Discrimination sensitivities and identification patterns of vowel quality and duration in German /u/ and /o/ instances

Fabian Tomaschek*a, Hubert Truckenbrodtb, Ingo Hertrichc

*Corresponding author: fabian.tomaschek@uni-tuebingen.de, Tel: +49 7071 29 73965
*aQuantitative Linguistics, Department of Linguistics, University of Tübingen, Wilhelmstraße 17, 72074 Tübingen, Germany
*bCenter for General Linguistics, Berlin, Germany
*cHertie Institute for Clinical Brain Research, University of Tübingen, Hoppe-Seyler-Straße 3, 72072 Tübingen, Germany

Abstract
The German vowel length contrast uses duration and quality (tense vs. lax) as phonetic cues. So far, the relative importance between these two cues has been assessed by identification tests. The present study extends this work by including discrimination tests assessing discrimination sensitivities to temporal and spectral changes in /u/-/o/ vowels embedded in disyllabic stimuli. Results favour vowel duration as the primary cue for vowel discrimination. Temporal discrimination showed a sensitivity maximum at the short-long boundary. Spectral discrimination did not reveal any sensitivity to lax-tense changes but only to changes between vowel types such as /u/ and /o/.
It is concluded that vowel duration is used as the primary cue whereas vowel quality, the secondary cue, supports the contrast by increasing the perceptual distance between short and long vowels.

Keywords
German, vowels, tenseness, quality, length, perception, discrimination

cite as:
1.0 Introduction
The German vowel system is based on an interaction between vowel duration and vowel quality. Whereas in the low /a/-/a:/ the distinction is based mainly on duration, these two cues interact in high and mid vowels. In the vowel triangle, short vowels are produced more centralized, long vowels are more peripheral (Figure 1a) (Sendlmeier & Seebode, 2006). In previous studies it has been shown that at the perceptual level, vowel duration is the primary acoustic cue in distinguishing low short from low long vowels; vowel quality is primarily used to identify the length contrast in mid and high vowels (Bennett, 1968; Sendlmeier, 1981; Weiss, 1974). These perceptual patterns have been found using identification tests. However, studies investigating the perceptual characteristics of native and non-native contrasts showed that discrimination patterns are important in assessing the perceptual status of contrasting cues (MacKain, Best & Strange, 1981; Ylinen, Shestakova, Alku & Huotilainen, 2005). Only a native, phonemic contrast exhibits a peak in sensitivity at the category boundary. This is why we investigated discrimination sensitivities to changes in vowel duration and vowel quality, in addition to the traditional identification patterns. Sensitivities were measured by a stair-case A-X discrimination test assessing just-noticeable-differences (Lapid, Ulrich & Rammsayer, 2008). Our continuum extended between short and long, as well as /u/ and /o/ vowels. Our hypotheses for the identification test are illustrated in Figure 1b-e. Given the vowel distribution in production (Figure 1a), the category boundary between the /u/ vowels and /o/ should be located diagonally in a two-dimensional duration-frequency continuum (Figure 1b). The relative importance between the acoustic cues affects the course of the short-long boundary. 1) If quality was the only cue taken into account, then the short-long boundary should have a horizontal course. 2) If duration was the primary cue, the short-long boundary should be vertical. 3) If both acoustic cues are relevant, the short-long boundary should be diagonal.

Before presenting the hypotheses for the discrimination tests, a point of clarification must be made. We are not investigating perceptual patterns such as categorical perception (Liberman, Harris, Hoffman & Griffith, 1957) or the Perceptual Magnet Effect (Diesch, Iverson, Kettermann & Siebert, 1999). It has repeatedly been shown that categorical perception depends on the discrimination task (Gerrits & Schouten, 2004; Massaro & Cohen, 1983; Schouten, Gerrits & Hessen, 2003). Rather, in line with Ylinen et al. (2005) and MacKain et al. (1981), we want to know if phonetic change indicates phonemic distinction. If so, it will exhibit an increase in discrimination sensitivity at the category boundary. In this sense, if the durational or the qualitative contrast is used phonemically, it will show a maximum in discrimination sensitivity, i.e. a minimum in just-noticeable-differences x. Such a pattern has already been shown for the German /a/-/a:/ contrast (Tomaschek, Truckenbrodt & Hertrich, 2011). If this is not the case, no JND minimum will be found (see Figure 1c).

