Behavioral and neural correlates of vowel length in German and of its interaction with the tense/lax contrast.

Dissertation

zur Erlangung des akademischen Grades

Doktor der Philosophie

in der Philosophischen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Fabian Konstantin Tomaschek

aus Königshütte / Polen

2013

Gedruckt mit der Genehmigung der Philosophischen Fakultät

der Eberhard Karls Universität Tübingen

Dekan: Prof. Dr. Jürgen Leonhardt

1. Berichterstatter: PD Dr. Ingo Hertrich

2. Berichterstatter: Prof. Dr. Wolfgang Sternefeld

Tag der mündlichen Prüfung: 16.07.2013

Universitätsbibliothek Tübingen, TOBIAS-lib

I hereby declare that I have produced the work entitled: “Behavioral and neural correlates of vowel length in German and of its interaction with the tense/lax contrast”, submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making false declaration under oath is punishable by a term of imprisonment up to three years or by a fine.

Tübingen, ___________________ _____________________ *Shiva Kebeka*

Deutsches Vokalgedicht

Hörst du das Dröhnen im All ?
Hörst du den Einen Vokal?
Hörst du, wenn ich gähne,
den Laut, den /a/ ich nenne?

Wenn wir Vokale erträumen,
wo sind sie? In welchen Räumen?
Die kurzen in der kleinen Kammer,
beim Zentrum, im Urvokalzimmer,
aber es ist ein Jammer,
ihre Färbung ist nur ein Schimmer.

Die Langen sind in der Peripherie
Als Reinform strahlen aus der Ferne.
Wir würden sie produzieren gerne
Aber wir ahnen schon,
sie sind eine Illusion.
Sie sind in der Peripherie
erreichen tun wir sie nie.

Es ist der Weisheit Kern
wie im Leben, so beim Vokal:
Das Schöne bleibt fern
das Nahe ist fahl.

Table of Contents

List of original publications

Study I was published in

Tomaschek F, Truckenbrodt H, Hertrich I. (2011). German Vowel Quantity: Categorical Perception or Perceptual Magnet Effect? In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII),* pp. 2002-5. Hong Kong.

I prepared, submitted and revised the manuscript with the support of PD Dr. Ingo Hertrich.

Study II was published in

Tomaschek F, Truckenbrodt H, Hertrich I. (2013). Neural processing of acoustic duration and phonological German vowel length: Time courses of evoked fields in response to speech and nonspeech signals. *Brain & Language,* 124:117-31

I prepared and submitted the manuscript with the support of the co-authors and revised it with the support of PD Dr. Ingo Hertrich.

List of publications in preparation

Study I and III are in preparation as

Tomaschek F, Truckenbrodt H, Hertrich I. (submitted to the Journal of Phonetics). The identification and discrimination characteristics of German vowel length and its interaction with tenseness.

Study IV is in preparation as

Tomaschek F, Truckenbrodt H, Hertrich I. (Draft). The sensitivity to acoustic and phonologic information: a priming study.

Personal contribution to the studies

Study I and II

I planned the study, designed the experimental setup with the help of my coauthors and developed routines for the data analysis with the help of PD Dr. Ingo Hertrich. I collected the data and performed the data analysis.

Study III and IV

I planned the study and designed the experimental setup with the help of PD Dr. Ingo Hertrich. I collected the data, adapted and performed the routines for data analysis.

Study V

Study V was a term project by Anna Gastel and Benjamin Layer. I helped to plan the study and to perform the analysis. The data was collected by Anna Gastel and Benjamin Layer. The presented text represents the abstract written by Anna Gastel and Benjamin Layer, successfully submitted to the Phonetik und Phonologie 8 in Jena, 2012.

Abstract

The purpose of the present doctoral thesis was to investigate whether and how the theoretical account of the German vowel length contrast is mirrored by psychoacoustic characteristics and neural correlates. German vowels are contrasted by vowel duration (+/- long) that interacts with vowel quality (+/- tense).

The questions arise i) which phonetic cue, i.e. duration or quality, is used primarily for the phonological distinction between short and long vowels ii) and to what extent the theoretical account is mirrored in the psychoacoustic and neural reality. This aim was accomplished by combining behavioral psychoacoustic tests and whole-head magnetoencephalography (MEG).

Mainly two behavioral tests were performed in various experiments in order to assess the effects of vowel duration and quality on the perception using temporal and spectral continua: an identification test to assess the category boundary and an adaptive discrimination test, which tests the sensitivity to temporal and spectral change. It was found that the vowel length contrast matched the psychoacoustic criteria for categorical perception: Characteristics exhibited a sharp boundary between category short and long. In addition, identification response times were slower for durations at the boundary in contrast to within-categorical durations. Finally, the discrimination sensitivity to temporal change increased at the boundary as opposed to within-categorical durations.

An analysis of the relationship between duration and quality found that listeners have an ambiguous spectral area whose identification depends on vowel length: Short instances from that area were perceived as high back *lax* vowels (e.g. [ʊ]), long instances as mid back *tense* vowels (e.g. [oː]). This result was interpreted to be the effect of a strong magnet effect in case of short vowel durations of the high back vowel /u/: Sensitivity to spectral change decreased significantly the lower and shorter the vowel was, extending the vowel space for /u/. Simultaneously, differences between lax and tense /u/ instances were neglected.

The findings of the behavioral tests were supported by two neural experiments in the MEG. The first neural experiment investigated whether and how the category boundary was reflected by auditory evoked responses. At this aim, a passive listening task was performed using a vowel duration continuum embedded in a disyllabic nonsense word. It was found that each syllable of the word was reflected by a typical M50/M100 deflection. The phonological boundary between category short and long, however, was by a reflected neural correlate that emerged between the first and second syllable which has so far not been described. The new correlate occurred for categorically long vowels only, indicating additional neural activity for phonologically long vowels.

The second neural experiment set out to show whether and to what extent the auditory system is sensitive to minimal changes in vowel quality. A priming study was performed for this purpose, examining the effects of within-categorical and across-categorical quality changes in a probe word on the M100 in response to a target word. By these means it was found that the auditory cortex can differentiate between lax and tense instances during the processing of vowel identity, even in short vowels.

By taking the findings in the behavioral and neural experiments into consideration, it was concluded that vowel duration is the primary distinctive phonetic cue for the phonological vowel length contrast in German. Vowel quality serves as a secondary – supporting – cue which enhances the perception of categorical durations as short or long vowels.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Ingo Hertrich, who guided me through my dissertation stage with his incredible ability to combine knowledge from theoretical and practical linguistics, from neurology and sometimes even from philosophy. His gentle and patient way to show directions and solutions as well as to motivate me made the science of language a very special experience. Furthermore, I would like to thank Prof. Dr. Hubert Truckenbrodt whose cooperation made it possible to obtain funding for the DFG project (Priority Program 1234, project: HE 1573/4-1, AOBJ: 568283), which this thesis was part of.

My thanks go as well to Paolo Bernadinelli, Erick Cortez, Krunoslav Stingl and Mathias Witte for their frequent and elaborated help with programming, analysis, technical and methodological matters.

The manuscript in its current form would not have been possible without the proof-reading and detailed commenting on theoretical and practical questions by Sascha Kuhn, Henrik Nieman, Timo Röttger and Ida Zündorf. Thank you very much. Especially Anja Wühle invested a lot of effort to read and comment the manuscript twice. I owe you one and bonne chance à Paris!

Finally and most importantly, I am deeply grateful to my wife and best friend, Sabine Tomaschek. She supported me during this exciting time with love, care, patience, interesting discussions and repeated proof-readings.

Abbreviations

C	Consonant
CB	Category boundary
CI	Confidence interval
Δ	Difference / slope in the mixed-effect model
EEG	Electroencephalography
EMF	Evoked magnetic field
ISI	Inter-stimulus-interval
JND	Just-noticeable-difference
LH	Left hemisphere
MEG	Magnetoencephalography
MMN	Mismatch negativity
nAmp	Nanoampere
PME	Perceptual Magnet Effect
RH	Right hemisphere
RT	Response times
SOA	Stimulus onset asynchrony
SQUID	Superconducting quantum interference devices
STG	Superior temporal gyrus
TWI	Temporal window of integration
V	Vowel (short)
VV	Vowel (long)

Introduction

In the contemporary study of language perception, the phonetic speech signal constitutes the center point of several descriptive levels (Figure 1). Theoretical linguistics try to set up formal and logical systems which describe and make predictions about the surface structures of a language such as the phonetic form, the syntactic structure as well as the semantics of the speech signal. These linguistic descriptions, to which phonology is counted in the present doctoral thesis, can be regarded as descriptive grammars of a language. The validity of these grammars can and has been examined in many studies which have used several psychoacoustic as well as neural measures. Such measures not only provide insight about how language is processed. They allow us, first, to infer the cognitive representation(s) of speech. And second, they validate formal descriptions or, in case of falsification, adjust them in accordance to the results. The present doctoral thesis is considered to cover these three descriptive levels, as the assessment of the psychoacoustic characteristics shown during the perception of the German vowels and the investigation of its neural processing were driven by formal descriptions of the German vowel system.

Modern Standard German, a member of the West Germanic branch of the Indo-European language family, shows a large and complex vowel system which has contrastive length. As such, it belongs to the 10% of the world's languages which differentiate minimal pairs by lengthening or shortening the vowel (homepage: UPSID). In German, the length contrast is signaled by means of two correlating cues: duration and quality. Short vowels are *lax*, long vowels are *tense* (e.g. *Mitte* [mɪ.tə] 'center' vs. *Miete* [miː.tə] 'rent'). However, the correlation between the duration and quality cue is not consistent because of two points. First, vowel length is contrastive only in stressed syllables.

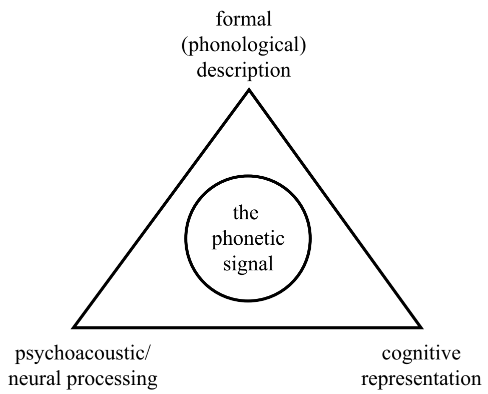

Figure 1: Schematics of the three levels that describe the phonetic (speech) signal and which were addressed in the present doctoral thesis.

The change in vowel quality, however, is functional in both, stressed and unstressed syllables. Second, the two vowel pairs [a-aː] and [ɛ-ɛː] are differentiated mainly by duration, not by quality. These inconsistencies have lead to rich and partly dogmatic discussions in the phonological and phonetic literature about how this contrast has to be described formally and what is the hierarchy of these two cues. Particularly, which is the phonetic cue that differentiates primarily between short and long vowels and can therefore be considered the underlying phonological feature and which cue can be regarded as the secondary cue because it is predictable.

Linguistic theory argues that the phonetic inventory of a language, i.e. the sounds that it uses, is not an arbitrary construct but structured into a phonological system, a framework that represents varying acoustic cues by which speech sounds are contrasted as features. Such features, like vowel length and vowel tenseness (representing the change in vowel duration and vowel quality, respectively), are stored in a *mental lexicon.* They represent two kinds of information: motor commands for speech production as well as acoustic templates for speech perception. Recent findings of behavioral and neural studies have shown that this theoretical assumption is reflected by cortical structures in the human brain.

Such phonological representation of contrastive sounds change the psychoacoustic characteristics of perception such as categorization patterns or discrimination sensitivity. These psychoacoustic characteristics provide information about how the acoustic cues used for a contrast are weighted during perception. That information can be in turn interpreted in terms of a hierarchy between the formal phonological features and their representation in the mental lexicon.

The purpose of the present doctoral thesis was to study the psychoacoustic characteristics and the neural processing of the German vowel length contrast in order to evaluate the contemporary theoretical assumptions about the German length contrast. It investigated how the perceptual mechanisms that cope with auditory spectro-temporal information process the two phonetic cues, vowel duration and vowel quality, and integrate them into the perception of categorical vowel length. Ideally, the investigation should reveal a hierarchy between the cues and therefore the features in the phonological system of German vowels.

The present doctoral thesis extended previous investigations on the perception of German vowels that used identification tests to a range of discrimination tests. These tests assessed the temporal and spectral discrimination sensitivity, providing further insights about the perception of this contrast. Additionally, magnetoencephalographic (MEG) experiments have been performed, investigating how the contrast was processed at the neural level.

The present doctoral thesis consists of two parts. Part one provides the theoretical and methodological background for the five studies conducted in the course of the thesis and presents a summary of all five studies performed during this thesis.

Part two contains the detailed manuscripts of the studies, as they were planned for submission in international journals. The manuscripts present a full description of the investigation and analysis methods as well as the statistical results, which were omitted in part one. Each study can be read on its own which is why some figures and tables presented in part one recur in part two.

Theoretical and practical background

Distinction between phonetics and phonology

Before presenting the case of the vowel length contrast in German a short distinction between phonetics and its formal description – phonology – has to be made. *Phonetics* is concerned with different acoustic cues of the speech signal. For example, i) one of the cues which indicates the distinction in voicing of a plosive consonant (e.g. [t] or [d]) is the duration of the voice onset time (VOT) between the burst and the vowel (Pisoni & Lazarus, 1974a). Its acoustic duration can vary, e.g. depending on language (Flege, Munro & MacKay, 1995; Lisker & Abramson, 1964) or speaking rate (Flege, 1988). In addition, voicing is supported by cues like the vowel/VOT duration ratio (Kohler, 1979) or the F1 onset frequency (Jiang, Chen & Alwanc, 2006). ii) In the coronal consonant [d], the form of the transition in the second formant depends on the height of the following vowel (e.g. [da] or [di] see Pompino-Marschall (2003, P. 161); Strange (1987)). Phonetics is thus concerned to describe the acoustic cues which are used to produce a linguistic contrast. Their processing has been investigated in many psychoacoustic as well as neural studies (additionally to the studies cited above: e.g. Gerrits & Schouten (2004); Lotto, Kluender & Holt (1998); Massaro & Cohen (1983); Poeppel, Guillemin, Thompson, Fritz, Bavelier & Braun (2004); Schouten, Gerrits & Hessen (2003a); Toscano, McMurray, Dennhardt & Luck (2010)).

The formal description of a language's phoneme inventory and its rules, *Phonology,* neglects such acoustic differences as long as they are non-distinctive at the perceptual level (e.g. [da] vs. [di]). It takes only those differences into account which are distinctive (e.g. [ta] vs. [da]). One way to describe distinctive sounds – *phonemes* – is the use of formal features, in this example [+/- voiced] (Hall, 1992; Wiese, 2000). A difference is made between phonemes which are *predictable* and whose occurrence at the phonetic surface can be derived by a set of rules and those phonemes which are not predictable and stored in the mental lexicon (Kenstowicz, 2004; Wiese, 2000, P.151ff.). The latter are called *underlying* phonemes. A standard example are the dorsal fricatives [χ/ç] in German words like *Buch* [buːχ] 'book', *Bücher* [byːçɐ] 'books'. The uvular [χ] occurs only after non-front vowels ([u,o,a]) whereas the palatal [ç] occurs elsewhere (after front vowels [i,e,y]; at the beginning of a word like in *Chemie* [çemi] 'chemistry'). Hence, they occur in complementary distributions. Since the uvular [χ] is more restricted with respect to its possible locations than the palatal [ç], it is regarded as an *allophone* of the underlying [ç]. Its occurrence is predictable (i.e. 'after non-front vowels') and therefore rule-based (Wiese, 2000, P. 209ff.; 2011, P. 54ff.). Taking this example into account, the distinction between predictable and non-predictable information is expanded with ease to features. In a radically underspecified system such as the FUL model (Lahiri & Reetz, 2002), those features which cannot be predicted are represented in the mental lexicon, i.e. they are *underlying* (Wiese, 2000). Features which are predictable are not represented in the mental lexicon but are regarded as *underspecified* and can be derived by rules. The validity of this theoretical account has been shown in several neural studies (Eulitz & Lahiri, 2004; Eulitz & Obleser, 2007; Friedrich, Lahiri & Eulitz, 2008). Regarding the topic of the present thesis, the question arises which feature is predictable in order to differentiate between short and long vowels and which has to be stored in the mental lexicon? This question will be elaborated in the next section.

The phonetics of German vowel length

German belongs to a group of languages which uses vowel duration in a contrastive way. Two phonetic cues are used to produce the contrast: vowel duration and vowel quality (Braunschweiler, 1997; Sendlmeier & Seebode, 2006; Sendlmeier, 1981; Weiss, Dressler & Pfeidder, 1977). In the present doctoral thesis, *phonetic* will refer to all changes in the speech signal that are produced by the human speaker; *acoustic* will refer to all changes in the speech signal that are manipulated by the experimenter or that refer to an analytic, physical view.

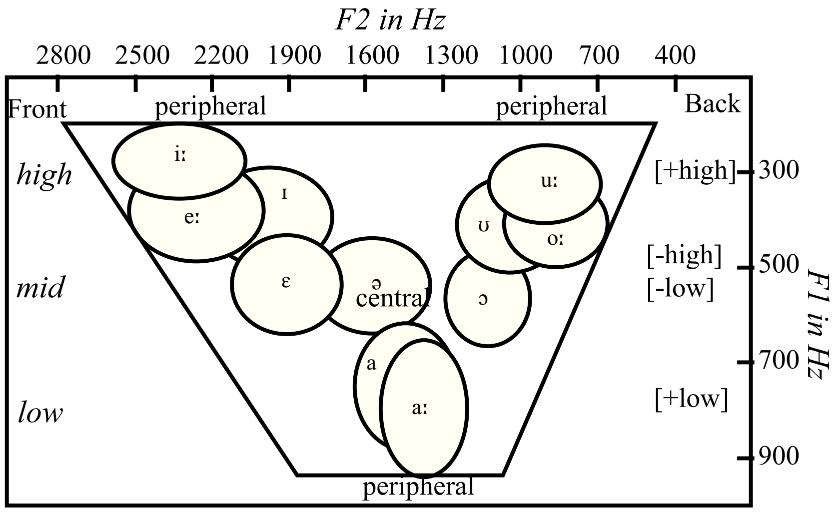

Figure 2: German vowel space, kindly provided by Walter Sendlmeier (2006). The first formant represents the vertical, the second the horizontal dimension.

Phonetic vowel duration is manifested on a temporal continuum between short and long duration. Its realization is affected by word stress (de Jong, 2004), phrase accent (Heldner & Strangert, 2001), speaking rate (Hirata, 2004), consonant voicing (Braunschweiler, 1997) as well as by the amount of syllables in a word (Lindblom, 1968). In spite of this variance, the contrast shows on average an approximate ratio of 1:2 between short and long vowels (Antoniadis & Strube, 1984; Braunschweiler, 1997; Heike, 1972; Meyer, 1904). This phonological *length* distinction is differentiated formally by the [+/- long] feature.

Vowel quality describes a phonetic distinction between two instances from one vowel type, e.g. /i/ or /u/. Two instances, which differ in quality, vary with respect to their position in the trapezoid vowel space, which represents a cross-section of the mouth. These two positions are called lax and tense. However, two contrary definitions can be found in the literature. According to Heidolph (1981, p. 907), tenseness correlates with tongue height. Tense vowels are produced with a closed, higher tongue position; lax vowels are produced with a opened, lower tongue position. According to Pompino-Marschall (2003, p. 227), tenseness correlates with the distance of the tongue with respect to the center in the vowel space. A lax vowel is central, i.e. it is articulated nearer toward the center of the vowel space. A tense vowel is peripheral, i.e. it is articulated further away from the center (Figure 2). The present thesis adopts the position of Pompino-Marschall (2003). The phonological *tenseness* distinction between lax and tense vowels will be expressed by the [+/- tense] feature. Due to the commonness of the termini, *lax/tense* describe the qualitative change that arises by different articulatory positions of the vowels within the vowel space. The terminology does not refer to the idea that the tongue muscle is more tense during the production of tense in contrast to lax vowels, as has been described elsewhere in the literature (Moulton, 1962a). The nomenclature used in the present thesis is presented in Table 1. Examples of words containing lax and tense vowels are provided in Table 2.

The phonology of German vowel length

Regarding the phonological representation of the contrast the question arises which of the two features is underlying and which is predictable. Two absolutely contradictory accounts have been presented in the phonological literature.

a) Tenseness is underlying, length is predictable (Moulton, 1962a; Reis, 1974).

b) Tenseness is predictable, length is underlying (Hall, 1992; Ramers, 1988; Vater, 1992; Wiese, 1988; 2000).

Both positions rely on different kinds of arguments. These will be presented in detail in the following.

Position a) relies on two arguments: i) the distribution of the length and tenseness contrasts to stressed and unstressed words and ii) the complementary distribution of vowel quality and consonant voicing. Vowel length is contrastive only in the stressed syllable, whereas in the

	Articulation	**Acoustics / Phonetics**	**Phonology**
temporal dimension	duration short vs. long	duration short vs. long	length [- long] vs. [+long]
spectral dimension	centralization central vs. peripheral	quality lax vs. tense	tenseness [-tense] vs. [+tense]

Table 1: The nomenclature for the description of German vowels used in the present thesis

unstressed syllable it is not functional. There are lax and tense vowels in unstressed syllables, but no short and long vowels (Table 2). Position a) argues that vowel length can be derived by rule 1: An underlying tense vowel becomes long when it is located in a syllable that is assigned word stress; otherwise it stays short.

Rule 1: vowel [+tense] → [+long] / stressed, else [-long]

The rule functions as can be shown in the example *Kritik* [kʁi.’tiːk] ‘criticism’ vs. *Kritiker* [‘kʁiː.ti.kɐ] ‘critic’. The vowel in the unstressed syllable in *Kritik* [kʁi.’tiːk] is tense. When stress falls onto the first syllable [kʁi] during the derivation process, as in *Kritiker* [‘kʁiː.ti.kɐ], it becomes long. Regarding lax vowels, the example *Dogma* [‘dɔg. ma] ‘dogma’ – *Dogmatiker* [dɔg.’maː.tɪ.k.ɐ] ‘dogmatist’ is presented. The vowel in the syllable [dɔg] is lax and stays short independently of whether the syllable is stressed or unstressed.

A second argument for position a) is based on a trend regarding the obstruent following the vowel in German disyllabic trochaic words. When one assigns the features “voiced = lax”, and “voiceless = tense” to the obstruent’s voicing contrast, there is a strong tendency in German that these features are distributed complementarily within the syllable (Kleber, John & Harrington, 2010; Lenerz, 2000; Reis, 1974). Tense vowels are followed by lax (voiced) consonants, lax vowels by tense (voiceless) consonants. Hence, the underlying quality specification of the vowel determines the voicing of the consonant. Examples are *niesen* [niː.zən] ‘to sneeze’, *labern* [laː.bɐn] ‘to babble’, for a long vowel followed by a voiced obstruent and *missen* [mɪ.s.ən] ‘to miss’, *plappern* [pla.p.ɐn] ‘to chatter’ for a short vowel followed by a voiceless obstruent. However, there are examples which violate this complementary distribution (e.g *Miete* [‘miː.tə] ‘rent’, *Pike* [‘piː.kə] ‘pike’ and *Ebbe* [ɛ.b.ə] ‘low tide’, *Widder* [vɪ.d.ɐ] ‘Aries’). Since there is only a small number of these words, Lenerz (2000, P. 198ff.) regards these counter examples as exceptions which have to be stored in the lexicon.

Position b) relies on two arguments: i) There are vowels which are contrasted only by length and ii) vowel length interacts with the syllable structure in German. It argues that tense vowels can be derived by Rule 2:

Rule 2: vowel [+long] → [+tense] / stressed, else [-tense]

Vowels which are contrasted only by means of length are [a - aː] and [ɛ - ɛː]. The [a - aː] contrast is an acceptable argument. It operates mainly on the length dimension because the formant frequencies of the two instances overlap strongly. See Figure 2. (Hoffmann, 2011; Sendlmeier et al., 2006; Weiss, 1974; Weiss, 1978). However, the [ɛ - ɛː] contrast is a weak argument. Its functionality is highly discussed in the literature. [ɛː] has been characterized as lax (Ramers & Vater, 1995; Wiese, 2000), as tense (Becker, 1998; Kohler, 1995) or it has even been excluded from the German vowel system (Hakkarainen, 1995).

The inclusion of [ɛː] into the German vowel system gives rise to minimal pairs with /e/ sounds that are contrasted by tenseness ([eː - ɛː]) such as *Seele* [zeː.lə] ‘soul’ – *Säle* [zɛː.lə] ‘halls’, *Beeren* [beː.ʁən] ‘berries’ – *Bären* [bɛː.ʁən] ‘bears’ (Hakkarainen, 1995; Wiese, 2000). However, [eː] and [ɛː] are contrasted mainly in a small number of pairs in Southern German dialects. E.g. *Bär* ‘bear’ and *sägen* ‘to saw’ are pronounced [beː.ʁən / zeː.gən] in northern and [bɛː.ʁən / zɛː.gən] in southern varieties of German (see Pompino-Marschall (2003P. 264) for the geographic distinction of northern and southern German dialects). In Modern Standard German, the contrast [eː - ɛː] is restricted to the morphophonological alternation between verbs of the 1st person singular indicative and conjunctive (*sehe* [zeː.ə] ‘I see’ vs. *sähe* [zɛː.ə] ‘I would see’). Hoffmann (2011, P. 60) shows that [ɛː] overlaps strongly with [eː] in informal speech but is distinguished from [eː] in formal speech. Becker (1998, P. 12) argues that [ɛː] is the result of a reinterpretation of orthographical conventions. The morphological umlaut from [a] to [e] (e.g. in *Tal* [taːl] ‘valley’ – *Täler* [‘teː. ɐ] ‘valleys’) is coded by the letter <ä> to indicate that it was derived from <a> and not from <e>, which would represent its correct phonetic form. The reader interprets <ä> as different from <e> and produces the lower vowel [ɛː]. It seems doubtful that in everyday speech the triple contrast [ɛ - ɛː - eː] is maintained, since there seems to be a tendency in German that the synthetic conjunctive is replaced by an analytic form in colloquial speech (*ich sähe* → *ich würde sehen* (Lotze & Gallmann, 2008)).

Furthermore, [ɛː] is not derived from [ɛ] during the derivation process which is why two possibilities remain: i) [ɛː] has to be excluded from the German vowel system and regarded as a dialectal variety (Hakkarainen, 1995) or ii) it must be specified with regard to length and tenseness (Hall, 1992; Wiese, 2000).

However, position b) does not rely only on the [a - aː] and [ɛ - ɛː] contrasts, but also on phonotactic regularities in the syllabic structure that are argued to arise from the quantity distinction. In contrast to a syllable containing a long vowel or a diphthong, a syllable containing a short vowel can have one additional consonant. This regularity can be explained in terms of syllabic constituents − slots− which function as placeholders for phonetic segments. In German, the syllabic rhyme consists of three slots which can be occupied by vowels (V) or consonants (C). A *short* segment occupies one slot, a *long* segment occupies two slots. Formalizing a long vowel as VV and a short vowel as V, it becomes obvious that syllable rhymes like [VVC], and [VCC] are equivalent in their amount of slots (Wiese, 1988, P. 62; 2000, P. 38). Additional consonants which are adjoined to the VVC / VCC rhyme *must* only be coronal obstruents (i.e. [t] or [s] (Hall, 1992; Wiese, 1988; 2000)). The addition of coronal obstruents to the full syllable is allowed furthermore only in word final position. This leads to the phonotactic restriction:

Rule 3: $VVC + C^{coron.\ obstr}$. or $VCC + C^{coron.\ obstr.}$

See examples in Table 3. Hall (1992, P. 28) uses the phonotactic restriction in Rule 3 to argue that vowel duration is the underlying feature. This is explained by means of a word that does not exist in German. If quality was the primary feature (e.g. [i]) and the vowel was not specified for duration, words like **vielp* [fiːlp] could exist in German. Figure 3 illustrates how the word **vielp* would be derived assuming that vowel quality or vowel quantity is the underlying feature. Before any phonological rule is applied, the underlying word is syllabified by means of syllabification rules, which will not be further discussed here (Hall, 1992, P.

