produced with one's own motor system as compared to sounds that have simply been perceived [7]. Also, spoken words tend to be remembered better than words that are mouthed without sound [8], or listened to without movement [9]. In SLA research there have been studies that show how perceptual learning improves with length of residence and with the amount of L2 usage [10]. However, it has been difficult to prove whether improvement really was due to pronunciation training or simply due to interaction in an L2 environment. Eskenazi [11] proposed that listening to minimal pairs might lead to an improvement of both perception and production, and several other studies examine the effect of perceptual training on how well learners can distinguish between vowels [12], [13]. The overall results show that the ability to perceive differences between trained vowel pairs generally improved. Strange [10] reports that Japanese learners of English were successful in perceiving the difference between /l/ and /r/ after training perception only. However, there was no improvement in production.

The central question that arises from these research findings is "whether perceptual training can affect production, so that training efforts could result in positive transfer from one modality to the other" [13, p. 1034]. L2 sounds that do not exist in the speaker's L1 are often easier to perceive (and thus easier to categorize) than sounds or sound contrasts that are very similar to those of the L1 system [14, p. 203]. However, the categories a speaker might then build for the new sounds are still rather different from a native speaker's category. Flege explains this difference by saying that speakers try to preserve the contrasts they acquired. He gives evidence for the fact that categories for L1 and categories for L2 sounds "exist in a common phonological space" [4, p. 242]. In learners, an L2 vowel may be "deflected" away from an L1 vowel and so differ from a native speaker's category for the L2 vowel sound" (ibid). In the context of his Speech Learning Model (SLM), Flege calls this process 'category dissimilation' [15, p. 470], as opposed to 'category assimilation' which suggests that learners who do not manage to perceive a difference between an L1 sound and a similar L2 sound will ascribe an L2 sound to an existing L1 category or create one single category for two L2 sounds that are wrongly perceived as the same. In a learner who is not able to perceive two sounds as phonetically distinct, category formation will be blocked, which might result in the inability to differentiate between two sounds, also on the level of production. From this assumption follows another thesis: "perception leads production in L2 speech learning" [16, p. 2]. This means that an L2 sound can never be produced more accurately than it is perceived (and represented in the mind of the learner). There are other theories that suggest it is the other way around and production leads perception or that they are simultaneous (e.g. Best &

10.21437/Interspeech.2015-253

Tyler's Perceptual Assimilation Model (PAM) [17]). Logically, there are four scenarios: 1) learners neither perceive nor produce the L2 contrast; 2) learners are able to both perceive and produce the target contrast; 3) a contrast is perceived but not produced correctly; 4) the contrast is evident in production but not in perception [18, p. 82].

3. Study: training perception and improving perception and production

German speakers often seem to have problems in making a clear distinction between the two English front vowels /e/ and /æ/, /æ/ being a phoneme that does not exist in standard High German and thus leading learners to wrongly assume allophonic contexts or produce a vowel that lies between the two English phonemes. Generally, English native speakers' position for /æ/ in the oral cavity is significantly lower than in German learners [19]. The goal of this study was to determine if the subjects' level of production improved after training only perception. Two hypotheses were formulated:

- (1) German learners of English who have problems perceiving a contrast between the vowels /æ/ and /e/ will also have difficulties in pronouncing them distinctly.
- (2) Training only the perception of this vowel contrast will lead to an improvement of the same learners' distinction in the production of the given vowels.

3.1. Methodology

15 students (8 female, mean age 16.5) participated in the study. All of them had studied English for at least six years. Subjects took four tests (two pre-tests and two post-tests) and received perception training before taking the post-tests. The pre-tests consisted of one perception test and one recording of each student. Then there was a 30-minute computerized perception training session, followed by the two post-tests. Subjects were recorded with Edirol R-09HR High-Resolution WAVE/MP3 Recorders (24-bit/96kHz). Stimuli for the perception tests were recordings of two native speakers of Standard British English, one female and one male.

3.1.1. Pre-test 1: Production task

Subjects were instructed to read out a list of 30 monosyllabic English words with a CVC structure. 16 of the words contained either /æ/ (8) or /e/ (8), eight of which formed minimal pairs, such as <dead> and <dad>. The other eight only contained the said vowels without forming minimal pairs (e.g. <sell>, <ham>). Directly after having read the word list, subjects read out a short text, which was later used to compare words spoken in isolation and in connected speech. The experiment started with a production task and not with a perception test because hearing the English sounds in the perception test might have an influence on production.

