PRODUCTION OF VOWEL CONTRASTS IN NORTHERN STANDARD GERMAN AND AUSTRIAN STANDARD GERMAN

Hans Georg Piroth, Peter Skupinski and Bernd Pompino-Marschall

Humboldt-Universität zu Berlin, Berlin, Germany
hans.georg.piroth@hu-berlin.de

ABSTRACT

The inventory of 15 German full vowels realized by speakers of Northern Standard German (NSG) and Austrian Standard German (ASG) is investigated acoustically. The vowels were produced in phonotactically regular nonsense words (bVp and dVt environments) embedded in a sentence frame.

The analyses revealed that the main difference between both varieties can be found in a different acoustic realization of the tense-lax contrast in high (especially front) vowels. With respect to spectral as well as temporal structure ASG lax high vowels are very much closer to their tense counterparts than in NSG. For the tense-lax differentiation here the greater long/short ratio as well as a higher f_0 for lax vowels in ASG are much stronger cues than in NSG.

Keywords: vowel systems, tenseness, Standard German varieties.

1. INTRODUCTION

Phonemic vowel systems as cognitive systems can be characterized as n-dimensional spaces of a categorical nature. Established phonological dimensions in the description of vowel systems are tongue height and tongue position, perceptually based on scales of *sonority* and *chromaticity*, and acoustically signaled – in a first approximation – by first and second formant frequency, respectively. In the system of Cardinal Vowels according to Jones [4] these dimensions – together with the intervening articulatory dimension of lip rounding – constitute a vowel space in form of perceptually equidistant vowel exemplars also characterized by maximal *dispersion*. The smallest systems seem to be 3-vowel systems employing 'primary cardinal' contrasts, e.g. /i – a – u/, /e – a – o/. A majority of the world's languages shows vowel inventories of five items, e.g. /i – e – a – o – u/ and lacking central vowels (cf. [6]). Larger vowel systems (n >> 5) normally involve additional parameters (e.g. length, nasality etc.). Such complex vowel systems can be subdivided into subsystems serving as "natural classes" with respect to phonological rules. Thus, in the feature system of SPE [2] the feature [±tense] is provided as an additional manner of articulation feature. This feature is claimed to be necessary besides the cavity features, i.e. the tongue-body features [±high], [±low] and [±back] and the feature [±round], e.g. to differentiate between modern German vowel pairs.

This study aims at carrying-on the effort to recover the set of necessary and sufficient description parameters for the Standard German vowel system. For this purpose two regional varieties – Northern Standard German as spoken in the northern part of Germany (NSG) and Austrian Standard German as spoken in the Vienna region of Austria (ASG) – are acoustically analyzed and compared. Both systems contain the same number of vowel categories, yet differing in the realization of the tense/lax contrast.

2. METHOD

2.1. Subjects and material

To yield a sufficiently large corpus of vowel utterances 8 native speakers of NSG (4 male, 4 female, mean age 32.5 yrs, s = 7.68) and 8 speakers of ASG (4 male, 4 female, mean age 36.13 yrs, s = 12.10) were recorded under laboratory conditions (at 16 bit resolution, 22050 Hz sampling rate).

Each speaker produced 14 repetitions of all 15 German full vowels [iː, yː, eː, øː, aː, oː, uː, ɪ, ʏ, ɛ, œ, a, ɔ, ʊ] in bVp and dVt contexts embedded in a sentence frame *"Ich habe ___ gesagt" ("I said ___")*. The selection of these contexts is owed to the fact that the obligatory process of final obstruent devoicing does not allow voiced stops in syllable-final position. The stimuli were presented in randomized order in a PowerPoint prompting sequence. After the production of a sentence the next prompt was delivered by the experimenter with a minimum interval of 3s. The test stimuli were given in German orthography.

2.2. Acoustic analyses

The recorded material was segmented and labeled with the PRAAT 4.6 speech analysis software [1]. Vowel durations were determined as the intervals between start and end of the voiced part of F2. Formant frequency measurements were done by means of an LPC analysis (10 coefficients, 25ms analysis window in 5ms steps with a pre-emphasis of 6 dB/octave above 50 Hz) on the down-sampled signal (11025 Hz for female, 9200 Hz for male voices). Target frequencies for F1, F2 and F3 were calculated as the mean formant frequencies between the 40% and the 60% point of vowel duration in the resulting PRAAT formant object. Fundamental frequencies for each vowel were determined as mean f_0 values over the complete vowel duration in a PRAAT pitch object (10ms window). In this way the first 10 analyzable productions of each vowel in each context were analyzed per speaker. In a few cases formant values had to be manually determined after inspection of the spectrogram. Thus, in total a corpus of 4800 vowels analyzed for duration, f_0, F1, F2 and F3 was made available.

