The Spectral Large-Scale Integration Mechanism in Perceptual Classification of Second Language Vowels*

Ewa Jacewicz
University of Wisconsin-Madison

1 Introduction

One of the goals of the substance-based approach to predicting vowel inventories of human languages is to provide information about the mechanism of speaker-listener interactions. Since, in Lindblom's words, "in any scientific field the question of independent motivation is the crucial one in the search for explanations" (1986a:17), seeking the mechanism or the mental process that potentially answers the question of how phonological contrast comes to existence is fully justified.

How does it happen that the vowel contrast is being acquired and the speaker-listener remains sensitive to its manifestation in even the most rapidly spoken utterances? Also, if a certain mechanism is indeed utilized in the process of shaping the native language (L1) vowel inventory, does it operate in the same way in the acquisition of a second language (L2) vowel system by adult learners? Such questions are of particular importance in testing whether phonological constructs such as distinctive features are active in the acquisition of phonology of a language or whether the acquisition of contrasts is independently motivated by processes underlying speaker-listener interactions in that the sounds of a language must sufficiently differ from each other in order to be perceived and acquired.

The vowel-internal representation still awaits explanation and further work is needed to investigate whether the subsegmental featural units of description function in any way in both vowel category formation and acquisition of contrast. This paper examines the operation of a process that uses vowel formant frequencies as perceptual cues to vowel quality as one possible avenue of investigation of how phonological contrast is perceived and acquired.

The existence of a large-scale integration (LSI) process in the auditory system has been demonstrated by Schwartz and Escudier (1989). As shown in their identification and discrimination experiments, spectral integration occurs at the non-peripheral intermediate level of processing and plays a role in perceptual classification of a vowel. Its operation provides a cue for a phonetic decision, a cognitive process which is believed to take place at the central level in the auditory path. The assumption that an LSI mechanism plays a part in vowel identification is to be traced back to earlier experiments by Delattre et al. (1952) which first pointed to an auditory averaging of two formants close in frequency. Subsequent experiments along this line by Carlson et al. (1970, 1975) introduced the concept of the effective second formant (F2'). More precise formulation of the integration process is particularly due to the further work by Chistovich and

CLS 34: The Panels (1998): 277-291.

Lublinskaya (1979) and Chistovich et al. (1979). As observed by the latter and, independently, by Bladon (1983), the "critical distance" where vowel formant energy undergoes a spectral integration does not exceed 3.5 Bark. In Stevens' words (1997: 503), "one can conclude that when two spectral prominences are separated by less than about 3.5 Bark, there is some kind of auditory integration of the two prominences, whereas each prominence maintains a separate auditory representation when the separation is greater than 3.5 Bark".

The concept of critical distance served further as a basis for the perceptual model of human vowel recognition proposed by Syrdal and Gopal (1986) who observed that formant distance measures between F2 - F1, F3 - F2, F4 - F3, F4 - F2, and F1-F0 for American English vowels in the Peterson and Barney (1952) study conform to the 3 Bark criterion. Based on whether Bark-difference values were within or exceeded the 3 Bark distance, they found the F1-F0 dimension corresponding to vowel height and the F3-F2 dimension to the front-back distinction in American English vowels. Thus, the 3-Bark critical distance is shown to play a role in achieving stability in vowel recognition within a whole vowel system of a language.

1.1 The LSI in predicting vowel systems

Dispersion Theory (Liljencrants & Lindblom 1972, Lindblom 1986a, b, 1990), seeks to define the general principle of perceptual contrast which is believed to be an independent motivation for shaping vowel inventories of human languages. According to the later versions of the theory, perceptual contrast needs to be sufficient in order to operate in real vowel systems. However, computer simulations which used the sufficient contrast algorithm still did not generate the predicted vowels within 3 to 7 vowel systems (in UPSID database, Maddieson 1984). The aim of the Dispersion-Focalization Theory (DFT) is to address this problem by introducing a criterion of stability in the perceptual domain within a vowel system. As summed up by Schwartz et al. (1997b:259-60), "The principle of the Dispersion-Focalization Theory (DFT) is precisely to define an energy function consisting of the sum of two terms, namely a structural dispersion term based on inter-vowel perceptual distances and a local focalization term based on intra-vowel perceptual salience, which aims at providing perceptual preferences to vowels showing a *convergence between two formants.*"