41ff.; Wiese, 2000, P. 49ff.). In case tenseness is specified in the mental lexicon, the /i/ occupies one syllabic slot. The bilabial /p/ can be assigned to the syllable structure since it occupies an unrestricted slot and not the one which is restricted to coronal obstruents. It thus does not violate rule 3. Once each segment is sequentially assigned to a slot, syllabification is

	[VVC]	**[VCC]**
mono-morphemes	*Bahn* [baːn] 'railway' *viel* [fiːl] 'much' *Reim* [ʁaɪm] 'rhyme'	*Bank* [baŋk] 'bank' *Film* [fɪlm] 'movie' *gelb* [gɛlp] 'yellow'
poly-morphemes	*reimt* [ʁaɪmt] 'to rhyme 2nd pers. sing.' *bahnt* [baːnt] 'to channel 2nd pers. sing.'	*filmt* [fɪlmt] 'to film 2nd pers. sing.' *Films* [fɪlms] 'movie gen. sing.'

Table 3: Monomorphemic and polymorphemic examples with a short vowel (V) and a long vowel (VV). The amount of consonants (C) is complementary.

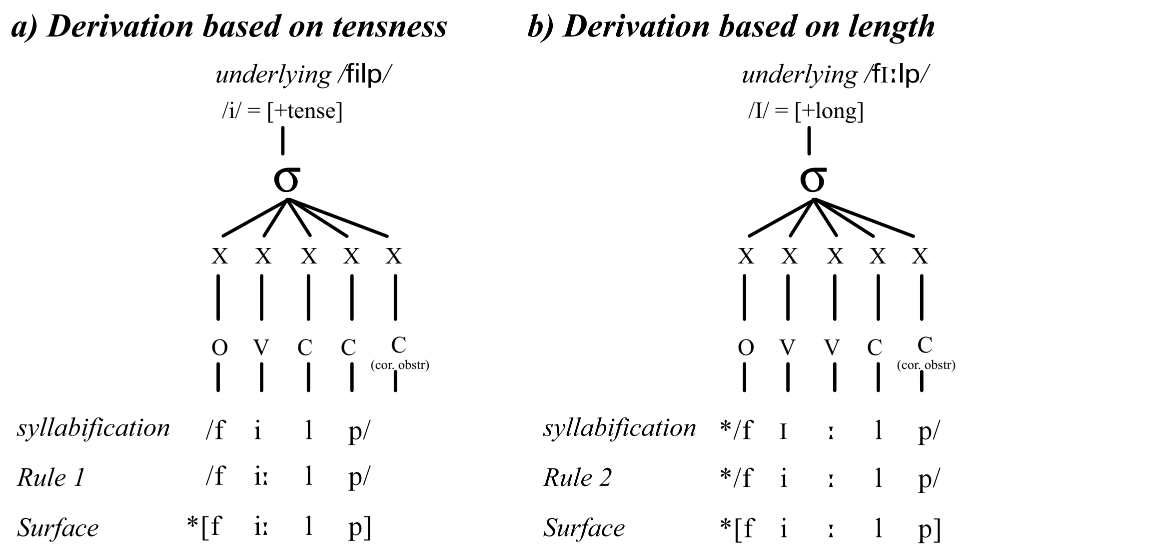

*Figure 3: Derivation of the vowel based on whether tenseness or length is the underlying feature. *[fi:lp] has a consonant cluster in the coda that violates German phonotactics. a) Tenseness allows such an illegitimate cluster, whereas b) length does not, showing that length has to be the underlying feature.*

completed and the phonological rules can be applied, in this case rule 1, rendering the vowel long. The surface structure of the underlying /filp/ would be *[fiːlp]. In case length is specified in the mental lexicon, the /ɪ/, which is unspecified for tenseness, occupies two syllabic slots. /p/ has to be assigned to the slot which is restricted to coronal obstruents. Once each segment is sequentially assigned to a slot, syllabification is completed and rule 2 applies, rendering the vowel tense. The surface structure of the underlying /fIlp/ cannot be [*fiːlp], since it violates rule 3.

Four additional remarks have to be made. i) Rule 2 has to apply prior to stress assignment in order to guarantee that vowels in unstressed syllables can have lax and tense vowels (see Table 2). After stress assignment, underlying long vowels are shortened in unstressed syllables (Hall, 1992, P. 32).

ii) Table 2 neglects the distribution of short and long vowels in compound words. Compounds like *Bahnhof* [‘baːn.hoːf] ‘train station’ have phrase stress which overrides the primarily assigned word stress at the surface level. Phrase stress is assigned by the compound stress rule (Sternefeld, 2007, P. 10; Wiese, 2000, P. 296ff). This is why compounds can have long tense vowels in unstressed syllables.

iii) There is another phonotactic argument that vowel duration has to be the underlying feature (Becker, 1998; Hall, 1992). Although the argument is motivated in production and not perception it will be presented in the following for the sake of theoretical background. Regarding the monosyllabic examples in Table 2 one can see that short vowels cannot occur in open syllables, i.e. syllables without a closing coda consonant. Long vowels can. Taking this complementary distribution into account one can infer the *minimal* stressed syllable in German: a nucleus with two positions: X^1X^2. A long vowel occupies both slots (X^1X^2), a short vowel only one (X^1). Since both slots *have* to be occupied in German, a consonant is assigned to the second slot (X^2). Figure 4a illustrates how phonetic segments are connected to the intrasyllabic structure in monosyllabic words. The intrasyllabic structure is represented by a tree in which the minimal nucleus is governed by the syllable to which an onset and a coda can be connected. In this sense, not vowel length by itself is contrastive but the syllable structures VV vs. VC. In unstressed vowels the additional slot (X^2) is deleted when it is not occupied by a consonant (Figure 4b, Becker (1998)).

	stressed		*unstressed*	
	lax	*tense*	*lax*	*tense*
monosyllabic	*weg [‘vɛk]* *Schiff [‘ʃɪf]* *Block [‘blɔk]* *dünn [‘dʏn]* *Rost [‘ʁɔst]*	*Tee [‘teː]* *sie [‘ziː]* *schief [‘ʃiːf]* *Bahn [‘baːn]*	Ø	Ø
polysyllabic	*Dogma [‘dɔg. mɑ]* *Doktor [‘dɔk.tɔɐ]* *Impuls [ɪm.'pʊls]* *Galopp [gɑ.'l.ɔp]* *bette [‘bɛ.t.ə]* *Hölle [‘hœ.l.ə]* *dünner [‘dʏ.n.ɐ]*	*spuken [‘ʃpuː.kən]* *Doktoren [dɔk.’toː.ʁən]* *Höhe [‘hø.ə]* *fehlte [‘feːl.tə]*	*Impuls [ɪm.'pʊls]* *Doktor [‘dɔk.tɔɐ]* *Doktoren [dɔk.’toː.ʁən]* *Kollege [kɔ.’l.eː.gə]* *blockieren [blɔ.’k.iː ʁən]*	*Kritik [kʁi.'tiːk]* *kritisieren [kʁi.ti.‘ziː.ʁən]* *Propan [pro.'pɑːn]* *legal [le’.gɑːl]* *total [to.’tɑːl]* *Idee [i.’deː]* *real [ʁe.’ɑːl]*

Table 2: Word examples. In polysyllabic words, the syllable with the vowel in question is underlined. Word stress is marked with an apostrophe. The syllable boundary is marked with a point. Ambisyllabic consonants, i.e. consonants shared by two syllables, are flanked by two points.

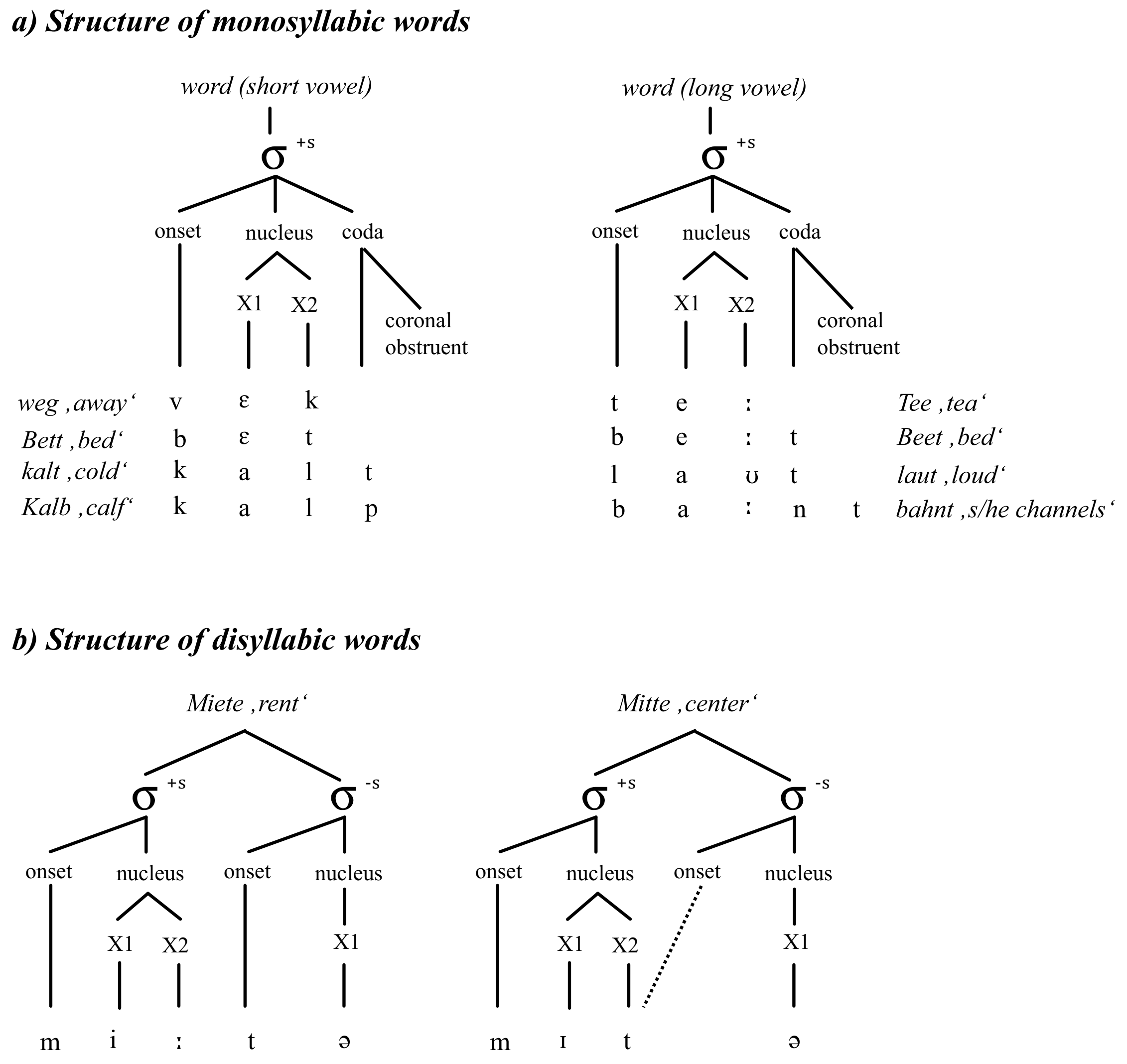

*Figure 4: Syllable structure in German monosyllabic **(a)** and disyllabic **(b)** words. The syllable consists of an obligatory nucleus that contains two constituents (X1X2) which have to be occupied either by a long vowel or by a short vowel + consonant. **b)** In disyllabic words the former results in a tautosyllabic intervocalic consonant, the latter in an ambisyllabic consonant, i.e. the consonant is shared by both syllables.*

The importance of the minimal syllable becomes evident by taking disyllabic words into consideration. An underlying long vowel occupies both slots (X^1X^2) in the syllable, which is why there is no slot left for an additional consonant. The vowel is assigned [+tense] by rule 3 and the intervocalic consonant is connected to the following syllable (like /t/ in *Miete* [ˈmiː.tə] ‚rent'). An underlying short vowel occupies only one syllabic slot (X^1). It is assigned [-tense] by rule 2. Since it has to be obligatorily occupied, the second slot (X^2) becomes occupied by the intervocalic consonant during the derivation (Wiese, 2000, P. 36). A single intervocalic consonant is assigned to both syllables, i.e. it becomes ambisyllabic (like the /t/ in *Mitte* [ʻmɪ.t.ə] ‚rent'. Ambisyllabicity is indicated by the two embracing points). Figure 4b illustrates the intrasyllabic structure in disyllabic words.

Since all vowels are specified for length in the mental lexicon, and since single consonants after a short vowel become ambisyllabic, ambisyllabicity is possible after both stressed and unstressed vowels (Wiese, 2000, P. 36). After the application of rule 3, long vowels, however, are shortened in unstressed syllables (/l/ in *Kollege* [kɔ. ʼl. eː. gə] ‚colleague' is ambisyllabic; /p/ in *Propan* [pro. ˈpɑːn] ‚propane' is tautosyllabic. See Table 2 for further examples). Since tenseness is predictable in both, stressed and unstressed vowels, it is always allophonic and does not have to be stored as a feature in the mental lexicon.

Hall (1992) and Wiese (2000) argue that the intrasyllabic structure is assigned to the underlying segments during the derivation process. (Becker, 1998) and (Lenerz, 2000) argue the other way round: Syllable structure is lexically specified and vowel length as well as quality are assigned during the derivation. In a nutshell: The syllabic position X^2 is specified in the mental lexicon to whether it has to be occupied by a consonant or by a vowel. The first results in a short lax vowel, the latter in a long tense vowel. This assumption, called *syllable cut*, is production driven. It regards length and tenseness as the result of articulatory movements which are based on the lexicalized syllable structure.

Recently, a third phonological view has been proposed by Caratini (2007). Using statistical and diachronical methods, she argued against an ambisyllabic representation of vowel length. In her point of view it represents a "circular representation" as ambisyllabicity is derived from vowel length, vowel length is derived from ambisyllabicity. Rather, the vowel length contrast constitutes a complementary duration contrast with the following consonant, as it is found in Swedish or Norwegian (Strangert, 2001). At the underlying phonological level, the short vowel is followed by a consonantal cluster or geminate (i.e. VCC); the long vowel is followed by a single consonant (i.e. VVC). She shows that this specification is a remnant from Old and Middle High German and is still mirrored in modern German orthography, where a short vowel is followed by two consonantal letters in contrast to a long vowel which entails a single consonantal letter (e.g. *Mitte* 'center' vs. *Miete* 'rent'). Since German developed into a language which forbids consonantal geminates, these underlying segments are deleted at the surface level.

Summary of the phonological arguments

The above presented positions and arguments can be summarized as follows:

Position a: Vowel quality (i.e. [+/- tense]) is the phonological feature that contrasts between short and long vowels and which is stored in the mental lexicon; vowel length is assigned during the derivation based on rule 1. The arguments for this position are that quality is functional in stressed and unstressed syllables whereas vowel length is functional only in stressed syllables. Position a) assumes that the vowels /a ɑ e ɛ i ɪ o ɔ u ʊ y ʏ ø œ/ are the categorical prototypes represented in the mental lexicon, which become important in the experimental section.

Position b: Vowel length (i.e. [+/-long]) is the feature that contrasts between short and long vowels and which is stored in the mental lexicon; vowel quality is assigned during the derivation based on rule 2. The argument for this position are the regularities regarding ambisyllabicity and the complementary amount of phonetic segments which can be assigned to a syllable. Position b) assumes that the vowels /A A: E E: I I: O O: U U: Y Y: Ø Ø:/ are the categorical prototypes in the mental lexicon (vowels are written in capitals to indicate that they lack a tenseness specification).

The validity of these two positions was investigated in the course of the present doctoral thesis.

The psychoacoustic characteristics of phonological contrasts

In general, phonological accounts like those presented for German vowel length assume that phonetic contrasts are represented in the mental lexicon by forms which are less specified in terms of the amount of contrastive features/cues than the phonetic surface forms. These phonological accounts provide grammars that describe what rules are applied to the alleged underspecified representation during derivation at whose end it is attributed the full set of cues needed for production.

Finding the psychoacoustic characteristics such as identification patterns, discrimination sensitivity or goodness judgments can assess the importance of a specific cue during speech perception that is used for a contrast and as such to infer its representation in the mental lexicon. The following section will summarize the findings about which contrasts and cues exhibit which perceptual characteristics and how these are supposed to be represented in the mental lexicon. The assessment of such characteristics for the German vowel length contrast will be used to infer which phonetic cue can be interpreted as the predictable and which as the unpredictable feature in the phonological description of the vowel system.

One of the early findings about speech perception was that speech contrasts exhibit *categorical perception* (CP). This has been shown especially for consonantal contrasts. i) The duration of VOT, indicating plosive voicing (Lisker et al., 1964), is highly sensitive to speaking rate (Flege, 1988). In addition, voicing is supported by cues like the vowel/VOT duration ratio (Kohler, 1979) or the F1 onset frequency (Jiang et al., 2006). ii) The slope of the transition in the second formant indicating the coronal consonant [d] depends on the height of the following vowel (e.g. [da] or [di] see Liberman, Harris, Hoffman & Griffith (1957); Pompino-Marschall (2003); Strange (1987)).These variances are reduced to an either-or decision during perception. I.e. there is no discrimination sensitivity between phonemes with different acoustic forms from one category. This reduction is reflected by identification judgments that only one type of [b] or [d] (place of articulation) and [b] or [p] (voicing) is perceived although acoustically varying tokens are presented. Two acoustic forms can only be differentiated when they stem from the opposite sides of the category boundary, meaning that they have to be two different phonemes (Liberman et al., 1957; Studdert-Kennedy, Liberman, Harris & Cooper, 1970).

There are findings, however, that vowels do not show such characteristics as attested for the perception of consonants (Gerrits et al., 2004; Iverson & Kuhl, 1993; Kuhl, 1991). Rather, vowels from one category that vary in their spectral acoustic form can still be discriminated. However, sensitivity to spectral changes is not the consistent within the entire category but shows a *Perceptual Magnet Effect*: It decreases from the edge of the category towards the *prototype*. The prototype is the acoustic form that yields the best goodness ratings in comparison to other acoustic forms from the same category (Iverson et al., 1993; Kuhl, 1991).

The strict view that in categorical perception there is no discrimination within a category has been critically questioned. Generally, sensitivity to change and its increase to across-categorical items has been assessed by having subjects indicate whether or not they perceive a difference between two subsequent items (AX – task, will be explained at a later point of this text) (Macmillan & Creelman, 2004). Massaro et al. (1983) introduced a new discrimination test in which subjects had to indicate where on a continuum between the endpoints of a contrast they would locate a perceived item. By these means they could show that fine acoustic differences were perceived, even for items stemming from a voiced-voiceless continuum, as Hanson (1977) could already show. Hence, the strict account of categorical perception is not valid. Recently, Toscano et al. (2010) replicated this finding in a neuroimaging study. In their review of several studies on identification and discrimination, Schouten et al. (2003a) argued that a finding of categorical perception is based on a bias of the subject. If a test is used in which subjects tend to rely on their phonological representation in the mental lexicon in order to discriminate between items, categorical perception is highly probable to be found. If, however, the test is constructed that this is not possible, no categorical perception can be found. In this sense, Gerrits et al. (2004) showed that the AX-task increases the probability that subjects refer to phonological labels and as such show higher discrimination sensitivity to across-categorical than for within-categorical items. As already argued above, categorical perception has been assumed for consonant but not for vowel perception. However, Gerrits et al. (2004) showed categorical perception for vowel perception, which supported their view that this perceptual characteristic depends on the method.

This finding is of great importance for the present doctoral thesis. It shows that the presence of a phonological category changes discrimination performance in the AX-task. This has already been shown by Miyawaki, Strange, Verbrugge, Liberman, Jenkins & Fujimura (1975), who investigated the perception of the American /l-r/ contrast by American and Japanese native speakers. The latter group, for whom this contrast is not phonological, did not show any increase for across-categorical items. Regarding vowel length, Ylinen, Shestakova, Alku & Huotilainen (2005) have shown that speakers of a language with a phonological length distinction (Finnish) showed such an increase, speakers of a language lacking the distinction (Russian), did not.

Previous behavioral studies on vowel length and tenseness

The peculiar interaction of vowel duration and vowel quality in German makes German native speakers interesting subjects for studies on how temporal and spectral cues are processed during speech perception. Behavioral studies that used non-native vowels showed that German native listeners are highly sensitive to changes in vowel duration (Bohn & Flege, 1990) and vowel quality (Escudero, Benders & Lipski, 2009). The processing of one cue affects the processing of the other: Lax/more central vowels have a higher probability to be categorized as short than tense/more peripherial vowels (Gussenhoven, 2007; Lehnert-LeHouillier, 2010; Meister & Meister, 2011). This effect is not reciprocal regarding vowel duration. When a long tense vowel is shortened acoustically, not only the perceived quality of the vowel can be changed but also the category: Mid tense long [oː, eː] are perceived as high lax short [ɪ, ɛ] (Bennett, 1968; Lindner, 1976; Sendlmeier, 1981; Weiss, 1974).

The picture is more complex regarding the weighting of vowel duration and quality as cues for the phonological length contrast. The results of fifty years of perceptual studies on German vowel length can be summarized in the following sentence: "The importance of the duration cue is inversely proportional to the distance between the qualities of a given vowel pair" (Bennett, 1968, P. 65). In other words: High and mid vowels (/i/, /u/) have a large spectral distance between the short and long instances which is why vowel duration is less important; the low vowel /a/ shows a small spectral distance between the short and long instances, which is why vowel duration is more important (see Figure 2 for comparison of the distance as well as (Hoffmann, 2011) and (Sendlmeier et al., 2006)).

It has repeatedly been shown that the identification and discrimination patterns of vowels such as /i/ and /e/ show characteristics which can only be explained if one assumes that they have a phonological representation in the mental lexicon (Diesch, Iverson, Kettermann & Siebert, 1999; Iverson et al., 1993; Kuhl, 1991; Lotto et al., 1998). So far, however, no study could show such characteristics for the perception of lax and tense vowels. This is the point, on which the present thesis focuses. If it could be shown that the lax and tense vowel show effects of a category boundary insofar as discrimination of across-categorical items is increased, the contrast could be stated as phonological. If no such finding will be found, the contrary should be the case.

Processing of speech in the brain

As already mentioned in the introduction, the studies in the present doctoral thesis investigated – besides their psychoacoustic characteristic – how the phonetic cues used to indicate the German vowel length contrast are processed at the cortical level of perception. The following section summarizes recent findings about how the speech signal and the segments it consists of are processed in the brain.

Articulating a sentence results in a continuous signal in which no distinct segments or word boundaries are indicated. However, the auditory system in the brain is able to extract words from this continuous signal using various pre-lexical, word-form and semantic mechanisms (McQueen, 2007) and cues such as stress patterns (Jusczyk, Houston & Newsome, 1999) or phonotactic regularities (McQueen, 1998). The auditory system is also able to further divide words into consonantal and vocalic segments although they are connected with each other by segment dependent transitions (Strange, 1987) and, furthermore, although they mutually affect their acoustic shape due to coarticulation (Strange & Bohn, 1998). This means that, on the one hand, the auditory system can provide a macroscopic perception of speech (i.e. words, sentences) which results in a cognitive analysis in terms of semantic meaning. On the other hand it can analyze the continuous signal in a microscopic way resulting in distinct and categorical information.

In addition to psychoacoustic techniques, the complex process of speech perception in the brain has been revealed by means of different neuroimaging techniques. It could be shown that, at an initial stage, the continuous information about the speech signal is represented on tonotopic (i.e. frequency) and amplitopic (i.e. loudness) maps in the primary auditory cortex, which are located at the posterior supratemporal plane in the left and right hemisphere (Formisano, Kim, Di Salle, van de Moortele, Ugurbil & Goebel, 2003; Hickok & Poeppel, 2000; Pantev, Hoke, Lehnertz & Lütkenhöner, 1989a; Pantev, Hoke, Lutkenhoner & Lehnertz, 1989b; Rademacher, Morosan, Schormann, Schleicher, Werner, Freund & Zilles, 2001). However, the hemispheres differ with respect to the type of information they process preferentially at *higher* processing levels. The right hemisphere is specialized for the processing of spectral information, the left hemisphere for the processing of temporal information (Obleser, Eisner & Kotz, 2008; Zatorre & Belin, 2001). This specialization seems to result from different sizes of *temporal windows of integration* (TWI). A TWI can be seen as the perceptual resolution of the auditory system, analogous to pixels in the digital graphic processing which code visual information. (Yabe, Winkler, Czigler, Koyama, Kakigi, Sutoh, Hiruma & Kaneko, 2001b) have shown that varying information within a TWI is processed as an unitary cognitive event. In the left hemisphere, the TWI is small (25 – 50 ms) and is used to process information with a high temporal but a low spectral resolution such as consonant information. In the right hemisphere, the TWI is large (150 – 200 ms), accounting for information with a low temporal but a high spectral resolution (Boemio, From, Braun & Poeppel, 2005; Poeppel, 2003). The high spectral resolution is used to process spectral changes such as pitch or the extraction of vowel identity (Obleser et al., 2008; Zatorre et al., 2001). This matches the findings of Wallace & Blumstein (2009) who have shown that the optimal vowel duration to extract the vowel identity is 150 ms. The information is worse represented, the shorter the vowel gets. Several studies have also shown that the long TWI seems to be used to integrate and structure speech to chunks with durations that match the average size of a syllable (Boemio et al., 2005; Luo & Poeppel, 2007).

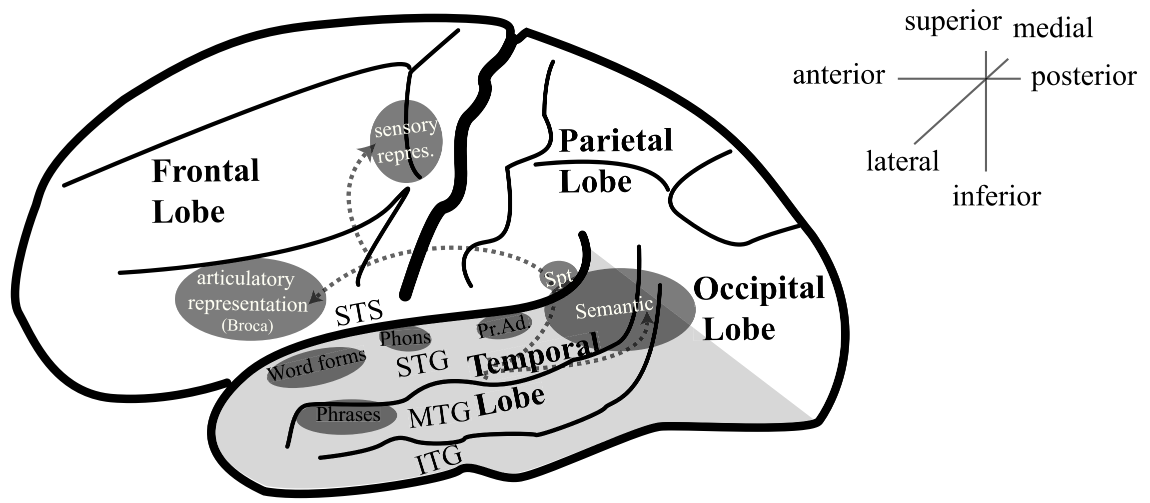

Figure 5: Schematics of the left hemisphere. The grey-shaded area represents the temporal lobe which contains the primary and secondary auditory cortices. Locations of areas are indicated for word forms, phonemes and phrases. Legend: STG = superior temporal gyrus. MTG = middle temporal gyrus. ITG = inferior temporal gyrus. Pr. Ad. = primary auditory cortex. Spt = Sylvian parietal-temporal.
Figure according to DeWitt & Rauschecker (2012); Poeppel & Idsardi (2011).

It has been advanced that at this point, the speech signal is segmented into distinct events such as syllables (Luo et al., 2007; Mehler, Dommergues, Frauenfelder & Segui, 1981) or single segments (Gutschalk, Micheyl, Melcher, Rupp, Scherg & Oxenham, 2005; Yabe et al., 2001b). This is accomplished by a neural mechanism which compares the amount of the spectral separation between sequential segments (Rose & Moore, 2000; Yabe et al., 2001b). Sequential information which has a spectrally small separation is integrated into an unitary event. Information that has a spectrally large separation is segregated as distinct events.