In a second step all frequency values were normalized to reduce inter-individual (especially gender) differences. This was achieved by applying the vowel-extrinsic and speaker-intrinsic version of the Lobanov algorithm as provided by NORM [7] given in Equation (1)

(1) $$F_{n[V]}^{N} = (F_{n[V]} - Mean_n)/s_n$$

where $F_{n[V]}^{N}$ is the normalized value for $F_{n[V]}$ (i.e., for formant n of vowel V). $Mean_n$ is the mean value for formant n and s_n is the standard deviation for the n^{th} formant over all repetitions of all vowel categories of the individual speaker. f_0 values were treated accordingly.

3. RESULTS

3.1. Raw data

The mean duration of stressed long tense vowels is 146ms for the NSG corpus, 182ms for ASG. Stressed short lax vowels show a duration of 79ms on the average for NSG and of 75ms for ASG, resulting in long/short ratios of 1.8 (NSG) and 2.4 (ASG). Table 1 gives the raw f_0, F1, F2, F3 frequencies as well as the durational values for the individual vowel categories.

Besides the trivial anatomy-based facts that female speakers show (i) higher f_0-values and (ii) an F1/F2-plane shifted towards higher frequencies, f_0 decreases with vowel height. On the other hand, vowel duration increases when vowel height is lowered as well as from front to back.

3.2. Normalized data

To allow a gender-independent view of the positioning of the vowel categories in the F1/F2-plane Figure 1 shows the Lobanov-normalized formant values for the individual vowel categories as an F1/F2-plot (NSG vs. ASG) rescaled to Hz-like values according to Equation (2).

Table 1: Average fundamental and formant frequencies [Hz] and vowel durations [ms] (standard deviations in parentheses) of Northern Standard German (NSG) and Austrian Standard German (ASG) vowels pooled over bVp/dVt contexts for male/female speakers.