Unlike Lindblom's dispersion theory whose aim is to generate vowel systems based on auditory spectral distances, the DFT is data-driven in that it first computes 33 prototypes based on trends in vowel system inventories observed in the UPSID database.[1] Formant-based distances are used in computations. Then, a given vowel system is generated using the vowels selected among the prototypes which minimizes the total energy (summing the dispersion term and the focalization term). For the present purpose it must suffice to say that the introduced focalization component plays a crucial role in the simulations in that it defines the perceptual stability of a given vowel in the system. The focalization term minimizes the energy of a given system with vowels having "focal points", i.e., convergence of F1- F2 as in back rounded vowels, F2-F3 as in front rounded vowels, or F3-F4 as in front unrounded /i/. The DFT makes use of the LSI process and the 3.5 Bark critical distance measure in that the perceptual integration mechanism reduces variations around the formant convergence (within the 3.5 Bark) making it possible to maintain perceptual contrast between focal points such as in /i/ (F3-F4 convergence) and in /y/ (F2-F3 convergence).[2]

The greater adequacy of simulations with the UPSID database within the DFT framework which makes use of the LSI mechanism,[3] the long history of studies on the critical distance criterion as well as the powerful explanation of the Peterson and Barney (P&B) data in terms of the operation of the LSI by Syrdal and Gopal (S&G) serve as a basis for the present study which sought to determine whether the LSI process is active in perceptual classification of L2 vowels and, as such, whether it can be defined as a mechanism used in the acquisition of a spoken language.

2 Methodology

2.1 Stimuli

The vowels selected for investigation in the study were four German vowels: back rounded lax /ʊ/ and /ɔ/, and front rounded lax /Y/ and /œ/. The choice of the vowels was dictated by the following criteria:

(a) The German vowels were non-native segments for native speakers of American English (AE) learning German as their L2 who served as subjects for the study.

(b) The back German /ʊ/ and /ɔ/ have their counterparts in AE vowel system although their phonetic realization is substantially different: AE /ʊ/ is far more fronted than the standard German variant, and AE /ɔ/ exhibits an unknown in the German vowel spectral change with prolonged duration which classifies it as a long variant (Crystal & House 1988). As a result, AE speakers are not familar with the lax vowel contrast /ʊ/-/ɔ/ along the back peripheral dimension.

(c) The front rounded /Y/ and /œ/ represent the non-peripheral dimension in the acoustic vowel space which is not utilized in AE. Thus, both vowels are novel segments for an AE speaker.

For perceptual testing, all four vowels were embedded in each of the two consonantal contexts d_t and t_p. The naturally produced tokens were modelled after German disyllabic infinitives as they represent the tendency in the German lexicon for higher preference of disyllables over monosyllables (Kohler 1995). The following tokens were used in the perception test: ([dʊtən], [dɔtən], [dYtən], [dœtən]) and ([tʊpən], [tɔpən], [tYpən], [tœpən]) which can be represented in the German spelling as: (*dutten, dotten, dütten, dötten*) and (*tuppen, toppen, tüppen, töppen*). All tokens were nonsense words in German.

The choice of natural speech and contexts was motivated by the specific statement made in S&G (1986:1094): "(...) while specific bark-difference values for a vowel may vary significantly within speakers across speaking rates and segmental and prosodic contexts, the binary critical distance classifications appear to be very robust. Because of the extreme acoustic variability of speech, robust context- and rate-independent normalization may be best achieved through phonetic feature classification." It was assumed that both contextual and speaker variability which affect vowel formant frequencies and, consequently, the Bark-difference level would be a good measure of listeners' sensitivity to the critical distance criterion and to the operation of the LSI in perception of the naturally spoken language. Since the previous experiments used synthetic isolated vowels and the data from the P&B study were based on the neutral context /h_d/, the need for controlling for vowel context before making any claims with regard to the extend of the context-independent normalization is next logical step in pursuing the search for an LSI mechanism.

2.2 Speakers and recording procedure
Two carefully selected native speakers of German recorded the stimuli for perceptual testing. It was expected that the inter-speaker variability would affect subjects' normalization procedure which, in turn, would shed some light on their sensitivity to the Bark-difference distance. The choice of the speakers was particularly dictated by the /ʊ/-/ɔ/ contrast. In standard northern German, both vowels have their F1-F2 convergence within the 3.5 Bark whereas in AE, the F1-F2 convergence in /ʊ/ far exceeds the critical distance measure ranging between 4.4 -5.0 Bark (based on the P&B data). Since the subjects of the study were L2 learners of German, the German /ʊ/ could in their judgment collapse with /ɔ/ which, in turn, could affect their classification of the front rounded vowels. To avoid the potential problems due to the transfer effects or L2-L1 vowel assimilation, both speakers were selected from the southern dialects of German.
Speaker A (female) was from Austria (from the area of Graz), currently professor of German at the University of Wisconsin-Madison. Speaker B (male) came from Switzerland (from the area of Lucern) and was graduate student at the same university. Both speakers spoke standard German of those regions and in their production, the /ʊ/ was considerably fronted giving the F2-F1 Bark-difference well above 3.5 Bark which was comparable with the AE /ʊ/.