3.1.2. Pre-test 2: Perception task

After the production task, participants completed a computer based identification test in which they had to assign perceived vowels to one word of a minimal pair. Tokens were presented to the participants over headphones. The test consisted of fifty words including six non-words, as well as distracters. Non-words were used to test whether categorical perception would be different where the participants' attention might be focused on acoustic rather than lexical differences. Non-words

sometimes enhance the ability to perceive L2 sounds from different categories [20].

3.1.3. Perception training

Perception training consisted of a 30-minute computer-based self-study session. Subjects saw a list of 70 minimal pairs (including distracters) and were instructed to look at the orthographic realizations of the members of a minimal pair and then listen to each of them in turn by clicking on the words. They could click on the same word repeatedly and choose the sequence for themselves in order to compare vowels and improve their ability to distinguish between /æ/ and /e/. Immediately after the training session, two post-tests were conducted using the same methods and technology as the pre-tests with partially different stimuli.

3.2. Analysis

3.2.1. Analysis of the production tasks

The recordings were entered into PRAAT (42 tokens_15 speakers = 630 tokens; 51 were eliminated due to insufficient recording quality). Formant values for F1, F2 and F3 were determined from the most stable point in the middle of the vowel. All formant values (in Hz) were then copied into a NORM-form in tab-delimited text files [21]. Bark values were used in all instances where different speakers' vowel productions are compared, which is why F3 values had to be integrated.

3.2.2. Analysis of the perception tests

Test results were compared and examined in terms of the following parameters: (1) changes in individual performance of each student (difference between perception pre-test and perception post-test), (2) difference between perception pre-test and perception post-test for the entire group and (3) correlation with production results.

3.3. Predictions

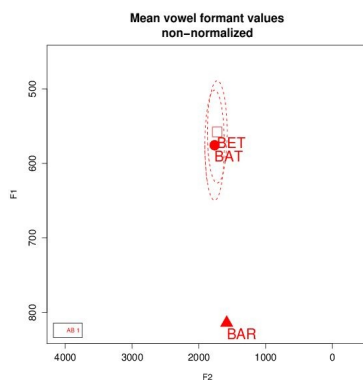
According to the learning models SLM [4] and PAM-L2 [22] there is a strong likelihood that the English vowel /e/ will be classified as 'similar' to the German vowel /ɛ/ by native speakers of German and that learners do not create a different category for this vowel. Subjects would thus transfer the German vowel category for /ɛ/ into English and would not produce the English vowel /e/ in a native-like way. PAM predicts that if listeners perceptually assimilate two L2 sounds in one phonetic category, they will be perceived and pronounced as homophones. This is predicted for the vowel /æ/, which will likely be perceived and produced close to English /e/. Therefore, F1 and F2 values are predicted to be similar for /æ/ and /e/, reflecting the inability to differentiate between the two vowels. It is assumed that greater perceptual accuracy can be achieved through training, and that this will eventually lead to changes in production. With reference to studies that have shown the effectiveness of perceptual training [13] [16], it is predicted that participants will score higher in both the perception and the production post-test. This is due to new categories for the English vowels that will have been established during perceptual training. The distance between produced vowels is predicted to be greater in the post-test, and development of a category for /æ/ separate from /e/ will be reflected in higher F1 values. However, it is also

possible that only one area, perception or production, improves significantly. From such one-sided results it could be followed that one area, perception or production, leads the other in phonological acquisition.

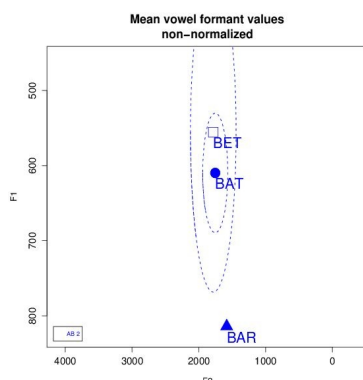
3.4. Results

3.4.1. Results of the production tasks

Of the fifteen speakers, nine showed an improvement in the discrimination of the two target vowels. Graphs 1 and 2 exemplify the difference between the two vowels' formant frequency values in the pre-test as opposed to the post-test in one speaker. The vowel /a:/, produced by a native speaker of Standard British English, was inserted as a reference point (BET, BAT and BAR represent the vowels /e/, /æ/ and /a:/). The graphs show that mean formant values and their range of dispersion move away from each other as the result of differences in articulation. As predicted, the vowels were produced more distinctly, with values for F1 increasing for the vowel /æ/. A higher F1 mirrors increased openness in the production of the vowel, and thus /æ/ is produced closer to its actual position in the English vowel chart, /æ/ being an open front vowel. This can also be seen in the distance to the reference vowel, with /æ/ now slightly approaching /a:/.



