PROCESSING GERMAN VOWEL QUANTITY: CATEGORICAL PERCEPTION OR PERCEPTUAL MAGNET EFFECT?

Fabian Tomaschek[a], *Hubert Truckenbrodt*[b] *& Ingo Hertrich*[a]

[a]Hertie Institute for Clinical Brain Research, University of Tübingen, Germany;
[b]Center for General Linguistics, Berlin, Germany
fabian.tomaschek@uni-tuebingen.de; ingo.hertrich@uni-tuebingen.de;
truckenbrodt@zas.gwz-berlin.de

ABSTRACT

The German vowel system shows a complex structure based on the interaction between vowel duration and formant structures between short and long cognates. This leads to the question how vowel duration is processed. The perception of vowel duration in German native speakers was tested by an identification test, a goodness rating and an adaptive discrimination test. The test results show a sharp boundary between the short and the long category. Furthermore, the category border was characterized by bad goodness ratings in comparison to within-category stimuli and a maximum in discrimination performance regarding subtle durational differences. These results meet the criteria of both the perceptual magnet effect and categorical perception.

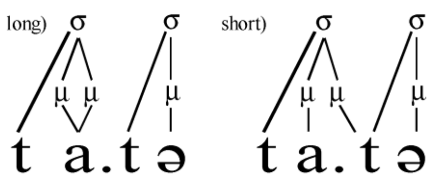
Keywords: German, vowels, quantity, categorical perception, perceptual magnet effect

1. INTRODUCTION

Previous studies on phonetic perception reported two different processing mechanisms: "categorical perception" for consonant processing [15] and the "perceptual magnet effect" for vowel frequency perception [8, 9]. Categorical perception is defined as showing a sharp categorical boundary between two phonological categories with high discriminative sensitivity but no sensitivity to change within the category. The perceptual magnet effect is defined as discrimination performance showing an inverse correlation with categorical goodness ratings: The further away an item is located in relation to the category center, the better the discrimination. Most importantly, discrimination within a category is still possible. Hence, the magnet effect describes the internal structure of a category from its prototypical center toward the periphery. Importantly, tests for categorical perception usually do not include goodness ratings, ignoring perceptual variability within categories. By contrast, tests for the perceptual magnet effect mostly omit an identification task, ignoring the particular perceptual processes at category boundaries. Therefore a direct comparison of studies on categorical perception and of the perceptual magnet effect seems to be difficult.

In the present experiment the processing of vowel duration in German native speakers is studied. The German vowel system has a distinction between long and short vowels that is largely constrained to stressed syllables. Furthermore, vowel quantity interacts with some aspects of vowel quality: High vowels are either long and tense or short and lax. Only the low vowel /a/ exhibits a purely durational contrast [1], which was the motivation to investigate the durational contrast of /a/ versus /a:/ in the present study. In phonological theory, vowel length contrast, called quantity, is modeled by associating a vowel quality representation for a short vowel with one length unit (called mora), and with two moras for a long vowel (Fig. 1) [17]. To test the psychophonological mapping of a linearly changing continuum onto the contrasting phonetic surface, an identification test 'IT' was used to find the boundary between category long and short; a goodness rating 'GR' was used to find the structural organization of the categories and an adaptive discrimination test 'DT' (standard < comparison) was used to find both the sensitivity to change and the just-noticeable-difference (JND) for vowel duration along the continuum.

Figure 1: Phonological Representation of Vowel Duration.



2. METHODS

2.1. Subjects

20 native speakers of German were recruited and paid for their participation (10 males, 10 females. Mean age: 27.4 years, SD = 4 years). All subjects were speakers of Standard German, coming from all over Germany.

2.2. Stimuli

The trochaic nonsense word /tatə/ was used as stimulus. The use of the disyllabic nonsense word was motivated by three aspects. First, it avoided a lexical bias, i.e. a perceptual preference for the item that is more familiar to the subjects. Second, it represents a trochaic foot, which is the common metrical structure of German [4]. Third, the invariant duration of the second syllable reduces the possibility of a perceptual confounding between phonological vowel length and the hearer's representation of the speech tempo along the physical continuum [7].

During synthesis all phonetic segments were controlled, varying only the duration of /a/ in steps of one pitch period (~10ms), producing a continuum from category short to long. In the IT /a/ duration was 50ms and 187ms. In the GR two sets were used: one for category long and one for category short with vowel durations of 21ms-128ms and 79ms-187ms, respectively. In the DT standard duration was 50ms-158ms and potential comparison durations for the adaptive procedure ranged between 21ms-305ms.

2.3. Procedures

Subjects were seated comfortably in a sound attenuated chamber. Stimuli were presented via Sennheiser HD 201 headphones at an intensity agreeable for the subject. All three tests were controlled by a MATLAB (version R2009a, Mathworks) procedure running on a laptop (Acer Extensa 7630EZ) including both stimulus presentation and acquisition of behavioral data. No training phase was offered. No repeated listening option was available.

In the IT, subjects had to categorize all items from the continuum 10 times to category long or short. Additionally, response times (RT) from the offset of the stimulus were measured.