The speech signal’s spectral and temporal information is conveyed from both hemispheres to the superior temporal lobe in the left hemisphere, which responds preferentially to speech sounds (Hickok et al., 2000; Poeppel et al., 2011). It can be assumed that the speech information is still continuous at this stage (Toscano et al., 2010). By overlaying the results of more than 100 neuroimaging experiments onto a single neurocortical map, DeWitt et al. (2012) have shown that the superior temporal lobe is divided into different areas, each specialized for recognition of patterns in phonemes and integrating them into words as well as phrases. The more anterior a specific area is located in the temporal lobe, the more complex is the representation of speech (Figure 5. The temporal lobe is indicated by the grey shaded area.).

The secondary area which is responsible for phoneme processing constitutes an initial stage of phonological processing. It has recently been found that vowels are represented on a cortical map that mirrors a feature specification. Front vowels are processed in a more anterior, back vowels in a more posterior position in the superior temporal lobe (Obleser, Lahiri & Eulitz, 2004). This arrangement, however, might be interpreted to represent solely a tonotopic organization, since the front-back dimension is acoustically represented by one spectral cue, namely the second formant (see Figure 2). Front vowels have a high second formant, low vowels have a low second formant. The argument for a mapping in terms of features is provided by vowels which are contrasted in rounding, a phonological feature that affects the second and third formant. Rounded vowels have a lower second formant in contrast to unrounded vowels. The third formant, however, is inconstantly affected: It increases in back vowels but decreases in front vowels (Scharinger, Idsardi & Poe, 2011). In spite of the inconsistent spectral changes in the third formant, the mapping of rounded vowels onto the cortical map in the temporal lobe is consistent: Rounded vowels are processed in a more inferior position; unrounded vowels are processed in a more superior position. This finding indicates that the mapping of phonetic information to neural cohorts mirrors phonological rather than acoustic categories.

The information extracted in these secondary auditory areas is conveyed into other cortical areas by means of two *paths*, i.e. synaptic connections (Hickok & Poeppel, 2004; Poeppel et al., 2011). A *ventral path* projects to areas in the posterior inferior temporal lobe and superior temporal sulcus (Figure 5). These areas interpret the acoustic signal in terms of semantic information and will not be of any concern in the present thesis. A *dorsal path* projects spectro-temporal information to frontal regions containing somatosensory and secondary motor cortices, including the Broca area. Here, phonetic information is assumed to be interpreted in terms of motor programs that represent articulation gestures (Liberman & Mattingly, 1985). The dorsal path is furthermore assumed to constitute the cortical network of the phonological working memory (Baddeley, 1992; Wilson, 2001). The network, mediated by an small area in the posterior Sylvian fissure at the parietal–temporal boundary (Figure 5: Area Spt, Hickok, Buchsbaum, Humphries & Muftuler (2003)), repeatedly transforms the information about the speech signal from an auditory representation to a motor representation which is used to simulate a sensory representation of the articulatory movements. By these means, the loop i) temporally stores the phonological information of a speech signal and ii) rehearses the articulatory gestures to produce it. The latter function is regarded as the mechanism which is used for the acquisition of the first language and new vocabulary (Baddeley, 2003; Koelsch, Schulze, Sammler, Fritz, Müller & Gruber, 2009).

Previous studies on the neural correlates of vowel length

The neural processing of duration in speech signals has been studied primarily by means of the mismatch negativity (MMN), a neural response that is elicited preattentively, i.e. before conscious perception, when in a stream of frequent (standard) stimuli a rare (deviant) stimulus is detected that represents a change in acoustic, phonetic, phonological etc. information (May & Tiitinen, 2010; Näätänen, Paavilainen, Rinne & Alho, 2007). By these means it was found that shorter vowels produce higher neuronal activity than longer vowels (Inouchi, Kubota, Ferrari & Robert, 2003; Inouchi, Kubota, Ohta, Shirahama, Takashima, Horiguchi & Matsushima, 2004; Jaramillo, Alku & Paavilainen, 1999). The enhancement might be produced due to a worse extraction of vowel identity in shorter vowels, as was shown by Wallace et al. (2009). Nenonen, Shestakova, Huotilainen & Näätänen (2003) and Ylinen, Shestakova, Huotilainen, Alku & Näätänen (2006) have shown that changes in vowel duration produce larger MMNs in speakers of a quantity language, i.e. a language using a length distinction, in contrast to speakers whose language lacks one. They argue that the enhancement mirrors the former’s sensitivity to duration changes because they often occur in their native language. Additionally, these speakers have access to durational prototypes which additionally enhance their sensitivity. Speakers of non-quantity languages do not have such prototypes. This was particularly shown by Nenonen et al. (2003) in a study with nonspeech signals. Finnish native speakers, i.e. who know a phonological length distinction, were more sensitive to duration changes in nonspeech signals than second-language speakers of Finnish.

Regarding the neural processing, vowel length seems to differ from other contrasts such as place of articulation ([ba] vs [da]). To date, since a MMN has been found only when the deviant represents an across-categorical but not a within-categorical change, these exhibit characteristics of categorical perception (Dehaene-Lambertz, 1997; Phillips, Pellathy, Marantz, Yellin, Wexler, Poeppel, McGinnis & Roberts, 2000). Vowel duration, however, seems to be processed in terms of acoustic rather than categorical perception because no differences in neural activity between within-categorical and across-categorical changes could be found (Hisagi, Shafer, Strange & Sussman, 2010; Ylinen et al., 2006). Rather, the difference between phonological length and acoustic duration changes is reflected by lateralization. Phonological length is processed bilaterally as well as preferentially in the left hemisphere. Non-phonological changes are processed preferentially in the right hemisphere (Inouchi et al., 2003; Kasai, Yamada, Kamio, Nakagome, Iwanami, Fukuda, Itoh, Koshida, Yumoto, Iramina, Kato & Ueno, 2001; Minagawa-Kawai, Mori & Sato, 2005).

Methods used in the present thesis

The goal of the present doctoral thesis was to investigate the psychoacoustic characteristics and the neural mechanisms that represent spectro-temporal processing in order to validate or falsify the formal phonological description of the German vowel length contrast. Cognitive processing was investigated using traditional behavioral methods such as the identification and discrimination test. Neural processing was studied by using magnetoencephalography (MEG). How these techniques work will be presented in the following section.

Psychoacoustic tests

As the presence of a phonological category has been shown to affect the psychoacoustic characteristics of a listener, i.e. his or her perceptual performance (Eulitz, Diesch, Pantev, Hampson & Elbert, 1995; Kuhl, 1991; Liberman et al., 1957; Pisoni, 1973b), three behavioral tests were used in the present study in order to assess the characteristica with regard to the vowel length contrast: the identification test, the discrimination test and the goodness rating. The stimuli for these tests are items that are randomly chosen from an acoustic continuum that spans between the two contrastive phonemes (see below).

The *identification test* forces the subject to categorize a single item, which was chosen randomly from the continuum, to a category. By repeating this procedure several times with all items from the continuum, the temporal or spectral point in the continuum can be found that represents the boundary between the contrastive phonemes (Macmillan et al., 2004).

The *discrimination test* assesses the sensitivity to change along the continuum. This is performed by forcing the subject to indicate whether or not they perceived a difference between two presented items (AB). Generally, the two items are the juxtaposed acoustic forms or those which are separated by one step (Macmillan et al., 2004). This procedure assesses the sensitivity to change, but not the just-noticeable-difference, which is needed to perceive a change (Lapid, Ulrich & Rammsayer, 2008). This is why a staircase procedure (**AX**-task) was used in the present study, in which the acoustic form of the second item (**X**) was changed for the next presentation of the first item (**A**). Subjects had to indicate whether or not they perceived **A** and **X** as same (in a spectral or temporal way). Depending on their answer, the acoustic difference was made either larger or smaller (see studies for details). By these means the actual acoustic form in **X** ‘hovered’ around the just-noticeable-difference which was found by mathematical procedures explained in part II of the present thesis.

The *goodness rating* assesses the acceptability of the acoustic information in a given item from the continuum as a representative form of a certain category (e.g. short). In contrast to the previous two tests, subjects had to provide a relatively continuous judgment that ranges between 'very good' and 'very bad'.

The present studies used different sets of speech and nonspeech stimuli which were synthesized using the synthesizer described in Hertrich & Ackermann (1999); 2007) (stimuli description see below). Synthesizing the speech stimuli was preferred over recording a real speaker and changing the signal acoustically in order to be able to control the spectral and temporal information contained in the stimuli. The speech stimuli in all studies were nonsense words in order to exclude any accidental lexical bias for one or the other word. In contrast to previous studies, which used open monosyllables (Lehnert-LeHouillier, 2010) or even single vowels (Escudero et al., 2009), the speech stimuli consisted of two syllables. This guaranteed that shortening the vowel in the first syllable did not violate German phonotactics, which prohibit short vowels in open syllables (Becker, 1998). The second syllable always contained a schwa, which was not changed in duration. This was assumed to provide a stable model of speech tempo (see Hirata (2004) for speech tempo effects).

Magnetoencephalography (MEG)

In addition to behavioral studies, cortical processing of spectro-temporal information in the auditory stimuli was investigated by the use of MEG. MEG was chosen because of its high temporal resolution in the range of milliseconds that enables it to record electrophysiological activity evoked by temporally fast changes as they are present in the speech signal (see Figure 5, which shows the spatio-resolution of MEG in comparison to common neuroimaging and electrophysiological techniques such as functional magnetic resonance imaging −fMRI−, near infrared spectroscopy −NIRS−, positron emission tomography −PET−, electroencephalography −EEG− and) (Rugg, 1999). Furthermore, the MEG has a high spatial resolution which makes it possible to localize the center of the neural activity elicited during the perception of auditory signals, i.e. the *equivalent current dipole*, by means of mathematical fitting methods (see Figure 7a for a dipole location in the auditory cortex. Sams, Kaukoranta, Hämäläinen & Näätanen (1991)). In auditory studies, a dipole provides information about the strength of an *auditory evoked response* as a function of time (see Figure 7d for a stereotypic response in the auditory cortex). Such a time course generally consists of a deflection around 50ms (M50) and a deflection around 100ms (M100) after stimulus onset which represent a maximum of neural activity (Hertrich, Mathiak, Lutzenberger & Ackermann, 2000; Hillebrand & Barnes, 2002). The M50 has been shown to reflect the initiation of signal processing (Lijffijt, Lane, Meier, Boutros, Burroughs, Steinberg, Gerard Moeller & Swann, 2009), whereas higher cortical processes like the extraction of vowel identity were shown to be reflected in the amplitude and latency of the M100 (Poeppel, Yellin, Phillips, Roberts, Rowley, Wexler & Maranth, 1996; Roberts, Flagg & Gage, 2004a; Scharinger et al., 2011).

The MEG is an imaging technique that records magnetic fields around the human skull, which are produced by neuronal currents when the dendrites of around 50000 pyramidal neurons are synchronously activated in response to auditory input (see schematics in Figure 7b&c, Hämäläinen et al. (1993); Heinze et al. (1999)). These magnetic fields are ten thousands to a million times weaker than the magnetic field of the earth (Figure 6b). Since the strength of a magnetic field is inversely proportional to the square of the distance to its source, and since the neuronal magnetic fields are so weak, the recorded signal might be disturbed by other magnetic sources. This is why the recordings have to be performed in a magnetically shielded chamber in order to prevent artifacts.

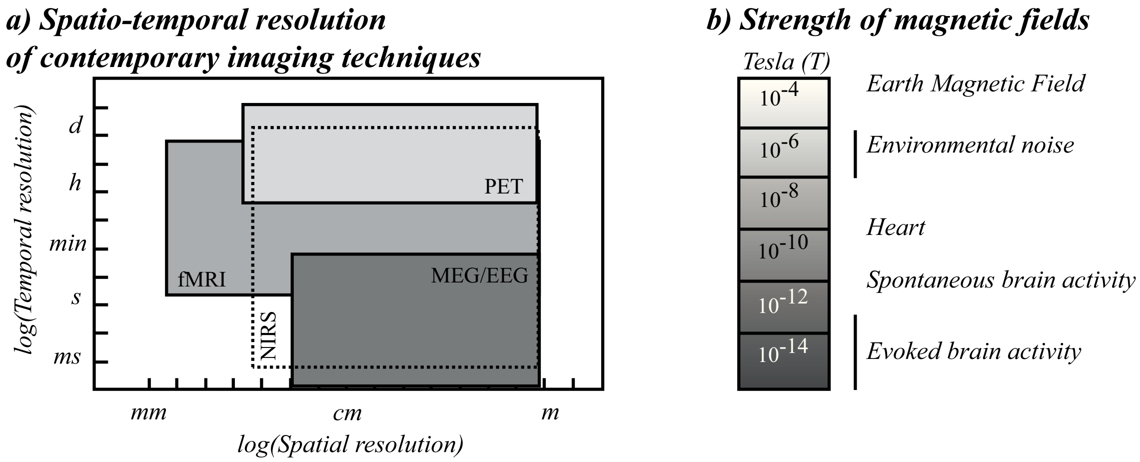

*Figure 6: **a**) spatio-temporal resolution of the neuroimaging and electrophysiological techniques (Figure according to Strangman, Boas & Sutton (2002)). **b**) Strength of the magnetic fields. Kindly provided by Hubert Preissl and Christoph Braun (MEG Center, Tübingen).*

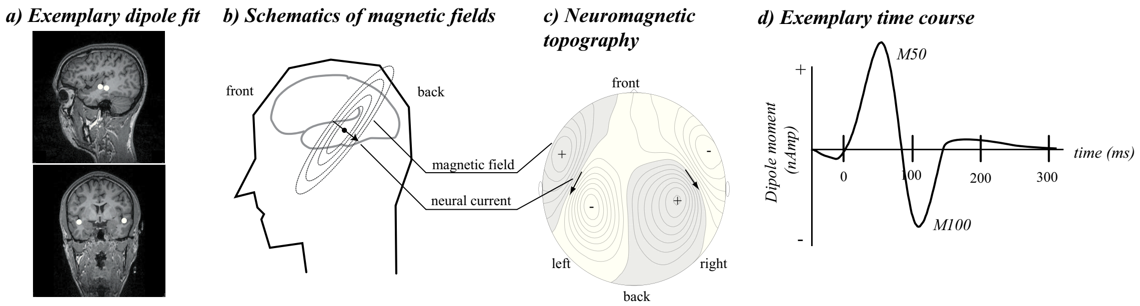

*Figure 7: **a**) Dipole location representing the source of evoked responses in the auditory cortex.*

***b**) Schematics of weak magnetic field in the brain that can be measured with the MEG. **c**) Typical neuromagnetic topography in response to an auditory stimulus for one single time point. **d**) Stereotypic evoked response to auditory stimuli. Figures according to Hämäläinen, Hari, Ilmoniemi, Knuutila & Lounasmaa (1993); Heinze, Münte, Kutas, Butler, Näätänen, Nuwer & Goodin (1999).*

Summaries of the studies

Whereas the theoretical literature assumes that the length contrast is differentiated in the mental lexicon by solely one feature – length or tenseness – phonetic studies show that the importance of each cue depends on vowel height. In the following, the five studies performed in the course of the present doctoral thesis will be summarized. Their goal was to elaborate the discrepancy between phonological description and phonetic information by expanding the range of behavioral methods used so far and investigating the neural processing of vowel length by means of MEG. The summaries concentrate on the results about the interaction between vowel duration and vowel quality and its implications on the lexical representation and its formal description of German vowels. The detailed statistical analyses and statistics are postponed to the experimental section.

The sections are structured as follows: The underlying hypotheses are presented for each study. After a description of the stimuli is provided, the issue of behavioral and neural methods is addressed. Finally, the results are summarized and discussed. Studies I & II & V addressed the question of duration processing. They concentrated on the vowel /a/ in which the spectral information of the short and long instance overlaps strongly and therefore omitted the effects of vowel quality (Figure 2). The effects of vowel quality were investigated in study II & IV, addressing the question how vowel duration and vowel quality interact.

Summary of Study I

Hypotheses

Study I investigated the psychoacoustic characteristics of vowel duration. It expanded the previous studies, which used solely identification paradigms, by assessing the just-noticeable-differences as a function of duration and performing goodness ratings for each category. Study I was based on two assumptions. i) Smits, Sereno & Jongman (2006) found that when a duration category is formed, temporal categorization is accomplished by comparing the perceived segment with respect to a category boundary. Taking this into account, it was hypothesized that if the just-noticeable-differences (JNDs) for duration changes did not cross the category boundary vowel duration would be perceived continuously. If temporal JNDs crossed the category boundary, vowel duration would be perceived categorically. ii) Ylinen et al. (2005) have shown that a lexicalized category changes the discrimination characteristics as a function of duration. The sensitivity is increased at the category boundary, but decreased within the category. If vowel length is lexicalized, such a sensitivity change should be found.

Furthermore, if duration is lexicalized, identification should result in a steep category boundary. (Pisoni & Tash, 1974b) have shown that identification response times are faster for clear within-categorical items than for ambiguous items from the category boundary. If duration was lexicalized, a similar result should be found in the present study. Finally, in line with Kuhl (1991), prototypical (i.e. clear within-categorical) items should have good goodness judgments in contrast to untypical items.

Stimuli

Study I used a duration continuum of the vowel /a/ in steps of ~10ms, which was embedded in the disyllabic context /t_tə/. The structure of the stimuli is illustrated in Figure 8a.

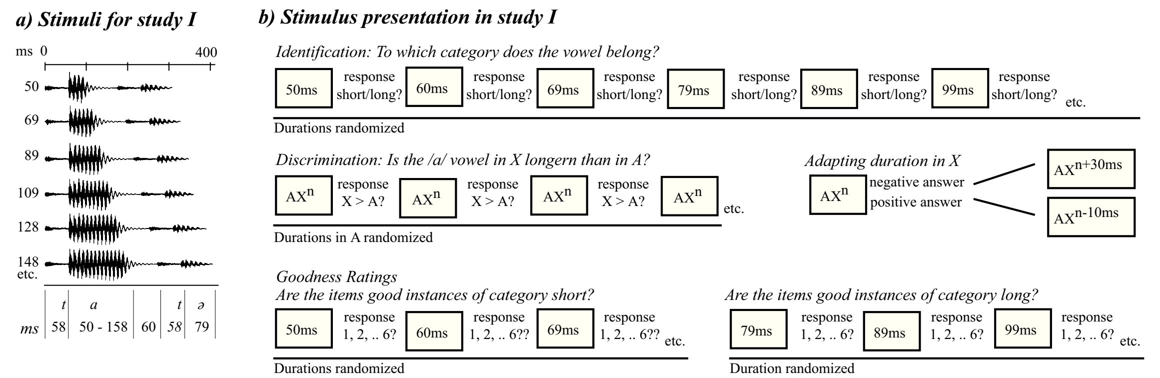

Figure 8: Experimental design in study I. ***a)*** *Waveforms and temporal structure of the stimuli.* ***b)*** *Stimulus presentation in all three behavioral tasks.*

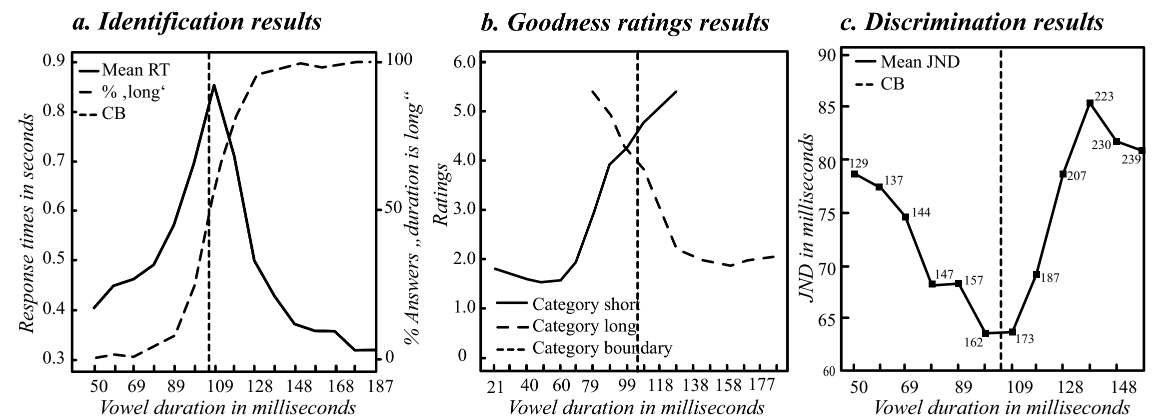

Figure 9: Results from Study I. ***a)*** *Identification results including the mean identification curve (% 'long'), the mean response times (RT) and the category boundary (CB) as found by the arcus-tangens-fit.* ***b)*** *Mean group averages for the goodness rating. Sets were rated as though the vowel belonged to category short or category long.* ***c)*** *Mean just-noticeable-difference (JND). The numbers next to the squares indicate the vowel duration (in milliseconds) in the comparison. By means of Student's T-test it was found that in the category short the vowel durations in the comparisons were significantly longer than the category boundary (p< 0.008, Bonferroni corrected).*

Methods

20 subjects participated in three behavioral tests. Figure 8b illustrates the stimulus presentation in study I. The /a/ duration continuum was subject to an *identification test*, which demanded a categorization of the vowel in the first syllable of the perceived word to either short or long. Furthermore, a *goodness rating* was performed that used two continua, one for category short and one for category long with clear within- and across-categorical instances. Subjects were asked to indicate how well the perceived instance represented category short or category long by means of 1 = very good till 6 = very bad. Finally, a staircase *discrimination test* demanded a judgment whether or not in a two word sequence (*standard-comparison sequence*: /tatə/ - /ta:tə/) the /a/ vowel in the comparison was longer than in the standard.

Results and discussion of study I

The identification test found that the category short and long were separated by a steep category boundary (Figure 9a). The location of the boundary was supported by bad goodness ratings. Clear within-categorical durations yielded good ratings (Figure 9b). Just-noticeable-differences (JNDs) in the category short increased to such an extent that only across-categorical discrimination was possible (Figure 9c). In this sense, vowel duration was not perceived in a continuous, but a categorical way. Furthermore, the discrimination sensitivity increased toward the category boundary which is reflected by decreasing JNDs; it decreases within the categories as reflected by increasing JNDs. Since identification response times replicate the findings of Pisoni et al. (1974b), – slower response times at the category boundary but faster for within-categorical items (Figure 9a) – this result supports the assumption that vowel length has a representation in the mental lexicon. Additionally, response times were faster in category long than in category short when measured from stimulus offset. This indicates some kind of trigger during the perception of the vowel which indicates the change from category short to category long. Such a result is in line with the results of Smits et al. (2006), who have found that the categorization of duration is accomplished by a comparison of perceived items to a representation of a category boundary.

Summary of study II

Hypotheses

Study I showed that vowel duration in disyllabic words – at least in the case of /a/ – is perceived categorically. Using the MEG, study II investigated to what extent the processing of categorical vowel length is reflected in the neural activity. This was done by comparing the evoked responses to the duration continuum from study I (speech stimuli) and a continuum that consisted of nonspeech stimuli which were equivalent in their temporal structure to the speech stimuli. The study was based on three assumptions:

i) An auditory evoked response not only consists of the M50 and M100, as explained in the introduction, but also of later deflections (around 200ms) which reflect for example the extraction of multiple frequencies from the signal (Kuriki, Kanda & Hirata, 2006), phonological or the semantic processing (around 400ms) of the speech input (Connolly & Phillips, 1994; Kujala, Alho, Service, Ilmoniemi & Connolly, 2004; Newman, Connolly, Service & McIvor, 2003). In addition, several studies have found that the syllabic patterns of speech are mimicked by neural activity (Abrams, Nicol, Zecker & Kraus, 2008; Ahissar, Nagarajan, Ahissar, Protopapas, Mahncke & Merzenich, 2001; Hertrich, Dietrich, Trouvain, Moos & Ackermann, 2012; Luo et al., 2007). Taking these findings into account, it was hypothesized that the later components (M200, M400) would not be elicited but *each* syllable, be it speech or nonspeech, would elicit a typical auditory M50/M100 evoked response.

ii) The perception of two identical items which are perceived in a sequence results in an attenuation of the neural activity to the second item in comparison to the first item (Cadenhead, Light, Geyer & Braff, 2000; Light & Braff, 2001; Mathiak, Ackermann, Rapp, Mathiak, Shergill, Riecker & Kircher, 2011). This phenomenon is the result of a comparison process between the first and the second item. When the characteristics of the second item match the expectations built up by the first item, the second stimulus is *gated* and the perceptual system needs less neural activity to process the second one (Lijffijt et al., 2009). The reduction has been observed during the perception of auditory stimuli (Bergerbest, Ghahremani & Gabrieli, 2004; Brown & Hagoort, 1993; Radeau, BEsson, Fonteneau & Castro, 1998) of visual stimuli (Gruber, Malinowski & Müller, 2004; Gruber & Müller, 2005a; Gruber, Klimesch, Sauseng & Doppelmayr, 2005b) but also during tactile stimulation

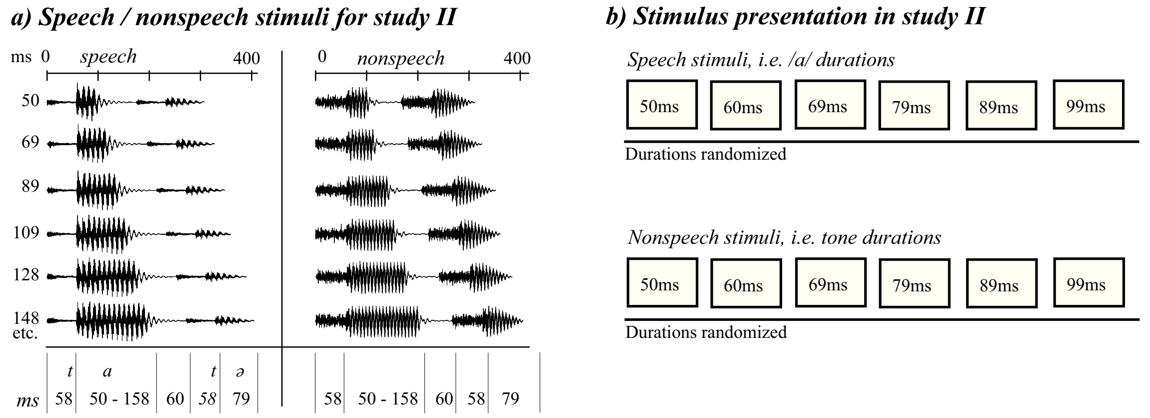

Figure 10: Experimental design in study II. ***a)*** *Waveforms and temporal structure of the speech and nonspeech stimuli.* ***b)*** *Stimulus presentation in the speech and nonspeech condition.*

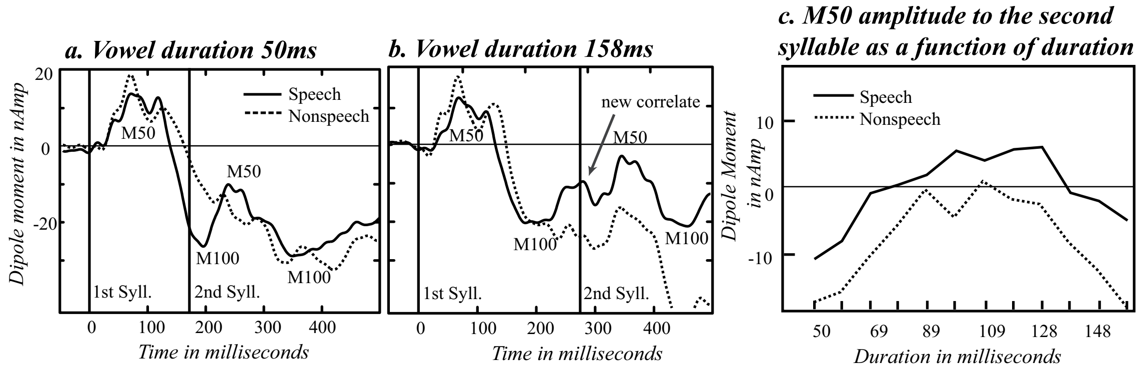

Figure 11: Results from Study II. Time courses (TCs) of dipole moments in the speech and nonspeech condition. The onsets of the first and second syllables are indicated, as well as the most prominent deflections. Only the results for the shortest and longest duration from the continuum are presented. ***a&b)*** *TCs for the two endpoints of the tested continuum (50 ms and 158 ms). The TCs for all vowel durations can be seen in part II.*

Remarks: ***a)*** *Unlike in the nonspeech condition, each syllable produces a distinct M50/M100 response in the speech condition.* ***b)*** *Distinct M50/M100 responses to each syllable can be seen in both conditions. Additionally, a new correlate occurred in the speech condition.* ***c)*** *M50 amplitudes as a function of vowel duration in the speech and of tone duration in the nonspeech condition. M50 amplitudes increase toward the temporal location of the category boundary and decrease afterwards in both conditions.*

(Wühle, Mertiens, Rüter, Ostwald & Braun, 2010). Importantly, it occurs at processing stages as early as the M50 (Lijffijt et al., 2009; Mathiak et al., 2011). Since study I showed that vowel length has a lexical representation, it was assumed that this representation would build up temporal templates, which are used to process vowel duration. The perceived vowel duration would either match or not match the build-up expectations. This match/no-match should be reflected by neural activities as early as the M50 in response to the second syllable since, at this point, duration measurement should be terminated. It was expected that within-categorical durations would be gated, resulting in lower M50 amplitudes in response to the second syllable. By contrast, untypical durations located near the category boundary would result in higher M50 amplitudes. No such gating mechanism was expected with respect to the evoked responses in the nonspeech condition since no categorical templates could be built up.

iii) In study I it was concluded that the categorization to short and long was bound to a category boundary. Taking this into account, it was hypothesized that the 'temporal crossing' of the category boundary by the perceived vowel duration should be reflected by an additional neural correlate that occurs between the first and the second syllable for categorically long vowels. By contrast, nonspeech duration should not have any lexicalized category boundary which is why no additional correlate was expected.

