

AUTHENTICITY OF PRONUNCIATION IN NATURALISTIC SECOND LANGUAGE ACQUISITION: THE CASE OF VERY ADVANCED LATE LEARNERS OF DUTCH AS A SECOND LANGUAGE

Theo Bongaerts, Susan Mennen & Frans van der Slik

Abstract. The article reports on the fourth study in a series of four designed to test the prediction that a nativelike accent is unattainable for those who start to acquire an SL after the close of the critical period. Sentences read out by late learners, who acquired Dutch in an immersion setting, were rated for accent by native speakers of Dutch. The results from this study, in combination with those from three previous studies, suggest that, in spite of the claims of the critical period hypothesis, late learners can achieve a nativelike accent in an SL, and that a combination of input, motivational, and instructional factors may compensate for the neurological disadvantages of a late start.

1. Introduction

In 1990 Long published a review of the literature on age-related differences in second language acquisition (SLA). He concluded that the empirical evidence available to date provided firm support for the notion of a biologically- or rather neurologically- based critical period for SLA, according to which nativelike attainment would be beyond the reach of learners who started to acquire an SL after they had reached a given age. Long (1990:280) summarized his findings as follows: 'Native-like morphology and syntax only seem to be possible for those beginning before age 15' and 'The ability to attain native-like phonological abilities in an SL begins to decline by age 6 in many individuals and to be beyond anyone beginning later than age 12, no matter how motivated they might be or how much opportunity they might have'. It should be noted, however, that Long also hinted at the possibility that the lack of evidence against the existence of a critical period in the studies surveyed by him might be due to the fact that none of these studies had specifically targeted very advanced SL learners. He suggested, therefore, that future studies should include 'the very best SL learners' (Long 1990:281).

Since the publication of Long's review article, several researchers have taken up Long's suggestion and conducted studies which included very advanced late learners in their designs. Interestingly, such studies have tended to identify learners who performed at nativelike levels on the tasks that they were set. For example, in the domain of morphosyntax there are studies by Birdsong (1992, 1997), Van Wuijtswinkel (1994), White &

Genesee (1996) and Cranshaw (1997), in which at least some SL learners demonstrated native-like competence (for a recent review, see Birdsong 1999).

Pronunciation has tended to occupy a special position in the discussion about a critical period for SLA. For one thing, it has been claimed to be the first aspect of language to be affected by a critical period (Long 1990). For another, it has been argued that it is the **only** aspect that is subject to critical period constraints, as it is 'the only aspect of language performance that has a neuromuscular basis', requires 'neuromotor involvement' and has a 'physical reality' (Scovel 1988:101). However, it looks as if in this domain, too, it is not impossible for late learners to attain a nativelike level of proficiency, as has been demonstrated in a series of studies conducted at the University of Nijmegen (Bongaerts, Planken & Schils 1995, Bongaerts, Van Summeren, Planken & Schils 1997, Palmén, Bongaerts & Schils 1997; for a summary, see Bongaerts 1999). The key subjects in these studies were very advanced Dutch learners of English or French who had started to learn these languages as foreign languages in Dutch pre-university education around the age of twelve, had subsequently read English or French at institutions of tertiary education in the Netherlands and were employed as lecturers or professors in departments of English and French of a Dutch university or teacher training institute at the time of the experiment. In each of these studies at least some learners were identified whose pronunciation was judged to be indistinguishable from that of native speaker controls. We claimed that this result argued against the existence of an absolute biological barrier to the acquisition of a nativelike pronunciation of a foreign language and suggested that the combination of a very high motivation, intensive training in both the perception and the production of target language sounds, and ample exposure to the target language might have been responsible for the nativelike achievement of these late learners.

The main aim of the present study was to find out whether the conclusions of the earlier studies could be generalized to uninstructed learning contexts and to other L1-L2 pairings than Dutch-English and Dutch-French. In this study, the key subjects were advanced learners of Dutch from a variety of L1-backgrounds who had settled in the Netherlands after the age of twelve and had acquired Dutch mainly in a naturalistic, non-classroom setting.