Both subjects recorded at least six repetitions of each token embedded in the carrier sentence "Sag _____ für mich" ("Say ____ for me"). The subjects were instructed to speak at their normal speech rate maintaining the same intonation pattern throughout the sentences. The speech samples were then digitized and stored on separate files of the CSpeech 4.0 computer program for further analysis. The recordings were performed in the UW-Madison Phonetics Laboratory using standard laboratory equipment.

Next, all tokens were edited out of the carrier sentence and underwent a series of measurements to select one token of each type for use in the perception test. The criteria for the final selection of the tokens were: VOT, duration of the target vowel, overall duration of the token, formant frequencies and fundamental frequency of the target vowel. The amplitude was adjusted after the final selection of the tokens. Two identical perception tests were then prepared, one with the tokens recorded by Speaker A, and one with the tokens recorded by Speaker B. There were 64 stimuli in each test (each of the eight tokens was repeated for eight representations) which were arranged in a random order with an interstimulus interval of 3 seconds. The testing stimuli were first prepared on longer CSpeech files and then transmitted to the tape for perceptual testing.

2.3 Acoustic analysis of the target vowel

The target vowel in the selected testing stimuli ranged between 78-82 msec in total duration. There was a short steady-state portion between the initial and final formant transitions ranging between 8-11 msec at which frequency measurements of F1, F2, and F3 were taken. Given our present state of the knowledge with regard to the invariance problem, it was assumed that the steady-state portion of the vowel was still the best interval where the transformation from acoustic to Bark-difference values could reliably be performed.[4] If the hypothesis of further transformation from Bark-difference to phonetic features as proposed by S&G is correct, it is unlikely that listeners track and compute the Bark-difference values throughout the entire formant trajectories. The steady-state portion of the vowel was already affected by the final transitions which was evident in formant measurements. The formant values in Herz were transformed into the Bark scale using the formula developed by Schroeder, Atal and Hall (1979) which was also used in the computation of prototypes in the Schwartz et al. 1997b study:

$$F_{Bark} = 7\ ArcSh(F_{Hz}/650).$$

Tables 1a (Speaker A) and 1b (Speaker B) show the measurents of formant frequencies in both Herz and Bark for each target vowel in production of both speakers.

		Formant freq. in Hz			Formant freq. in Bark		
Context	Vowel	F1	F2	F3	F1	F2	F3
d_t	/ʊ/	517	1529	3187	5,10	11,14	16,05
	/Y/	474	1981	2885	4,74	12,83	15,37
	/œ/	603	1938	2885	5,81	12,69	15,37
	/ɔ/	711	1443	3252	6,62	10,77	16,19
t_p	/ʊ/	560	1184	3122	5,46	9,53	15,91
	/Y/	495	1830	2735	4,92	12,31	15,01
	/œ/	689	1615	2735	6,46	11,49	15,01
	/ɔ/	711	1184	2993	6,62	9,53	15,62

Table 1a. Formant frequencies in Hz and in Bark for the target vowels in tokens recorded by Speaker A.

The Bark-difference measures between F2-F1 and F3-F2 in the tokens used in the study were compared with the prototypical values computed in the Schwartz et al. study, with the AE values for AE variants of the vowels /ʊ/ and /ɔ/ reported in S&G (in neutral context), and with the transformed from Herz- to Bark-scale values for northern German vowels (in neutral context) reported by Jørgensen (1969) for speaker NB.[5]

		Formant freq. in Hz			Formant freq. in Bark		
Context	Vowel	F1	F2	F3	F1	F2	F3
d_t	/ʊ/	366	1249	2067	3,76	9,86	13,12
	/Y/	366	1658	2304	3,76	11,66	13,85
	/œ/	517	1357	2261	5,10	10,38	13,72
	/ɔ/	560	1034	2433	5,46	8,71	14,21
t_p	/ʊ/	345	1077	2067	3,56	8,95	13,12
	/Y/	366	1615	2196	3,76	11,49	13,52
	/œ/	538	1335	2239	5,28	10,27	13,65
	/ɔ/	581	991	2606	5,63	8,46	14,68

Table 1b. Formant frequencies in Hz and in Bark for the target vowels in tokens recorded by Speaker B.

Table 2 summarizes the Bark-difference measures.

			In the study		Prototypical		A. English		German	
Vowel	Context	Speaker	F2-F1	F3-F2	F2-F1	F3-F2	F2-F1	F3-F2	F2-F1	F3-F2
/ʊ/	d_t	A	6.04	4.92	2.50	8.11	5.0	5.5		
	t_p	A	4.07	6.39						
	d_t	B	6.10	**3.26**			4.4	5.2	3.63	6.75
	t_p	B	5.39	4.17						
/ɔ/	d_t	A	**4.15**	5.43	2.50	6.04	2.5	7.1		
	t_p	A	2.91	6.10						
	d_t	B	3.25	5.50			2.1	6.8	3.35	5.61
	t_p	B	2.83	6.22						
/Y/	d_t	A	8.09	2.54	8.54	1.54				
	t_p	A	7.39	2.70						
	d_t	B	7.90	2.18					7.16	2.74
	t_p	B	7.73	2.03						
/œ/	d_t	A	6.88	2.68	5.60	3.30				
	t_p	A	5.03	**3.52**						
	d_t	B	5.28	3.34					5.59	3.25
	t_p	B	4.99	3.38						

Table 2. Bark-difference measures for formant frequencies of vowels used in the study. Bark-differences less or more than the critical distance of 3.5 Bark in the relevant F2-F1 and F3-F2 dimension are bold-faced. See text for further details.

