DEVELOPMENTAL STAGES AND VARIABILITY IN THE ACQUISITION OF SECOND LANGUAGE SEGMENTS AND PROSODY

Christiane Ulbrich

Philipps University Marburg
christiane.ulbrich@uni-marburg.de

ABSTRACT

The perceived impression that a spoken utterance is not native is due to the culmination of a combination of segmental and suprasegmental phonological and phonetic characteristics entrenched in the speakers native language (L1) or it results from different types of interplay between those characteristics of the L1 and the second (or foreign) language (L2). The present paper reports findings of a detailed phonetic analysis of German L2 utterances produced by native speakers of English of different proficiency levels. The results show that the mastery of segments and prosody is not a progression of one and the other but rather that the mastery of individual characteristics of either level follows a structured sequence as predicted under the assumption of language learning as a self-organising dynamic process. The results are best interpreted in adaptation of the framework of the Dynamic Systems Theory (DST).

Keywords: second language acquisition, prosody, segments, production, foreign accent

1. INTRODUCTION

Deviations for a target language in the process of foreign language learning are the reason for a foreign accent (FA). Whilst in many traditional approaches to language learning and teaching pronunciation has been largely neglected, in today's society, with online meetings or job interviews and the employment of foreign-born workers with different educational and professional qualifications, adequate oral communication skills become increasingly important. A FA is usually perceived as a barrier by the foreign language speaking individual and has been shown to consciously or unconsciously influence employment-related decisions [18] even though employment strategies underly policies insuring racial, ethnical and national equality (amongst other). But what constitutes a FA? Most research has focused on the contribution of segments to the production and perception of FA [e.g. for vowels: 1, 8, 9, 16, 17, 19, 38] and for consonants: 7, 10, 13, 14, 29, 31, 32] and therefore unsurprisingly, current L2 speech acquisition models (e.g. Flege's Speech Learning Model [15], Best's Perceptual Assimilation Model [2, 3, 4], Kuhl's Native Language Magnet [23], Eckman's Markedness Differential Hypothesis [12] or Major's Ontogeny Phylogeny Model [25]) account for segmental errors and predict learning difficulty or ease of L2 speech on the basis of segmental proximity of and interference between L1/ L2 sounds. More recently however, phonological and phonetic components of prosody either in separation or in combination with segments have been examined in the effort to better understand the nature of the contribution of prosody and it's interaction with segments in the L2 acquisition process [33 for an overview]. The contradicting nature of these previous findings may be due to the difficulties to capture and model intra- and inter-speaker variability. The issue has been addressed from different perspectives [37 for an overview] in order to identify the different causes of variability in L2 production and perception. Linguistic approaches in the Chomskyan tradition of Universal Grammar try to abstract away from variation whereas sociolinguistic approaches focus predominantly on external sources [27] and psycholinguistic approaches consider the variability as a product of an intrinsic property of dynamic self-organising systems [11]. A large body of research was dedicated to the identification of (universal) stages in the development of L2. However, researchers have not reached a consensus due to the complexity and interaction of factors influencing L2 acquisition and only recently the variability in itself is considered a useful source of facts and information [30]. The framework of DST is a line of research that attempts to describe and model the development and the interaction of linguistic sub-systems by looking at patterns of variability [36]. This approach is adopted in a detailed phonetic analysis of spoken L2 German by native speakers of English. Previous findings suggest that the mastery of segmental and prosodic characteristics interacts in the acquisition process. In a perception study of German native speakers of L2 English a correspondence of segments and prosody was observed on the basis of FA ratings obtained from native speakers of English. Here, prosody was found to influences the degree of perceived FA only when corresponding with the

perceived FA on the segmental level [Ulbrich & Mennen, 2015].

2. METHOD

The paper presents a detailed analysis of L2 German produced by L1 speakers of American English (AmE) in order to understand the intricate interplay between segmental and prosodic characteristics and to identify developmental patterns in the acquisition of L2 speech.

2.1. Material

The speech material of this study consists of a set of nonce words (1) to be produced in isolation and (2) to be inserted into a visually presented context. The nonce words were two- and three-syllabic with a constant /CV:CV:/ and /CV:CV:CV:/ syllable structure respectively. The two syllabic words were introduced as names for individuals (not included in the analysis of the present paper) and the three-syllabic words as names of locations, such as a castle, a park, a lake etc.). We used voiced consonants to make sure that a continuous intonation contour can be produced in order to allow for the observation of prosodic characteristics.