In addition, the JND magnitude indicates to what extent sensitivity to change is present. If within-categorical discrimination is possible (importantly, in the context of disyllabic words), JNDS for the respective cues will be small enough so that item X remains on the left side of the boundary (X1 in Figure 1d.top). If only across-categorical is possible, JNDS will be so large that item X is located beyond the boundary (X2 in Figure 1d.top). The first will result in a negative difference from the category boundary (difference = location item X – location boundary), the second in a positive difference (Figure 1d.bottom).

2.0 Subjects and Stimuli
20 native speakers of German were paid for their participation (10 males, 10 females. Mean age: 26.4 years, SD = 8.2 years). All subjects were right handed and speakers of Standard German, coming mainly from the south of Germany. All had normal hearing as confirmed by an audiometric screening test.

As in our other studies (Tomaschek et al., 2011; Tomaschek, Truckenbrodt & Hertrich, 2013), the stimulus material comprised synthesized instances of trochaic nonsense words (/g_bɒ) and was drawn from a two-dimensional continuum between /u/-/o/ and short – long. Stimuli were synthesized using the formant synthesizer of (Hertrich & Ackermann, 1999). Vowel quality between /u/ and /o/ was changed in the F1 dimension (Hose, Langner & Scheich, 1983). Formant transitions were implemented in the initial 20 ms of each vowel. Stimulus design is
illustrated in Figure 1e. In order to superimpose a trochaic stress pattern (Grabe, 1998), a minimal intonation was implemented (1st syll: F0 = 100 to 110 Hz; 2nd Syll: 105 to 80 Hz). No critical effects of the rise on the categorization were assumed as F0 changes do not affect the length categorization (Lehnert-LeHouillier, 2007). Besides these changes, no other temporal (Lehtonen, 1970) or spectral dynamics (Nearey & Assmann, 1986) have been included. For the stair-case discrimination test, vowel durations (VD) were available up to 536 ms and F1 frequencies up to 730 Hz.

3.0 Tests and Results
Subjects performed three discrimination tests and the identification test (in that order) in a sound proof chamber, equally distributed on two days. Before each test, they performed a short training session in the presence of the first author.

3.1 Identification test – Design and Results
The identification test covered the entire 2D continuum as illustrated by Figure 1b. Stimuli were played 10 times in pseudo-randomized order. Subjects categorized the perceived words as [gu:b@], [gu:b@], [go:b@] (sampa transcription). Like in Tomaschek et al. (2011), category boundaries were calculated by fitting the logistic psychometric function to identification probabilities (Lapid et al., 2008, function 1).

Boundaries were analyzed in the following subsets: 1) the /u/-/o/-boundary in category short (VD < 100ms); 2) the /u/-/o/-boundary in category long; 3) the short-long-boundary in the /u/ space (F1 < 340 Hz) and 4) the short-long-boundary in the /o/ space. Subsets were analyzed by mixed-effects models (lme4 package in R version 2.15.2 (R_Development_Core_Team)) using an interaction between vowel category (/U/-/u:/, /O/-/o:/) and the course of the boundary as a function of duration/F1. Subjects were included as random effects. The interaction was included as random slopes, accounting for repeated measures (Schielzeth & Forstmeier, 2009).