2. Design of the study

2.1. Speakers

Two groups of speakers participated in the experiment:

- Group 1 consisted of ten native speakers of Dutch (mean age 33), who served as controls. They had all completed secondary education and

had graduated from or were still students at various institutions of tertiary education at the time of the experiment.

- Group 2 consisted of thirty advanced learners of Dutch as an L2 (mean age 40), who had settled in the Netherlands between the age of 11 and 34 (mean age on arrival 21). The participants in this group were native speakers of eleven different languages. The majority, sixteen, were speakers of German. There were also three speakers of English, two of French and two of Spanish, and one speaker of each of the following seven languages: Armenian, Berber, Czech, Greek, Italian, Swedish and Turkish. Their level of education was comparable to that of the speakers of group 1.

2.2. *Speech samples*

All speakers read out the following ten sentences, which contained multiple of examples of all but the most marginal Dutch phones:

- (1) Alle exemplaren van de dichtbundel zijn uit de handel genomen.
- (2) Mijn vader werkt als administratief medewerker bij de schouwburg in Goes.
- (3) Nooit eerder was er in Flevoland zo'n mooie schoonheidskoningin gekozen.
- (4) Een krassend krijtje op een schoolbord krijgt zelfs de vervelendste leerling stil.
- (5) De kinderen aten pannenkoeken tijdens het verjaardagsfeestje.
- (6) Onze adjunct-directeur vindt het leuk om naar de bioscoop te gaan.
- (7) Vlak voor de opvoering werden de instrumenten voor de laatste maal gestemd.
- (8) De deuk in de auto van de pastoor was schijnbaar door niemand veroorzaakt.
- (9) Literatuurliefhebbers komen tijdens de boekenweek aan hun trekken.
- (10) De digitale revolutie geeft leraren behoorlijk het nakijken.

2.3. *Judges*

The speech samples were rated by 21 native speakers of Dutch (mean age 43), who were resident in the Netherlands at the time of the experiment. They had all completed tertiary education. Eleven of them were teachers of Dutch as an L2. They were all experienced judges. The remaining ten judges lacked expertise in Dutch as an L2, nor had they had any formal

training in phonetics and linguistics. They can, therefore, be characterized as inexperienced judges.

2.4. *Rating procedure*

For each judge a tape was prepared which contained 10 sets of speech samples, each of which consisted of one sentence spoken by all 40 speakers. The 10 sets were presented to the judges in different orders and within each set the order of the speakers was randomized. The judges were asked to rate all 400 samples for foreign accent, using a Dutch version of a five-point scale, which ranged from 1 (very strong accent: definitely non-native) to 5 (no foreign accent at all: definitely native). The judges were told that they would hear sentences spoken by native and nonnative speakers of Dutch, but they were not informed about the proportion of native and nonnative speakers.

3. Results

To get a first impression of the data, we calculated the ratings assigned to each speaker, averaged across 10 sentences and 21 judges. These ratings are displayed in Table 1.

This table shows that the native speakers of Dutch were given ratings ranging from 4.00 to 4.91 (group mean: 4.73) and that the ratings assigned to the nonnative speakers ranged from 1.70 to 4.59 (group mean: 3.50). Note the rather low mean rating (4.00) for one of the native speakers (6). We will return to this later.

Before further analysing the differences between the ratings assigned to the native and the nonnative speakers, we first wanted to establish whether we could pool the data from the experienced and the inexperienced judges or whether we would have to analyse the data from each of the two groups of judges separately.