NSG/male	**f_0**		**F1**		**F2**		**F3**		**Duration**	
[iː]	127	(26)	246	(35)	2151	(138)	3083	(188)	118	(20)
[eː]	116	(19)	316	(31)	2092	(130)	2709	(142)	151	(24)
[ɛː]	110	(19)	504	(38)	1794	(121)	2522	(132)	176	(26)
[aː]	115	(28)	706	(70)	1227	(95)	2484	(134)	184	(25)
[oː]	118	(20)	345	(30)	622	(72)	2560	(168)	158	(28)
[uː]	126	(25)	290	(26)	705	(107)	2258	(180)	136	(25)
[yː]	125	(25)	250	(23)	1697	(194)	2132	(208)	130	(23)
[øː]	115	(18)	327	(27)	1432	(127)	2164	(156)	158	(23)
[ɪ]	132	(47)	349	(35)	1749	(118)	2449	(125)	66	(10)
[ɛ]	122	(35)	482	(46)	1680	(112)	2445	(120)	84	(14)
[a]	124	(41)	622	(56)	1295	(144)	2423	(108)	88	(17)
[ɔ]	128	(56)	495	(32)	997	(123)	2433	(139)	84	(14)
[ʊ]	135	(49)	366	(26)	956	(195)	2459	(186)	72	(12)
[ʏ]	129	(41)	340	(41)	1455	(151)	2296	(187)	75	(12)
[œ]	130	(59)	457	(32)	1383	(113)	2325	(130)	89	(16)
ASG/male	**f_0**		**F1**		**F2**		**F3**		**Duration**	
[iː]	143	(36)	252	(30)	2131	(64)	3044	(114)	175	(52)
[eː]	128	(34)	304	(28)	2177	(114)	2760	(198)	209	(58)
[ɛː]	123	(35)	466	(42)	1904	(131)	2475	(111)	227	(58)
[aː]	116	(34)	697	(56)	1154	(63)	2386	(95)	248	(63)
[oː]	129	(36)	353	(29)	646	(67)	2443	(96)	226	(60)
[uː]	144	(38)	293	(38)	671	(85)	2364	(230)	203	(61)
[yː]	144	(37)	259	(29)	1632	(126)	2059	(186)	200	(57)
[øː]	129	(35)	319	(27)	1405	(126)	2109	(126)	228	(58)
[ɪ]	152	(34)	301	(41)	1983	(83)	2680	(193)	64	(14)
[ɛ]	131	(32)	456	(36)	1809	(112)	2443	(135)	82	(17)
[a]	124	(32)	673	(59)	1188	(81)	2351	(96)	94	(20)
[ɔ]	131	(33)	506	(41)	875	(110)	2435	(99)	86	(18)
[ʊ]	154	(34)	338	(37)	873	(181)	2347	(129)	69	(14)
[ʏ]	151	(36)	314	(42)	1520	(152)	2127	(160)	71	(13)
[œ]	131	(32)	474	(41)	1370	(119)	2202	(150)	89	(20)
NSG/female	**f_0**		**F1**		**F2**		**F3**		**Duration**	
[iː]	222	(11)	251	(26)	2559	(183)	3503	(186)	113	(22)
[eː]	210	(10)	373	(22)	2585	(189)	3245	(211)	144	(29)
[ɛː]	197	(8)	626	(64)	2109	(147)	2920	(137)	167	(29)
[aː]	193	(6)	835	(95)	1364	(103)	2856	(169)	180	(34)
[oː]	209	(11)	433	(47)	849	(96)	2904	(180)	160	(34)
[uː]	223	(12)	317	(53)	815	(102)	2862	(116)	130	(33)
[yː]	220	(9)	257	(28)	1883	(149)	2568	(245)	123	(31)
[øː]	209	(11)	390	(32)	1623	(133)	2538	(207)	156	(39)
[ɪ]	221	(11)	416	(29)	2048	(225)	2864	(233)	63	(15)
[ɛ]	204	(9)	584	(40)	2020	(136)	2910	(220)	85	(18)
[a]	201	(8)	771	(92)	1452	(149)	2837	(343)	91	(18)
[ɔ]	205	(8)	583	(49)	1087	(94)	2807	(215)	90	(21)
[ʊ]	220	(12)	421	(39)	1014	(167)	2876	(164)	71	(18)
[ʏ]	219	(10)	403	(34)	1618	(222)	2643	(210)	68	(16)
[œ]	205	(8)	546	(44)	1601	(141)	2634	(230)	88	(22)
ASG/female	**f_0**		**F1**		**F2**		**F3**		**Duration**	
[iː]	218	(31)	264	(20)	2571	(77)	3515	(146)	128	(35)
[eː]	197	(27)	362	(35)	2589	(82)	3251	(212)	153	(42)
[ɛː]	186	(29)	618	(64)	2052	(118)	2833	(164)	176	(49)
[aː]	177	(26)	903	(84)	1388	(83)	2780	(94)	191	(41)
[oː]	199	(31)	442	(47)	761	(109)	2974	(161)	161	(49)
[uː]	223	(36)	326	(49)	804	(122)	2847	(157)	143	(43)
[yː]	222	(33)	281	(31)	1828	(158)	2524	(124)	132	(40)
[øː]	198	(31)	393	(52)	1629	(97)	2535	(94)	158	(43)
[ɪ]	233	(36)	350	(63)	2318	(106)	2924	(155)	58	(8)
[ɛ]	193	(27)	549	(52)	2059	(88)	2819	(169)	77	(10)
[a]	185	(30)	846	(61)	1434	(99)	2709	(139)	86	(9)
[ɔ]	193	(29)	591	(45)	1054	(112)	2886	(168)	79	(13)
[ʊ]	231	(33)	410	(61)	942	(169)	2902	(136)	61	(10)
[ʏ]	229	(35)	362	(60)	1732	(128)	2585	(156)	60	(9)
[œ]	197	(32)	548	(48)	1613	(75)	2617	(119)	81	(12)