As expected, the /ʊ/ in the production of both speakers did not show the F1-F2 convergence within the 3.5 Bark in any of the contexts. However, the d_t context additionally fronted it in Speaker B production which resulted in the F2-F3 convergence of 3.26 Bark being a focal point of the front rounded vowel. It will be of particular interest if subjects of the study would classify the fronted /ʊ/ as /Y/ and, if this was the case, would they further detect the difference between the "proper" /Y/ in this environment (F2-F3 = 2.18 Bark). Two additional measures exceeded the 3.5 Bark distance in Speaker A production: in /ɔ/ (d_t context, F1-F2 = 4.15) and in /œ/ (t_p context, F2-F3 = 3.52). Similarly, it will be of interest if the identification pattern would reflect these out of range distances.

2.4 Subjects and experimental procedure

51 subjects participated in the study. 43 of the subjects were native speakers of American English. They differed in their experience with German as an L2 ranging from beginning undergraduate students to German instructors at the UW-Madison. Based on their German background, they were divided into three groups: Beginner, Intermediate, and Advanced. The characteristics of all subject groups are summarized in Table 3. In addition, 8 native speakers (NS) of German served as a control group. The German subjects were university exchange students who arrived in the US two month prior to their participation in the experiment. All of them were from central Germany and spoke English at different proficiency levels.

Group	No. of subjects	Years of Education in German	Years of Residence in a German-speaking Country
Beginner	19 (M12, F7)	0.4 (0.1)	0.0 (0.0)
Intermediate	12 (M6, F6)	3.9 (1.6)	2.2 (3.5)
Advanced	12 (M6, F6)	18.5 (10.7)	2.9 (1.9)

Table 3. Characteristics of L2 subjects. Standard deviations are in parentheses.

The testing procedure took place in the Phonetics Laboratory where all subjects were individually tested. Since a matching experiment cannot prove the existence of an LSI mechanism (Schwartz & Escudier, 1989), the subjects identified the vowels in an open task not being explicitly asked to match them with any L1 or L2 vowel category. They were instructed to identify the vowel in the first stressed syllable of each token according to their own judgment and to write it down on a testing form using German orthographic symbols. All subjects were familar with German orthography and were able to assign a symbol to a particular spoken vowel. Each subject was tested at the same comfortable listening distance from two loudspeakers. After responding to the first test (Speaker A), the subjects was given a five-minute break after which the second test (Speaker B) was presented.

3 Results

The two perception tests were performed to investigate whether L2 learners of German were sensitive to the critical distance of 3.5 Bark between the F2-F1 and the F3-F2 convergence. The variation of Bark-difference measures due to the consonantal context and to the inter-speaker variability served as determinants of whether the LSI process was active in perceptual classification of L2 vowels. Three groups of L2 learners and one group of NS served as subjects to investigate whether and how the perceptual reaction changes at different stages of L2 development and how it compares with vowel classification by the NS.

In general, ANCOVA showed significant effects of the factors context, speaker, and experience with L2 on identification of most of the vowels. Length of residence in a German-speaking country (included as a covariate) did not have statistically significant effect on vowel perception by L2 learners. Specifically:

(a) context effect: highly significant main effect in /Y/ ($F(1,156)=115.181$, $p<.001$) and /ɔ/ ($F(1,156)=123.215$, $p<.001$); in the case of /ʊ/, the interaction context*language experience was strong and significant ($F(2,156)=104.798$, $p<.001$) which overrode the main effect ($p=.1$); not significant in /œ/ ($p=.067$).

(b) speaker effect: highly significant main effect in /œ/ ($F(1,156)=17.424$, $p<.001$) and /ɔ/ ($F(1,156)=104.926$, $p<.001$); two interactions speaker*context were marginally significant for /ɔ/ ($F(1,156)=28.488$, $p=.031$) and /ʊ/ ($F(1,156)=27.68$, $p=.034$), and the interaction speaker*language experience was slightly significant in /œ/ ($F(2,156)=16.71$, $p=.04$).

(c) experience with L2: highly significant main effect in /Y/ ($F(2,156)=49.048$, $p<.001$) , /œ/ ($F(2,156)=139.5$, $p<.001$), and /ɔ/ ($F(2,156)= 51.046$, $p<.001$); not significant in /ʊ/ ($p=.518$).