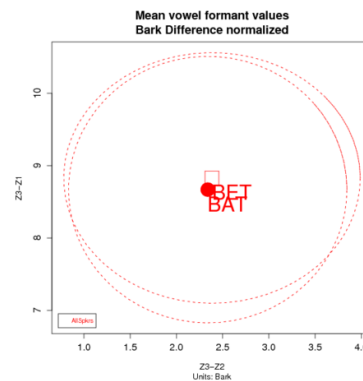
Graph 1: Mean formant values for speaker AB (Hz) before training



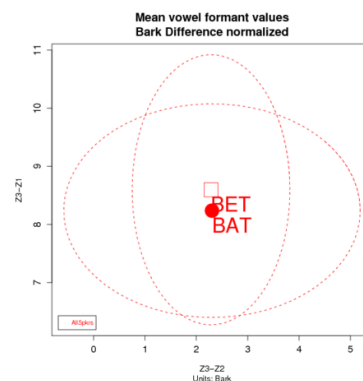
Graph 2: Mean formant values for speaker AB (Hz) after training

In three speakers' productions, the distance between the target vowels remained almost unchanged. In three speakers, the values changed negatively. This could mean that category formation did not take place in those speakers' systems. The English vowels were still perceived and thus produced as members of either a common L1/L2 category (one-category

assimilation) or a shared L2 category. If this is the case, it does not matter whether the differences between the vowels produced in the pre-test were slightly greater, because they were also mere examples for members of one and the same category. On the basis of the mean values for /e/ and /æ/ of each speaker (in Bark), differences between their productions of /æ/ and /e/ (F1 values only) were computed separately for each speaker. From these results, the difference between the two vowels of the pre-test was contrasted with the difference between the vowels in the post-test, ultimately showing the distance of the two vowels in the pre-test compared to the distance between the vowels in the post-test for each speaker. A group mean was determined for those values. This new value, the mean difference of the difference, is $d = 17.63$. The fact that d is positive (<0) means that the difference between the two vowels in the post-test was higher than their difference in the pre-test. Thus, speakers generally improved their distinction of the two vowels (regarding F1 frequencies only). A t-test showed that this change was significant ($n=15$; $t=1.94$; $p<.05$). Results indicate that the majority of the participants produced the two vowels more distinctly after training.



Graph 3: Mean formant values for all speakers (Bark) before training



Graph 4: Mean formant values for all speakers (Bark) after training

3.4.2. Results of the perception tests

Error rates before and after training were computed for each student. Where students had chosen 'don't know', this was counted as an error, because it was assumed that the vowel could not be identified correctly. Overall performance of the participants in the perception test improved slightly after training as compared to before training but this was not significant ($t=.6$; $p >.05$).

3.4.3. Relation between perception and production

In a first correlation test, only perception and production results of the pre-tests were compared. Results of a statistical correlation test (Pearson) implied a weak correlation ($r=-.34$), not significant at the .05 level. Students who performed weakly in the perception test thus did not necessarily show bad results in the production test and vice versa. The same test was conducted for the post-tests, and again there was no correlation ($r=-.09$). To see whether training improved perception and production and whether these improvements correlated for the single students, another correlation test was performed, taking into account the difference between pre- and post-test. Again, no significant correlation could be shown ($r=.02$). Perception and production post-tests show an overall improvement, though only significant for production. Despite these positive findings, the single participants did not improve in perception when they improved in production and vice versa.

4. Discussion

4.1. Discussion of production results

It was predicted that the difference between English /æ/ and /e/ would not be established in the learners' systems before training as an instance of one-category assimilation. This was supported by the findings of the first production task that showed an overall close resemblance between the two vowels. The second production test revealed that the overall distance between the two vowels was greater than before. This confirms the predictions of the SLM [4], which assumes that category formation leads to a better ability in distinguishing between two target language sounds. Nevertheless, the distance between the two vowels in the overall learner productions of the post-test was still much smaller than the distance between the same vowels in a native speaker. This suggests that training was helpful but not entirely sufficient in increasing all speakers' ability to distinguish between the vowels. More training would be necessary, though it cannot be expected that native-like category formation would take place in German students. It has been shown that even experience only led to a slight improvement of these abilities [12] and a longer period of repeated training sessions could only improve them to a certain extent [23, p. 58].