(2a) $$F1[Hz'] = 250+500\,(F^{N}_{1} - F^{N}_{1MIN})/(F^{N}_{1MAX} - F^{N}_{1MIN})$$
(2b) $$F2[Hz'] = 850+1400\,(F^{N}_{2} - F^{N}_{2MIN})/(F^{N}_{2MAX} - F^{N}_{2MIN})$$
(2c) $$F3[Hz'] = 2300+1000\,(F^{N}_{3} - F^{N}_{3MIN})/(F^{N}_{3MAX} - F^{N}_{3MIN})$$

where F^N_n (with $1 \le n \le 3$) is the normalized formant value for a particular vowel, F^N_{nMIN} and F^N_{nMAX} are the minimum and maximum normalized formant values over all repetitions of all vowel categories of all speakers of the variety concerned. f_0 is rescaled in an analogous way by means of Equation (3).

(3) $f_0[Hz'] = 100+50\ (f_0^N - f_0^N{}_{MIN})/(f_0^N{}_{MAX} - f_0^N{}_{MIN})$.

Figure 1: Lobanov-normalized F1/F2-plots (re-scaled to Hz´-values, 1σ-ellipses) for (a) Northern Standard German (NSG) and (b) Austrian Standard German (ASG) vowels pooled over bVp and dVt contexts based on 8 speakers of each variety.

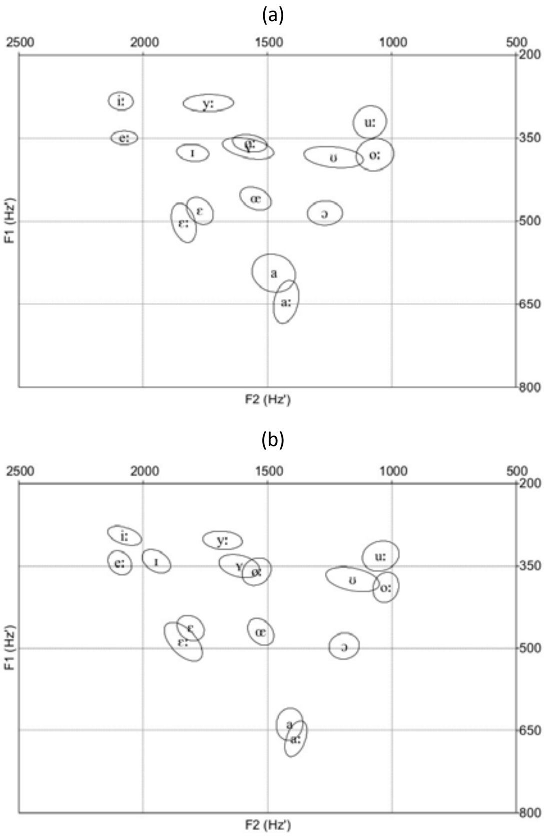

Figure 1 shows the positions of the vowel categories in the F1/F2-plane. For both varieties /o/, /uː/ and /ʊ/ in the low F1/low F2 area are very close. The same is true for /ɛː/ and /ɛ/ in the lower mid front area and for the low vowels /aː/ and /a/. In some cases the 1σ-ellipses even slightly overlap. Also /ʏ/ and /øː/ overlap in the F1/F2-plane although belonging to different vowel qualities. Slight differences between the last two categories can be revealed by taking F3 into account. Short (lax) vowels are more centralized than long (tense) ones. This tendency is found for lax /ɛː/ too. An exception seems to be tense /øː/ that exhibits F2-values close to /ʏ/ (at least in NSG) and /œ/ (in both varieties). This might be due to a stronger lip rounding in /øː/ than in /yː/. In general, F3 decreases with vowel height and in the front-to-back direction and is lower for rounded than for unrounded front vowels.

The most striking differences between both varieties appear in the front/high part of the vowel space with /ɪ/ being strongly shifted towards /iː/ in ASG as opposed to NSG and /ʏ/ and /yː/ showing the same tendency.

In order to evaluate these results statistical analyses were performed using a 2x2x2-factorial GLM analysis of variance in SPSS [8] on the factors VOWEL (tense, lax), CONTEXT (bVp, dVt) as within-subject, and VARIETY (NSG, ASG) as between-subject factor on each pair of tense vs. lax vowels (/ɛː/ was excluded from this analysis, since the Standard German vowel system is asymmetrical in containing this vowel as the only one that is long and lax besides the low vowel pair /aː a/). In the statistical evaluation this latter pair is treated as [±tense] (in contrast to most phonological descriptions of German that describe it as [–tense; ±long]). Dependent variables were mean f_0^N, mean $F1^N$, mean $F2^N$, mean $F3^N$, and vowel duration.