Stimuli

Additionally to the same /a/ continuum in the /tatə/ context as in study I (vowel durations ranging from 50 ms to 158 ms), study II used a tone continuum which was embedded in a 'disyllabic' nonspeech context. The nonspeech stimuli were equal to the speech stimuli with regard to their temporal structure but not with regard to their spectral structure, so that they could not be perceived as speech. The structure of the nonspeech stimuli can be seen in Figure 10a.

Methods

Study II consisted of a passive listening paradigm recording the auditory evoked responses to the stimuli in the MEG. Using the same subjects as in study I, it was possible to compare changes in the auditory evoked responses to the temporal location of the short-long boundary. Figure 10b illustrates the stimuli presentation in study II. Subjects had to listen passively to six blocks of stimuli. The odd numbered blocks contained the speech stimuli, in which /a/ durations were randomized. The even numbered blocks contained the nonspeech stimuli, in which the tone durations were randomized.

Results and discussion of study II

The results of study II can be summarized to three points:

i) In the speech condition, each syllable was reflected by a distinct M50/M100 response (Figure 11a&b; only the endpoints of the continuum are presented). This was not the case in the nonspeech condition: Short stimuli did not elicit a distinct M100 in response to the first nor a distinct M50 in response to the second syllable (Figure 11a). The missing M100/M50 responses became distinct for tone durations longer than 79 ms, i.e. for an onset of the second syllable at 211 ms (not shown in Figure 11. See Figure II.4 in part II).

The differences in M50/M100 deflections between speech and nonspeech are interpreted in the light of the theory of “asymmetric sampling in time” (Boemio et al., 2005; Poeppel, 2003), which assumes that auditory input is processed in terms of different temporal integration windows (short: 20-50 ms, long: 150-200ms). They show that the duration of the integration window was different between the conditions. In the nonspeech condition, the entire long TWI must have been used to integrate the auditory signal. Otherwise the syllables in both conditions would have been perceived as distinct, producing distinct evoked responses. A response was elicited when the onset of the second syllable exceeded the long TWI. In the speech condition, however, the duration of the integration window must have been shorter, because the second syllable elicited a distinct M50 for all vowel durations.

ii) In the speech condition, the amplitude of the M50 in response to the second syllable was enhanced at the category boundary (study I: 105.9 ms), but attenuated within the categories. This resulted in a bell-shaped course of the M50 amplitudes as a function of vowel duration (Figure 11c). Behavioral response times in the identification task from study I (Figure 9a) correlated with the *M50* amplitude in response to the second syllable. This finding supports the interpretation that the M50 in response to the second syllable reflects 1) stimulus categorization as well as 2) a gating mechanism which matches the incoming vowel duration to temporal templates which were created by a lexical length representation.

However, a similar bell-shaped pattern was also found in the nonspeech condition (Figure 11c). This is surprising, because no durational templates could be build up in the nonspeech condition because it lacks a lexical representation of duration. This finding implies a different explication of the results. Several studies have shown that auditory input triggers neural oscillations whose frequency mimic the rhythmical structure of the stimuli (Cummins, 2009; Nozaradan, Peretz, Missal & Mouraux, 2011; Will & Berg, 2007). Additionally, it has been shown that neural oscillations in the alpha band (~10 Hz) affect the amplitudes of evoked responses such as the M50 (Barry, Rushby, Johnstone, Clarke, Croft & Lawrence, 2004; Fellinger, Klimesch, Gruber, Freunberger & Doppelmayr, 2011; Stefanics, Hangya, Hernádi, Winkler, Lakatos & Ulbert, 2010). Such a connection has also been found in study II: Alpha oscillations affected the M50 in response to the second syllable in both conditions (the details of the analysis are skipped here due to their complexity; see part II). Taking this into account, the changes in the M50 amplitude in the speech condition might not reflect a gating mechanism. They rather reflect a mechanism which responds to the metrical structure of the stimuli.

iii) Finally, an additional neural correlate has been found, which emerged between the first and second syllable when vowel duration exceeded category short, located at around 200ms from stimulus onset (Figure 11b). Since this neural correlate appeared only in the speech condition and became distinct only for long durations, an interpretation that the deflection represents a new correlate reflecting vowel length was favored over an interpretation that it represented a M200 deflection. This is because a M200 has been shown to reflect memory access (Menning, Roberts & Pantev, 2000) and the signal's spectral complexity (Shahin, Roberts, Miller, McDonald & Alain, 2007) in both, speech and nonspeech condition, which was not the case in the present study.

The correlate is thus seen to categorize the additional vowel duration as phonologically 'long'. It is concluded that the categorization was based on a definition of the category boundary that, based on the given phonetic context and speech tempo, was derived from a phonological representation. As such, the new correlate supported the conclusion of study I that vowel duration is categorical.

Summary of study III

Hypotheses

The results of study I suggest that the length categories are defined by a category boundary which was reflected by a neural correlate in study II. So far, such a neural correlate has not yet been described in the literature. However, study I and II concentrated only on the duration cue, leaving aside the quality cue, which is supporting the length contrast in all non-low vowels. This is why study III expanded the scope of cues to quality.

Previous studies have shown that the importance of the duration cue correlates inversely with vowel height: The higher the vowel, the less important is duration (Bennett, 1968; Escudero et al., 2009; Lehnert-LeHouillier, 2010; Sendlmeier, 1981; Weiss, 1974). However, these studies investigated the weighting of the duration and quality cue solely by means of identification tests, but omitted to assess the temporal and spectral discrimination sensitivities. Since speech perception has to be regarded as an interaction between identification and discrimination (Livingston, Andrews & Harnad, 1998; Smits et al., 2006), study III performed a series of temporal and spectral discrimination tests in addition to the identification test. Study III investigated the psychoacoustic characteristics of the interaction between vowel duration and vowel quality as acoustic cues for phonological vowel length. The hypotheses underlying study III were:

i) If the findings of study I&II are correct, vowel duration should be perceived categorically even when contrasting high vowels. This should be reflected insofar as temporal just-noticeable differences become so large that only across-categorical duration changes are perceived.

ii) Smits et al. (2006) showed in a learning study that spectral contrasts are categorized with respect to a lexicalized boundary separating the two categories. If lax and tense vowels had a lexical representation, their boundary should be reflected by an increase in spectral discrimination sensitivity. If no increase is observed, tenseness would not be regarded as having a lexical representation.

iii) The courses of the boundaries between the categories serve as a supporting evidence for phonological representation by reflecting the stability of either cue. In this sense, quality – represented by acoustic changes of F1 –will be regarded as the more important cue than duration if it affects the location of the short-long boundary

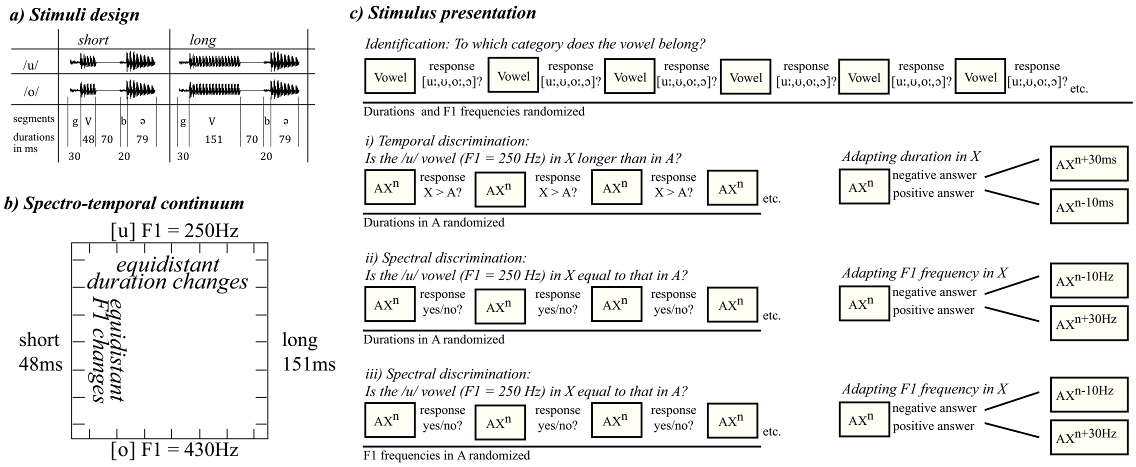

Figure 12: Experimental design in study III. a) Waveforms and temporal structure of the stimulus material. V = the vowel in question. b) Representation of the two-dimensional spectro-temporal continuum used for identification and discrimination. c) Stimulus presentation in all behavioral tasks.

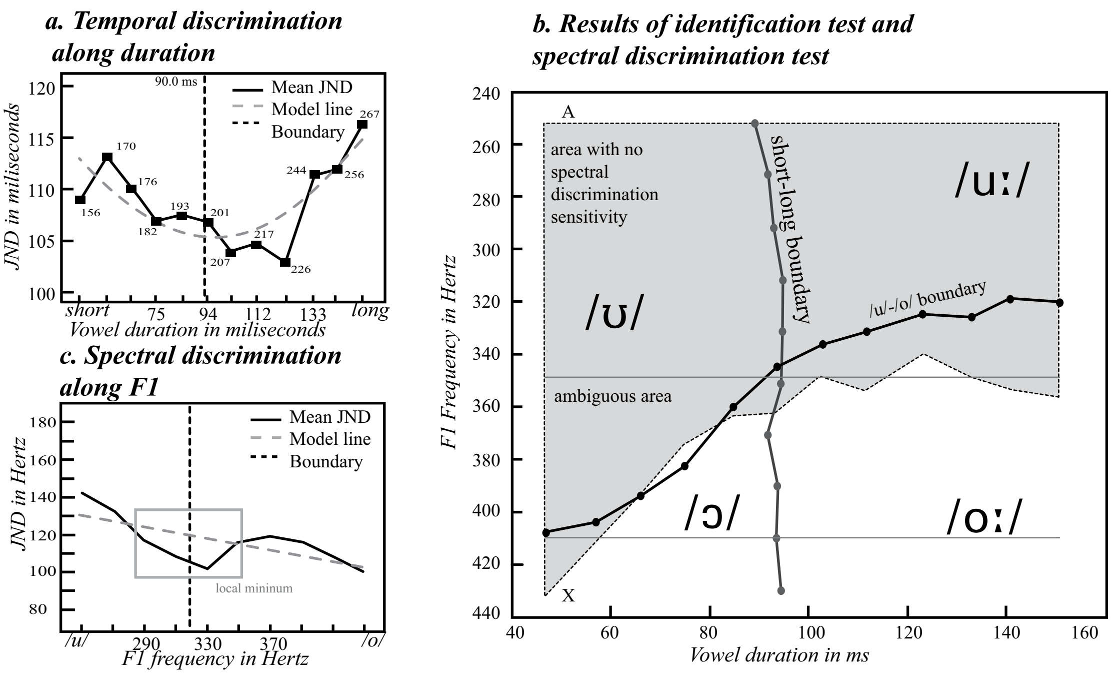

Figure 13: Results from Study III. ***a)*** *Temporal just-noticeable-differences (JND) along the duration continuum. Numbers indicate the vowel duration in the comparison.*

b) *The identification results indicate the boundary locations between short and long as well as between /u/ and /o/. The ambiguous spectral area indicates, where [ʊ] and [oː] are discriminated on the basis of vowel length. The grey-shaded area represents the vowel space where no discrimination between adjacent items was possible. The upper edge of the grey-shaded area represents the continuum in A, the lower edge represents the F1 frequency in X.*

c) *Spectral JNDs along the /u/-/o/ continuum.*

and duration does not affect the quality boundary. However, duration will be regarded as more important if it affects the quality boundary and quality in turn does not affect the duration boundary.

Stimuli

Study III used a new set of stimuli which covered a two dimensional continuum between [u] and [o] as well as between short and long in the disyllabic context /g_bə/ (Figure 12a&b). The duration continuum consisted of items which differed by ~10ms. The spectral continuum was created by changing the first formant from /u/ to /o/ in steps of ~20 Hz.

Methods

Study III consisted of a series of behavioral tests (Figure 12c). The *identification test* was the same as in study I. Subjects had to categorize to which category the vowel in the first syllable belonged – [ʊ], [uː], [ɔ] or [oː]. Regarding the staircase *discrimination tests*:

i) The first discrimination test was equal to the discrimination test in study I, testing the temporal discrimination sensitivity as a function of duration. Subjects had to indicate whether or not the /u/ vowel in X was longer than in A.

ii) The second discrimination test assessed the spectral discrimination sensitivity as a function of duration. A duration continuum between /ʊ/ and /uː/ with a stable F1 = 250 Hz was used. Subjects had to indicate whether the vowel in X was equal to the vowel in A.

iii) The third test assessed the spectral discrimination sensitivity as a function of frequency. It used the F1 frequency continuum between /uː/ and /oː/ with a stable vowel duration of 151 ms. Subjects also had to indicate whether the vowel in X was equal to the vowel in A. Figure 12b indicates as well how the phonetic information in X was adapted depending on the given answers. Like in the temporal discrimination test, the frequency in X 'hovered' around the just-noticeable-difference for spectral change and was found by mathematical procedures explained in part II of the present thesis.

Results and discussion of study III

The results in study III replicated the findings in study I that vowel duration is perceived categorically and seems to have a phonological representation in the mental lexicon. This interpretation was based on three findings.

i) Vowel durations in X had to be located in the category long in order to be perceived as 'longer' when they were compared to durations in the category short. This replicated the findings in study I. In line with the findings of Gerrits et al. (2004); Schouten et al. (2003a), vowel duration was perceived categorically in disyllabic words.

ii) Ylinen et al. (2005) showed that speakers of a language which uses vowel length phonologically exhibited a discrimination maximum at the short-long boundary, in contrast to speakers who lack such contrast in their native language. Study III found that temporal JNDs showed a minimum at the short-long boundary (Figure 13a), replicating the findings of study I. In this sense, vowel length is regarded to have a phonological representation in the mental lexicon.

iii) The category boundary between /u/ and /o/ was strongly affected by vowel duration in contrast to the effects which had F1 frequency on the short-long boundary (Figure 13b). The shorter the vowel, the higher the F1 frequency at which the boundary was located. The location of the /u/-/o/ boundary changed to such an extent that a spectral area was found, whose categorization depended on phonological length ('ambiguous area' in Figure 13b). This is in line with previous findings in the literature (Sendlmeier, 1981; Weiss, 1974). Hence, vowel duration was a stronger cue for the length contrast than vowel quality. This is also supported by the spectral JNDs. These are indicated by the grey shaded area in Figure 13b. The lower edge of the grey shaded area represents the F1 frequency in X when the vowel was perceived as not equal to the vowel in A. Subjects obviously did not discriminate between quality changes within the /u/ category. It seems as though only across-categorical quality changes were discriminated, i.e. they discriminated only between /u/ and /o/. Since these vowels have a lexical representation, one can assume that this status is reflected by the local decrease in the spectral JNDs (Figure 13c) at the /u/-/o/ boundary.

However, the picture regarding the quality cue seems to be more complex than the discrimination test indicates. In contrast to the discrimination test, the identification test could show that subjects have nevertheless a minimal sensitivity to quality. F1 frequency significantly affected the short-long boundary between [ʊ] and [uː] (The skewed vertical line in Figure 13b). In line with Lehnert-LeHouillier (2010), lower F1 frequencies (i.e. rather tense /u/ vowels) had a higher probability to be identified as "long" than higher F1 (i.e. rather lax /u/ vowels).

Study III shows that, on the one hand, subjects lack the ability to discriminate vowel quality in disyllabic words. On the other hand, fine phonetic detail can affect their categorization. An explanation of how to explain these contradictory findings will be postponed to the global discussion. First, the study will be presented which assessed the spectral sensitivity at the neural level.

Summary of study IV

Study III showed contradictory results. On the one hand, vowel quality affected the location of the short-long boundary, indicating that German native listeners seem to be sensitive enough to quality changes. On the other hand, no spectral discrimination sensitivity was found. Study IV investigated the latter point by assessing the sensitivity to within-categorical quality changes at the neural level using the MEG. This was done by means of the *priming paradigm,* which will be explained before presenting the hypotheses.

Methods

The priming paradigm investigates in a probe–target sequence how the information contained in the probe affects the processing of the target (e.g. in the sequence [gɔbə], [guːbə]). This is assessed by comparing the behavioral/neural responses to the target from a test condition to the responses in a control condition. If neural responses are attenuated and response times are reduced, the information in the probe *facilitates* the processing of the target. If neural responses are enhanced and response times are increased, the probe *inhibits* the processing of the target (Bergerbest et al., 2004; Goldinger, 1998; Zwitserlood, 1996). The reduction in neural activity is explained by the brain's capacity to differentiate between relevant and irrelevant incoming information and as such adapt its sensitivity to auditory input (Boutros & Belger, 1999). In study IV, the test condition and the control condition were played to the subjects twice: i) Once in a behavioral identification task, where they had to indicate whether the target contained an [uː] or an [oː] and ii) once in a passive listening task in the MEG (Figure 14b) recording the neural responses to the target.

Stimuli

For the test condition, a subset of five disyllabic /g_bə/ stimuli was, used which was taken from the spectro-temporal continuum in study III (Figure 12b). Three stimuli were used as probes, containing a short vowel (48 ms) and differing in the F1 frequency ([u] = 250 Hz, [ʊ] = 370 Hz, [ɔ] = 540 Hz). Two stimuli were used as targets, containing a long vowel (151 ms)

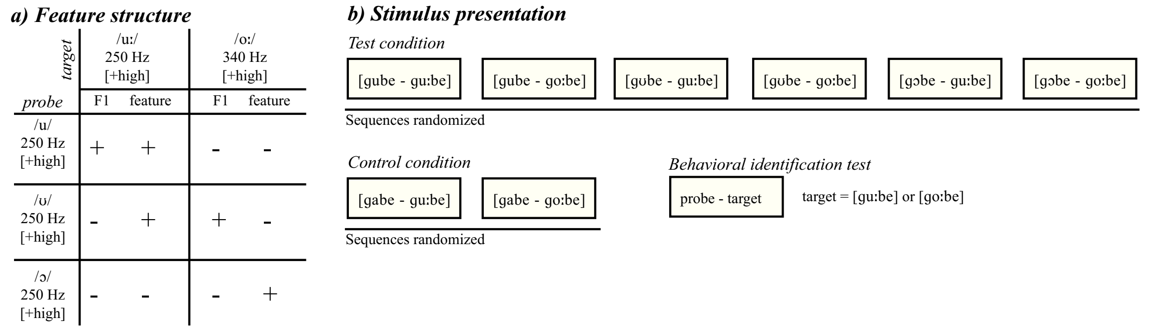

Figure 14: Experimental design in study IV. ***a)*** *Acoustic and phonological match (+) / no-match (-) constellations between the probe and the target stimulus.* ***b)*** *Stimulus presentation in the MEG experiment and in the behavioral identification test.*

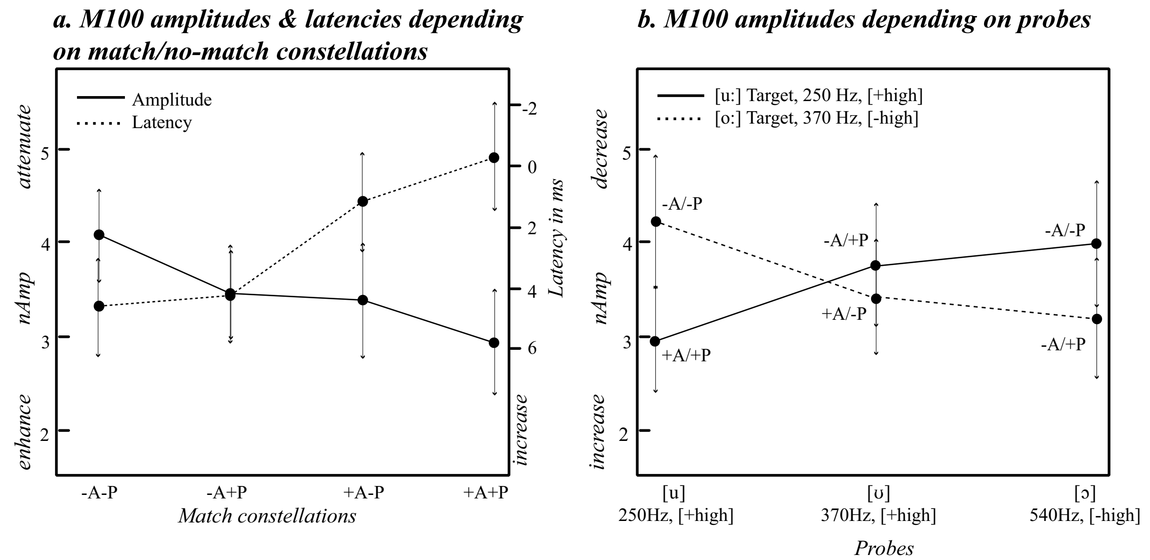

Figure 15: Results from Study IV. Whiskers indicate the confidence interval. The plots show the mean M100 amplitudes and latencies of the target words containing [uː] or [oː] primed by the probes containing [u], [ʊ] or [ɔ]. Acoustic and phonological information is included in the plot. ***a)*** *M100 amplitudes and latencies in the target depending on the match/no-match constellation between probe and target. Bottom of the figure indicates increase in latency/amplitude, top of the figure indicates a reduction of latencies/amplitude. M100 amplitudes increase with a full match, latencies decrease.* ***b)*** *M100 amplitudes in the target depending on the probes.*
Legend: A = acoustic information, P = phonological information, '+' = match, '–' = no-match.

and differing in the F1 frequency ([uː] = 250 Hz, [oː] = 340 Hz). The stimuli for the probe and the target did or did not match with regard to their phonological height feature [+/- high] or in their quality information (see Figure 14a for the information distribution in the stimuli). In the control condition the two targets from the test condition (long [uː] and [oː]) were preceded by a control stimulus [gabə] that was not equal to the target neither with respect to the phonological feature nor to the spectral information. Recently, it has been shown that the processing of spectral information contained in a vowel is reflected by the latency and generator location of the M100 (Obleser et al., 2004; Roberts et al., 2004a; Scharinger et al., 2011). Since facilitation/inhibition has been found to be reflected by the strength of neural activity, the sensitivity to within-categorical quality changes was assessed by investigating the effects of the probe onto the amplitude and the latency of the M100 in response to the target. Priming effects were calculated by subtracting the M100 amplitudes and latencies assessed in the control condition from the test condition.

Hypotheses

Previous studies investigated to what extent the amount of phonetic overlap between probe and target (i.e. the number of segments) facilitated/inhibited the processing of the target (Dumay, Benraïss, Barriol, Colin, Radeau & Besson, 2001; Fear, Cutler & Butterfield, 1995; Slowiaczek, McQueen, Soltano & Lynch, 2000). Study IV expanded this approach by two points: i) It concentrated on solely one segment in a disyllabic nonsense word, whereas the remaining structure of the word was equal; ii) it investigated the effects of a phonological feature additionally to acoustic overlap. The experiment design used in study IV is presented in Figure 14b. The questions were:

i) How do the priming effects of an acoustic and phonological overlap differ? To what extent will these two kinds of information increase/inhibit each other?

ii) Will the targets be differently processed depending on whether the probe contains tense [u] and lax [ʊ]? If yes, a sensitivity to within-categorical changes can be assumed. If not, the difference between lax and tense would be regarded as already leveled out at early processing stages.

ii) Which information shows the stronger priming effect, acoustic or phonological information? This was investigated by means of the effects of the lax [ʊ] vowel onto the [uː] and [oː] targets as the probe is equal in its feature information to [uː], but equal in its acoustic information to [oː]. The results to this question reveal whether the M100 already reflects categorical vowel or still continuous vowel information.

Results and discussion of study IV

In the behavioral identification test, the acoustic and phonological overlap between probe and target had no significant effect on the response times. This is why the presentation of these results is postponed to the experimental section in part II of the thesis.

The acoustic and phonological overlap, however, had significant priming effects on the M100 in response to the target (Figure 15a), although these effects differed between latency and amplitude. M100 amplitudes correlated with the amount of acoustic and phonological overlap, indicating an *inhibition* effect due to the overlap. This result contradicts the findings of Dumay et al. (2001); Fear et al. (1995 and); Slowiaczek et al. (2000), who have shown facilitation effects due to an overlap. Nevertheless, inhibition effects, called *negative priming*, have been shown in a paradigm where subjects are asked to actively ignore the probe and respond to the target (Tipper, 2001). In this paradigm, it is assumed that negative priming results from a "tagging" of the information in the probe as to be ignored. In order to delete the tag, more neural activity is necessary, which results in the observed inhibition effect. This was probably the case in the present study. Subjects had to ignore the probe and identify the vowel in the target in the behavioral task, which was performed before the MEG experiment. It might be the case that they transferred this strategy to the passive listening task in the MEG.

However, M100 latencies correlated inversely with the amount of acoustic and phonological overlap. They became shorter the more information probe and target shared, indicating a *facilitation* effect of overlap. Hence, the presence of an acoustic or phonological information reduced the processing time but increased the neural activity to process the target. So far, such a contradictory result was not presented in the literature, which is why an explanation of this effect is left open for future studies.

Regarding the question on the spectral sensitivity, study IV has shown that the tense [u] and lax [ʊ] differed in their priming effects on the amplitudes of the M100 in response to the [uː] and [oː] target (Figure 15b). The effect, however, was not significant when tested pair-wise. Two reasons, which interact with each other, might be possible for this result. It has been shown that attention increases neural activity (Näätänen, Paavilainen, Titinen, Jiang & Alho, 1993). However, the present study used a passive listening task whereas priming experiments generally use a task in which subjects have to respond actively to the stimuli (Bonte & Blomert, 2004). Furthermore, the representation of fine phonetic detail diminishes as a function of time (Pisoni et al., 1974a), which, regarding the temporal distance between the vowels in question, might have affected the results. It was concluded that, at levels which concern the extraction of vowel information, the auditory system is sensitive enough to within-categorical changes between lax and tense vowels. This distinction, however, is not used for the phonological processing of vowel but used as a supporting cue for vowel length, which will be considered in the general discussion.