The results are: i) A significant interaction (\( \Delta_{\text{Category}} \times \text{Slope} = -1.51199, t = -7.11 \)) indicates that the slope of the /u/-/o/ boundary is steeper in category short (slope = -1.8 Hz/ms, t = -9.4) than in category long (slope = -0.3 Hz/ms, t = 6.5). ii) F1 frequency affected the short-long boundary in the /u/ space (slope = 0.07 ms/Hz, t = 5.8), but not in the /o/ space (slope = 0.008 ms/Hz, t = 0.46). The difference between the slopes was significant as indicated by the interaction (\( \Delta_{\text{Category}} \times \text{Slope} = 0.06, t = 2.5 \)). The location difference between the two boundary endpoints in the /u/ space (\( \Delta_{\text{F1}} = 250-310\text{Hz} = 5.6, CI = 7.8 – 3.3, \text{Student’s T-Test, } t = 5.2, p < 0.001 \)).

3.3 Discrimination tests – Design and Results
In the stair-case discrimination tests, two words were presented in an A-X sequence testing the temporal/spectral discrimination sensitivity for changes in the respective vowel. Discrimination sensitivity is shown by means of just-noticeable-differences (JND), a value that indicates the perceptual distance which is needed to perceive a difference between item X and item A. The lower the JND, the higher the discrimination performance. In order to acquire temporal and spectral JNDs, durations/frequencies in item X were changed depending on the subject's answer (exact procedure explained below. See also Figure 1 in Lapid et al. (2008), for an example of how X 'hovers' around the JND). The following tests were performed:

1) Item A was drawn 30 times from a durational /u/-/u:/ continuum with a fixed F1 frequency of 250 Hz (12 steps à ~10ms). Subjects performed two tests. a) In the Temp/VD condition, they indicated whether vowel duration in item X was longer than in item A. After a positive answer, the duration in X was shortened (-10 ms). After a negative answer, it was lengthened (+30ms). JNDs were calculated by fitting the logistic psychometric function to the probability of positive answers as a function of vowel duration for each item A and subtracting the duration in A from the curve's 75% intercept (Lapid et al., 2008; Tomaschek et al., 2011). b) In the Spectral/VD condition, subjects indicated whether vowel qualities where the same in item X and A. After a positive answer, F1 frequency was increased (+30 Hz). After a negative answer, F1 was decreased (-10 Hz). The probability of negative answers was used for calculating JNDs.

2) Item A was drawn 30 times from a F1 /u/-/o:/ continuum with a fixed vowel duration of 151 ms (10 steps à 20 Hz). In the Spectral/F1 condition, subjects performed a spectral
discrimination test as described above in 1a. The probability of negative answers was used for calculating JNDs. Hence, three sets were discriminated: a) Temp/VD, b) Spectral/F1, c) Spectral/VD.

Temporal and spectral JNDs are presented in Figure 2b. Since the JND values were not normally distributed (Shapiro-Wilk-test: W > 0.9, p<0.001), medians are presented instead of means. The statistical significance of the present results was assessed by fitting linear models to the JND courses. Table 1 presents the parameters for each model in the sets a – c. Significant values are marked with asterisks.

1) The Temp/VD JNDs do not exactly follow the hypothesized course. Rather, the minimum is located on the right side of the short-long boundary (compare Figure 1c and 2b, line A). JNDs as a function of duration are best represented by a positive square model with all parameters being significant (Table 1a). The minimum of the square curve is located at 96.3 ms in the near proximity of the short-long-boundary for 250Hz (89.7 ms). Item X’s location with respect to the short-long boundary is illustrated in Figure 2c, line A. A linear model indicates that differences are all significantly positive (Intercept = 59.2 ms, slope = 1.13 ms/ms, p < 0.001).