4.2. Discussion of perception results

Overall performance in the perception test only improved slightly and was not significant. This is interesting because perception is the feature that was actually trained. It is possible that the minimal pairs heard during training did not affect results of the perception post-test because the production task preceded the post-test and students might have been influenced by their own productions rather than by what they heard in the training. Another explanation is that category formation, contrary to the predictions made above, did not take effect in perception. A third explanation would be that subjects concentrated on a difference in length in order to distinguish between the said vowels (vowel-length is also what most students said when asked what they thought the experiment had been about). The presence of distracter-pairs such as <fit> and <feet> may have inclined them to expect the /æ/-/e/ difference to be based on length as well.

4.3. Implications for the perception-production relationship

It was predicted that participants would score higher in both post-tests. However, this was only significant for the production task. As results did not improve significantly in the second perception test, it can be assumed that learners were not able to establish new categories for the English vowels and category assimilation persisted even after training. It seems odd, then, that production results did improve after training. This is where questions of sequence in the perception production relationship come in. Out of the three possibilities (1) perception leads production, 2) production leads perception, 3) both are acquired simultaneously, Flege supports the first one whereas PAM can be brought into connection with the third [24, p. 57]. The above results, however, seem to support the second theory, which has often been declined [5]. If it is true that learners can produce contrasts before they are able to perceive them, it is possible to say that production leads perception. This hypothesis is in part supported by studies by Smith [17], Kosky & Boothroyd [25], and Parsloe [26], which feature similar results, although in connection with hearing impaired people. In spite of this apparent support for the "production-first"-hypothesis, care must be taken because the above results are not sufficient to say that the learners really could produce the contrast better than they heard it. The different learners showed great variation, some producing the contrast better after training, others hearing it better without showing better results in the production task. The fact that production improved after training only perception supports the hypothesis of an interdependence of perception and production, which was not considered obvious, as the two processes – one involving motor control and the other auditory processing – may or may not be controlled by different mechanisms" [5, p. 231].

5. Conclusion

The first hypothesis was confirmed only in part. Students who performed weakly in the perception tests did not necessarily have more problems in the production tasks. The fact that some of the participants showed high scores in production and low scores in perception and the other way around led to the conclusion that no assumption can be made regarding the sequence in which perception and production are acquired. Students might simply differ in the way they acquire new categories, some being better with perceptual learning, others with motor learning to improve the distinction of contrasts. The second hypothesis, however, was confirmed by the results of the study. The prediction had been made that after training perception the speakers would show an increased ability in producing the two target vowels as distinctly different from one another. This was the case, as formant values for F1 were significantly different for the two vowels in comparison to their distance in the pre-test. Hence, the perception training was successful in improving production abilities, which confirmed the findings of Wang et al. [13], and Aliaga-García & Mora [16]. Having shown that training only perception results in an improvement on the level of production, the results of this study can be seen as a piece of evidence that the two processes are closely linked, and, even though suggestive, these findings may be helpful in finding out more about the perception-production relationship in FLA.