Summary of Study V

Hypotheses

Study V investigated i) how flexible the boundary between category short and long was with regard to speech tempo and ii) whether the identification of vowel length could be affected by a preceding length feature (Hall, 1992). This was investigated by means of a priming paradigm in which probe and target did or did not match with respect to their length feature.

It was hypothesized that the speaking rate of the probe would affect the category boundary (Hirata, 2004). The location of the boundary should be shifted to an earlier position in the fast speaking rate than in the normal rate. Concerning priming, if vowel duration was processed in terms of length features and if these features were stored long enough, two effects on the identification of the target are possible. If the presence of the length feature in the probe imposed an inhibition effect on the identification of the target, response times should be slower in case of a match in contrast to a no-match constellation. If the presence of the length feature imposed a facilitation effect on the identification, response times were hypothesized to be faster in case of a match in contrast to a no-match constellation.

Stimuli

Study V used a priming paradigm. In the test condition, two sets of stimuli were used. As probes, a set of twenty real disyllabic words of the structure CVCV was recorded; ten had a phonologically short [a] such as *Krabbe* ‘crab’*, Karre* ‘gun’*, Waffe* ‘weapon; ten had a phonologically long [a:] in the first syllable such as *Jahre* ‘year, pl.’*, Kater* ‘tomcat’*, Name* ‘name’. Each word was recorded with two different speaking rates: a slow rate of 80 words per minute and a fast rate of 120 words per minute, controlled by means of a metronome. As targets, a continuum of [a] durations was embedded in the disyllabic nonsense word [t_kə] by shortening the [a] pitch-period wise in a recording of [taːkə]. As control condition, each target was preceded by a short sine wave of 100 ms.

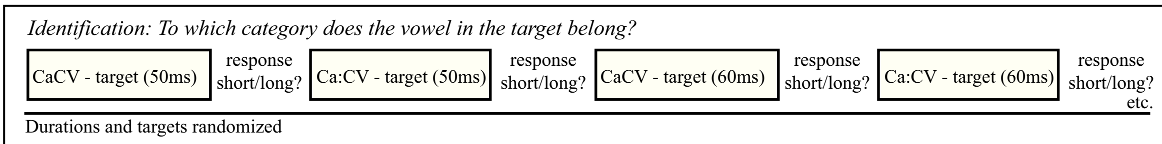

Figure 16: Stimulus presentation in study V.

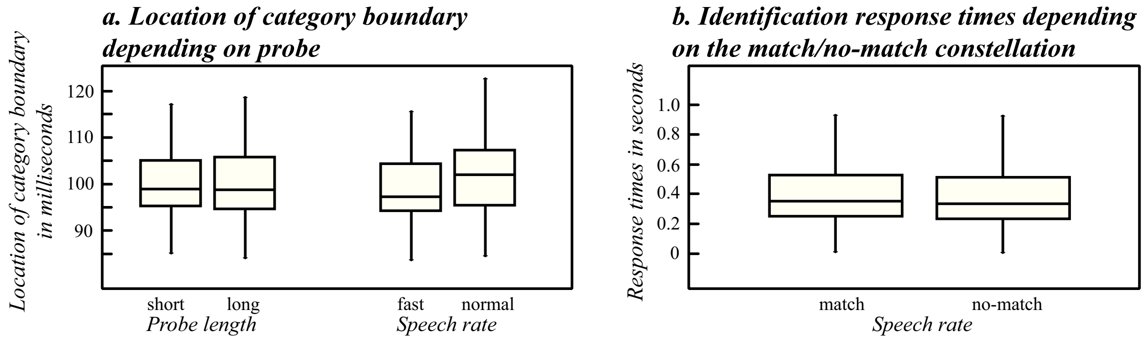

Figure 17: Results from study V. Whiskers indicate standard error, boxes the confidence interval and lines in the boxes the group mean.

Methods

Subjects had to identify whether they perceived [taːkə] or [takə] in the target in a slow and a fast speaking rate condition. The experiment design is illustrated in Figure 16.

Results and discussion of Study V

The results in Study V have shown that the location of the category boundary differed significantly between fast and normal speaking rate. It was located at an earlier time point when identification was performed with the fast speaking rate than with the normal speaking rate. The location of the boundary was not affected by the phonological match/no-match constellation (Figure 17a). Regarding the priming constellations, no significant differences between the phonological match/no-match constellation could be found (Figure 17b). This is why the results do not support any of the two hypotheses. Two possible explanations arise for this finding:

i) Vowel duration was not processed in terms of phonological features, which is why no priming effect was found. The lexical representation of the category boundary created perceptual durational templates for each category which either were or were not matched but did not affect later processing. These templates should be calculated anew for each syllable depending on speaking rate.

ii) After the comparison with the durational templates, vowel duration was processed in terms of phonological features. These are necessary since they are distinctive with respect to the perception of vowel height, as it has been shown in study III. However, the feature did not leave a trace that was strong enough to affect the processing of later segments (see McClelland & Elman (1986) and their TRACE model, which states that each processed information is stored for later access). Why that may be cannot be answered in the present study.

General discussion

The German vowel system contrasts vowels by means of length, which is cued by different durations and quality, i.e. the vowels' relative position towards the center in the vowel space. The aim of the present thesis was to investigate which phonetic cue – vowel duration or vowel quality – is used primarily during the perception of the German vowel length/tenseness contrast and can therefore be interpreted as the lexicalized feature in the phonological system. To this aim, the contrast's psychoacoustic characteristics and neural correlates were assessed. The results of the five studies can be summarized and interpreted as follows:

i) Perception of vowel duration

It could be shown by means of behavioral tasks that the vowel length contrast in German exhibited a steep boundary between category short and long. The discrimination of duration change in the category short was not possible, only across-boundary duration changes were discriminated. Accordingly, it can be concluded that the perception of vowel length was categorical (Liberman et al., 1957). As it has been shown that whether a continuum exhibits categorical or continuous perception depends on the discrimination task (Gerrits et al., 2004), one could argue that this outcome was probably biased by the present paradigm, i.e. the discrimination of vowel duration embedded in a disyllabic word using a stair-case procedure. Especially the **AX** paradigm, which was used in the present studies, increases the probability to find categorical perception. This was shown by Schouten et al. (2003a), who argue that subjects refer to their lexicalized contrastive categories while accomplishing this task, in order to make sure that they are absolutely correct when they judge **A** and **X** as different. Nevertheless, vowel duration in everyday speech has often to be categorized in di- and polysyllabic words, which is why the present paradigm is considered to mimic reality very well. Hence, the present results support the importance of the duration cue. However, it is not sufficient to conclude that length has a lexical representation.

It has been shown in the literature that, when duration change does not concern phonological contrasts, the discrimination sensitivity along a duration continuum is generally linear. The fraction between the duration in the **A** and in the **X** item remains constant (for a definition of **A** and **X** see methods in study I). This relation between **A** and **X** is known as the Weber's Law. The effect is that the just-noticeable-differences (JND) increase in proportion to the increasing duration in question (Dehaene, 2003; Rammsayer, 2010). This was not the case in the present study. The discrimination tests found that the temporal JNDs decreased toward the category boundary and increased within the categories (Figures 9c & 13a). They thus violated

Weber's Law. Accordingly, the non-linear change mirrored an increase of temporal sensitivity toward the category boundary. As was also shown by Ylinen et al. (2005), such a finding could be regarded as an indicator for a phonological category in the mental lexicon. In line with Pisoni (1973b), this status was supported by identification response times which were larger at the category boundary than within the categories (Figures 9a).

The question arises whether the representation refers to a categorical prototype, as has been shown for vowel categories such as /i/ and /e/ (Diesch et al., 1999; Kuhl, 1991) or is bound to a category boundary. Smits et al. (2006) have convincingly shown in a series of learning experiments (using nonspeech stimuli) that the identification of length categories is accomplished by a reference to a category boundary as well as by a comparison of duration exemplars. Taking this finding into account, one might hypothesize how this type of perception is implemented in speech perception. The category boundary might be specified in terms of a rule like '*category short = vowel duration shorter than B ms, else long*' where B represents temporal location of the category boundary. This rule produces a durational template for category short in analogy to the phonological time slots (Figure 4) (Becker, 1998). Depending on whether one slot or two slots are occupied, a short or a long vowel is perceived, respectively. This interpretation was supported by study II. It was shown that the opening of the second time slot for long vowels was reflected by an additional neural correlate, called *DP* (Figure 11b). This correlate could have been elicited by neurons which are specialized for the processing of categorical vowel length. Equivalent neurons have been found in a learning experiment with rhesus monkeys performed by Leon & Shadlen (2003). Importantly, the threshold representing the change from category short to long is not fixed but can be adjusted anew, depending on external factors such as speaking rate (Hirata, 2004) or the vowel's consonantal context (Lehtonen, 1970). Such an interpretation is supported by the finding in study V, where the category boundary could be shifted to different temporal locations depending on the speaking rate.

ii) Perception of vowel quality

Study III has shown that the discrimination between the lax and tense instances of the /u/ vowel was not possible (Figure 13b). Rather, subjects showed a discrimination sensitivity solely to across-categorical changes between /u/ and /o/. This finding is interesting insofar as vowels have been shown to allow within-categorical discrimination (Kuhl, 2004). The present results could be biased by the **AX** paradigm as explained for temporal discrimination, increasing the probability that subjects made their decision by referring to a lexicalized representation of the vowels (Gerrits et al., 2004; Schouten et al., 2003a). Since subjects referred to /u/ and /o/ as categories and not to lax /u/ and tense /u/, it can be assumed that there is no phonological representation in the mental lexicon of the tenseness contrast. Furthermore, the spectral discrimination along the /u/-/o/ continuum showed an increase in sensitivity at the /u/-/o/ boundary, obviously reflecting the lexical status of these vowels. No such local minimum could be found in the /u/ vowel space at the alleged lax-tense boundary (Figure 13b).

Although the discrimination test indicated that subjects are not sensitive enough to spectral changes between lax and tense vowel quality, the identification results partly contradict this finding. F1 frequencies significantly affected the category boundary between short [ʊ] and long [uː] insofar as /u/ instances with lower F1 frequencies (i.e. tense) had a higher probability to be attributed to the category long than /u/ instances with higher F1 frequencies (i.e. lax). Study IV supported this finding, i.e. that the auditory cortex is sensitive enough to such within-categorical spectral changes. It was shown by means of a priming experiment that the M100 amplitude and latency in the long tense /u/ target was differently affected by the lax and tense /u/ instance in the probe (Figure 15b). This is in line with Toscano et al. (2010) who have shown different M100 amplitudes for within categorical VOT changes, concluding that continuous information of phonological contrast is available to the auditory cortex.

If there is no lexical specification of the tenseness contrast to which subjects can refer during perception, why should the auditory system be sensitive to within-categorical spectral changes? This question can be answered taking into account the results of Gussenhoven (2007) and Meister et al. (2011). It has been known for long that the vowel's intrinsic duration inversely correlates with vowel height, i.e. lower vowels are longer than higher vowels (Antoniadis et al., 1984). Gussenhoven (2007) found that human perception compensates for this correlation. The auditory system adjusts to the perception of duration so that vowels with different height are perceived as equally long (or short). The inverse conclusion is that when a low and a high vowel are equal in their duration, the low vowel is perceived as *shorter*. Meister et al. (2011) replicated these findings by showing that high vowels had a lower probability to be perceived as short than low vowels: The short-long boundary occurred at an earlier temporal location for high vowels than for low vowels. The identification of vowel duration is thus affected by spectral information. In the vowel length contrast, high short vowel instances have a higher F1 than long instances. This trade-off could be used to increase the perceptual distance between categorically short and long vowels. This is why, at the neural level, where the spectral information is extracted, the auditory system has to be sensitive enough to perceive the difference between lax and tense vowels, although there is no categorical specification. Accordingly, the distinction between lax and tense vowels in stressed syllables is allophonic. It is a secondary cue, which increases the categorical perception of vowel length.

iii) Vowel length as the primary feature

In study III, an ambiguous F1 frequency area was found in which short vowels were perceived as [ʊ] and long vowels as [oː] (Figure 13b). This finding reproduced the studies of Bennett (1968); Sendlmeier (1981); Weiss (1974). In contrast to these authors it was concluded that duration is the primary cue, since the length distinction is necessary not only to differentiate short vowels from long vowels, but also to differentiate high vowels from mid vowels (Figure 2). However, Bennett (1968); Sendlmeier (1981); Weiss (1974) argued that vowel quality and not vowel duration is used primarily to differentiate short vowels from long vowels such as [ɔ] and [oː]. How can the present findings be brought in line with the interpretation of these authors? Figure 18a shows a basic representation of the spectro-temporal vowel space. It is a schematic illustration of Figure 13b, which does not take into account the continuous effect of vowel duration on the /u/-/o/ boundary.

The latter authors have argued that quality and not duration is the primary distinctive cue, because lengthened instances of [ɔ] (i.e. [ɔː]) were not identified as long [oː] but still as short [ɔ] instances. How can the present results be reconciled with the findings of Bennett (1968); Sendlmeier (1981); Weiss (1974). Both, Bennett (1968, P. 69) and Sendlmeier (1981, P. 297) show that shortening a mid tense vowel, e.g. [o:], changes not only the length category but also the vowel category into a high lax vowel, e.g. [U]. This finding has been well replicated in study III. However, all three authors show that the reverse is not possible, i.e. lengthening a short lax mid vowel, e.g. [O], does not result in [o:]. These instances were still attributed to [O]. The reason becomes obvious considering two points. First, unlike with [U], there is no long vowel category which overlaps with /O/. This is at least true at the phonological level. There is phonetic evidence that the vocalization of [r] in 'vowel+r' sequences provokes a lengthening and an increase in F1 of the respective vowel. In case of [O] and [o:] this would result in instances of [O:] (compare with Figure 2) (Dittrich & Reibisch, 2006). Unfortunately, Dittrich et al. (2006) have not tested the /o/ category. This is why, secondly, the aforementioned authors did not offer a possible answer to which [O:] could be attributed to. Instead, subjects attributed all lengthened [O] instances to [O]. This is why Bennett

(1968); Sendlmeier (1981); Weiss (1974) concluded that it is vowel quality which is the primary cue to differentiate between short and long vowels.

iv) A minimal model of length and tenseness processing

The results of the experiments that were conducted in the course of the present thesis are incorporated into a minimal model of vowel perception (Figure 18b). Based on a basic representation of the vowel space found in study III (Figure 18a), it is assumed that vowel categorization is rule-driven. The model consists of a set of rules which calculate the expected durational and spectral boundaries.

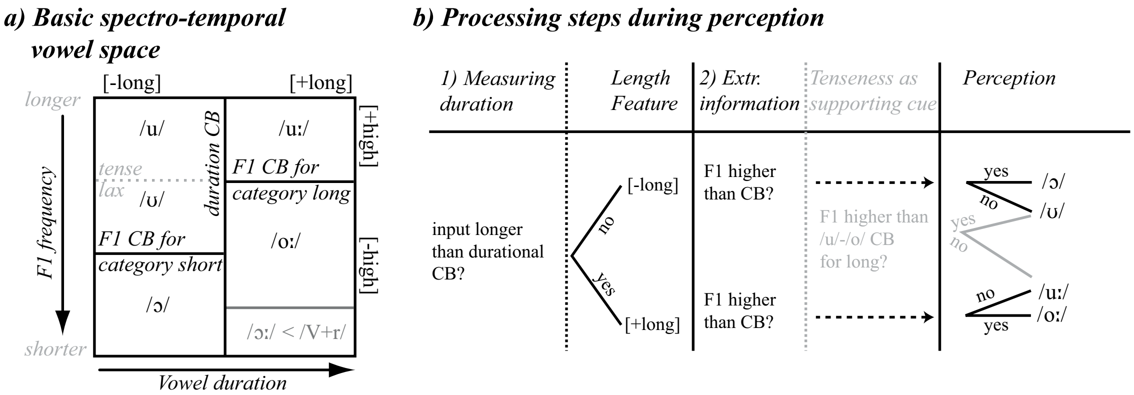

Figure 18: a) Basic spectro-temporal vowel space used to model the perception of short and long vowels in (b). The category boundaries (CB) between the vowel categories and between the duration categories are based on statistical language learning. The alleged category boundary between lax and tense vowels is inserted in grey. Additionally, the long lax vowel /ɔː/ is included. See text for explanation. b) The steps during the processing of vowel quality and duration. 1) Based on the duration measurement, different spectral CBs are assumed for the vowel categorization. 2) Depending on the categorical length the spectral information vowel is differently categorized. In case duration cannot be measured properly, tenseness functions as a supporting cue.

The phonological short-long boundary that represents a durational threshold is transformed into a perceptual durational template that represents a short (one time slot) or a long vowel (two time slots). The physical size of the time slots is calculated anew depending on the consonantal context (Lehtonen, 1970) and speech tempo as shown in Study V. The feature [-long] is set when the vowel is integrated into one time slot, [+long] when the vowel occupies two time slots. Due to its categorical status, the length feature can be used to adapt the [+/-high] boundary in the spectral vowel space. This accounts for the ambiguous area found in study II (Figure 13b). Accordingly, the perceived vowel is the result of 'yes-no' decisions during processing (black lines and black writing in Figure 18b) and not of a direct perception of the physical reality.

The question arises how vowel quality in terms of lax and tense is represented in the model. Since it is regarded as a secondary, supporting cue to vowel duration, since it affects the processing of duration (Gussenhoven, 2007; Meister et al., 2011), it should be included. In the light of the present results, the tenseness feature [+/-tense] functions as surrogate for [+/-long] when the length cannot be differentiated correctly. Such a situation occurs when speaking rate is increased to such an extent that the phonetic instances of short and long vowels overlap, as has been shown by Hirata (2004). This, of course, is somehow contra-intuitive, because the extraction of vowel information becomes worse with shorter vowel duration (Wallace et al., 2009), which furthermore significantly decreases when speaking rate increases (Flege, 1988; Lindblom, 1963). Regarding the present results it is not possible to answer this question. Importantly, the [+/- tense] feature still refers to [+/-long]. In order to account for its allophonic status, the [+/- tense] feature does not have an own lexical representation, but refers to the lexical boundary between /u:/ and /o:/. When F1 frequency is higher than the /u:/-/o:/ boundary, the feature is set to [-tense], resulting in the perception of a "short" vowel. When F1 frequency is lower, the feature is set to [+tense], resulting in the perception of a "long" vowel.

Conclusion and outlook

The present thesis investigated the perception of the duration and quality cue in the German vowel length contrast. Based on the results of the experiments conducted in the course of this doctoral thesis it is concluded that vowel length has a phonological representation in the mental lexicon whereas vowel quality serves as a supporting phonetic cue enhancing the perception of categorical length.

The question remains whether the [+/-tense] feature is necessary for production. The present proposal follows the phonological literature (Hall, 1992; Ramers, 1988; Wiese, 2000) in assuming that the [+/-tense] feature is filled in during the phonological derivation. Its phonetic equivalent is centralization. [+/- tense] can be derived by means of two rules:

a) V → [+tense] / [+long], else [-tense]

b) V → [+tense] / __$, else [-tense]

Rule (a) is based on Hall (1992) and Ramers (1988) according to whom a long vowel is attributed the feature [+tense] and is therefore centralized at the surface. Rule (b), where $ indicates the syllable boundary, is based on Becker (1998) who argues that tenseness is a phonetic reflect of syllable structure. Open syllables become tense, closed syllables become lax. In the end, the articulatory result is equal to rule (a). However, the choice between the two rules is left open.

Furthermore, since the present thesis just scratched the tip of the iceberg regarding the interaction between the processing of temporal information and the processing of spectral information future research can address the following points.

1. Given that at the phonetic level the surface form of the length contrast consists of short lax and long tense vowels, the question arises what is the nature of the lexical representation of vowel length. In other words: is it a segmental or a syllabic contrast? This question could be addressed with speech production studies.

2. Study II found an additional neural correlate that emerged for categorically long vowels. This finding and its current interpretation has to be investigated in future experiments. Learning studies represent one approach in which native speakers of a language lacking an phonological length contrast could be trained to discriminate categorical vowel length.

3. The most intriguing results were found in study IV. Priming affected the M100 amplitude and latency in a contradictory way. Further studies should investigate whether priming with a single segment results in an inhibitory or a gating effect, as well as the function of attention, the number of segments, the temporal distance and the inter-stimulus time between the segments in question.

Experimental section

The following section reports all five studies conducted in course of the present doctoral thesis. The section presents the specific theoretical background, experiment design, methods, analysis and discussion for each study in more detail than the foregoing section.

Study I: The psychoacoustic characteristics of German vowel duration

Introduction

The German vowel system

The phonological distinction between short and long vowels in German is important for phonotactics and syllabification (Hall, 1992), as well as the correct assignment of phonemes to intrasyllabic slots of stressed syllables (Becker, 1998). Unlike the /a/ - /a:/ contrast, where the phonetic difference is more or less restricted to vowel duration since short and long cognates overlap strongly in formant values (Hoffmann, 2011), quantitatively contrasting mid and high vowels exhibit an interaction between vowel duration and spectral quality (Hall, 1992). Short vowels are more centralized, long vowels are more decentralized in the vowel space. Concerning the distribution of these vowels, the length contrast is limited to stressed syllables whereas in unstressed syllables the durational contrast is largely neutralized or non-existent. De-stressed cognates of non-low vowels, though, may still exhibit some differences in vowel quality, i.e. centralization. This gave rise to the assumption that these differences, irrespective of the length aspect, refer to a lexical specification of vowel *quality* (Reis, 1974). However, phonotactic restrictions due to vowel *length* (Becker, 1998; Hall, 1992), the allophonic distribution of (physically) short lax and tense vowels in unstressed syllables (Ramers, 1988), as well as the exclusively durational /a/-/a:/ contrast (Hoffmann, 2011) suggest that *duration* rather than quality (i.e. in terms of tenseness), is the most consistent and reliable aspect of vowel length in the German vowel system. Thus, in order to avoid a phonological over-specification, length-associated differences in vowel quality should be considered as secondary modifications that do not rely on lexical phonological specification.

The present study does not primarily intend to solve this phonological dispute since it omits qualitative influences on perception. It rather investigated the psychoacoustic characteristics of the German /a/-/a:/ contrast in order to evaluate the phonological status of the duration cue. As such it can be seen as a basis on which further studies on the German length-quality contrast can be performed. In terms of Massaro et al. (1983) the question arises whether the linear changing duration cue will "... result in [...] continuous or discrete changes along a perceptual dimension" as assessed by psychoacoustic measurements.

Parameters for phonological contrasts

The mechanism responsible for the perception of a contrast, i.e. the mapping of an acoustic cue to a phonological category, has been described as Categorical Perception (Goldstone & Hendrickson, 2009; Liberman et al., 1957) which states that an item from a linearly changing continuum is attributed to either one or another category. The discrimination along the continuum is governed by identification, insofar as within a category no discrimination is possible. Only items from different categories can be discriminated. This traditional concept of categorical perception has been questioned in several studies since within-categorical discrimination has been shown even for stop consonants (Carney, Widin & Viemeister, 1977; Hanson, 1977). Furthermore, it has been shown that the traditional experiment design comparing identification and discrimination performances cannot reveal how speech is categorized, since discrimination results are influenced by test person's "subjective criteria" (Ylinen et al., 2005) during the discrimination task (Gerrits et al., 2004; Massaro et al., 1983; Schouten, Gerrits & van Hessen, 2003b). However, in line with Ylinen et al. (2005), the analysis of the discrimination performance along the continuum, especially at the category boundary, can be used to show whether a phonological category exists in the first place. As Ylinen et al. (2005) stated it: "[...] the listener's utilization of their subjective criteria should improve discrimination performance at the category boundary if the categories are accessed". Ylinen et al. (2005) have shown that the phonological status of vowel duration in Finnish produces an increase in discrimination performance at the category boundary in native speakers of Finnish, but not in native speakers of Russian (which do not have phonological vowel length). Hence, unlike phonological contrasts, non-phonological contrasts will not increase the discrimination performance at the category boundary (Flege & Munro, 1994; Miyawaki et al., 1975; Pisoni et al., 1974a; Pisoni et al., 1974b; Stevens & Blumstein, 1978; Xu, Gandour & Francis, 2006; Ylinen et al., 2005).

Whereas the discrimination performance supports the location of a phonological boundary, goodness ratings reveal the structure of the category itself (Diesch et al., 1999; Iverson et al., 1993; Kuhl, 1991; Kuhl, 2004): The best rated item is seen as the category center, i.e. the prototype, which structures its surrounding psycho-physical space into good and bad areas. In this sense, a contrast is seen as phonological, when it exhibits a sharp category boundary concomitant with an increase in discrimination performance as well as a worsening of goodness ratings.

Spectral and durational aspects of long and short vowels in German

So far, no consensus was found in the literature concerning the phonological status of either acoustic cue. For vowel quality, and thus against vowel duration, as a lexically specified feature speak the findings of Bennett (1968), Weiss (1974; 1978) and Sendlmeier (1981). Using identification tests, they have shown that non-low vowels seem to be categorized mainly on the basis of quality. Nevertheless, vowel length cannot play a completely subordinate role, since the distinction between /a/-/a:/ is performed on the basis of duration. As such, Heike (1970} has found that the location of the category boundary was supported by a test which could be interpreted as a goodness rating. *Acceptance*, as he named it, dropped toward the category boundary. The change in quality due to shortening affects the position of *high short lax* vowels to such an extent that they overlap strongly in formant frequencies with *mid long tense* vowels (Hoffmann, 2011). Although the distinction between these vowels is not based on a primary phonological distinction, since it involves two features – vowel height and vowel length –, the *mid long tense* vowel /oː/ and the *high short lax* vowel /ʊ/ are differentiated mainly on the basis of duration (Bennett, 1968; Heike, 1970; Weiss, 1974).

The present study

The present study is part of a research project on German vowel length, experimentally testing the phonological status of vowel duration and vowel quality in German. In order to minimize potential interactions of vowel quality and vowel duration, the present study only considers the durational contrast of /a/ vowels. An identification test was used to find the category boundary. An adaptive discrimination test was performed to further evaluate whether putative length categories affect the mapping of linear changing duration to the psycho-physical duration space. By means of a discrimination test the durational just-noticeable-differences (JND) were assessed (Lapid et al., 2008), i.e. the threshold when one vowel is perceived as longer from another. Finally, a goodness rating was performed to find the areas of good durational instances which might be interpreted as durational prototypes. Pisoni et al. (1974b) showed that, irrespective of the absolute acoustic differences, behavioral response time (RT) is shorter when subjects have to categorize within-categorical items as compared to items near the category boundary. Thus, besides goodness ratings, RTs in the identification test represent an additional parameter for examining the space of perceptual categories

(Näätänen, Lehtokoski, Lennes, Cheour, Huotilainen, Iivonen, Vainio, Alku, Ilmoniemi, Luuk, Allik, Sinkkonen & Alho, 1997).

Taking into account the findings of the results of Miyawaki et al. (1975), Flege et al. (1994) and Näätänen et al. (1997), a non-phonological duration contrast will be mirrored in a *linear* mapping of the linearly changing duration to the psycho-physical vowel space. Linear mapping should be reflected by just-noticeable-differences increasing linearly with increasing vowel durations. As such they will match Weber's Law (Dehaene, 2003). The category boundary between short and long should not be steep, i.e. it will change linearly from one end of the continuum to the other mirrored by non-changing response times and goodness ratings.

Should the contrast be governed by categories and as such be phonological, linear vowel duration will be mapped onto the psycho-physical vowel space in a non-linear way. The category boundary will be steep insofar that in the continuum there will be clear categorical areas and a small area with changing identification probability. The category boundary forces an increase in discrimination performance, prolongs response times and worsens goodness ratings.

Methods

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. All had normal hearing as confirmed by an audiometric screening test.

Stimuli

The stimulus material comprised instances of the trochaic nonsense word /tatə/, which were drawn from a continuum in which only the vowel /a/ was varied in duration. The use of the disyllabic nonsense word was motivated by three aspects. Firstly, it avoided a lexical bias, i.e. a perceptual preference for the item that is more familiar to the subjects. Secondly, it represents a trochaic foot, which is the most common metrical structure of German (Féry, 1996) and respects the phonotactic restrictions of short vowels (Becker, 1998; Hall, 1992). Thirdly, 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 (Hirata, 2004); see Gottfried, Miller & Payton (1990) for speech tempo effects.