The analysis shows highly significant main effects ($p<.001$) in all pairs of VOWEL for $F1^N$ and for all $F2^N$ except /øː/-/œ/ ($p=.063$) and at least significant main effects ($p<.05$) for $F3^N$ in all contrasts, revealing that tense vs. lax vowel pairs are qualitatively different in all three formant dimensions. The factor CONTEXT yields significant main effects for $F1^N$ and $F2^N$ for all vowel contrasts except for $F1^N$ in /aː/-/a/ where $F1^N$ is highest, showing a nearly consistent influence of consonantal place of articulation on the formant target positioning. For $F3^N$ these effects are significant only for front vowels. A nearly consistent main effect of VARIETY for all formants is found only in the case of /iː/-/ɪ/ ($p<.01$ for $F2^N$ and $F3^N$, $F1^N$ shows only a trend). VOWEL*CONTEXT interactions are not found for /øː/-/œ/, but are else consistent for $F2^N$ in general, and lacking in $F1^N$ and $F3^N$ only for back vowels. These effects occur, because context-dependent differences in formant target frequencies are usually greater for lax than for tense vowels. On the other hand, CONTEXT does not interact with VARIETY except for one marginal case. There are significant interactions of VOWEL*VARIETY for /iː/-/ɪ/ and /yː/-/ʏ/ in all formants (with $F3^N$ failing to reach significance level for the rounded vowels but being close to it). This result reflects the finding that both so-called "lax" high front vowels in ASG are markedly shifted towards their tense counterparts. No systematic three-fold interactions on formant frequencies were detected. f_0^N provides consistent significant effects of CONTEXT for all vowel pairs and a highly significant main effect ($p<.001$) of

VOWEL for the /aː/-/a/ contrast in which no marked tenseness opposition is assumed as well as a significant main effect ($p<.05$) for high vowel pairs. Both factors do not interact for f_0^N. VARIETY delivers a slightly significant effect ($p<.05$) for the /yː/-/ʏ/ pair and a very significant interaction ($p<.01$) with VOWEL for both high front vowels as well as a slightly significant one for /uː/-/ʊ/ in which the tense/lax difference in the F1/F2-plane is minimized in ASG. For these pairs f_0 contributes more to the tense-lax distinction in ASG as opposed to NSG. The remaining interactions are not significant.

Concerning vowel duration, separate GLM analyses were calculated for the individual vowels. As vowel quantity correlates to the tense/lax opposition, the normal distribution condition could not be satisfied if data were pooled over tense/lax vowel pairs. Instead, GENDER (male, female) had to be introduced as a factor in the GLM analysis, since duration values were not normalized. Vowel duration consistently shows significant main effects for CONTEXT except for /ʏ/ where the duration difference is too small to reach significance. Vowels are longer in dVt context where the tongue is involved in stop production than in bVp context where the articulators for consonant and vowel production work independently. Only a single isolated main effect of VARIETY (/iː/) and no one of GENDER could be encountered. Interactions were not significant except for four marginal cases.

To determine the contribution of the parameters measured on the acoustic *gestalt* of the individual vowel categories discriminant analyses were calculated in SPSS [8] to retrieve the confusion matrices of the classification results. When all variables ($F1^N$, $F2^N$, $F3^N$, f_0^N and vowel duration) are fed into the analysis, classification is very sharp (94.8% for NSG, 95.3% for ASG). Even when relying on $F1^N$ and $F2^N$ only, the classification results reach 80.4% (maximum 98.8% for /ɛː/, minimum 53.8% for /ʏ/) in the NSG corpus and 78.4% for ASG (range 99.4% for /ɔ/ to 53.8% for /ʏ/). In this analysis for NSG /iː/ and /ɪ/ never were confused, /ʏ/ was misclassified as /yː/ in only 1.9% of all cases. Most classification errors occurred between /ʏ/ and /øː/ and between /ɛː/ and /ɛ/ as could be expected after inspection of Figure 1. In contrast, the analysis for ASG shows misclassifications of /iː/ for /ɪ/ (6.3%) and vice versa (8.1%) as well as /yː/ for /ʏ/ (16.9%) and /ʏ/ for /yː/ (7.5%). Inclusion of $F3^N$ into the analysis does not contribute much to the classification (NSG: 84.4%, ASG: 79.3%), but adding duration is more important (NSG: 94.6%, ASG: 92.7%) than adding f_0^N (NSG: 86.3%, ASG: (87.5%).