2) The picture is different considering the spectral discrimination performances. Spectral/F1 JNDs decrease significantly with increasing F1 frequencies (Figure 2b – line B), best represented by a significant negative linear model (Table 1b). Next to the /u/-/o/ boundary, spectral/F1 JNDs deviate in a U-shape from the model results. This local increase in sensitivity at the F1 frequencies 290 Hz, 310 Hz and 330 Hz might have emerged under the influence of the /u/-/o/ boundary. However, only the JND for 310 Hz differed significantly from the model line as tested by a Mann-Whitney U-test for non-normally distributed data (u = 41, p = 0.02). After a Bernoulli-correction (N = 3, p < 0.017), this difference failed significance. Item X’s difference with respect to the /u/-/o/ boundary is illustrated in Figure 2c, line B. A linear model indicates that differences are all significantly positive (Intercept = 64.2 Hz, slope = 0.78 Hz/Hz, p < 0.001).

3) Spectral/VD JNDs increase significantly with decreasing vowel durations (Figure 2b, line C). Their course is best represented by a positive square model (Table 1c), showing a rightward shift with respect to the short-long-boundary (minimum located at ~122 ms). Item X’s difference to the /u/-/o/ boundary is illustrated in Figure 2c, line C. A linear model indicates that differences are not significant at the beginning (Intercept = -6.1, p = 0.1; Hz, slope = 0.3 Hz/ms, p < 0.001). Only differences for durations longer than 100 ms become significant (Student’s T-tests, t > 20, p < 0.01, not Bernoulli-corrected), i.e. those located in category long.

4.0 Discussion

The aim of the present study was to investigate the perceptual relevance of phonetic vowel quality and duration as cues for a distinction between German short lax and long tense vowels. The contrast was investigated by means of an identification test and temporal and spectral discrimination tests using stimuli from a finely graded two-dimensional duration and F1 continuum.

The identification test showed that vowel duration and frequency have different effects on the location of the /u/-/o/ and short-long boundary (Figure 2a). The course of the /u/-/o/ boundary was affected significantly in both length categories. It was steeper in the short category than in the long category. The course of the short-long boundary was significantly affected by F1 frequency in the /u/ space, but not in the /o/ space. These findings partially reproduce the findings of (Lehnert-LeHouillier, 2010), who have shown that the short-long category in German native speakers was affected by vowel quality. Considering our initial hypothesis about how cue weighting would divide the two-dimensional vowel space (Figure 1b), the present identification results (Figure 2a) represent an intermediate pattern between the diagonal and the vertical short-long boundary. Hence, vowel quality, at least in the domain of F1, was not used as the main cue distinguishing short lax from long tense vowels, since vowel quality was stronger affected by duration than vice versa. Rather, vowel duration was more important.

The difference in /u/-/o/ boundary slopes between the length categories indicates that F1 frequency was mainly used to distinguish /U/ from /O/ in category long. In category short, both cues interacted to distinguish /U/ from /O/. This interpretation is in line with findings by
Escudero & Boersma (2004). The use of both cues for the /U/-/O/ distinction is probably the reason why the /O/-/o:/ boundary was not affected by F1 frequency: it competed with the /U/-/O/ boundary. By contrast, the /U/-/u:/ boundary could be shifted toward the shorter category by lower F1 frequencies since no competing category was present. Finally, the results reveal a spectral area that is ambiguous with respect to the identification of the vowel height category (Figure 2a): qualitatively identical instances are perceived as /U/ when short, as /o:/ when long. Three staircase A-X discrimination tests have been performed in order to assess temporal and spectral discrimination sensitivities along a duration and a F1 frequency continuum. The results in the temporal discrimination along the duration continuum (condition Temp/VD) replicated the findings in Tomaschek et al. (2011). JNDs decreased towards the category boundary, but increased toward the category edges (Figure 2b). Apparently, the present JND minimum is not located exactly at the category boundary. Unfortunately, an explanation for this rightward shift cannot be provided with the available data. Nevertheless, we argue that the JND minimum is a result of the boundary between category short and long. This can be seen by comparing the present discrimination pattern with discrimination patterns of duration in nonspeech signals. For example, Creelman (1962) has shown that sensitivity to temporal changes decreases with longer signals (i.e. JNDs increase). Hence, JNDs positively correlate with signal duration. By contrast, the present results show a significant reduction in JNDs. The only suitable explanation for this reduction is that JNDs are affected by some kind of (phonemic) boundary between category short and long. This interpretation is in line with (1981) and Ylinen et al. (2005). The boundary is also supported by the magnitude of temporal JNDs, indicated by the difference between vowel duration in item X and the short-long boundary (Figure 2c). Item X had to be located well beyond the category boundary, specifically in category short, in order that subjects perceived a difference between item A and item X.