3.1. *Experienced versus inexperienced judges*

In order to examine the differences and similarities between the ratings assigned by the two groups of judges, we first calculated the euclidian distances between the rating patterns of all 21 judges. Each pattern contained 400 ratings, one for each speaker-sentence combination. Next we devised an artificial pattern, which we termed **strict**. This pattern was constructed by calculating, for each of the 400 speaker-sentence combinations, the mean and the standard deviation across all 21 judges, and then defining the pattern **strict** as representing an imaginary judge whose ratings are 1.5 standard deviations below the mean. This artificial pattern **strict** was used as the reference point for calculating the differences between the rating patterns of the two groups of judges. The average

Table 1: Mean ratings for all speakers averaged across 10 sentences and 21 judges

Native speakers		Non-native speakers					
Speaker	Mean	Speaker	Mean	Speaker	Mean	Speaker	Mean
1	4.86	11	1.99	21	2.04	31	4.24
2	4.83	12	4.45	22	3.07	32	3.56
3	4.78	13	4.41	23	4.59	33	4.58
4	4.76	14	2.96	24	3.82	34	3.64
5	4.58	15	2.46	25	4.34	35	4.28
6	4.00	16	2.07	26	3.87	36	2.96
7	4.91	17	3.89	27	4.42	37	2.90
8	4.80	18	3.27	28	3.01	38	1.70
9	4.87	19	4.41	29	4.07	39	3.97
10	4.86	20	2.98	30	4.33	40	2.66
Group mean:				Group mean:			
4.73				3.50			

distance of the experienced judges' rating patterns from the pattern **strict** was 2.81 ($SD = 1.13$), compared to 1.84 ($SD = 0.95$) for the inexperienced judges. Application of the Mann-Whitney test revealed a significant difference between the ratings assigned by the two groups of judges ($z = -2.11$; $p = 0.035$, two-tailed). In view of the above, we decided to analyse the ratings assigned to the speakers for each of the two groups of judges separately.

3.2. Ratings assigned to the native speakers and the learners of Dutch by the experienced judges

In order to examine the differences between the ratings assigned to the two groups of speakers by the experienced judges we adopted a similar procedure. We calculated the euclidian distances between the patterns of ratings assigned to the 40 speakers, each pattern comprising 110 ratings, one for each judge-sentence combination. We also defined an artificial pattern **max**, which represents an imaginary speaker who has only received ratings of 5. The average distance from **max** was 0.36 ($SD = 0.34$) for the native speakers and 1.94 ($SD = 1.16$) for the learners of Dutch. As was to be expected, the difference between the two groups turned out to be highly significant: application of the Mann-Whitney test resulted in a z of -4.19 ($p < 0.001$, two-tailed).

However, as our main aim was to find out whether or not at least some of the learners had received ratings which were comparable to those assigned to the native speakers, we focussed on individual speakers in our

Table 2: Standard scores for nativelikeness for all speakers (except speaker 6), based on ratings from the experienced judges

Native speakers		Non-native speakers					
Speaker	Z	Speaker	Z	Speaker	Z	Speaker	Z
1	-0.48	11	16.60	21	16.37	31	2.68
2	0.63	12	1.26*	22	9.49	32	7.75
3	-0.15	13	2.31	23	0.42*	33	1.43*
4	0.12	14	10.23	24	5.70	34	7.36
5	2.14**	15	13.69	25	2.82	35	3.45
7	-1.00	16	15.80	26	5.05	36	11.33
8	0.46	17	4.71	27	2.26	37	9.87
9	-0.72	18	8.04	28	10.02	38	18.50
10	-1.01	19	1.56*	29	3.47	39	4.88
		20	9.67	30	4.48	40	12.13
Group mean: 0.00		Group mean: 7.44					

Note: * = learners with $z < 2$ (nativelike). ** = native speaker with $z > 2$.

subsequent analysis of the data. In this analysis, we adopted the criterion of nativelikeness proposed by Flege, Munro & MacKay (1995). In their study, with Italian learners of English, learners with a mean rating that was within two standard deviations of the mean rating assigned to the native speakers were considered to have achieved a native(like) pronunciation of English. Our analysis revealed that no fewer than eight learners met Flege et al.'s ' $z < 2$ ' criterion of nativelikeness. There was one native speaker (speaker 6, see also Table 1) who did not meet the criterion. This speaker had been given a mean rating of 4.35¹ by the experienced judges, the next lowest mean rating assigned to a native speaker being 4.65. As this is a very low rating to be given to a native speaker (see Bongaerts et al. 1997) and as, in a group of ten speakers, the data of one 'outlier' can have a strong effect on the strictness of the criterion of nativelikeness adopted in our experiment, we decided to repeat the analysis, this time excluding the data from the 'outlying' native speaker. The results of this analysis are presented in Table 2.