3.1 Test 1: Speaker A

Table 4 summarizes the results obtained in Test 1 recorded by Speaker A. The identification of the vowels is given by percentage correct. The vowels in parentheses represent at least 50 percent of all substitutions made by the subjects for an intended vowel category. The bold-faced data correspond to the bold-faced Bark values reported in Table 2. The results are discussed below for each vowel separately.

/ʊ/: In both contexts, there was no clear focal point which would perceptually mark the vowel as back. In the d_t context, the F1-F2 convergence far exceeded the 3.5 Bark distance (6.04 Bark) while in the t_p context (which considerably lowered F2), the Bark-difference was closer to the critical distance but still above it (4.07 Bark). Native speakers were sensitive to the differences between both contexts: the vowel in d_t sounded too fronted for the German /ʊ/ and was frequently misidentified as /Y/ whereas in t_p, the vowel reached the ceiling identification as /ʊ/. It seems that in the case of an ambigious formant configurations, both variants of the same vowel were perceived relative to each other. The same tendency was observed in L2 learners. In d_t, the identification of the vowel as /ʊ/ decreased with language experience being substituted by /Y/ and the t_p context showed exactly the reverse order. Recall that the interaction context*language experience was highly significant. Interestingly, the less experienced subjects substituted the vowel /ʊ/ by /ɔ/ in the t_p context which indicates that the back vowel contrast /ʊ/-/ɔ/ was poorly perceived when both vowels showed similar Bark-difference level.

		Subject groups			
Vowel	Context	Beg	Int	Adv	NS
/ʊ/	d_t	63	40 (Y)	34 (Y)	54 (Y)
	t_p	14 (ɔ)	41 (ɔ)	77 (Y)	100
/ɔ/	d_t	**45** (ʊ)	**57** (ʊ)	**44** (ʊ)	**79** (a)
	t_p	81	88 (ʊ)	92 (ʊ)	97
/Y/	d_t	37 (ʊ)	33 (I)	61 (I)	80 (I)
	t_p	53	56 (ɔ)	81 (I)	95
/œ/	d_t	19 (ɛ)	34 (ɛ)	54 (ɛ)	95
	t_p	**6** (ʊ)	**22**	**27** (ʊ)	**60**

Table 4. Results of Test 1 (Speaker A)

/ɔ/: The context played an important role in classification of the vowel in that it created two different patterns of responses. In d_t, the low correct identification (around 45%) generally did not improve with language experience being also low among the native German speakers (79%). The substitution pattern showed a high level of confusion with /ʊ/. In the t_p context, the identification was very high already among the beginners (81%), improved with language experience, and reached 97% by native speakers. A closer look at vowel formants made it clear that the identification pattern was linked with the critical distance: in the d_t context, the F2-F1 convergence exceeded the 3.5 Bark (4.15 Bark) while in the t_p context it properly stayed within its range (2.91 Bark).

/Y/: In the case of this vowel, the results are interesting for two reasons. First, since in both environments the vowel falls entirely within the 3.5-Bark critical distance with regard to the F3-F2 convergence, high identification rate and uniformity of judgments across both contexts are to be expected. However, there is a reason why this conclusion may not be correct. In the d_t context, the 'perceptual position' of /Y/ could be occupied by the unusually fronted /ʊ/ which was misidentified as /Y/ by a large number of subjects. Two possibilities arise with regard to the latter: the /ʊ/-/Y/ contrast will not be perceived in the d_t context, or it will be perceived by 'placing' the /Y/ in more fronted position relatively to the fronted /ʊ/. The identification data show that this was indeed the case. Subjects in the Beginner group had difficulties in perceiving the /ʊ/-/Y/ contrast as seen in high substitutions of /Y/ by /ʊ/. However, more advanced learners and native speakers detected the contrast by identifying /Y/ as /I/ whose focal point is also the F3-F2 convergence within the 3.5 Bark. As already reported, language experience was a significant factor in identification of /Y/. In the t_p context, the data show a higher rate of correct classification.

/œ/: In general, the identification of the vowel was low among the L2 learners. However, in the t_p context, its correct classification was especially low reaching 27% in the Advanced and 60% in the NS groups. Again, closer look at the F3-F2 convergence shows that the Bark-difference slightly exceeded the 3.5-Bark critical distance (3.52 Bark) which caused not only the low identification rate but the confusion of the vowel with /ʊ/ by some of the subjects. In the d_t context, the substitution pattern shows its misidentification as /ɛ/ which can be attributed to its proximity to /ɛ/ in terms of the F3-F2 convergence.