Segmental level: The realisation of /r/ differs in German and English [5, 24]. English speakers produce an approximant [ɹ] while speakers of Northern Standard German (NSG) produce a voiced fricative in syllable initial position [ʁ]. Stimuli are designed to contain Vs and Cs most similar in German and English, i.e. [i:, u:, n, m] or to contain /r/ as a sound that differs in English and German.

Table 1: corresponding and deviant C used in the stimuli-design

corresponding	deviant	
English/German	English	German
[n], [m]	[ɹ]	[ʁ]

Table 2: Examples and number of stimuli in the two conditions used in the stimuli-design

condition	number	examples
1	48	[mu:ni:mi:] or [nu:ni:nu:]
2	48	[ʁu:ni:mi:] or [ʁi:mu:mi:]

The combinatory possibilities of the six syllable types [ni:, nu:, mi:, mu:, and, ʁi:, ʁu:] adds up to 219. In 48 of the target words deviant C [ʁ] occurs. In order to limit the number of target words with corresponding C and V, all words with subsequent identical syllables (e.g. [mi:mi:mi:] or [nu:nu:nu:]) were excluded, so that a total of 96 words were used.

Prosodic characteristics: In a comparative analysis of utterances produced by native speakers of German and AmE, German speakers were shown to realised a so called continuation rise (CR) final position of an intonational phrase (IP) as L*H % whereas AmE produced an (L+H*)L-H%. In other words, Germans produce a simple rise followed by a default boundary tone and AmE speakers produce an "explicit rise in the boundary constellation itself (low phrase accent followed by high boundary tone: L-H%)". A rise-fall-rise pattern occurs if the rise is preceded by a rising nuclear pitch accent [21]. In other words, peak alignment is expected to be considerably earlier in AmE compared to German. Stimuli were embedded into carrier sentences in order to allow for the elicitation of a continuation rise.

Speakers and recordings: We recorded a total of thirty female speakers of L1 American English born and raised in and around Wisconsin and aged between 20 and 29 (average age 23) and different levels of L2 proficiency in German. Ten (beginners = B) of the subjects had only recently arrived in German and received less then a year of instructed German and were recorded prior to their German intensive training. Ten subjects (intermediate = I) had received more than one years of German training, had lived in Germany for at least one year and had completed the intensive training. The remaining ten subjects (advanced = A) had been learning German as a foreign language for at least seven years prior to their arrival in Germany, had completed the intensive training and lived in Germany for at least three years. We also recorded ten age matched native speakers of German (as controls = C). All subjects were unpaid and had normal hearing and vision.

Data elicitation: Recordings took place in a sound attenuated cabin directly onto a macintosh computer via a Beyerdynamic microphone (MC 930). Subjects were placed in front of a computer-screen. The data elicitation took place in two phases:

Phase 1: single word production

Phase 1 consists of two recordings. Initially participants were asked to read the unknown single words appearing individually on screen as names for an individual or a location shown in a word-picture-pair (WPP). Subjects were instructed to stress the first syllable of each word. The word was presented in orthographic spelling, the initial syllable was written in capital letters and bold face. Since the production of long vowels was aimed at, <*h*> or double-Vowel <*uu*> or <*ie*> were used to indicate vowel length, e.g. <*MUUnimi*> or <*Riemuhnie*>.

2. Phase: words in context

In phase 2 participants were presented with the same WPPs, whereby the WPP of the individual appeared always on the top and two WPPs of locations on the bottom of the screen. Each three-syllabic WPP was shown twice, once on the right-hand side of the screen and once on the left-hand side. Subjects were instructed to insert the three variables into a fixed context sentence: *<X geht zum Schloss Y oder zum Schloss Z>* (X is going to the castle Y or to the castle Z). Information structure requires a prenuclear pitch accent on X, a continuation rise on Y and a nuclear sentence accent on Z. Randomisation was applied for the combination of individual and locations as well as for trial sequence in the two recordings for each experimental phase per subject.