From Table 2 it can be concluded that there are still 4 learners,

¹ It is not clear to the authors why this native speaker of Dutch was given such low ratings compared with the ratings assigned to the other native speakers who participated in the experiment. This speaker, a woman who was 22 years old at the time of the experiment, was born and raised in the south east of the Netherlands. In a background questionnaire, which she completed before the start of the experiment, she self-reported that she did not have a strong regional accent. The two first authors, who listened to recordings of the sentences spoken by her, also judged her pronunciation to contain no more than a slight regional colouring. We do not know on what grounds the judges rated her pronunciation so low, as, for reasons of time, they had not been asked to comment on the ratings they gave.

marked with an asterisk in the table, who met the criterion (12: $z = 1.26$; 19: $z = 1.56$; 23: $z = 0.42$; 33: $z = 1.43$). There is also one native speaker, marked with a double asterisk, who just failed to meet it (5: $z = 2.14$).

Next we applied the same procedure to the ratings which the individual speakers received for each of the ten sentences separately, again excluding the data from native speaker 6 from our calculations. This analysis revealed that of the four learners who met the criterion in the overall analysis, one (23) reached it on nine out of ten sentences, two (12 and 33) on eight sentences and one (19) on seven sentences. In comparison, the native speaker (5) who had failed to meet the criterion in the overall analysis, did not reach the criterion on five sentences, and another native speaker (8) did not meet it on one sentence.

3.3. *Ratings assigned to the native speakers and the learners of Dutch by the inexperienced judges*

We repeated the entire procedure for the ratings assigned by the inexperienced judges. This time the average distance from **max** was 0.55 ($SD = 0.48$) for the native speakers and 2.31 ($SD = 1.15$) for the learners, the difference between the two groups of speakers again being highly significant (Mann Whitney: $z = -4.28$, $p < 0.001$, two-tailed). In our analysis of the ratings of individual speakers, using the ' $z < 2$ ' criterion of nativelikeness, we decided to ignore the data of speaker 6, for reasons explained in the previous section. This 'outlying' native speaker received a mean rating of no more than 3.62 from the inexperienced judges, the next lowest rating assigned by them to a native speaker being 4.50.² The results of this analysis are given in Table 3.

Table 3 shows that there are two learners, marked with an asterisk, whose standard scores fall within the ' $z < 2$ ' criterion (23: $z = 1.96$; 33: $z = 1.57$). These two participants were among the four learners who also passed the test of nativelikeness with the experienced judges. As before, native speaker 5 did not meet the criterion ($z = 2.41$). Application of the same procedure to the ratings which the individual speakers received for each of the ten sentences separately revealed that the two learners (23 and 33) who had met the criterion of nativelikeness in the overall analysis, also met it on seven out of ten sentences. In comparison, native speaker 5 did not meet the ' $z < 2$ ' criterion on three sentences, and another native speaker (4) failed to reach it on one sentence.

If we combine the results of the analyses of the ratings assigned by the experienced and the inexperienced judges, we can conclude that we have identified at least two highly successful late learners of Dutch who performed, on the sentence reading task used in our study, within the

² If we had included the data from this native speaker in the analysis, no fewer than ten learners would have met the criterion of nativelikeness.