3.2 Test 2: Speaker B

Table 5 presents the summary of results obtained in Tests 2 recorded by Speaker B. The identification of the vowels is given by percentage correct. As in Table 4, the vowels in parentheses represent at least 50 percent of all substitutions made by the subjects for an intended vowel category. The bold-faced data correspond to the bold-faced Bark values reported in Table 2.

		Subject groups			
Vowel	Context	Beg	Int	Adv	NS
/ʊ/	d_t	**53**	**41 (Y)**	**25 (Y)**	**8 (Y)**
	t_p	50	54 (Y)	66 (Y)	80 (Y)
/ɔ/	d_t	57	75	99	100
	t_p	72	94	99	100
/Y/	d_t	34 (ʊ)	57	46 (I)	88 (I)
	t_p	53 (ʊ)	69	84 (I)	100
/œ/	d_t	22	39	73	100
	t_p	18	38	78	100

Table 5. Results of Test 2 (Speaker B)

The focus of this section is particularly on the differences and similarities of the results with the Test 1 data.

The most drastic difference can be observed in the identification of /ʊ/ in the d_t context, where in Test 2, the correct classification of the vowel progressively decreased with language experience reaching only 25% in the Advanced and 8% in the NS group. Although the /ʊ/ was fronted in the southern German of Speaker B similarly to the German spoken by Speaker A, the d_t context additionally raised its F2 causing the F3-F2 convergence to fall within the 3.5-Bark critical distance (3.26 Bark) which is a focal point of the front rounded vowel. Thus, the more experienced German learners were correct by classifying the vowel as /Y/ based on their standard German judgments. However, the speaker also produced the /Y/ which existed and functioned in his dialect unaffected by the frontness of /ʊ/. What was the perceptual reaction of listeners to his /Y/ in the d_t environment? As in Test 1, the subjects perceived both vowels relative to each other. First, they had difficulties in perceiving the /Y/-/ʊ/ contrast (Beginner) and, once they have it detected, they tended to label /Y/ as /I/ and /ʊ/ as /Y/. Similar "vowel shift" was observed in Test 1. In the t_p context, the /ʊ/ sounded still ambiguous to more experienced subjects who did not perceive it clearly as /ʊ/ and even NS reached only 80 % of accuracy.

A second difference between the results in both tests is found in /ɔ/ in the d_t context. Since the Bark-difference was in its proper range, the identification of the vowel was higher and it was not confused with /ʊ/ as in the Test 1 data. One more difference can be attributed to the critical distance criterion. Unlike in Test 1, the identification of /œ/ in the t_p context increased with language experience and showed similar developmental pattern with the vowel in the d_t context. Bark-difference measures indicate that both vowels fell well within their proper range in both F2-F1 and F3-F2 convergence zones.

4 Discussion

The question to be addressed now is whether the overall pattern of results gives support to the existence of an LSI process. The subset of vowels investigated in the study was naturally produced by two speakers of two varieties of southern standard German. Additionally, the vowels were embedded in two environments which, affecting their formant frequencies, allowed us to observe subjects' perceptual reaction and sensitivity to contextual variability of a vowel. The subjects identified the vowels in an open task according to their own judgments. There was a clear pattern of identification which showed that whenever the speaker exceeded the 3.5 Bark critical distance in a particular formant convergence zone, the listener was sensitive to it. This sensitivity progressively increased as the L2 learners gained experience with German. Results did demonstrate that the LSI mechanism was at work in perceptual identification of the vowels.

The present data cannot be analyzed from the perspective of the L2-L1 category assimilation as outlined in Best's Perceptual Assimilation Model (Best 1994) without conducting further discrimination experiments. The goal of the present experiment was different too in that it aimed to explain the very basis for labelling of the discriminated contrast. The question asked was not whether the sound was 'same' or 'different', or was a better or worse exemplar of a given vowel category. Rather, the subject was asked to tell exactly what was the sound she was hearing in her L2. Based on the data, it seems less likely that the subjects in this study referred to their L1 at all. If this was the case, why would the /ʊ/ show such a low level of identification given that the AE /ʊ/ is more fronted than that in standard German and, additionally, does not have clear formant convergence? Did the /ʊ/ in the present study not meet the criteria of the AE /ʊ/? Moreover, why would both contexts yield such a diverse level of identification of the same category /ʊ/ and why would the subjects label it as /Y/ in one environment and as /ɔ/ in the other? The case of /ʊ/ clearly shows that L2 vowels may not be perceived with reference to L1 if presented within a sub-system and not as an isolated vowel contrast. Kingston et al. (1996) concluded along similar lines: "Our adult American English listeners may (...) approximate the German vowel categories by responding to the physical dissimilarity among the German vowels themselves rather than by gauging the physical similarity between the German and American English vowels". The present results support this observation.