In the GR items in both sets had to be rated 5 times (1 = very good exemplar, 6 = very bad exemplar).

In the DT each standard was played 30 times with its potential comparison. Subjects had to discriminate in the sequence “standard < comparison” whether comparison was longer than the standard. The procedure was insofar adaptive that the duration of the comparison varied with respect to the answers given by the subject: After three sequent positive “comparison vowel is longer than standard” answers, the duration of the comparison was decreased by one pitch period (~10 ms). After one negative answer, i,e, "comparison is *not* longer than standard", the duration of comparison was increased by one pitch period. The adaptive procedure was motivated by two reasons: First, it considerably reduced the amount of trials needed for an exhaustive study. Second, in addition to testing the sensitivity for small durational differences along the continuum, as does the traditional test, the adaptive discrimination test actually determines the just-noticeable-distance (JND) between two vowels in subsequent test items.

2.4. Analysis

The individual percental results from the IT were fitted with an arcus-tangens curve as a function of vowel duration, extracting exact numerical data. The category boundary was defined as the x-axis value in milliseconds at which the arcus-tangens curve crossed the 50% line on the ordinate scale. Individual RTs were fitted with a Gaussian curve as a function of vowel duration with the curve’s maximum defined as the RT maximum. In the GR, individual ratings were fitted with the Gaussian-function as well and the individual intersection between both curves was calculated. Furthermore, ratings for vowel duration intersecting between both sets were cut out and the residual ratings were concatenated to one large set, used for correlation analysis between DT and GR. In the DT for each subject and each standard the weighed percentages of positive answers were fitted with the arcus-tangens curve as a function of comparison duration. The JND was defined as the durational standard-comparison difference at which 70% of positive responses were obtained.

3. RESULTS

Fig. 2 shows the category boundary and the mean RT along the tested continuum. RT is highest close to the category boundary and decreases in the categories. In the identification test, the group mean value for the category boundary, as found with the arcus-tangens-function was 105.9 ms (SE = 1.7ms). Pearson's product-moment shows a significant positive correlation between individual category boundaries and RT maxima (r = 0.65, $F_{(1,18)} = 13.45$, $p < 0.01$).

Fig. 2 shows also the mean just-noticeable-difference (JND) along the continuum. JND within category short decreases linearly the nearer an item is located toward the category boundary, with a clear minimum at category boundary. A square group regression analysis of normalized JNDs yielded a significant square component ($F_{(2,225)} = 15.4$, $p < 0.001$) with a minimum located at 98.0ms.

Figure 2: Mean Response Times in Identification and Just Noticeable-Difference (JND)

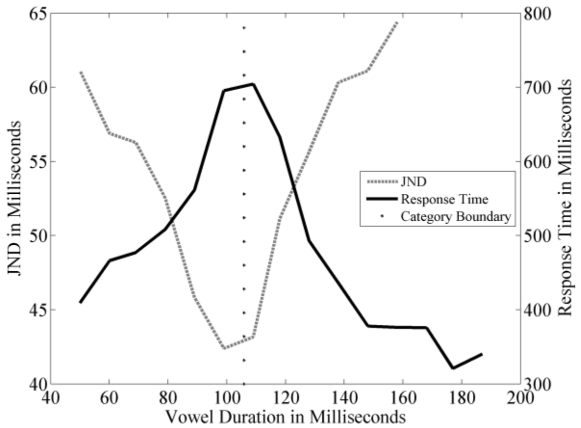
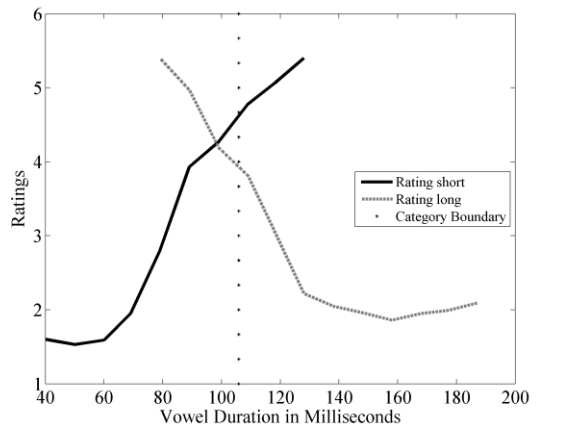


Fig. 3 shows the mean goodness ratings along the tested continuum for both sets. Ratings worsen toward the category boundary, but stabilize in the set short for vowel durations shorter than 60ms and in the set long for vowel durations longer than 128ms.

Figure 3: Mean Goodness Ratings



In the goodness rating, mean point of intersection between the fitted curves in the rating set long and rating set short is 97.7ms (SE = 2.5ms), yielding a significant correlation between individual category boundaries and individual points of intersection (r = 0.5, p = 0.01).

Pearson's product-moment yielded a significant inverse correlation between goodness ratings and discrimination performance (r = -0.33, t = -5.2, $p < 0.001$).