In order to exclude any accidental variations of natural speech, these items were synthesized by concatenating noise segments with vowel segments. The principal validity of synthesized stimuli regarding phonetic/phonological research questions has been shown in several studies (Ainsworth, 1972; Blomert & Mitterer, 2004; Bunton & Story, 2009; Sams, Aulanko, Aaltonen & Näätänen, 1990). Stimuli were synthesized using the formant synthesizer of Hertrich & Ackermann (1999; 2007) which computes formants as amplitude- and phase-modulated sinusoids on the basis of the F0 frequency. Aperiodic speech segments, which represent fricatives or stop consonant bursts, are generated by adding a random parameter into the formants' phase progression. Formant values and the temporal aspects of the stimuli were approximately derived from a real recording of the first author, analyzed with PRAAT (Boersma & Weenink, 2009). Segment durations and the vowel formant frequencies can be seen in Figure I. 1. Formant transitions between the consonants and the /a/ vowel were implemented as reported in Strange (1987) and Strange & Bohn (1998) comprising a *rising* F1 (600 - 800 Hz) and *falling* F2 (1440 - 1240 Hz) and F3 (2700-2500 Hz).

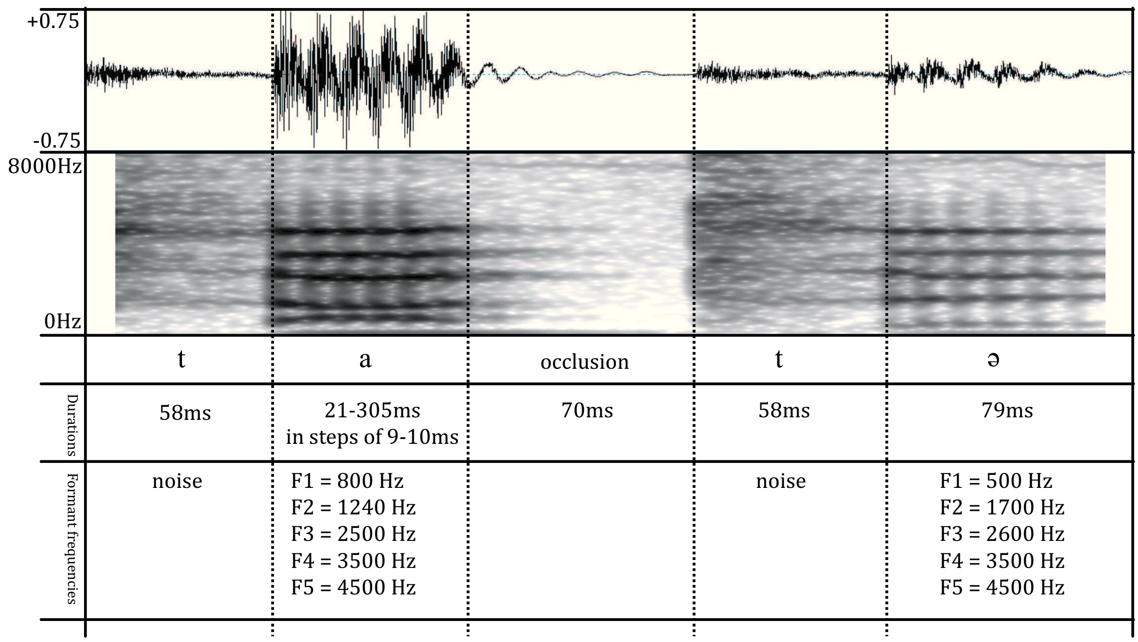

Figure I. 1: Spectrogram and waveform of the synthesized stimulus /tatə/ as shown by praat. The spectrogram is analyzed with a window length of 10 ms. The clipping of the edges in the spectrogram is a result of the analysis window in praat.

In order to increase naturalness and the stressed-unstressed relation between the syllables, some minimal intonation was implemented in the test stimuli comprising a pitch accent on the first syllable (F0 = 100 to 110 Hz across the entire vowel), followed by a terminal fall on the second syllable (105 to 80 Hz across the entire vowel). This is a frequent pattern for single-word utterances in German (Grabe, 1998). The different versions of the vowel /a/ differed in their number of pitch periods, i.e., units of 9 - 10 ms depending on pitch (110 - 100 Hz). The shortest /a/ vowel in the basic continuum, from which stimuli were drawn for the different tests, was 21 ms long, the longest 305 ms.

Procedure

Every subject participated in three tests during a single session in a sound proof chamber at the Department of Neurology of the University of Tübingen. The order of the tests were 1) a forced choice identification test, 2) a goodness rating and 3) an stair-case forced choice discrimination test. Since the subjects were not given any response about their identification and goodness rating and the continuum varied only in one dimension, no learning effect was expected during the identification task (see Livingston et al. (1998), who conducted category learning experiments with one- and multidimensionally varying items, and Ylinen et al. (2005), where the discrimination task followed the identification task). The number of items varied between the tests (see below) in order to save time. The authors were well aware of the arbitrary size of the tested sets. In any case, each set covered the needed range between the short and the long phonological category.

Subjects had to answer via keyboard. No training phase was offered, since it was assumed that German native speakers can categorize short and long vowels *per se*. No repeated listening option was available. Instructions were given twice: Once orally by the first author with the possibility to ask questions in case of misunderstanding and secondly via headphones before each test. Stimuli were presented via Sennheiser HD 201 headphones at an intensity agreeable for the subject. The entire test was controlled by a MATLAB (version R2009a, Mathworks) procedure running on a laptop (Acer Extensa 7630EZ) including both stimulus presentation and acquisition of behavioral data. Technical response time of the Acer keyboard was ca. 50 ms (SE = 0. 6 ms). The analysis of the behavioral data was performed with MATLAB, and the statistical calculations were done with the GNU R 2. 9. 2 statistical package.

Description and analysis of the identification test

In order to find the location of the boundary between category short and long in the duration continuum, subjects had to perform a two-alternative forced-choice identification test. To direct the subjects' attention to the phonological length distinction, the German words *Ratte* [ratə] “rat” and *Rate* [ra:tə] “installment” served as lexical examples when the test was introduced to the subjects. Subjects were instructed to push the button for short with their left hand, when they perceived an /a/ like in *Ratte* [ratə] ‘rat’ and the button for long with their right hand when they perceived an /a/ like in *Rate* [ra:tə] ‘installment’. The short-long continuum was covered by 15 steps with vowel durations ranging from 50 ms to 187 ms. Each stimulus was played 10 times in a pseudo-randomized order. Interstimulus time depended on response time (RT) ; a new stimulus was presented 1. 5 s after the response. RT was measured from the offset of the stimuli. Subjects were instructed to answer "concentrated and quickly", but were not informed that RT was actually measured. For each vowel duration, RT longer than three standard deviations from the group mean were excluded from calculations concerning the response times. These where 41 of 3000 values (0. 03%) in the entire group.

The individual category boundaries and reaction times along the continuum were determined by mathematical model fitting procedures. Each individual probability bin “*vowel was perceived as long*” was fitted by an psychometric logistic function by MATLAB’s least squares estimation in order to model the sigmoid shape of the identification results and to extract the individual category boundary (Lapid et al., 2008; Wichmann & Hill, 2001). The model in (1) was used:

(1) $$\Psi = \frac{1}{1 + e^{-(x-\alpha)/\beta}}$$

‘Ψ’ represents the probability of “*vowel was perceived as longer*”, ‘x’ represents the vowel duration, ‘α’ the curve’s midpoint (P = 50%) in the x-axis and ‘β’ the slope of the curve at that midpoint. Hence, parameter ‘α’ provided the location of the category boundary. See Figure I. 2a for an exemplary fit of one subject’s identification results.

In order to extract exact numeric values for the RT-maxima, the individual RTs were chosen to be fitted by a Gaussian function since it modelled best the rising-falling shape of the data curves (compare Figure I. 3a). The maximum value of the Gaussian fit was picked and interpreted as the RT maximum. These maxima were correlated with individual category boundaries to test whether the RT maxima are consistently located at the category boundaries.

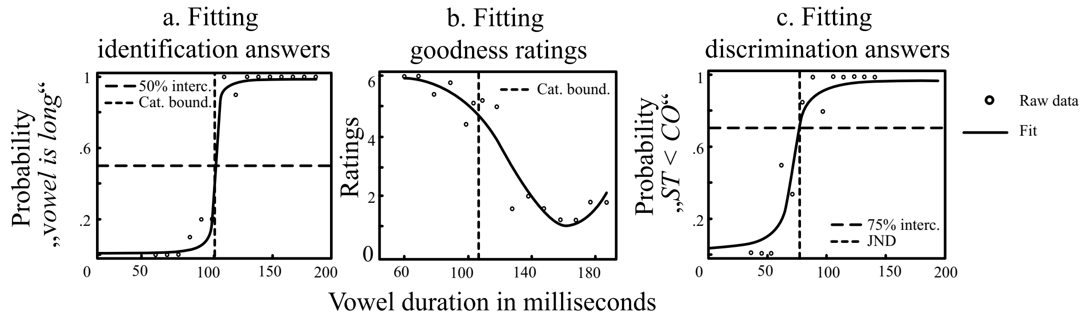

*Figure I. 2: **a & c)** Examples of the psychometric logistic function fit to raw data for the extraction of precise numerical data for the 50/75% y-axis intercept for identification and discrimination results. **b)** Example for the Gaussian model fit to goodness ratings. The circles indicate the raw data for one random subject. The black curve represents the model fit.*

Description and analysis of the goodness ratings

In order to find the prototype of each category and to evaluate the goodness profile across the durational continuum within each category, two sets of stimuli were used. One set had to be judged as though /a/ belonged to the category short; the second set was judged as though it belonged to category long. Vowel duration ranged from 21 ms to 128 ms (goodness of short vowels) and from 79 ms to 187 ms (goodness of long vowels). Both sets consisted of 12 stimuli each, including within-categorical items as well as some items beyond the category boundary. These, of course, were expected to be rated as bad. In each set every stimulus was played 5 times in pseudo-randomized order resulting in 2*12*5 = 120 trials. Using the German school mark system (1 = very good, 6 = very bad), subjects had to rate the goodness of each exemplar according to their native speaker's intuition, i.e., as though the perceived vowel was produced by a non-native speaker. As in the identification task, a new stimulus was presented 1. 5 s after an answer was given. One subject was excluded from calculations for the category long, since he or she stereotypically rated "5" for all items.

The aim of the analysis was i) to find the prototypical category center which is defined as the item with the best mean goodness rating and ii), whether the worsening ratings support the category boundary. For this, the individual results had to be interpolated. This was done by fitting an upside-down flipped Gaussian curve to individual mean ratings in each category set (see Figure I. 2b for an exemplary fit in the set long). The curve was chosen since it modeled best the quasi up-side-down bell-like shape of the results, in comparison to e.g. a square or a sigmoid curve. The asymmetric shape of the curve is seen as a result of the limited range of the stimuli. Furthermore, a Gaussian curve was already used to model the exemplar distribution of a category (Feldman & Griffiths, 2007) which can be reflected in goodness ratings. To analyze goodness ratings with respect to the category boundary, the point of intersection between the curves was calculated and correlated with the individual category boundary for each subject.

Description and analysis of the discrimination test

The discrimination performance along the continuum is crucial for analyzing the impact of the categories upon the mapping of a continuum onto the psycho-physical space. Subject had to decide whether in the word sequence "**A-X**", the /a/ vowel in X was longer than the /a/ in A. A discrimination test using only two subsequent items reduces the memory load (Pisoni et al., 1974a). Subjects were instructed to push the "yes" button when they perceived /a/ in X as longer than in A and "no" when both were perceived as equally long or when the /a/ X was perceived as shorter. Suitable examples were given. Offset to onset time between A and X was 500 ms. Interstimulus time depended on response time, a new stimulus was presented 1. 5 s after the response.

The discrimination set consisted of 12 A stimuli with /a/ durations ranging from 50 ms to 158 ms which were played in pseudo-randomized order. A staircase procedure was performed for each of the A (Lapid et al., 2008). The procedure was as follows: After one positive answer, the duration of X was decreased by one pitch period (~10 ms), i.e. duration was made more similar. To increase the probability that subjects will perceive a duration difference after one negative answer, the duration of X was increased by three pitch periods after one negative answer. Possible X duration ranged from 21 ms to 305 ms. Based on a preliminary study, the test started with a difference of y = ~ 70 ms for each pair. The test ended when each A stimuli was played 30 times.

The staircase 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 to small durational differences along the continuum, the adaptive discrimination test actually determines the just-noticeable-difference (JND) for each vowel in the continuum (Lapid et al., 2008). JNDs reflect the value which is needed in order to perceive a difference between two subsequent items.

The aim of the test was to find the just-noticeable-differences (JND) along the duration continuum. This was done by fitting the answer probability "*vowel was longer in X than in A*" by the psychometric logistic function as a function of X duration for each A duration. An example of the fit for a single subject and the vowel duration of 128 ms for A is shown in Figure I. 2c. Due to the adaptive procedure, the actual data concentrated around the perceptual threshold, which is why the margins of the scale were more or less underrepresented. Therefore, in order to fit the sigmoid curve more smoothly, two 0% values at the right and two 100% values at the left end of the abscissa were added. A probability of 75% was chosen as threshold for the just-noticeable-difference (JND), mirroring the ¼ to ¾ distribution of yes-to-no answers in the staircase test (Lapid et al., 2008) and was calculated by means of (2).

$$\text{(2)} \quad T = \alpha + (-\beta * \log(\frac{1}{P} - 1))$$

where ‘T’ represents the duration threshold, i.e. the duration when the X stimulus was perceived as longer and ‘P’ the probability of 75%. ’α’ and ’β’ were assessed by the fitting procedure. In two subjects, no 100% saturation was observed at the right margin of the tested continuum. This indicated that the actual vowel duration in X needed by the respective subjects to perceive a difference between A and X was not covered by the test materials. Hence, these instances were excluded from further calculations.

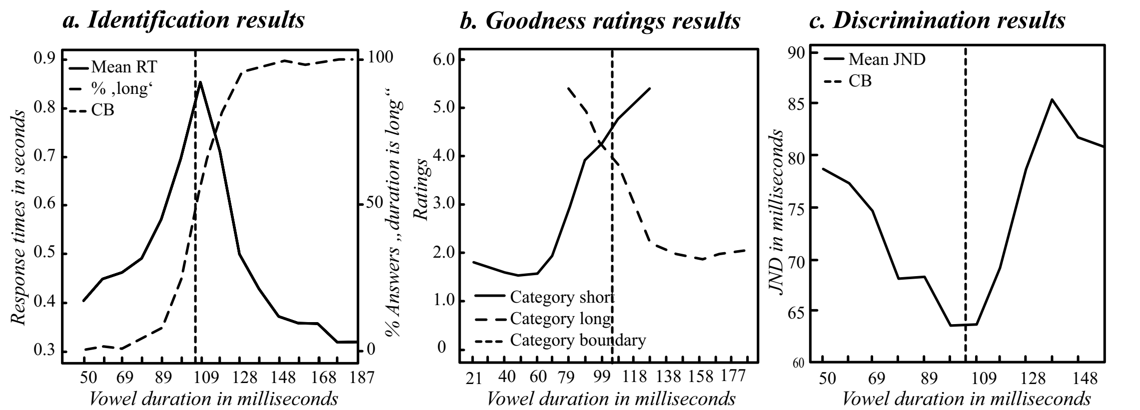

Figure I. 3: ***a)*** *Identification results including the mean identification curve, the mean response times and the category boundary as found by the psychometric logistic function.* ***b)*** *Mean group averages for the goodness rating. Sets were rated as though the vowel belonged to category short or category long.* ***c)*** *Mean just-noticeable-difference (JND).*

Results

By means of the identification test a boundary was found between category short and long at 105. 9 ms (95% confidence interval = 102. 2 – 109. 5 ms). As can be seen in Figure I. 3a, the boundary is steep insofar as there are clear areas attributed to category short or long and that the probability of perceiving a long vowel changes from ~10% to ~100% during 40 ms. In this sense, the flat areas reflect the phonological length categories.

The phonological status of the category boundary is supported by the course of the response times (RT) which exhibit a clear maximum at 103. 9 ms (95% CI = 100. 0 – 107. 8 ms). Individual RT maxima are temporally associated with individual category boundaries as found by a Pearson correlation analysis ($r = 0.65$, $t = 3.6$, $p < 0.01$).

The location of the category boundary is furthermore reflected by the course of the goodness ratings in the short and long set (Figure I. 3b). Ratings exhibit a stable area but become abruptly worse toward the category boundary. Rating minima, which could be interpreted as the category center, can be seen at 60 ms in category short and at 158 ms in the category long. However, both minima could not be confirmed statistically. The point of intersection between the ratings for short and long in the 'bad areas' serves as a measure how goodness ratings behave near the category boundary supporting indirectly its location. The point was found by calculating the intersection between the Gaussian models fitted to individual ratings. It is located at 97. 7 ms (95% CI = 92. 5 – 102. 9 ms). Although it deviates significantly in average by 8. 1 ms from the category boundary into the category short ($t = -2.7$, $p = 0.01$), the individual points of intersection are significantly correlated with the individual category boundary ($r = 0.5$, $p = 0.01$).

Figure I. 3c shows that course of the just-noticeable-differences (JND) for duration changes in a non-linear way insofar as there is a minimum at the category boundary. This minimum corresponds to a maximum in discrimination performance. The squared shape of JND course is supported by a squared linear model for by-subject normalized data ($F_{(2,218)} = 17.2$, $p < 0.001$), whose minimum is located at 97. 1 ms, i.e. in the proximity of the category boundary of 105. 9 ms. In this sense, German length matches the criterion for Categorical Perception. The present discrimination at 75% performances obviously violate Weber's Law as indicated by X/A ratios (Table I. 1) reflecting a non-acoustic discrimination. The ratio drops from 2. 5 at the left edge of the continuum to ~1. 5 toward the category boundary. Within long category, the ratio stays nearly constant at ~ 1. 5. Thus, the X/A ratio may represent an additional psychoacoustic characteristic for the phonological status.

A duration	50	60	69	79	89	99	109	118	128	138	148	158
ratio (X/A)	2. 5	2. 3	2. 0	1. 8	1. 7	1. 6	1. 5	1. 5	1. 6	1. 6	1. 5	1. 4
SE (X/A)	. 1	. 1	. 1	. 08	. 09	. 07	. 07	. 06	. 07	. 08	. 07	. 06

Table I. 1: Mean durational ratios between vowels in the A-X design.

Discussion

The aim of the present study was to investigate the psychoacoustic characteristics of the German duration contrast and whether putative duration categories affect the mapping of linearly changing vowel duration to the psycho-physical vowel space. The study excluded consciously the interaction of vowel quality with duration. The psychoacoustic characteristics were assessed by means of an identification test, goodness ratings, and an stair-case discrimination test. It was hypothesized that if the length contrast is phonological in German, it will influence in a non-linear way the perception of the linearly changing duration cue: The category should affect the perceptual parameters insofar as a sharp category boundary should separate clear categorical areas, which is concomitant with non-linear changes in response times, goodness ratings and discrimination performance (Flege et al., 1994; Iverson et al., 1993; Miyawaki et al., 1975; Näätänen et al., 1997; Ylinen et al., 2005). Regarding these criteria, the present results support the conclusion that there exist duration categories in German, i.e. vowel length is phonological (Becker, 1998; Hall, 1992): The length contrast is reflected by a steep category boundary concomitant with an increase in the identification response times and in the discrimination performances as well as worsened goodness ratings (Figure I. 3).

In the following, certain aspects of the results will be discussed. Concerning the findings in the identification test, the location of the category boundary at 105. 9 ms is probably bound to the particular consonantal context used in the present study such as manner (Lehtonen, 1970) and place of articulation (Antoniadis et al., 1984) as well as voicing (Braunschweiler, 1997; de Jong, 2004; Metz, Allen, Kling, Maisonet, McCullough, Schiavetti & Whitehead, 2006). With the perceptive system adapting to context-related influences on vowel durations, the category boundary can be expected to move along the continuum. Also, the virtual definition of speech tempo in the present study presumably affected the location of the category boundary (Gottfried et al., 1990; Hirata, 2004). It can be expected that e.g. with shorter schwa length, the perceived speaking rate would increase, resulting in a shift of the category boundary toward shorter vowel duration.

One could argue that the present category boundary results from a development of ad hoc categories divided at the midpoint of the set (as was shown by Smits et al. (2006)). While this might be true for the identification test, the increase of discrimination performance at the boundary (Figure I. 3c) strongly supports the assumption that the location of the boundary is phonologically determined. In particular, the discrimination maximum was located at the boundary although the sets used for identification and discrimination were not coherent in size. Furthermore, the JNDs in the present study violate Weber's Law (Dehaene, 2003) speaking against a purely acoustic discrimination and as such supporting the phonological status of vowel duration.

In the present study, no specific category center was found. One could argue that no concrete phonological prototype are to be assumed (Diesch et al., 1999; Kuhl, 1991). In this sense, the present results could be considered in analogy to a hypothesis that prototypical targets can be located outside of the actual physical/phonetic articulatory space (see Diesch et al. (1999)'s and Johnson, Flemming & Wright (1993)'s proposal with regard to vowel quality). In this sense, there are no concrete duration prototypes but virtual durational targets. However, the results might as well be interpreted in a different direction. In generative phonology, segment durations are defined by intrasyllabic positional slots (Becker, 1998; Hall, 1992; Wiese, 1988). In terms of vowel length, a short vowel occupies one slot, a long vowel two slots. During speech perception, linearly changing vowel durations were mapped onto syllabic position slots. The flat areas in both the goodness ratings and the identification test (Figure I. 3a&b) could therefore reflect the duration bandwidths which optimally occupy either one or two slots. In this sense, the categorical prototypes would not be specific points in time but durational bandwidths, one for each category. Outside of the flat areas, ratings worsened significantly toward the category boundary. Although the intersection points between the ratings of the two sets (category short and long) were temporally cued to the categorical boundaries, their location was shifted toward the short category. The shift might be induced by the incoherent sizes of the tested goodness and identification sets. However, as the size of the discrimination set was not congruent with the identification set and the discrimination peak was nevertheless located at the category boundary, the shift might also be the result of different acceptance for short and long vowels: The acceptance of short vowels with regard to their prolongation is lower than the acceptance of long vowels with regard to their shortening.

Such an asymmetry of goodness ratings would speak in favor of a markedness of short vowels (Becker, 1998). Furthermore it would reflect findings concerning speech tempo: Short vowels uttered with normal and fast speaking rate have to be clearly separated from long vowels which is why they do not overlap with long vowels (Hirata, 2004).

Regarding vowel length as phonological accounts for several phonotactic regularities in German, as the maximal number of elements a German syllable can contain is restricted to three (Hall, 1992; Wiese, 1988): 1) The overall amount of post-nucleus consonants in a German syllable depends on vowel length. A short vowel can be followed by maximally two consonants. A long vowel can be followed by only one consonant. 2) The velar nasal [ŋ] can occur only after a short vowel. Atlhough [ŋ] is a single consonant at the surface, analyzing it the sequenz /n+g/ underlyingly, it occupies two positional slots, restricting the vowel to one position.

Conclusion

Study I showed that the German length contrast between /a/ and /a:/ affects the psychoacoustic characteristics as assessed by an identification test, discrimination test and a goodness ratings. It showed i) a sharp category boundary which is ii) accompanied by a maximum in discrimination sensitivity and iii) a maximum in response times. This impact on the psychoacoustic characteristics reflected the vowel duration’s status as a category. As such, the present study supported the assumption that the vowel duration contrast in German is phonological.

Study II: Neural processing of German vowel length.

Time courses of evoked fields in response to speech and nonspeech signals

Introduction

Vowel duration

Distinctive vowel duration triggers a change in meaning by lengthening the vocalic element, e.g. German [ban] 'spell' vs. [ba:n] 'course'. Maddieson (1984) reports that probably 20% of the world's languages use duration in such a way. Recent behavioral studies have shown that listeners who use vowel duration in a contrastive, phonological way exhibit psycho-acoustic characteristics matching categorical perception – unlike listeners lacking phonological vowel length in their native language (Ylinen et al., 2005).

In a study of the characteristics of German vowel duration, Tomaschek, Truckenbrodt & Hertrich (2011) found a sharp boundary between short and long vowel durations at 105.9ms in a continuum, accompanied by an increase in discrimination performance (Figure II. 1d), while an assessment of just-noticeable-differences revealed that within-categorical discrimination was significantly attenuated. Although these findings suggest that vowel length in German matches the criteria for categorical perception, stating that an item from a linearly changing continuum is attributed to either one or another category (Liberman et al., 1957), to date, categorical perception has thought to be a property of consonants. In contrast, vowels exhibit Perceptual Magnet Effect (Kuhl, 2004): The similarity of juxtaposed items increases the nearer they are located toward the category center, and within category discriminations are still possible.

The present study aimed to investigate whether the results of Tomaschek et al. (2011) do indeed support the idea that vowel durations exhibit categorical perception by examining the neural correlates of the processing of phonological and acoustic durations. The study was based on two questions: Do the different vowel length categories correlate with different patterns of neural activity? Is the boundary between category short and long marked by distinctive neural signature?

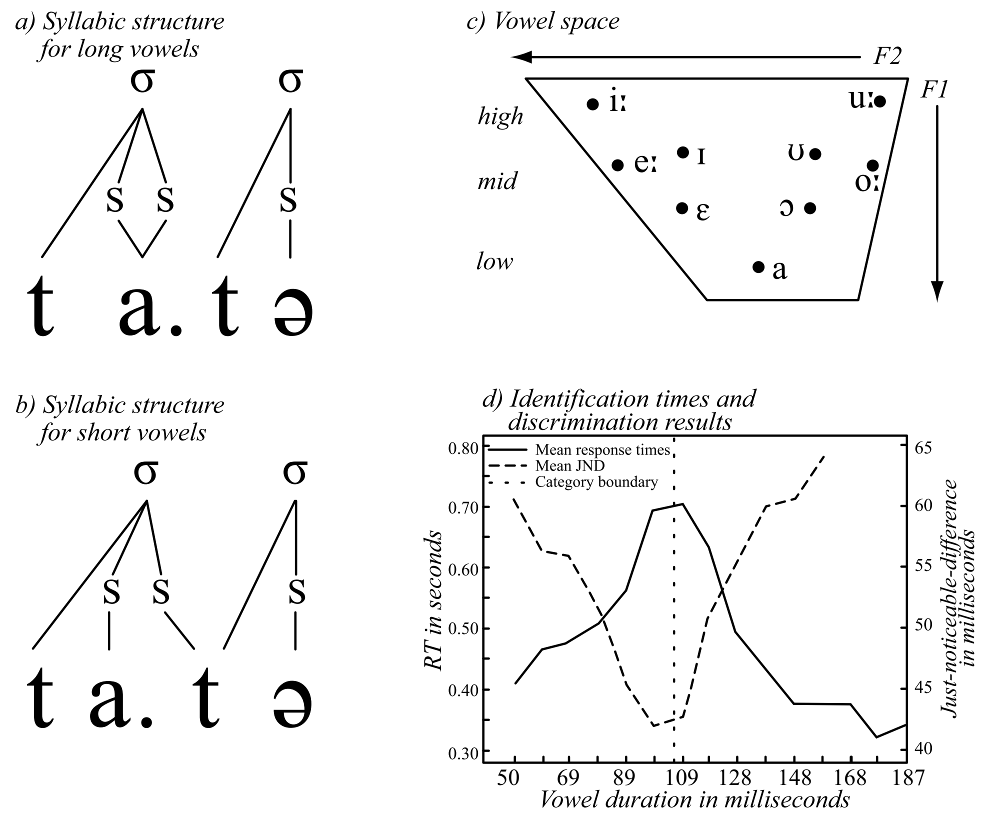

Figure II. 1: ***a&b)*** *Phonological representation of the German vowel duration contrast with a long vowel linked to two moras and a short vowel linked to one mora and the onset of the following syllable.* ***c)*** *Schematics of the German vowel space showing the centralization of short vowels. Short centralized vowels have higher F1 frequency than long peripheral vowels. /a/ does not show any significant quality change between short and long vowels.* ***d)*** *Results from Tomaschek et al. (2011) (Study I) showing the identification response times (vowel = long?) and just-noticeable-differences for duration change. Both measures show a peak at the category boundary (105. 9ms) matching the criteria of categorical perception*

Phonological vowel duration in German

In German, the phonological distinction between short and long vowels is important for phonotactics and syllabification (Becker, 1998; Hall, 1992). In formal phonology, this contrast can be described in terms of the amount of intra-syllabic constituents, – slots –, the vowel is linked to (Becker, 1998; Zec, 2007). A long vowel is linked to two slots while the following consonant just occupies the onset of the next syllable (Figure II. 1a). A short vowel is linked to one slot (Figure II. 1b). Since the second slot is a substantial part of the German stressed nucleus that must be occupied, the onset consonant of the following syllable becomes ambisyllabic: It is associated to both syllables. This status is reflected in the German ortography where a short vowel is followed by two consonant letters as in *Wasser [vase]* 'water'. In contrast to other Germanic languages like Swedish or Norwegian, German does not have long consonants (Hall, 1992; Lehtonen, 1970). Additionally to the length distinction, mid and high vowels exhibit an interaction between duration and *tenseness* (Hall, 1992; Ramers, 1988): Short and long instances have different positions in the vowel space (Figure

II. 1c). The short and long /a/ vowels, however, do not show large quality differences (Hoffmann, 2011; Sendlmeier et al., 2006).