The total number of nonce words recorded per subject were 96 in phase 1 (eight blocks with 12 items each) and 192 (96 as first location and 96 as second location) in phase 2 (sixteen blocks with 12 items each). After each block participants were allowed to pause and could continue the presentation by pressing a space key on a keyboard placed in front of them. In order to allow for the investigation of interactions between the segmental [ʀ] and prosodic (L*H vs LH*) characteristics, only the 96 productions of the first WPPs produced in phase 2 were analysed here.

2.2. Analysis

The sound files were digitised at 44kHz, 16 bits, mono-format for processing and analysis in PRAAT [6] Statistical analyses were carried out in R [28].

Acoustic measurements involved for the segment [ʀ] the proportion of friction as quantifiable in the Harmonic-Noise-Ration (HNR). The HNR was measured within the initial sound of the nonce word at the end of formant transition transition of F2 and F3 to the beginning of the steady vowel portion using "to harmonicity (cc)"analysis in PRAAT with a time step of 10 ms, minimum frequency of 75 Hz, a silence threshold of 100ms and 4.5 periods per window previously applied in the distinction of /r/ approximant and /r/-fricative [20]. For the prosodic level the timing and extend of the continuation rise (CR) as quantifiable in the f0 peak alignment and f0 range was measured adopting the parameter-based approach, PaIntE (Parametric Intonation Events) [26] previously shown to capture the phonetic dimension of intonational events [22]. In the original approach measurements include six parameters (height of the F0 peak, the temporal position of the peak in the accented syllable, the amplitudes and the steepness of both rises and falls.

The current analysis involved a comparison of the temporal position of the peak in the accented syllable. The time is normalized to the syllable lengths, i.e. if b = 0 the f0 peak is at the beginning of and if b = 1 f0 is at the end of the accented syllable if b=0.

3. RESULTS

A binary logistic regression was used as an initial step of the analysis to explore if HNR and CR become less salient with increasing mastery of an L2. The model thereby allowed for the estimation of the likelihood that the productions belong to the B, I, A, or C group of speakers. Two predictors (HNR and CR), were included in the constructed model which was statistically significant, $\chi 2(2,192)=.326,36$, $p<.0005$. This indicates that HNR and CR reliably predict the group of speaker for the individual productions. The model accounted for between 48% (Cox and Snell's R2) and 53% (Nagelkerke's R2) of variance and correctly classified 78% of the nonce word productions. Even though both HNR and CR were found to made a unique and significant contribution to the model, the better predictor was HNR with an odds ratio of 2.49 compared to CR with an odds ratio of 0.892. In addition we found differences in the contribution of HNR and CR between the 4 speaker groups. Table 3 illustrates the findings for HNR and CR in the four groups of speakers.

Table 3: absolute HNR and normalised timing of CR for the four groups of speakers (B, I, A, C) in phase 1 and phase 2

	HNR	CR-timing		
	HNR (db)	SD (db)	CR-timing	SD
B	11	2,23	0,237	0,023
I	15	4,80	0,228	0,03
A	16	1,96	0,514	0,13
C	20	1,35	0,859	0,089

Considering the averaged values obtained for the individual speakers two aspects attract attention.

(1) The degree of variability differs in HNR and CR. In HNR obtained values for I show more variability compared to all other groups (also noticeable in SD for the I group of speakers). The values obtained for the peak alignment (CR), however, show the greatest variability in the A group of speakers (reflected again in the relatively large SD).

(2) Comparing HNR and CR in the data of individual speakers it occurs that higher, i.e. more

'native-like' values for one of the measurements is often accompanied for a relatively low value of the other. For instance speakers I2 and I3 receive relatively high values for HNR but score lowest in CR. The reverse is the case in A6 who produced relatively late, i.e. native-like' f0 peaks but produced [ʀ] with a very low portion of friction, i.e. an approximant as typically produced in AmE rather than a fricative as produced in NSG.