4. CONCLUSION

With respect to their vowel systems, both standard varieties of German studied here show the expected universal tendencies of intrinsic articulatorily caused acoustic variation as well as the context-dependent coarticulatorily based ones, i.e. (i) vowel duration is intrinsically dependent on vowel height (and backness) as well as on consonantal context, and (ii) vowel intrinsic f_0 is also dependent on vowel height (and backness; cf. [9]). The latter dependency can only be stated for the [+tense] and [–tense] vowels as separate groups and not by vowel height *per se* (cf. [5]).

Although the oppositions within the quite complex system of German stressed vowels of both Standard varieties analyzed here are phonetically structured in a parallel manner, there are some remarkable variety-specific differences with respect to subsystems of the entire system: The most basic categorization of vowel systems by the phonological features [±high] and [±low] resulting in the three height categories 'high' [+high, –low], 'mid' [–high, –low], and 'low' [–high, +low] is universally primarily cued by the F1-frequency. This categorization with respect to vowel height in both Standard varieties of German is further enhanced by f_0 as a height cue: The subsystem of the 'mid' pairs of 'close mid' and 'open mid' vowels (i.e. /e/-/ɛ/, /ø-/œ/ and /o/-/ɔ/ that are distinguished by the feature [±tense]) show equal f_0 values each, irrespective of their F1-differences. Here the latter are the phonetic cues for [±tense] (i.e. higher F1-values for [–tense]). In the subsystem of the 'high' vowels there are clear variety-specific differences with respect to the phonetic cues for [±tense]: Whereas in NSG these vowels behave in parallel to the 'mid' vowels (i.e. higher F1-values for the [–tense] counterparts but equal f_0 values), in ASG the tenseness-cueing F1-difference is quite reduced (cf. [3]) with the [–tense] counterparts exhibiting higher f_0 values and the [+tense] counterparts showing longer durations. If one wants to posit a tenseness opposition for the 'low' vowel pair in German, the quite marginal cues would reside in small F1- and f_0-differences for both varieties in parallel.

In general it can be stated that ASG 'high' front vowel pairs are primarily distinguished by durational differences and thus behave similar to the 'low' vowels. Since 'high' vowels tend to pattern with 'low' vowels in ASG and with 'mid' vowels in NSG, this can be taken as a clear indication for postulating that 'high', 'mid' and 'low' vowels can be described as relatively independent subsystems within the rather complex German vowel system.

5. ACKOWLEDGMENTS

This research was supported by a German Research Council (DFG) grant (SPP 1234/1 PO 334/5) within the framework of the "Priority Programme: Phonological and Phonetic Competence: Between Grammar, Signal Processing and Neural Activity".

6. REFERENCES

[1] Boersma, P., Weenink, D. 2001. PRAAT, a system for doing phonetics by computer. *Glot International* 5(9/10), 341–345.

[2] Chomsky, N., Halle, M. 1968. *The Sound Pattern of English* New York: Harper & Row.

[3] Iivonen, A. 1987. Monophthonge des gehobenen Wienerdeutsch. *Folia linguistica (Acta Societatis Linguisticae Europaeae)* 21, 293–336.

[4] Jones, D. 1918. *An Outline of English Phonetics* Leipzig: Teubner, 1–221.

[5] Jørgensen, H. P. 1969. Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantfrequenzen. *Phonetica* 19, 217–245.

[6] Maddieson, I., Precoda, K. 1989. Updating UPSID. *UCLA Working Papers in Phonetics* 74, 104–111.

[7] NORM http://ncslaap.lib.ncsu.edu/tools/norm/index.php

[8] Norušiš, M. J. 1990. *SPSS Advanced Statistics User's Guide* Chicago: SPSS.

[9] Whalen, D.H., Levitt, A.G. 1995. The universality of F0 in vowels. *J. Phon.* 23, 349–366.