Table 3: Standard scores for nativelikeness for all speakers (except speaker 6), based on ratings from the inexperienced judges

Native speakers		Non-native speakers					
Speaker	Z	Speaker	Z	Speaker	Z	Speaker	Z
1	-0.61	11	18.64	21	18.08	31	4.12
2	-0.45	12	3.96	22	12.61	32	9.35
3	0.60	13	2.49	23	1.96*	33	1.57*
4	0.00	14	12.33	24	7.20	34	7.40
5	2.41**	15	15.65	25	3.04	35	3.79
7	-0.85	16	18.10	26	9.06	36	12.02
8	-0.48	17	7.20	27	2.63	37	14.33
9	-0.56	18	12.08	28	13.49	38	20.00
10	-0.05	19	3.70	29	6.38	39	6.85
		20	13.69	30	3.96	40	15.12
Group mean: 0.00		Group mean: 9.36					

Note: * = learners with $z < 2$ (nativelike). ** = native speaker with $z > 2$.

(lower) range of the native speaker controls. One of these learners was a native speaker of German and one was a native speaker of English. We have also identified two more learners, both native speakers of German, who performed at a nativelike level, if only the ratings assigned by the experienced judges are considered.

4. Conclusions and discussion

We believe that in this study, as in our earlier studies (summarized in Bongaerts 1999), we have demonstrated that it is not impossible for post-critical period learners to achieve a nativelike accent in a non-primary language, in spite of the alleged biological barriers. In our previous studies, the participants that beat the predictions of the critical period hypothesis were late Dutch learners of English or French as a foreign language, who were resident in the Netherlands and were, therefore, **not** immersed in the target language. They had all read English or French at Dutch universities, had received intensive instruction in the perception and production of target language speech sounds, were employed as lecturers of English or French at Dutch institutions of tertiary education, and considered it both personally and professionally to be very important to have an excellent pronunciation. It should be emphasized here that some of these learners had performed within the **upper** range of the native speaker controls in our studies. Let us now look at the profiles of the two star participants in the present study. Both had acquired Dutch in an

immersion setting since their arrival in the Netherlands. They had married Dutch women and had children whom they raised in Dutch, Dutch being the only language spoken at home. Both had been lecturers in their mother tongue at a Dutch university. At the time of the experiment one of them was retired and one was director of studies at a Dutch school of translation. One of them had received some instruction in the pronunciation of Dutch, the other one had received no such instruction at all. Like the participants in our earlier studies, both expressed a strong personal and professional interest in achieving a very good pronunciation. To sum up, the most successful learners in the present study were much more intensively exposed to natural, spoken target language input than were the participants in our earlier studies. On the other hand, they received no – or hardly any – special instruction in the pronunciation of the target language, whereas the subjects in our earlier studies were intensively drilled in pronunciation. The star participants in all our studies were highly motivated to achieve an excellent pronunciation.

In this respect it is interesting to note that, whereas the most successful participants in our earlier studies performed within the **upper** range of the native speaker controls, the star subjects in the present study performed within the **lower** native speaker range. We would like to suggest that such a result offers support for our earlier suggestion that, next to ample exposure to authentic target language input and a high motivation, intensive training in the perception and production of target language speech sounds might well be a very important determiner of post-critical period, nativelike achievement in the pronunciation of a nonnative language. Interestingly, a similar suggestion was made in a recent paper on ultimate attainment in the pronunciation of an SL by Moyer (1999). The results of Moyer's study led her to conclude that 'overt phonological instruction appears necessary for some learners to acquire native-level phonological production . . .' (Moyer 1999:99). For the importance of perception and production training see also Hammond (1995), Flege, Takagi, & Mann (1995) and Flege (1999) and the literature cited therein. The precise effects of such instruction on ultimate attainment would obviously need to be explored further in future research.