Data from native speakers of German in the present experiment show that dialectal differences (dialect-specific tuning) play a role in vowel identification. However, it is impossible to explain in general terms of frontness/backness of a given vowel within the system why the /ʊ/ in Test 2 was correctly classified only by 8% of the listeners in the d_t context, whereas its identification was much higher (80%) in the t_p context. Clearly, environment played an important role here. It can be inferred that F3-F2 convergence and the 3.26-Bark distance served as a perceptual focal point, a mark of a *front* rounded vowel which was audible enough to yield the pattern of 8% identifications of the vowel as back. This finding leads to the issue of preference of specific sound configurations in sound system of a language. The fact that the /d/- lax vowel- /t/ context is absent in lexical distributions in German may be dictated by the physics of speech and need for sufficient contrast in sound sequences used in the lexicon.

With regard to the suggestion in S&G's study that context-independent normalization occurs through phonetic feature classification, the present data showed that identification of the vowels is highly context-dependent in that the environment affects formant frequencies and, consequently, the Bark-difference level. The critical distance is not an independent criterion but an outcome of vowel perturbation caused by coarticulatory effects in natural speech. It seems that context plays a role in the normalization process although the exact way how it is put to use remains to be investigated.

Does the operation of an LSI process imply that an absolute invariance in the speech signal is needed to perform the Herz-to-Bark transformation in order to transmit the essential information from the acoustic signal to the phonetic code? Not necessarily. Ohala and Feder (1994:117) suggest that the influence of environment can be shaped by linguistic experience in that "(...) the final, more precise identification can be based on expectations of how already identified surrounding sounds would be likely to *color* the shape of the target segment", emphasis added. Although we can only speculate at present, it might be the case that the integration mechanism which operates at the intermediate level of signal processing provides but one piece of information which, acting as a "color", adds a necessary component to the identification of the vowel. The interpretation of vowel formant configurations by an LSI mechanism as a function of surrounding consonants may occur simultaneously with other modes of vowel processing and it is most likely that the final decision with regard to vowel identity and its position within the system occurs at a higher level of signal processing. The present state of the knowledge does not allow further speculation on this issue.

5 Conclusion

The observed pattern of identification leads to the conclusion that the speaker-listener mechanism is utilized in the same way in mentally shaping the native and non-native vowel systems as tested in perception of selected vowels by adults listeners. The results of the study cannot be understood without assuming the existence of an LSI process used in the auditory processing to transmit an essential information from the signal to the phonetic code.

Notes

* My thanks go to Joseph Salmons for his assistance and suggestions and to Jerzy Jacewicz for discussions on the physics of sound. All errors are my responsibility.

[1] See Schwartz et al. 1997a for a description of tendencies found in vowel systems.

[2] For a detailed discussion on whether LSI is present in human auditory system as well as for experiments, see Schwartz and Escudier (1987, 1989).

[3] Schwartz et al. (1997b: 282) note "(...) we must admit that knowledge of vowel representations in the auditory system is not sufficient to allow a completely deductive approach for the prediction of vowel systems." For the purpose of the present study, however, the data-driven approach provides a more desirable basis for proving the existence of the LSI process. The testing material consisted of natural speech stimuli which were "taken out" of existing vowel systems, i.e., two dialects of southern German. Thus, the data-driven predictions were checked against real data.

[4] This reasoning is in line with the (elaborated) target model of vowel perception which takes into account both speaker normalization and coarticulatory effects (e.g., Miller 1989, Syrdal 1985, Syrdal & Gopal 1986). The proponents of the dynamic specification approach showed that perceptually relevant dynamic information is specified over the initial and final transitions of a vowel surrounded by consonants as measured in CVC syllables (Jenkins & Strange 1987, Strange 1989, Verbrugge & Rakerd 1987, among others) and that the effect of transitions was evident in the so called silent-center paradigm (e.g. Strange 1989). However, there was also evidence that listeners were able to detect the influence of consonantal context when vowel transitions had been experimentally removed leaving the center only (Ohala & Feder 1994, Ohala, Riordan & Kawasaki 1978, Thorburn 1996).

[5] Although the F2-F1 Bark-difference in northern speaker's NB production exceeds the critical distance of 3.5 Bark, data in Kohler (1995) as well as the measurements in a forthcoming study by the present author indicate that the Bark-difference in a standard northern German variant of /ʊ/ is below the critical distance.

References

Best, C.T. 1994. The emergence of native-language phonological influences in infants: a perceptual assimilation model. In The development of speech perception: the transition from speech sounds to spoken words, ed. by J. Goodman and H.C. Nusbaum, pp. 167-224. Cambridge, MA: MIT Press.

Bladon, A. 1983. Two-formant models of vowel perception: shortcomings and enhancements. Speech Communication 2.305-313.

Carlson, R., B. Granström, and G, Fant. 1970. Some studies concerning perception of isolated vowels. Speech Transmission Laboratory: Q. Prof. Stat. Rep. 2-3.19-35.