The results in the spectral discrimination along the F1 continuum (condition Spectral/F1) do not match the hypothesized effect that JNDs would increase toward the category edges (Gerrits et al., 2004). Rather, spectral/F1 JNDs decrease with increasing F1 frequency (Figure 2b). The result is in line with findings on spectral discrimination in speech (Mermelstein, 1977) and in non-speech (Akin & Belgin, 2005; Dai & Micheyl, 2011): spectral JNDs correlate negatively with frequency between 20 Hz and 1000/2000 Hz (the right edge depends on the duration of the signal), i.e. they decrease toward 2000 Hz. Only at frequencies beyond 1000/2000Hz the correlation becomes positive. Although the overall Spectral/F1 pattern does not mirror the hypothesized pattern (Figure 1c), there is a minimal JND deviance next to the /u/-/o/ boundary. In line with Gerrits et al. (2004) we claim that this local deviance from the general course represents an increase in sensitivity due to the (phonemic) /u/-/o/ boundary. Comparable to the short-long boundary, the /u/-/o/ boundary is supported by its difference to item X (Figure 2c) which indicates that within-categorical discrimination was not present.

It is important to note that these results apply only to the context of disyllabic words, as it is well known that within-categorical discrimination of vowels is possible in contexts such as single vowel presentation. Given these results, discrimination in the present test was performed on a phonemic rather than acoustic basis, although subjects were specifically instructed to pay attention to minimal changes. From the articulatory point of view, the ‘short lax’ – ‘long tense’ contrast is reflected by a change in duration and centralization (Hoffmann, 2011; Sendlmeier et al., 2006). However, the present results do not indicate that subjects were sensitive to discriminate spectral changes within the /u/ space (Figure 2c). The sensitivity to spectral changes along the duration continuum (Spectral/VD) was inversely correlated with vowel duration. The shorter the vowel, the higher just-noticeable differences and thus the worse was sensitivity. This pattern is in line with findings on nonspeech discrimination, which show that the shorter the signal, the worse is frequency discrimination (Freyman & Nelson, 1986). It is also in line with Wallace & Blumstein (2009), who showed that the extraction of a vowel’s identity deteriorates at shorter durations. The optimal duration is ~150 ms. In addition, F1 values in item X were located in the near vicinity of the /u/-/o/ boundary, not within the /u/-space (Figure 2c). Hence, no within-/u/ discrimination was possible, which would have been necessary in order to classify the distinction between quality ‘tense’ and ‘lax’ as the primary cue.
Concluding, these results show that vowel duration is the more important cue for the length contrast. The question remains, 1) why short vowels are lower than long vowels and 2) how the quality cue is used in the contrast. A biomechanical explanation suggests that short vowels do not provide enough time for the tongue to reach its vocalic target, which is hypothesized to be located at the very periphery of the vowel space or even beyond it (Diesch et al., 1999; Johnson, Flemming & Wright, 1993). Therefore, short vowels are articulated more centrally. This is the case when vowels are shortened due to higher speaking rate (Flege, 1988; Lindblom, 1963). Another explanation is based on findings about the neural processing of vowels. It has been shown that higher vowel frequencies are processed faster than lower vowel frequencies (Poeppel, Phillips, Yellin, Rowley, Roberts & Marantz, 1997). This is probably due to a correlation between frequency and information density per time frame. One might therefore hypothesize that lowering the vowel and therefore increasing F1 frequency compensates for the lack of processing time by increasing information density. This makes perfect sense given that short vowels already transfer less information than long vowels (Wallace et al., 2009).