Another area of further research would be the role of typological distance between the mother tongue and the target language. In our earlier studies, participants were Dutch-speaking learners of English and French. The key subjects in the present study were learners of Dutch from eleven different L1-backgrounds, the majority being native speakers of German. Of the four most successful learners, three were native speakers of German and one was a native speaker of English. In the present study there were eleven learners who obtained mean ratings (averaged over all sentences and all judgements) of over 4.00, this being the lowest overall mean rating assigned to a native speaker of Dutch. Of these eleven learners,

eight were native speakers of German, one of English, one of French and one of Czech. In other words, with the exception of one learner, the eleven learners with the highest ratings spoke L1s which were typologically closely (German and English) to relatively closely (French) related to the target language Dutch. This suggests that in the domain of pronunciation, as in the domain of morphosyntax (see Kellerman 1995), typological proximity may be one of the determining factors of ultimate nativelike performance. This issue, too, would be a fruitful area for further research.

References

- BIRDSONG, D. 1992. Ultimate attainment in second language acquisition. *Language* 68, 706–755.
- BIRDSONG, D. 1997. Intransitivity and SE in French: Aspects of late L2 learnability. Paper presented at the Boston U. Child Language Development Conference.
- BIRDSONG, D. 1999. Introduction: Whys and why nots of the critical period hypothesis for second language acquisition. *Second language acquisition and the critical period hypothesis*, ed. D. Birdsong, 1–22. Mahwah, NJ: Erlbaum.
- BONGAERTS, T. 1999. Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. *Second language acquisition and the critical period hypothesis*, ed. D. Birdsong, 133–159. Mahwah, NJ: Erlbaum.
- BONGAERTS, T., PLANKEN, B. & SCHILS, E. 1995. Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis. *The age factor in second language acquisition*, ed. D. Singleton & Z. Lengyel, 30–50. Clevedon: Multilingual Matters.
- BONGAERTS, T., VAN SUMMEREN, C., PLANKEN, B. & SCHILS, E. 1997. Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition* 19, 447–465.
- CRANSHAW, A. 1997. A study of Anglophone native and near-native linguistic performance. Unpublished Ph.D. diss., U. de Montréal.
- FLEGE, J. 1999. Age of learning and second language speech. *Second language acquisition and the critical period hypothesis*, ed. D. Birdsong, 101–131. Mahwah, NJ: Erlbaum.
- FLEGE, J., MUNRO, M. & MACKAY, I. 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97, 3125–3134.
- FLEGE, J., TAKAGI, N. & MANN, V. 1995. Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech* 38, 25–55.
- HAMMOND, R. 1995. Foreign accent and phonetic interference: The application of linguistic research to the teaching of second language pronunciation. *Second language acquisition theory and practice*, ed. F. Eckman, D. Highland, P. Lee, J. Mileham, & R. Rutkowski Weber, 293–303. Mahwah, NJ: Erlbaum.
- KELLERMAN, E. 1995. Age before beauty: Johnson and Newport revisited. *The current state of interlanguage: Studies in honor of William E. Rutherford*, ed. L. Eubank, L. Selinker, & M. Sharwood Smith, 219–231. Amsterdam: John Benjamins.
- LONG, M. 1990. Maturational constraints on language development. *Studies in Second Language Acquisition* 12, 251–285.
- MOYER, A. 1999. Ultimate attainment in phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition* 21, 81–108.

- PALMEN, M.-J., BONGAERTS, T. & SCHILS, E. 1997. L'authenticité de la prononciation dans l'acquisition d'une langue étrangère au-delà de la période critique: Des apprenants néerlandais parvenus à un niveau très avancé en français. *Acquisition et Interaction en Langue Etrangère* 9, 173–191.
- SCOVEL, T. 1988. *A time to speak. A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.
- VAN WUITSWINKEL, K. 1994. Critical period effects on the acquisition of grammatical competence in a second language. Unpublished B.A. thesis, Department of Applied Linguistics, U. of Nijmegen.
- WHITE, L. & GENESEE, F. 1996. How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research* 12, 233–265.

Received September 13, 1999

Accepted January 20, 2000

*Theo Bongaerts
Susan Mennen
Frans van der Slik
Department of Applied Linguistics
University of Nijmegen
Erasmusplein 1
6525 HT Nijmegen
The Netherlands
email: th.bongaerts@let.kun.nl
email: f.v.d.slik@let.kun.nl*