6. References

- [1] M. Tatham and K. Morton, 2011, *A guide to speech production and perception*. Edinburgh: Edinburgh University Press, 2011.
- [2] R. Fox, "Individual Variation in the Perception of Vowels: Implications for a Perception-Production Link," *Phonetica*, no. 39, pp.1-22, 1982.
- [3] N. Bourguignon, S.R. Baum and D.M. Shiller, "Lexical-perceptual integration influences sensorimotor adaptation in speech," *Frontiers in Human Neuroscience*, no. 8, p. 208, 2014.
- [4] J. E. Flege, "Second Language Speech Learning: Theory, Findings, and Problems," in: *Speech Perception and Linguistic Experience: Issues in Cross- Language Research*, W. Strange, Ed. Timonium: York Press, 1995, pp. 233-277.
- [5] W. Barker and P. Trofimovich, "Perceptual paths to accurate production of L2 vowels: The role of individual differences," *International Review of Applied Linguistics in Language Teaching*, vol. 44, no. 3, pp. 231-250, 2006.
- [6] J.S. Perkell, F.H. Guenther, H. Lane, M.L. Matthies, E. Stockmann, M. Tiede and M. Zandipour, "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," *J. Acoust. Soc. Am.*, no. 116, pp. 2338-44, 2004.
- [7] C.M. MacLeod, N. Gopie, K.L. Hourihan, K.R. Neary, J.D. Ozubko, "The production effect: delineation of a phenomenon," *Journal of Experimental Psychology: Learning, Memory and Cognition*, no. 36, pp. 671-685, 2010.
- [8] S.E. Gathercole and M.A. Conway, "Exploring long-term modality effects: vocalization leads to best retention," *Memory & Cognition*, no. 16, pp. 110-119, 1988.
- [9] P.A. MacDonald and C.M. MacLeod, "The influence of attention at encoding on direct and indirect remembering," *Acta Psychologica*, no. 98, pp. 291-310, 1998.
- [10] W. Strange, "Learning Non-native Phoneme Contrasts. Interactions among Subject, Stimulus, and Task Variables," in: *Speech perception, production and linguistic structure*, Y. Tohkura et al., Eds. Tokyo: Ohmsha, 1992, pp. 197-221.
- [11] M. Eskenazi, "Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype," *Language Learning & Technology*, vol. 2, no. 2, pp. 67-76, 1999.
- [12] E. Simon and T. D'Hulster, "The effect of experience on the acquisition of a non-native vowel contrast," *Language Sciences*, no. 34, pp. 269-283, 2012.
- [13] Y. Wang et al., "Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training," *Journal of the Acoustical Society of America*, vol. 113, no. 2, pp. 1033-1043, 2003.
- [14] U. Gut, *Introduction to English Phonetics and Phonology*. Frankfurt a. M.: Peter Lang, 2009.
- [15] J.E. Flege, C. Schirru and I.R. MacKay, "Interaction between the native and second language phonetic subsystems," *Speech Communication*, no. 40, pp. 467-491, 2003.
- [16] A.C. Aliaga-García and J.C. Mora, "Assessing the Effects of Phonetic Training on L2 Sound Perception and Production," in: *Recent Research in Second Language Phonetics/Phonology: Perception and Production*, M. Watkins, A. Rauber and B. Baptista, Eds. Newcastle: Cambridge Scholars, 2009.
- [17] L.C. Smith, "L2 acquisition of English liquids: Evidence for production independent from perception," in: *The Past, Present, and Future of Second Language Research*, Bonch-Bruевич et al., Eds. Somerville, MA: Cascadilla Press, 2001, pp. 3-22.
- [18] F. Eckman et al., "Perception and Production in the Acquisition of L2 Phonemic Contrasts," in: *New Sounds: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*, 2007, pp. 189-198.
- [19] W. Barry, "Perception and Production of English Vowels by German Learners: Instrumental-Phonetic Support in Language Teaching," *Phonetica*, no. 46, pp. 155-168, 1989.
- [20] J.C. Mora. (2009). *Assessing L2 perceptual phonological competence: Methodological issues in assessing L2 perceptual phonological competence*. [Online], Available: http://www.phon.ucl.ac.uk/ptlc/proceedings/ptlcpaper_34e.pdf.
- [21] E. Thomas and T. Kendall: *The Vowel Normalization and Plotting Suite*. Available: <http://ncslaap.lib.ncsu.edu/tools/norm/>.
- [22] C. Best and M. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," in: *Second language speech learning: The role of language experience in speech perception and production*, M. Munro and O. Bohn, Eds. Amsterdam: John Benjamins, 2006, pp. 13-45.
- [23] G. Couper, "The short and long-term effects of pronunciation instruction," *Prospect*, vol. 21, no. 1, pp. 46-66, 2006.
- [24] R.C. Major, *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. London: Lawrence Erlbaum Associates, 2001.
- [25] C. Kosky and A. Boothroyd, "Perception and production of sibilants by children with hearing loss: A training study," *The Volta Review*, no. 103, pp. 71-98, 2003.
- [26] R. Parsloe, "Use of the speech pattern audiometer and the electropalatograph to explore the speech production/perception relationship in a profoundly deaf child," *International Journal of Language and Communication Disorders*, no. 33, pp. 109-121, 1998.