4. DISCUSSION

The results of the present study clearly meet the criteria for categorical perception. (1) In the identification test a sharp categorical boundary was found at 105.9 ms which is supported by both the goodness rating and discrimination test. Correlation analysis across subjects showed that individual RT maxima and category boundaries are co-located.

Furthermore, the significant correlation between individual category boundaries and individual points of intersection between the set long and set short in the goodness ratings and supports its location.

(2) The square regression analysis of the discrimination performance at group level was significant and found a JND minimum, hence, a sensitivity maximum, located in proximity of the category boundary, supporting the assumption of a more or less "hard-wired" structure underlying the phonological contrast between long and short quantity.

On the other hand, the present results meet the criterion for the perceptual magnet effect. The inverse correlation between goodness ratings and discrimination performance is significant.

These results provide further evidence that categorical perception and the perceptual magnet effect are not strictly distinct phenomena [13].

The identification test is *per definitionem* phonological since it forces the identification according to phonological categories. The discrimination test is *per definitionem* physical since it asks the subject to discriminate physically, not phonologically, between two adjacent items taken from a linearly changing continuum. The goodness rating combines the physical and phonological processing modes since it asks to judge the physical percept in relation to a "psychophonological" exemplar.

A complete description of the perceptual processes along a physical linearly changing continuum can be carried out only with those three tests together. In fact, different test methods such as identification, discrimination or goodness rating seem to induce different perceptual modes, which seem to be always active in speech perception.

1) A physical mode for purely acoustic perception and discrimination of noise and tone duration as well frequency characteristics [6, 10, 14, 16]. This mode applies in discrimination tests.
2) The present results for vowel duration and reports on the categorization of vowel quality [2, 3] as well the categorization of consonants for place of articulation [11] and voicing [5, 15] indicate a linguistic mode, which is purely phonological. This mode applies during speech perception and in identification tests.
3) Discrimination between subtle acoustic differences is not only possible within vowel categories [8] but also within consonant categories like voiced – voiceless plosives [5]. Though this is a linguistic mode, since distinct categories can be perceived, the perception of acoustic differences indicates a conjunction with the physical mode. This mode applies in linguistic discrimination tests as well in goodness ratings.

The phonetic contrast exemplified in the present study is analyzed differently in the phonology of different languages. In Japanese, which allows contrasting durations in the unstressed syllable, quantity is localized at the segmental, subsyllabic level [7]. The contrast in German is restricted to stressed syllables, only. It is localized at the prosodic, syllabic level, explaining quantity as a function of the word-internal consonant's lexically encoded position relative to the syllable boundary (Fig. 1) [1]. Estonian expresses quantity by means of the duration ratio between stressed and unstressed syllable, which is why the contrast is attributed to the suprasyllabic, foot level [12].

Hence, the contrast located at the phonetic surface can be attributed to different phonological levels.

5. REFERENCES

[1] Becker, T. 1998. *Das Vokalsystem der deutschen Standardsprache*. Lang, Frankfurt am Main: Peter Lang.

[2] Bohn, O.-S., Polka L. 2001. Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *Journal of the Acoustical Society of America* 110, 504-515.

[3] Cebrian, J. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics* 34, 372-387.

[4] Féry, C. 1996. German foot and word stress in OT. *Nordlyd* 24, 63-96.

[5] Hanson, V. 1977. Within-category discriminations in speech perception. *Perception & Psychophysics* 21, 423-430.

[6] Hellström, Å., Rammsayer T.H. 2004. Effects of time-order, interstimulus interval, and feedback in duration discrimination of noise bursts in the 50- and 1000-ms ranges. *Acta Psychologica* 116, 1-20.

[7] Hirata, Y. 2004. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32, 565-589.

[8] Kuhl, P.K. 1991. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Percept Psychophys* 50, 93-107.

[9] Kuhl, P.K. 2004. Early language acquisition: Cracking the speech code. *Nature reviews neuroscience* 5, 831-843.

[10] Kushnerenko, E., Ceponiene, R., Fellman, V., Huotilainen, M., Winkler, I. 2001. Event-related potential correlates of sound duration: similar pattern from birth to adulthood. *Neuroreport* 12, 3777-3781.

[11] Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A. 2005. Neural substrates of phonemic perception. *Cereb. Cortex* 15, 1621-1631.

[12] Lippus, P., Pajusula, K., Allik, J. 2009. The tonal component of Estonian quantity in native and non-native perception. *Journal of Phonetics* 37, 388-396.

[13] Lotto, A., Kluender, K., Holt, L. 1998. Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America* 103, 3648-3655.

[14] Mirman, D., Holt, L., McClelland, J. 2004. Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues. *Journal of the Acoustical Society of America* 116, 1198-1207.

[15] Pisoni, D., Tash, J. 1974. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics* 15, 285-290.

[16] Smits, R., Sereno, J., Jongman, A. 2006. Categorization of Sounds. *Journal of Experimental Psychology: Human Perception and Performance* 32, 733-754.

[17] Zec, D. 2007. The syllable. In de Lacy, P. (ed.), *Handbook of Phonology*. Cambridge: Cambridge University Press, 161-94.