Neural processing of acoustic signals

The present study investigated the neural processing of durations by means of auditory evoked magnetoencephalographic (MEG) responses. Several neuroimaging studies have shown that these reflect speech characteristics such as syllable onsets (Hertrich et al., 2012), pitch periodicity (Hertrich et al., 2000) and the onset structure of syllables (Kaukoranta, Hari & Lounasmaa, 1987). The amplitude and latency of auditory evoked responses such as the *P50, N100, P200* and their magnetic correlates *M50, M100,* M200 reflect different stages of auditory processing and different properties of the perceived signal.

P50/M50, which occurs around 50ms after the signal onset, reflects an initiation of signal processing (Lijffijt et al., 2009). When two identical signals are played consecutively within a short interval, the amplitude of the *P50/M50* to the second signal is significantly reduced, a phenomenon called *P50*-suppression (Cadenhead et al., 2000; Light et al., 2001; Mathiak et al., 2011). Since the characteristics of the first and second signal are equal, processing the second one appears to require less activity in the auditory system, possibly protecting higher cognitive functions from uncontrolled input (Lijffijt et al., 2009).

N100/*M100*, occurring around 100ms after stimulus onset, is sensitive to spectral and durational characteristics of speech and nonspeech input (Alain, Woods & Covarrubias, 1997; Čeponienė, Torki, Alku, Koyama & Townsend, 2008; Edmonds, James, Utev, Vestergaard, Patterson & Krumbholz, 2010; Hisagi et al., 2010; Vihla & Eulitz, 2003). The locus of *M100* generation is affected by vowel information such as vowel category and phonological features (Obleser et al., 2004; Scharinger et al., 2011). Roberts et al. (2004a) showed that *M100* latencies reflect phonetic aspects of the vowel, becoming shorter with increasing first formants (250 – 750Hz). *M100* latencies to speech signals cluster with respect to category affiliation of the signal, however *M100* latencies to nonspeech control signals with equivalent frequencies do not.

There is no consensus about the functional interpretation of the *P200/M200*. It seems to reflect tone duration (Alain et al., 1997), the harmonic complexity of the signal (Kuriki et al., 2006), categorical perception (Menning et al., 2000) and stimulus familiarity (Sheehan, McArthur & Bishop, 2005). Furthermore, the *P200/M200* response seems to be enhanced by discrimination training (Tong, Melara & Rao, 2009).

The processing of speech engages multiple brain areas. Auditory areas, located bilaterally in the superior temporal lobe (Alain et al., 1997; Zatorre et al., 2001), seem to 'chunk' auditory signals by integrating them into windows of short (20-40ms) and long (150-200ms) events (Boemio et al., 2005; Poeppel, 2003): Within a given temporal window, all events are integrated into one unitary perceptual event (Yabe et al., 2001b). Luo et al. (2007) found oscillations in the Theta band (4-8Hz, 250-125ms) during speech perception which have been thought to represent this integration mechanism, because the period duration of the Theta band approximately reflects the duration of syllables in speech (Greenberg, Carvey, Hitchcock & Chang, 2003).

Finally, the processing of durations differs depending on whether they represent a phonologically contrastive or a merely acoustic change. Phonological length contrasts are processed bilaterally, albeit preferentially in the left hemisphere. Acoustic duration differences, i.e. non-phonological contrasts, are processed preferentially in the right hemisphere (Inouchi et al., 2003; Kasai et al., 2001; Minagawa-Kawai et al., 2005).

The present study

The present study compared evoked magnetic fields (EMF) for disyllabic speech and temporally matched nonspeech stimuli with changing vowel and tone durations in order to investigate the differences of auditory and phonological processing of varying durations. The goal was to find a neural correlate reflecting the phonological boundary between category short and long with respect to its physical location.

So far, the neural processing of phonological vowel length has been investigated using mismatch-negativity, a neural correlate that is elicited by a deviant stimulus which is changed in spectral or temporal information and embedded in a sequence of standard stimuli (Näätänen et al., 1997). It has been found that listeners who have a native phonological vowel length distinction show enhanced neural processing as compared to listeners lacking one (Hisagi et al., 2010; Ylinen et al., 2006). However, no differences have been found between within- and across-categorical changes. To investigate the duration and category dependent changes in the EMF, the present study used the entire vowel (speech) and tone (nonspeech) duration continua, embedded into disyllabic trochaic carrier items which were presented in a passive listening task. (For the sake of convenience, the term ‘syllable’ will be used for both, the speech and nonspeech signal.)

The vowel /a/ was chosen in contrast to high vowels because the formant values between long and short phonological instances overlap strongly (Hoffmann, 2011), providing the opportunity to study the neural processing of phonological length without and spectral interference. Lehnert-LeHouillier (2010) have shown that the location of the short-long boundary is affected by spectral information, which in turn changes *N100*/*M100* latencies (Roberts et al., 2004a). It is important to note that the same subjects recruited in Tomaschek et al. (2011) participated in the present study, which is why their characteristics (category boundary at 105.9ms and identification response times) were used for the analysis and interpretation of the present MEG data.

Based on the P50-suppression paradigm (Cadenhead et al., 2000; Light et al., 2001; Mathiak et al., 2011) and the findings of Roberts et al. (2004a), it was hypothesized that the processing of categorical vowel durations in first syllables should be gated by underlying phonological durational templates. Gating should be reflected in the strength of the *M50* elicited by the second syllable, which at this point not only represents the initiation of the processing of the second syllable (Lijffijt et al., 2009) but also the processing of the measured vowel duration. In particular, it was expected that vowel durations representing *prototypical* exemplars would give rise to *M50* suppression effects in response to the second syllable. In this case, the onset of the second syllable would fall into an expected time interval. By contrast, in the case of ambiguous vowel durations, the onset of the second syllable would occur in an unexpected time point, violating the syllabic rhythm and increasing the strength of the *M50* to the second syllable. Since the processing of the nonspeech signal cannot refer to a categorical representation, no gating effects were expected in the nonspeech condition. However, durational effects might still be observed since the durational stimulus structure could interact with internal rhythms of the brain (Will et al., 2007). Taking into account the differences between phonological and acoustic processing, the *M50* to the second syllable should be lateralized to the left hemisphere in the speech, but to the right hemisphere in the nonspeech condition.

Finding that *M50* amplitudes increase and decrease across the continuum of vowel duration could thus be seen as the result of an enhanced or inhibited sensory gating mechanism for the processing of vowel duration. This would be consistent with the results of Tomaschek et al. (2011), who found that prototypical durations were reflected by faster identification response times in contrast to duration from next to the category boundary (Figure II. 1d) which is why it is expected to find a correlation between response times and M50 amplitudes.

Since the spectral information was not varied as a function of duration in both conditions, no effects on the *M100* latencies were expected. However, the *M100* strength to the second syllable could increase as a function of duration in both conditions because of sustained negativities which indeed increase with longer durations (Kushnerenko, Ceponiene, Fellman, Huotilainen & Winkler, 2001).

Considering the findings of Luo et al. (2007), the integration of duration should be reflected by a peak in the theta-band in both conditions. If that would be the case, the number of theta-cycles could be interpreted to represent the number of time slots the speech signal and accordingly the vowel was integrated into. Given this, it was hypothesized that the additional time slot for long vowels in contrast to short vowels should be reflected by an additional EMF component. As the nonspeech duration lacks a categorical representation, no particular effects of a category border were expected.

Methods

Participants

The same 20 native speakers of Standard German as in Tomaschek et al. (2011) participated in the present study (10 males, 10 females. Mean age: 27.4 years, SD = 4 years). They were paid for their participation. All were right handed and had normal hearing (as confirmed by an audiometric screening test). All participants gave written consent for their participation in the study, which was approved by the ethics commission of the University of Tübingen.

Stimuli

The stimulus material consisted of speech and nonspeech stimuli. The stimuli in both conditions were constructed so that they were equal in their temporal structure and differed in their spectral structure, which induced the perception as speech/nonspeech. Since the spectral information stayed equal along the continuum in both conditions but the duration in the first syllable changed, it was assumed that changes in the neuromagnetic responses as a function of duration could be connected to categorical/non-categorical perception of the signal.

The speech material was the same as in Tomaschek et al. (2011), comprising 12 versions of the trochaic nonsense word /tatə/ that differ in the duration of the vowel /a/. The use of disyllabic nonsense words was motivated by two aspects. First, it avoided a lexical bias, i.e. a perceptual preference for the item that is more familiar to the participants. Second, in the perception of vowel duration, the auditory system adapts to the speech environment (Luo et al., 2007) such as speech rate (Gottfried et al., 1990; Hirata, 2004). Therefore, the duration of the unstressed syllable /tə/ was kept constant to allow participants to use it for the creation of a model of speech rate. Although there is evidence that durations are perceived in a non-linear way, it was assumed that the use of linear changing durations would not affect the position of the category boundary (Smits et al., 2006).

In order to exclude any accidental variations of natural speech, the stimuli were synthesized. The specifications of the synthesizer can be found in Hertrich et al. (1999); and Hertrich et al. (2007). The duration continuum of the vowel in the first syllable ranged from 50ms to 158ms in steps of 9.5 - 10ms corresponding to the duration of single pitch periods (110 - 100Hz). Figure II. 2a shows the waveforms and the temporal structure of the stimuli.

Stop consonants /t/ were implemented by a noise burst and generating formant transitions in the speech signal in a similar way as reported in Strange (1987) and Strange et al. (1998) comprising a *rising* F1 (600 - 800Hz) and *falling* F2 (1440 - 1240Hz) and F3 (2700-2500Hz). The transition, however, could not give any information about vowel duration, since it was kept constant. Formant frequencies were F1 = 800Hz, F2 = 1240Hz, F3 = 2500Hz, F4 = 3500Hz and F5 = 4500Hz within the stationary part of vowel /a/. The schwa formant frequencies were F1 = 500Hz, F2 = 1700Hz, F3 = 2600Hz, F4 = 3500Hz and F5 = 4500Hz.

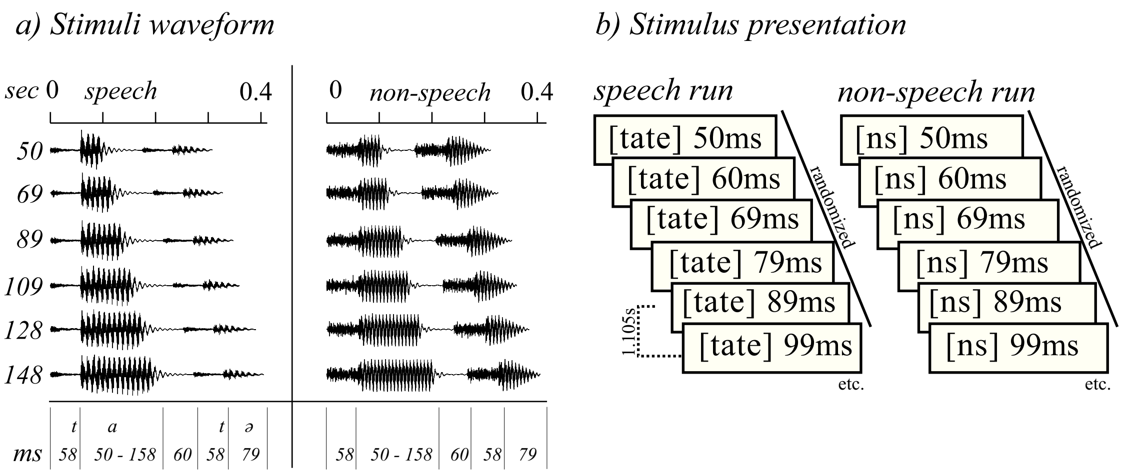

Figure II. 2: ***a)*** *The waveforms of the speech and structure matched nonspeech stimuli for every second stimulus duration. Segment durations are presented at the bottom.* ***b)*** *Schematics of the stimulus presentation in the MEG experiment. Durations were pseudo-randomized. Nonspeech stimuli were presented in the odd-numbered runs, speech stimuli in the even-numbered runs.*

Formant values were approximately derived from a real recording of the first author, analyzed with PRAAT (Boersma et al., 2009). Some minimal intonation was implemented in the test stimuli in order to increase naturalness in the speech stimuli, comprising a pitch accent on the first (stressed) syllable, followed by a terminal fall, which represents a frequent pattern for single-word utterances in German (Grabe, 1998). F0 of /a/ started with a rise from 100 to 110Hz over the entire first vowel /a/. During the final schwa F0 declined from 105 to 90Hz. No critical effects of the F0 rise on the EMF were assumed since it has been shown that F0 changes in the vowel do not affect the perception of vowel length in German and the location of the category boundary (Lehnert-LeHouillier, 2007), unlike changes in the spectral energy of the formants (Lehnert-LeHouillier, 2010). The rising pattern on the stressed vowel was preferred to a rising-falling pattern in order to avoid any interaction of perceived intonational variability with vowel length. Importantly, no additional dynamics were integrated in the speech stimuli (Lindblom, 1963).

In analogy to the disyllabic structure of the speech stimuli, each nonspeech syllable comprised a noise event that was followed by a periodic tone signal containing formant-like spectral modulations. For the sake of convenience, the periodic signal portions will henceforth be labeled “tone”. Synthesizing the nonspeech stimuli was favored over other techniques like spectral rotation in order to explicitly control the spectral and temporal structure of the stimuli. The nonspeech stimuli did not contain any spectral energy distributions that could have been interpreted in a phonetic way. In both nonspeech syllables, F0 and the five formants were kept constant in the entire stimulus in frequency and intensity: F0 = 150Hz, F1 = 500Hz, F2 = 700Hz, F3 = 900Hz, F4 = 1300Hz and F5 = 1500Hz, without any modeling of formant transitions.

Recording of neuromagnetic responses

For the MEG session, participants were comfortably seated in upright position, and encountered the stimuli in a passive listening task. An entire session was subdivided into six runs comprising 324 stimuli each. Even-numbered runs consisted of speech, odd-numbered runs of nonspeech stimuli (Figure II. 2b). Within each run the stimuli of the durational continuum (12 steps) were played in pseudo-randomized order with an onset-to-onset time interval of 1.105 sec. Depending on the vowel duration, the inter-stimulus-interval was 700-800ms. The interval was chosen so that the onsets of the stimuli would not get in phase with the 50Hz current rhythm. This ensured that each stimulus was perceived as independent, and that the sustained negativity after the foregoing stimulus would decrease (Kushnerenko et al., 2001). Due to the random presentation some speech stimuli might have elicited an across-category MMN (Ylinen et al., 2006). However, if one assumes an across-category MMN after three within-category repetitions, then only ~10% of the vowel durations were presented in such a constellation.

Each item from the continuum was played 83 times in total while participants looked at moving squares and circles projected at a screen in a distance of 50 cm. MEG recordings were taken using a 275-channel whole-head system (VSM MedTech Ltd., Coquitlam, Canada) in a sound-attenuated, magnetically shielded chamber. Stimuli were presented via insert earphones and plastic air tubes (Ear tone 3A system) at an agreeable intensity for the participants (ca. 70 dB). Stimulus presentation and synchronization of stimuli with MEG data was controlled by a custom-made computer program running on a DOS machine. MEG data were recorded at a sample rate of 1171.88Hz in epochs of 800ms duration including a 150-ms pre-stimulus baseline. The sampling rate was chosen for technical reasons.

Preprocessing of the MEG Data

For analysis, the individual neuromagnetic responses were fitted with two dipoles located in the left and right hemisphere after eye-blink rejection and pre-stimulus (-150ms – 0ms) baseline correction. About 15% of all trials had to be rejected due to eyeblink artifacts (an automatic procedure using subspace projection matched the measured data with the lead field of a prototypical eyeblink dipole structure; threshold = 70 nAm. see Hertrich, Mathiak, Lutzenberger & Ackermann (2002)). One subject was excluded from subsequent analysis due to a high eye blink rate (more than 1/3 of all trials).

Since individual dipole locations did not differ significantly between the conditions, neuromagnetic responses were averaged over the conditions and over the durations for individual dipole fit. This increased the signal-to-noise ratio and guaranteed that dipoles would capture the source variance of both conditions. A broad time window was chosen (50ms to 200ms) accounting for the variance of both M50 and M100 (see Mathiak, Hertrich, Lutzenberger & Ackermann (2000) who used a M50 dipole to account for mismatch negativities). Individual dipole sources in the left and right hemisphere were fitted using a spherical standard head model including all gradiometer sensors. Seed values were located in the bilateral auditory cortex based on previous studies (Hertrich et al., 2000). The bilateral dipoles accounted in average for 90% of the variance. The time course of left and right dipole strength was determined by means of subspace projection. This method provided the partial amount of overall activity for each sample point that can be accounted for by a given dipole structure (Hertrich et al., 2000). The polarity of dipole orientations was adapted to obtain a negative *M100* moment (Mathiak, Hertrich, Lutzenberger & Ackermann, 1999). The time courses were calculated for each duration, condition as well as subject and subsequently averaged across participants. Figure II. 3a shows group mean dipole locations including locations in MNI space, plotted to an exemplary anatomical display which was available for one subject. These were calculated by aligning the fiducials to the MRI anatomy.

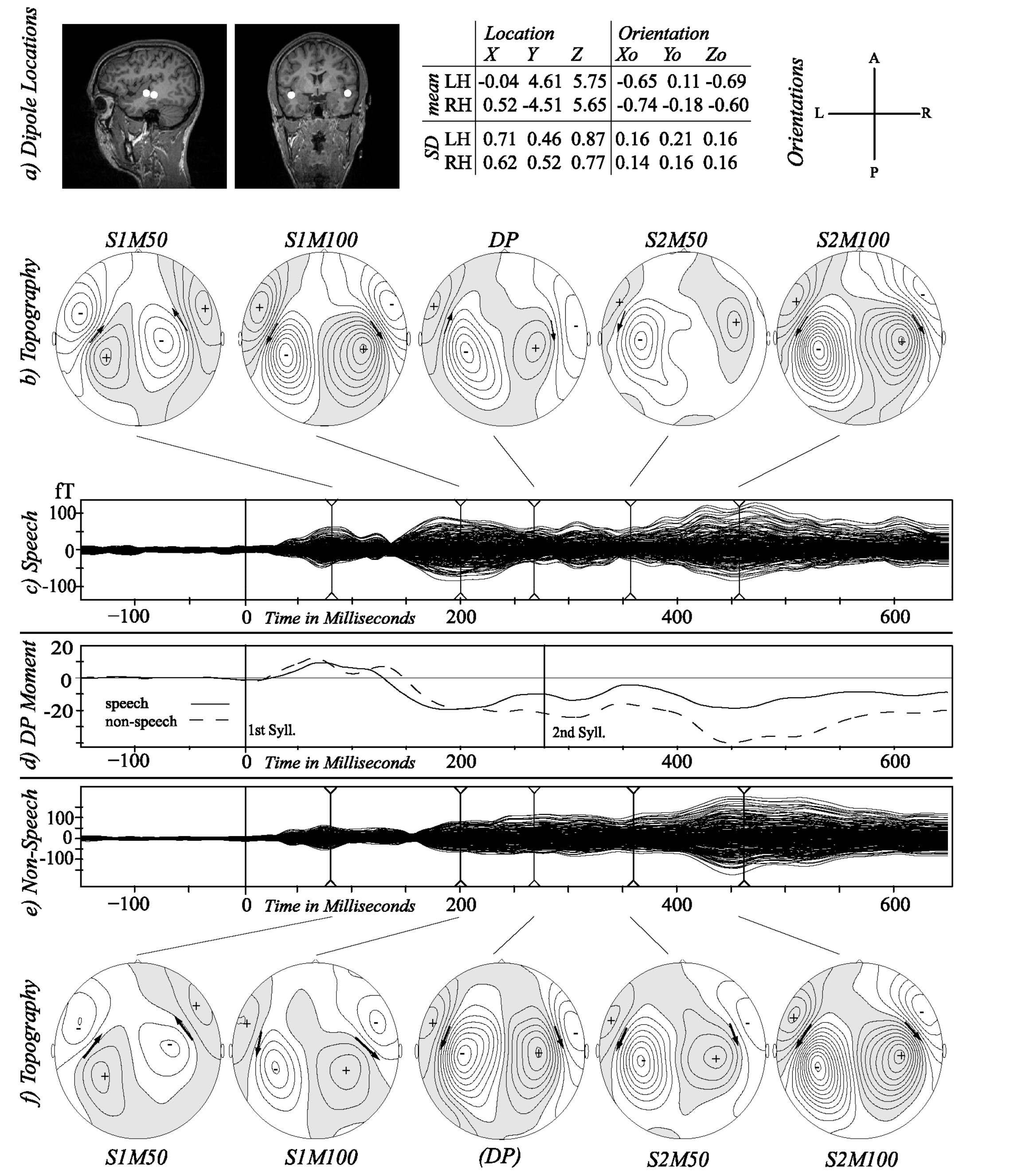

Figure II. 3: ***a)*** *Group mean dipole locations in the left and right hemisphere plotted in one exemplary subject, including their location and orientation in the CTF source space.* ***b&c)*** *Topographies and neuromagnetic responses for the stimulus with a vowel duration of 158 ms in the speech condition.* ***d)*** *Time courses of the dipole moment in the auditory cortex for the speech and nonspeech condition.* ***e&f)*** *Topographies and neuromagnetic responses for the stimulus with a vowel duration of 158 ms in the nonspeech condition. Arrows indicate the direction orientation of the dipole.*

Analysis of the auditory evoked responses

For the statistical analysis, mean dipole moment in five time windows was considered, covering the deflections which were assumed to be affected by duration. Analysis was performed by means of a mixed-effect model with the CRAN R statistic package, version 2.15 (R_Development_Core_Team, 2010). Mixed-effects models include random effects and fixed effects as well as random slopes and are able to handle empty cells. In contrast to the ANOVA analysis, mixed-effects models provide t-values but not F-values or p-values. These are associated to differences between factor levels or slopes for linear regressions as a function of covariates (Baayen, 2008). Generally, a difference/slope (Δ) can be seen as larger than zero and thus significant when $t > 2$.

The AMPLITUDES in the each time window were modeled by a triple interaction between CONDITION (speech and nonspeech), HEMISPHERE (LH, RH) and VOWELDURATION. Vowel duration was included as a covariate, accounting for linear and squared effects on the amplitude in some time windows (VOWELDURATION and VOWELDURATION ^2, respectively). Squared effects of vowel duration were included in order to model the hypothesized bell-shaped course of M50 amplitudes as a function of vowel duration based on considerations regarding P50-suppression/gating effects. PARTICIPANTS were included as a random factor in all models. The triple interaction was included as random slopes in order to account for repeated measures (Schielzeth & Forstmeier, 2009). Vowel duration was centered to a mean of zero in order to guarantee that the intercept was located at 0.

Analysis of alpha phase information

It has been shown that the amplitude of auditory evoked responses varies depending on the phase angle of alpha oscillations (Barry et al., 2004; Fellinger et al., 2011). In the present study, it was hypothesized that the *M50* in response to the second syllable could vary as a function of duration. Due to the varying onset time of second syllables relative to first syllables, the *M50* amplitude to second syllables could as well be affected by phase effects of alpha oscillations induced by first syllables. As a result, if the entire stimulus interval is taken as a window for spectral analysis, the effect of alpha phase on M50 amplitude could be expected to be stronger for the second as compared to the first syllable. Barry et al. (2004) used the slope of the alpha oscillations as a representative measure for the alpha phase (negative driving, positive driving). Since the power in the alpha band varied between participants in the present study, the grand average of 11.44Hz was chosen as the representative frequency (strongest activity in the alpha band. See results).

The slope of the alpha wave was determined by performing Fast-Fourier-Transform (FFT) on the time courses of the dipole moments for each trial after baseline correction (time window = 700ms including a 50ms pre-stimulus interval preventing the M50 to the first syllable from being cut off by the applied Hanning window). The FFT was performed on the first derivative of the time course in order to attenuate lower and enhance higher frequencies. The analysis was performed in the source space instead of the sensor space in order to reduce the amount of data and to increase the signal-to-noise ratio. Additionally, neural activity in the auditory cortex is better represented by the dipole because it is less affected by other (spontaneous) cortical sources than a single sensor (Hertrich, Mathiak, Lutzenberger & Ackermann, 2004; Ross, Borgmann, Draganova, Roberts & Pantev, 2000).

The slopes of the 11.44Hz sine wave were calculated for the location of the *M50* peaks in response to the first and second syllable for each stimulus duration by means of the first derivative of the sine function, using the power information and phase angle provided by the FFT. Mean *M50* peak locations were taken from the *M50* latency analysis. Figure II. 6a shows an alpha sine wave and its momentary slope. A negative driving alpha sine yields negative slopes (A), a positive driving alpha sine yields positive slopes (B). In order to analyze the effects of alpha slope on the *M50*, trial-based amplitudes were calculated in the time windows described above. In contrast to Barry et al. (2004) the polarity of the alpha wave cannot be interpreted in terms of cortical negativity/positivity since the interpretation of the neuromagnetic responses as ‘negative/positive’ is arbitrary. Therefore, no unit will be assigned to the slope. In order to exclude the simple explanation that the mean trial-based power-spectrum just represents the evoked responses, the mean power-spectrum was also calculated for the individual time courses based on the evoked magnetic fields after averaging across repetitions. If the average power-spectra of single trials are different from the spectra obtained from averaged trials, it follows that we have an indication that the measured alpha angle does not simply reflect stimulus-evoked activity.