Carlson, R., G. Fant, and B. Granström. 1975. Two-formant models, pitch and vowels perception. In Auditory analysis and perception of speech, ed. by G.Fant, and M.A. Tathem, pp. 55-82. London: Academic Press.

Chistovich, L.A., and V.V. Lublinskaya. 1979. The center of gravity effect in vowel spectra and critical distance between the formants. Hearing Research 1.185-195.

Chistovich, L.A., R.L. Sheikin, and V.V. Lublinskaya. 1979. 'Centers of gravity'and the spectral peaks as the determinants of vowel quality. In Frontiers of speech comunication research, ed. by B. Lindblom, and S. Öhman, pp. 143-158. London: Academic Press.

Crystal, T.H., and A.S. House. 1988. The duration of American-English vowels: an overview. Journal of Phonetics 16.263-284.

Delattre, P., A.M. Liberman, F.S. Cooper, and J. Gertsman. 1952. An experimental study of the acoustic determinants of vowel color: observations on one- and two-formant vowels synthesized from spectrographic patterns. Word 8.195-210.

Jenkins, J. J., and W. Strange. 1987. Identification of "hybrid" vowels in sentence context. Journal of the Acoustical Society of America 82.S82.

Jørgensen, H.P. 1969. Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantenfrequenzen. Phonetica 19.217-245.

Kingston, J., C. Bartels, J. Benki, D. Moore, J. Rice, R. Thorburn, and N. A. Macmillan. 1996. Learning non-native vowel categories. Amherst: University of Massachusetts, ms.

Kohler, K. J. 1995. Einführung in die Phonetik des Deutschen. Berlin: Erich Schmidt Verlag.

Liljencrants, J. and B. Lindblom. 1972. Numerical simulations of vowel quality systems: The role of perceptual contrast. Language 48.839-862.

Lindblom, B. 1986a. Phonetic universals in vowel systems. In Experimental phonology, ed. by J. Ohala and J. Jaeger, pp. 13-44. New York: Academic Press.

Lindblom, B. 1986b. On the origin and purpose of discreteness and invariance in sound patterns. In Invariance and variability of speech processes, ed. by J. Perkell, and D. Klatt, pp. 493-510. New Jersey: Lawrence Erlbaum.

Lindblom, B. 1990. On the notion of possible speech sound. Journal of Phonetics 18.135-152.

Maddieson, I. 1984. Patterns of sounds. Cambridge: Cambridge University Press.

Miller, J.D. 1989. Auditory-perceptual interpretation of the vowel. Journal of the Acoustical Society of America 85.2114-2134.

Ohala, J., and D. Feder. 1994. Listeners' normalization of vowel quality is influenced by 'restored' consonantal context. Phonetica 51.111-118.

Ohala, J., C.J. Riordan, and H. Kawasaki. 1978. The influence of consonantal environment upon identification of transitionless vowels. Journal of the Acoustical Society of America 64.S18.

Peterson, G.E., and H. Barney. 1952. Control methods used in a study of the vowels. Journal of the Acoustical Society of America 24.175-184.

Schroeder, M.R., B.S. Atal, and J.L. Hall. 1979. Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In Frontiers of speech communication research, ed. by B. Lindblom, and S. Öhman, pp. 217-299. London: Academic Press.

Schwartz, J.-L., L.-J. Boë, N. Vallée, and C. Abry. 1997a. Major trends in vowel system inventories. Journal of Phonetics 25.233-253.

Schwartz, J.-L., L.-J. Boë, N. Vallée, and C. Abry. 1997b. The dispersion-focalization theory of vowel systems. Journal of Phonetics 25.255-286.

Schwartz, J.-L. and P. Escudier. 1987. Does the human auditory system include large scale spectral integration? In The psychophysics of speech perception, ed. by M.E.H. Schouten, pp. 284-292. Dordrecht: Martinus Nijhoff Publishers.

Schwartz, J.-L., and P. Escudier. 1989. A strong evidence for the existence of a large-scale integrated spectral representation in vowel perception. Speech Communication 8.235-259.

Stevens, K.N. 1997. Articulatory-acoustic-auditory relationships. In The handbook of phonetic sciences, ed. by W.J. Hardcastle, and J. Laver, pp. 462-506. Oxford: Blackwell.

Strange, W. 1989. Dynamic specification of coarticulated vowels spoken in sentence context. Journal of the Acoustical Society of America 85.2135-2153.

Syrdal, A.K. 1985. Aspects of a model of the auditory representation of American English vowels. Speech Communication 4.121-135.

Syrdal, A.K., and H.S. Gopal. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. Journal of the Acoustical Society of America 79.1086-1100.

Thorburn, R. 1996. Vowel perception in consonantal context. Journal of the Acoustical Society of America 99.2589.

Verbrugge, R.R., and B. Rakerd. 1986. Evidence of talker-independent information for vowels. Language and Speech 29.39-57.