Figure 1: HNR (db) in B, I, A, and C

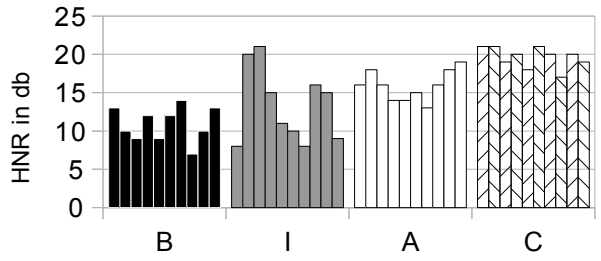
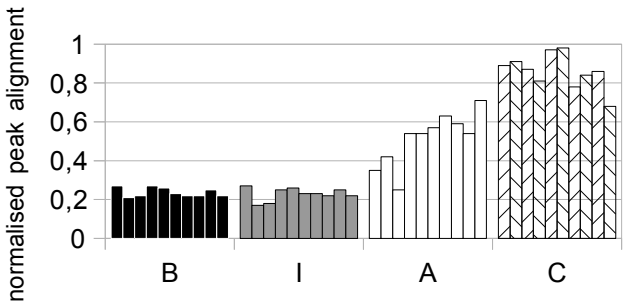


Figure 2: CR (normalised f0 peak alignment) in B, I, A, and C



4. DISCUSSION

The results suggest that contrary to the attempts to capture external factors as causes for variability in traditional approaches to L2 acquisition variability between learner groups and individual learners provides insight into developmental stages on the bases of the linguistic system itself. The findings of the phonetic analysis of selected segmental and prosodic characteristics produced by native and AmE-L2 speakers of German shows that the acquisition process changes with mastery of L2. More specifically the observed variability in HNR produced by beginners and intermediate L2 learners and the relatively 'native-like' production of HNR indicates an early development on the segmental level followed by a consolidation in higher L2 proficiency levels. The considerable jump between I and A in the realisation of more native#like peak alignment suggest a later development of the intonational level. It also seems noteworthy to point out that the identified stages cannot naturally account for other language pairings since similarities and differences vary between languages and even between varieties of languages [34, 35]. The relatively late acquisition of native-like prosody may be due to the complex role prosodic characteristics play in the interaction with other linguistic levels and it seem reasonable to study sequential interactions and developmental stages from the DST point of view [30, 37]. It remains to be noted that the present experimental setting only involved possible source for FA production on the segmental and the prosodic level so that it can be expected that other characteristics of the two levels of speech will pattern differently. In addition the role of the observed characteristics in the perception of FA has still to be evaluated. In general however, analyses of individual L2 speakers' productions lead to a better understanding of development, interaction of linguistic resources and thereby the development of individual feedback option in L2 learning.

5. REFERENCES

[1] Baker, W., Trofimovich, P. 2005. Interaction of Native- and Second-Language Vowel System(s) in Early and Late Bilinguals. *Language and Speech,* 48(1), 1-27.

[2] Best, C. 1995. A direct realist view of crosslanguage speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 171-204. Timonium, MD: York Press.

[3] Best, C. T., McRoberts, G. W., Goodell, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners' native phonological system. *JASA*, 109 (2), 775-792.

[4] Best, C. T., Tyler, M. D. 2007. Nonnative and second-language speech perception. Commonalities and complementarities. In: O-S. Bohn, M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege,* 13–34.

[5] Biersack, S. 2002. Systematische Aussprachefehler deutscher Muttersprachler im Englischen – Eine phonetisch-phonologische Bestandaufnahme (Systematic Pronunciation Mistakes of German Natives Speaking English). M.A. Thesis published in *FIPKM* 39, 37-130.

[6] Boersma, P. D. Weenink. 2003. PRAAT 4.1: *Doing phonetics by computer* <http://fonsg3.let.uva.nl/praat/>

[7] Bohn, O.-S., Steinlen, A. K. 2003. Consonantal context affects cross-language perception of vowels. In. M. J. Sol., D. Recasens J. Romero (Eds.), *Proceedings of the 15th ICPhS Barcelona.* 2289-2293.

[8] Bohn, O.-S., Flege, J. E. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception, *Applied Psycholinguistics,* 11, 303-328.

[9] Bohn, O.-S., Flege, J. E. 1992. The production of new and similar vowels by adult German learners of English, *SSLA*, 14, 131-158.

[10] Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and ?l/: IV. Some effects of perceptual learning on speech production. *JASA*, 101, 2299–2310.

[11] De Bot, K., Lowie, W., Verspoor, M. 2005. *Second language acquisition: An advanced resource book.* London: Routledge.