The final explanation aims at the processing level of vowel perception. Gussenhoven (2007) investigated the effect of vowel height onto the perception of duration. Due to biomechanical restrictions, low vowels are intrinsically longer than high vowels (e.g. /a/ vs. /i/). The perceptual system compensates this lengthening by subtracting the surplus duration from the lower vowel (or adding to the higher vowel), thus rendering both vowels equally long at the perceptual level (see Figure 2d for the schematics). Gussenhoven (2007) matched the durations of high and low vowels which in turn provoked that lower vowels were perceived as shorter. The perceptual system kept subtracting surplus duration from the lower vowel. Meister & Meister (2011) indirectly replicated these findings by showing that the short-long boundary occurred at shorter durations in high vowels than in low vowels, which is also the case in the present study. In this sense, lowering the short vowel reduces its perceptual duration, thus increases its distance to the long vowel and therefore supports the short-long distinction of the length contrast.

Acknowledgements
This study was funded by the German Research Foundation, Priority Program 1234, project: HE 1573/4-1, AOBJ: 568283. The present article was part of the first author’s dissertation. We want to thank Walter Sendlmeier for kindly providing Figure 1a. Furthermore our gratitude goes to Denis Arnold and Cyrus Shaoul for comments on the manuscript.

References


Sendlmeier W, Seebode E, Hessen Av. (2003). The end of categorical perception as we know it. *Speech Communication*, 41:70


Table 2
Parameters of the linear and squared model fits to discrimination results as shown in Fig. 3d-f. Asterisks indicate significance (p: * = 0.01, ** = 0.001, *** < 0.001).

<table>
<thead>
<tr>
<th>significance</th>
<th>Intercept</th>
<th>Linear parameter</th>
<th>Square parameter</th>
<th>minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>set a)</td>
<td>p = 0.01</td>
<td>134.3ms ***</td>
<td>-0.6JND/ms **</td>
<td>96.3 ms</td>
</tr>
<tr>
<td>set b)</td>
<td>p = 0.02</td>
<td>169.0Hz ***</td>
<td>-0.15JND/Hz *</td>
<td>-----</td>
</tr>
<tr>
<td>set c)</td>
<td>p &lt; 0.001</td>
<td>322.1Hz ***</td>
<td>-3.7 JND/ms ***</td>
<td>122.0 ms</td>
</tr>
</tbody>
</table>

Figure 1

a) Production vowel space. b) Hypothetical division of the duration-F1 space. Solid line: /u/-/o/ boundary. Dashed lines: short-long boundary → Horizontal: Quality is primary; Vertical: Duration is primary; Diagonal: Both are equal. c) Hypothetical just-noticeably-differences. Vertical dashed line = boundary. Diagonal line = no phonemic boundary. U-shaped line = phonemic boundary. d.top) Illustration of the location of item X with respect to the category boundary. X1: location of item X in case of more important cue. X2: location of item X in case of less important cue. d.bottom) Difference of X1/X2 with respect to category boundary. e) Stimuli design. Durations and F1 frequencies for /u/ and /o/ represent endpoints of the continuum.

Figure 2

a) Identification results: The vertical line represents the short-long boundary; the diagonal line the /u/-/o/ boundary. Frequencies from the ambiguous area between the two horizontal grey lines are perceived either as /u/ or /o/, depending on duration. b) Discrimination results: Solid lines = median just-noticeable-differences. Colour coding: A - black = temp/VD; B – dark grey = Spectr/F1; C – light grey = Spectr/VD. Dashed lines = model fit. Upper vertical dotted line = /u/-/o/ boundary; lower vertical dotted line = short-long boundary. c) Differences between the temporal/spectral location of item X and the respective category boundaries (Difference = location of item X – location of category boundary). Colour coding as in b. d) Schematics for effects of vowel height onto the perception of vowel duration.