[12] Eckman, F. 1996. A functional-typological approach to second language acquisition theory. In W. C. Ritchie & T. C. Bhatia (Eds.), *Handbook of Second Language Acquisition*. 195-211. San Diego: Academia Press.

[13] Escudero, P., Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. SSLA, 26, 551–585.

[14] Flege, J. E., Hillebrand, J. 1984. Limits of phonetic accuracy in foreign language production. JASA, 76, 708-721.

[15] Flege, J. E. 1995). Second Language Speech Learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in crosslanguage research.* 233-277. Timonium, MD: York Press.

[16] Flege J. E., MacKay, I. R. A., Meador, D. 1999a. Native Italian speakers' production and perception of English vowels. *JASA,* 106, 2973 – 2987.

[17] Fox, R. A., Flege, J. E., Munro, M. J. 1995. The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *JASA*, 97, 2540-2551.

[18] Hosoda, M., Stone-Romero, E. 2010. The effects of foreign accents on employment−related decisions, *Journal of Managerial Psychology* 25(2), 113 – 132.

[19] Ingram, J. C. L., Park, S.-G. 1997. Cross-language vowel perception and production by Japanese and Korean learners of English. *JoP*, 25, 343 – 370.

[20] Jaworski, S., Gillian, E. 2011. On the phonetic instability of the polish rhotic /r/. In: *Poznań Studies in Contemporary Linguistics,* 47(2), 380–398.

[21] Jilka, M. 2000. Testing the contribution of prosody to the perception of foreign accent. *New Sounds,* 4, 199-207.

[22] Jilka, M. 2007. Different manifestations and perceptions of foreign accent in intonation. In J. Trouvain, U. Gut (Eds.): *Non-Native Prosody - Phonetic Description and Teaching Practice.* 77 – 96, Mouton De Gruyter, Berlin.

[23] Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. Nelson, T. 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). P*hilosophical Transactions of the Royal Society B*, 363, 979-1000.

[24] Ladefoged, P. 2004. *Vowels and consonants: An introduction to the sounds of languages.* Oxford: Blackwells.

[25] Major, R. C. 2001. *Foreign Accent*. Mahwah, NJ : Lawrence Erlbaum.

[26] Möhler, G., Conkie, A. 1998. Parametric modelling of intonation using vector quantization. *Proc. of the 3rd ESCA Workshop on Speech Synthesis,* Jenolan Caves (Australia), 311-316.

[27] Piske, T., MacKay, I. R.A., Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *JoP* 29: 191-215.

[28] R Core Team 2013. R: *A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

[29] Sancier, M. L., Fowler, C. A. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *JoP*, 25, 421-436.

[30] Thelen, E., Smith, L. B. 1994. *A dynamic systems approach to the development of cognition and action.* Cambridge MA: MIT Press.

[31] Trofimovich, P., Gatbonton, E., Segalowitz, N. 2007. A dynamic look at L2 phonological learning: Investigating effects of cross-language similarity and input frequency. *SSLA*, 29, 407-448.

[32] Tsukada, K., Birdsong, D., Mack, M., Sung, H., Bialystok, E., Flege, J. E. 2004. Release bursts in English word-final voiceless stops produced by native English and Korean adults and children, *Phonetica,* 61, 67–83.

[33] Ulbrich, C., Mennen, I. in print. When prosody kicks in! The intricate interplay between segments and prosody in perceptions of foreign accent. *International Journal of Bilingualism.*

[34] Ulbrich, C. 2013: German pitches in English. Production and perception of cross-varietal differences in L2. *Bilingualism: Language and Cognition.* Volume 16. Special Issue 02, 397- 419.

[35] Ulbrich, C. & Ordin, M. 2014: Can L2-English influence L1-German? The case of post-vocalic /r/. *JoP* 45C, 26-42.

[36] Van Geert, P. 1994. *Dynamic systems of development: Change between complexity and chaos.* New York: Harvester.

[37] Verspoor, M., Lowie, W., Van Dijk, M. 2008. Variability in Second Language Development From a Dynamic Systems Perspective. *The Modern Language Journal,* 92, 214–231.

[38] Walley, A. C., Flege, J. E. 1999. Effect of lexical status on children's and adults' perception of native and non-native vowels. JoP, 27, 307-332.