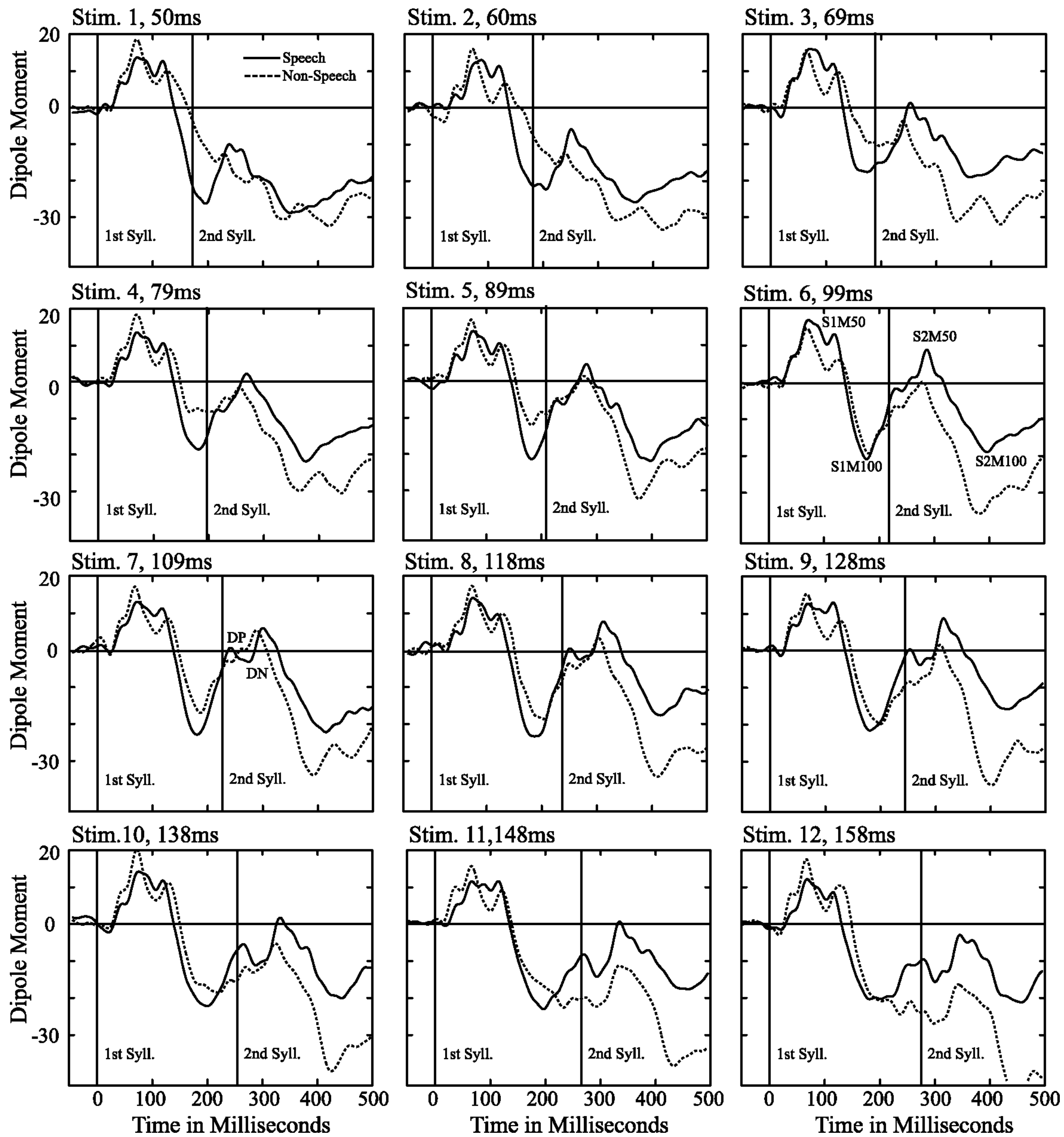

Figure II. 4: Time courses of the dipole moments for all durations (50ms - 158ms) in the speech (solid line) and nonspeech condition (dashed line), lowpass filtered at 40Hz for plotting. The onsets of the first and second syllable are indicated, as well as the labeling for the deflections in the respective syllables. Legend: S1M50, S1M100 = M50 and M100 to the first syllable; S2M50, S2M100 = M50 and M100 to the second syllable. DP, DN = additional deflection emerging for categorically long vowels.

Results

Latencies and dipole moments of evoked fields

Figure II. 4 shows the average time courses for all 12 speech and nonspeech stimuli. Left and right hemisphere were averaged for a better clarity of the presentation. Data were filtered with a 40Hz low pass for plotting. The time courses can be subdivided into two subsections that can be interpreted as responses to the first and the second syllable, respectively, as is shown e.g. for the stimulus 6 with a duration of 95ms. In most cases, each subsection shows a positive and a negative deflection. Considering the change of the time courses as a function of /a/ vowel duration (from stimulus 1 to 12), an additional peak between the responses to the first and the second syllable emerges at stimulus 7 (/a/ duration = 105ms), i.e., in case vowel duration reaches the category border towards the long category (Figure II. 4, stimulus 7).

In order to obtain quantitative parameters of MEG responses to the first and the second syllable, time windows were defined based on visual inspection of the time courses, considering also the latencies for the *M50* / *M100* reported in the literature (Alain et al., 1997; Eggermont & Ponton, 2002; Poeppel et al., 1996). The latencies for the *M50* deflections were looked for within a time window of 55ms – 85ms; latencies for the *M100* were looked for within a time window of 95ms – 200ms after the onset of each syllable. Regarding the latencies within these time windows, the location of the maximum / minimum amplitude was extracted.

The values in Table II. 1 show that the positive peaks were located in the time window of the *M50*; the negative peaks were located in the time window of the M100. Thus, *M50* / *M100* deflections to the first syllable were labeled *S1M50* and *S1M100*. The deflections to the second syllable were labeled *S2M50* and *S2M100*. The latencies of the two *S1M50* peaks show that the first peak, *S1M50a*, was timed to the onset of the first syllable; the second peak, *S1M50b* to the onset of the vowel / tone. The additional positive deflection between the peaks of *S1M50* and *S2M50* in the speech condition occurred only for durations longer than 105ms. In line with Kushnerenko et al. (2001), this deflection was labeled *DP* ‘duration sensitive positivity’. The negative deflection which separates *DP* from *S2M50* was labeled *DN* ‘duration sensitive negativity’.

Latencies in those windows which could have been affected by vowel duration, i.e. those later than *S1M50b* were submitted to a mixed-effects analysis with VOWELDURATION as a main effect. Random slopes for CONDITION, VOWELDURATION and HEMISPHERES accounted for repeated measures. For *DP/DN*, only values in the speech condition for durations longer than 105ms, i.e. those in the category long were taken into account. No significant effects of VOWELDURATION on the latencies in *S1M100*, *S2M50* and *S2M100* could be found. In *DP*, the linear slope was significantly positive, i.e. *DP* increased with increasing stimulus duration (slope = 0.9 nAmp/ 10ms, t = 16.3).

Dipole moments in the time windows

In Figure II. 3c&e, neuromagnetic responses evoked by test stimulus 12 (vowel/tone duration 158ms) are presented. Responses were low pass filtered at 40Hz for plotting. The positive *S1M50* and the negative *S1M100* and *S2M100* deflections showed typical magnetic field patterns (Hertrich et al., 2000; Mathiak et al., 1999). *S2M50* in both conditions and *DP* in the speech condition showed a topography resembling the *M100*. This was probably due to the sustained negativity after the first syllable which turned the subsequent EMFs negative. To assess statistically the effects of duration on the amplitudes, the following time windows were defined:

- *S1M50a*: 60ms – 80ms after first syllable onset,
- *S1M50b*: 105ms – 145ms after first syllable onset,
- *S1M100*: 170ms – 200ms after first syllable onset,
- *DP*: 30ms broad window in between of *S1M100* and *S2M50,*
- *DN*: varying width between *DP* and *S2M50,*
- *S2M50*: 55ms – 85ms after schwa onset,
- *S2M100*: 95ms – 125ms after schwa syllable onset.

Average dipole strength as a function of vowel duration within these time windows is shown in Figure II. 5a for the speech and in Figure II. 5b for the nonspeech condition. It can clearly be seen that amplitudes in *S1M50a* and *S1M50b* did not vary as a function of duration unlike in the remaining time windows. Amplitudes in time windows *S1M100*, *DP*, *DN*, *S2M50* and *S2M100* were submitted to a mixed-effects analysis. The results of the mixed-effects analysis are presented in Table II. 2, including only significant values and effects.

Although included in the model, no significant effects and interactions could be found regarding the factor HEMISPHERE. The following text refers to Table II. 2 and Figure II. 5.

	first syllable			**additional peaks**		**second syllable**	
what?	**1st pos.**	**2nd pos.**	**1st neg.**	**ad. pos.**	**ad. neg.**	**1st pos.**	**2nd neg.**
labels	**S1M50a**	**S1M50b**	**S1M100**	**DP**	**DN**	**S2M50**	**S2M100**
ns	66. 4 (SD = 7. 4)	125. 1 [68. 1] (SD = 7. 6)	173. 6 [116. 6] (SD = 10. 7)	---	---	67. 0 (SD = 8. 7)	168. 5 [111. 5] (SD = 11. 9)
sp	70. 3 (SD = 8. 1)	119. 1 [62. 1] (SD = 6. 4)	170. 6 [113. 6] (SD = 11. 1)	257. 5 [200. 5] (SD = 17. 3)	287. 3 [230. 3] (SD = 18. 2)	68. 4 (SD = 8. 2)	172. 1 [115. 1] (SD = 11. 8)

Table II. 1: Mean Latencies of the deflections in milliseconds as measured from the onset of the first / second syllable. Values in square brackets indicate measurements from the vowel / tone onset.
Legend: ns = nonspeech condition, sp = speech condition. pos. = positive, neg. = negative. ad. = additional.

		S1M100	***DP***	***DN***	***S2M50***	***S2M100***
a)	Condition (ns)	-5. 8 **			5. 5 ***	11. 2 ***
b)	VowelDuration (ns)	-3. 1 ***	-4. 1 ***	-2. 0 *	(1. 0)	(-4. 0)
c)	VowelDuration^2 (ns)	2. 3 ***	-5. 1 ***	-7. 4 ***	-6. 4 ***	-2. 1 *
d)	Cond (sp) ×VowDur	2. 9 ***	7. 8 ***	3. 4 ***	1. 5 *	6. 5 ***
e)	Cond×VowDur^2	-2. 1***	-2. 1 *			

Table II. 2: Results of the linear mixed-effects models in the time windows affected by duration. S1M100, S2M100 are negative deflections, i.e. an increase in strength is indicated by higher negative values; the others are positive; i.e. an increase is indicated by higher positive values.
*Main effects **(a-c)** show the mean amplitude (nAmp) and the mean slopes (nAmp / 10 ms) for the linear models (Amplitude as a function of duration) in the nonspeech (ns) condition. Interactions show **(d-e)** the slope differences between nonspeech and speech condition. Only significant effects and interactions are presented. Non-significant values are shown in (brackets) when needed for comparison. Asterisks indicate significance (* = t > 2, ** = t > 3, *** = t > 4).*

i) The significant main effect of CONDITION indicates that amplitudes were more negative/positive in *S1M100* and *S2M50* in the speech than in the nonspeech condition; *S2M100* amplitudes were more negative in the nonspeech than the speech condition.

ii) Significant interactions were found between vowel duration and the speech/nonspeech condition (CONDITION×VOWELDURATION and CONDITION×VOWELDURATION^2). Linear effects indicate a general trend of amplitude increase/decrease as a function of duration. Squared effects indicate whether amplitudes changed their increase/decrease pattern across the continuum. In *S1M100*, the negative linear effects differed between the conditions: In the case of speech, amplitudes stayed almost level, i.e. no effect of duration could be found whereas in the nonspeech condition, amplitudes increased as a function of duration. Regarding *DP*, both conditions showed a significant negative squared pattern, but differed in its steepness. The linear slope was negative in the nonspeech condition, but positive in the speech condition. In *DN*, the same holds true with regard to the linear trend, but speech and nonspeech did not differ with respect to the steepness of the squared pattern. In *S2M50*, speech and nonspeech showed an equal negative squared pattern, but differed in the slope of the linear trend: Amplitudes in the nonspeech condition stayed level, in speech they increased. Finally,, amplitudes in *S2M100* differed with respect to the linear slope of vowel duration: For speech it was positive (decreasing amplitudes as a function of vowel duration); for nonspeech it was negative (increasing amplitudes).

Missing S1M100 and S2M50 in the nonspeech condition

In the nonspeech condition, the *S1M100* and *S2M50* deflections were missing for the two shortest tone durations (50ms and 60ms in Figure II. 4, 1&2). The dipole moment was even more negative in *S2M50* than in *S1M100* for these durations (Figure II. 5b). As tested by a linear model, the linear effects of duration (50 – 99ms) yielded a negative slope in *S1M100* and a positive slope in *S2M50* (both $p < 0.01$). In the speech condition, no such interaction was observed. In Figure II. 5c, the difference between the dipole moment in *S1M100* and in *S2M50* is shown. For vowel durations shorter than 79ms the difference was significantly smaller than zero (paired Student's t-tests, $p < 0.001$), indicating that the two "syllables" did not elicit distinct *M50*/*M100* responses. By contrast, in the speech condition the difference was consistently positive across all 12 vowel durations, i.e., dipole strength in *S2M50* was always more positive than in *S1M100*.

The additional positive deflection between the two syllables for long vowels

Regarding the time course of dipole moment for vowel durations of 95ms and 105ms (Figure II. 4), it appears that in the speech condition a 'new' positive deflection emerged between *S1M50* and *S2M50,* which was absent for shorter vowel durations. No such deflection is visible in the nonspeech condition. The development of this additional deflection as a function of vowel duration is shown by the difference in dipole moment between *DP* and *DN* (Figure II. 5d). Negative values indicate that amplitudes in *DP* were more negative than in *DN* (= no distinct peak); positive/zero values indicate that amplitudes in *DP* were more positive than in *DN* (= distinct peak).

The differences were submitted to a mixed-effect analysis with a CONDITION×CATEGORY (short, long) interaction. Differences between *DP* and *DN* were pooled according to the short and long category(short category < boundary: 105.9ms < long category (Tomaschek et al., 2011)). Participants were included as random effects; CONDITION, CATEGORY and HEMISPHERES as random slopes accounting for repeated measures. CATEGORY missed a significant main effect ($\Delta_{(long/short)}$ = -1.8, t = -1.97); CONDITION yielded a significant main effect ($\Delta_{(ns/sp)}$ = -3.5, t = -3.0). The interaction which is shown in Figure II. 5e was significant ($\Delta_{(Cat\times Cond)}$ = 7.7, t = 7.9). The interaction plot shows that in the nonspeech condition, amplitudes in the *DP* window were more negative than in *DN* in both categories, i.e. no distinct peak emerged. In the speech condition, the amplitudes in the *DP* window were smaller in *DP* than in *DN* in category short, but larger/equal to in category long. This means that a new peak emerged.

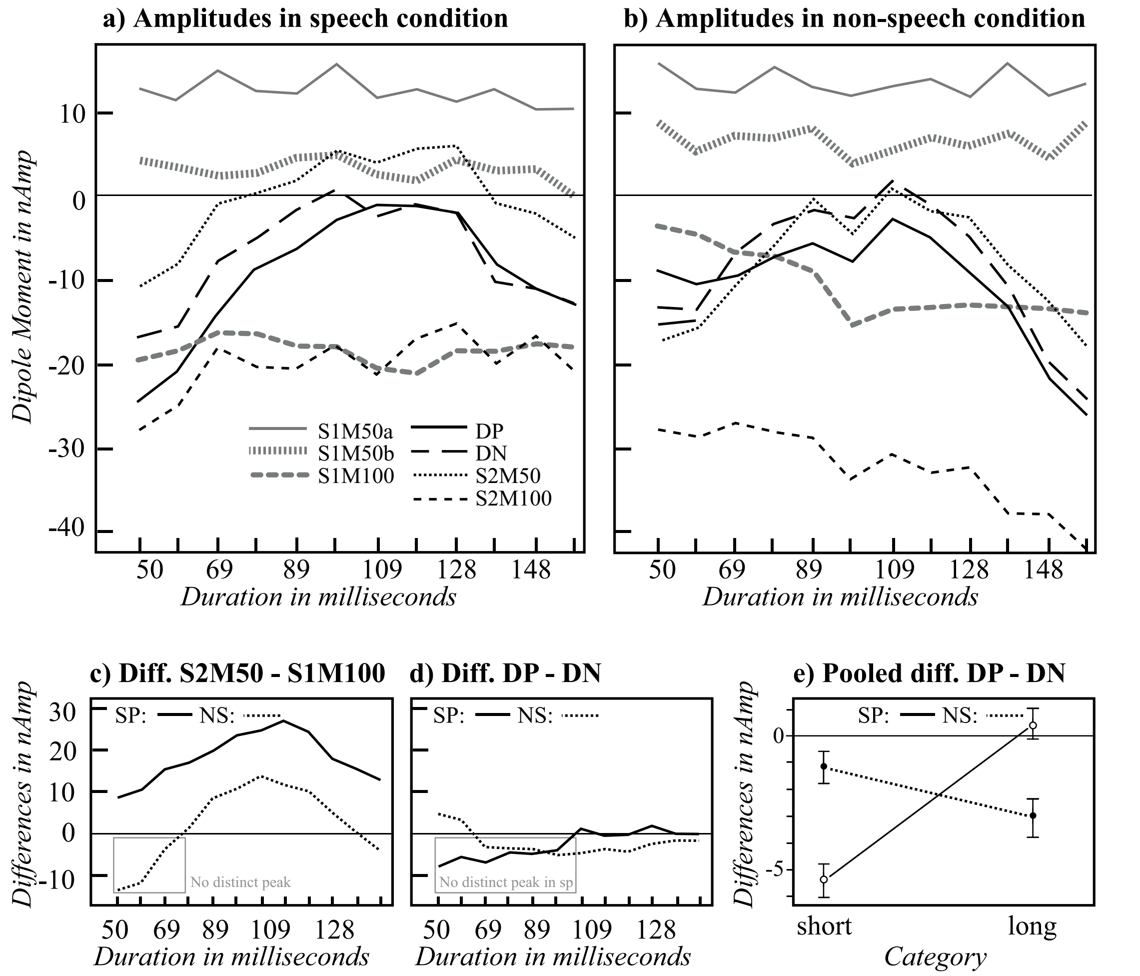

Figure II. 5: *Amplitudes as a function of duration in different time windows in* ***a)*** *the speech and* ***b)*** *the nonspeech condition.* ***c)*** *Amplitude differences in nAmp between the windows S2M50 and S1M100. Negative values indicate that both deflections are not distinct distinct in nonspeech.* ***d)*** *Differences between the windows DN and DP and* ***e)*** *interaction plot for the differences between DN and DP pooled according to the categories. Negative differences indicate that DP is NOT a distinct peak. At the category boundary (105. 9ms) the differences change from negative to negative/zero, indicating the emergence of DP.* ***Legend****: Diffs. = Differences, sp = speech, ns = nonspeech.*

Correlation between behavioral reaction time and S2M50

It was hypothesized that prototypical durations enhance the sensory gating of the duration processing as reflected by a suppressed amplitude of the *M50* in response to the second syllable (*S2M50)* concomitant with short response times during the identification test. By contrast, untypical durations could inhibit sensory gating and give rise to more effortful processing and longer response times. The relationship between the gating of the *M50* to the second syllable and identification response times was tested by means of a Pearson's correlation. A significant positive correlation was found between the identification response times from Tomaschek et al. (2011) and *S2M50* amplitudes which was averaged between the hemispheres ($R = 0.17$, $t = 2.5$, $p = 0.01$).

Alpha activity affecting M50 amplitudes

Figure II. 6b shows the mean power spectrum for the evoked responses after averaging across trials in the time domain. It does not resemble the trial-based spectrum (averaged in the spectral domain; Figure II. 6c). The first shows a peak in the theta band (4-8Hz), but lacks a distinct peak in the alpha band (8-10Hz) and a plateau in the beta band (13-20Hz). Taking these differences into account it was assumed that the trial-based spectrum reflects background and stimulus-induced activity rather than time-locked responses to stimulus onsets. A mixed-effects analysis was performed with the power of the frequency band centered at 11.44Hz (peak in the grand average spectrum) as the response variable and CONDITION, VOWELDURATION and HEMISPHERE as fixed effects and as random slopes to account for repeated measures. Since the model found only a significant main effect of CONDITION ($\Delta_{(ns/sp)} = 0.02$, $t = 2.6$), the grand average over trials, participants, hemispheres and durations is presented in Figure II. 6c. No significant interactions could be found.

The following analysis concentrates on the distinct peak at 11.44Hz which was considered as representative for the alpha band Figures 6d&e show the scatter plots of *M50* amplitudes in response to the first and to the second syllable as a function of alpha slopes, calculated for the time points of the peak of the *M50* in response to the first and second syllable (see Table II. 1 for mean peak latencies). *M50* amplitudes were subjected to a mixed-effect analysis using ALPHASLOPES as well as CONDITION as fixed effects and random slopes. *M50* amplitudes were significantly enhanced during negative and attenuated during positive alpha slopes. The strength of the effect differed between the syllables with the *S2M50* amplitudes being stronger affected (model slope = -41.6, t = -20.8) than *S1M50* amplitudes (model slope = -4.1, t = -2.3). A significant difference between the conditions was found only for the *S2M50* amplitudes ($\Delta_{(sp/ns)} = 6.2$, $t = 4.8$).

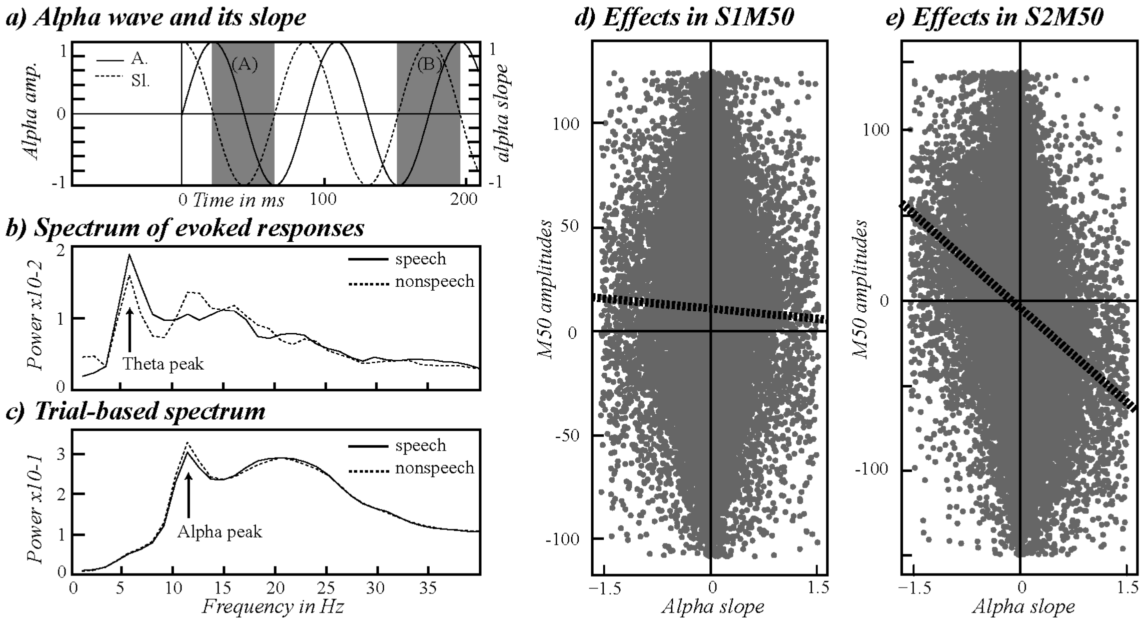

Figure II. 6: **a)** *Illustration of an alpha sine wave (A., solid line) as a function of time and its first derivative that reflects its slope (Sl., dashed line). In area A, the slope of the alpha wave is negative, in area B it is positive.* ***b)*** *Mean power-spectrum of the individual evoked responses which were averaged over the trials.* ***c)*** *Mean power-spectrum of the individual responses calculated on a trial basis.* ***d)*** *Scatter plot of the M50 amplitudes as a function of the alpha slope in the first syllable.* ***e)*** *Scatter plot of the M50 amplitudes as a function of the alpha slope in the second syllable. The dashed line represents the regression line. A negative alpha slope enhances, a positive alpha slope attenuates M50 amplitudes.*

Discussion

The present study investigated the way varying vowel durations are processed by native speakers of German, a language which uses vowel duration in a contrastive way. Disyllabic nonsense words /tatə/ and structure matched nonspeech stimuli were used in which vowel/tone duration in the first syllable was varied stepwise from short to long. The MEG results can be summarized to four points:

1) Varying vowel duration was reflected by amplitude changes of the *M50* in response to the second syllable. It increased for average vowel durations in contrast to vowel durations from the edges of the continuum.

2) Unlike in the nonspeech condition, an additional deflection (*DP*/*DN*) between the first and second syllable was observed for categorically long vowel durations in the speech condition.

3) Regarding short durations in the nonspeech condition, the responses to the two syllables were not as distinct as in the speech condition: The *M100* to the first and the *M50* to the second syllable did not appear as separate deflections.

4) Unlike in the speech condition, the *M100* to the second syllable continuously increased in amplitude along with longer tone duration in the nonspeech condition.

Latencies and deflection classification

Regarding peak latencies (Table II. 1), the two peaks of the *M50* to the first syllable represented onsets of the consonant/noise and the vowel/tone, respectively (Figure II. 4). The first negative peak was elicited ~110ms after the onset of the vowel and is thus the *M100* to the first syllable. Generally, a positive deflection elicited ~200ms after stimulus onset can be considered as the *P200/M200* (MEG: Ackermann, Lutzenberger & Hertrich, 1999; EEG: Čeponienė et al., 2008; Hisagi et al., 2010; Menning, Imaizumi, Zwitserlood & Pantev, 2002). In the present study, however, it was found that the second positive deflection (in the category short, third positive deflection in category long) was cued to the onset of the second syllable at 68.4ms in the speech condition and at 67ms in the nonspeech condition (measured from the acoustic onset of the second syllable).

These results match nicely the time window for the *M50* of other studies (Eggermont et al., 2002; Poeppel et al., 1996; Roberts & Poeppel, 1996). Therefore, the present paper considers the second positive deflection not as a *P200m* but as an *M50* to the second syllable. According to this logic, the second negative deflection represents the *M100*, which was elicited with a latency of 115.1ms after the onset of the second vowel and 111.1ms after the onset of the second tone. Regarding the differences in peak latency between the conditions, these could be attributed to different computational efforts needed to process speech and nonspeech signals (see Eulitz et al. (1995) and Tiitinen, Sivonen, Alku, Virtanen & Näätänen (1999) for latency differences). Additionally, the highest formant frequency in the speech signal was 4500Hz, whereas the highest formant frequency in the nonspeech signal was 1500Hz. Since Roberts et al. (2004a) have shown that M100 latencies increase with decreasing formant frequency, present latency differences might be attributed to the spectral differences between the conditions.

Amplitudes

The present study found stronger amplitudes of *M100* to the first syllable in the speech than in the nonspeech condition (Figure II. 5). This increase could indicate higher processing activities which possibly reflected the extraction of vowel identity from formant information (Obleser et al., 2004; Roberts et al., 2004a; Scharinger et al., 2011). Considering the phonotactic restrictions of German (i.e. no short vowels in open stressed syllables, see Becker (1998)), vowel duration was seen as the distance measured between the vowel's onset and the onset of the second syllable. Thus, the *M50* to the onset of the second syllable was hypothesized to reflect some aspects of the measured vowel duration. *M50* amplitude was stronger in the speech than in the nonspeech condition. The increase could also be attributed to the higher processing activity needed for top-down phonological categorization, which is in contrast to the nonspeech condition where the tone duration of the first syllable could not be categorized.

However, the evoked responses to the second syllable cannot be completely dissociated from the responses to the first syllable. In the literature, it was found that longer stimuli produce stronger *N100*/*M100* deflections as well as stronger sustained negativities. The increase in neural activity is probably due to the accumulation of neural pulses in the neurons responsible for duration measurement (EEG: Alain et al. (1997); Kushnerenko et al. (2001); Paavilainen, Jiang, Lavikainen & Näätänen (1993), Bendixen, Grimm & Schröger (2005); MEG: N'Diaye, Ragot, Garnero & Pouthas (2004); intracellular recordings: Leon et al., (2003)). Such duration effects might have affected the later components in the present study. Accordingly, *M50* deflections to the second syllable were attenuated in contrast to *M50* deflections to the first syllable in both conditions. In the nonspeech condition, the *M100* in response to the second syllable increased significantly with longer durations reflecting the effects of auditory duration processing (Figure II. 5b). The absence of such an increase in the case of speech stimuli, however, could indicate that the integration of the first syllable was terminated by segregation mechanisms as found by and Luo et al. (2007); Yabe et al. (2001b). This would make the system free for subsequent speech events which is reflected by the equal strength of the *M100* to the first and second syllable (Figure II. 4a).

Syllable and duration processing: Differences between speech and nonspeech

In the literature, it has been shown that the processing of duration is lateralized depending on whether the contrast is phonological (left hemisphere) (Kasai et al., 2001; Minagawa-Kawai et al., 2005) or non-phonological, i.e. acoustic (right hemisphere) (Kasai et al., 2001; Kirmse, Ylinen, Tervaniemi, Vainio, Schröger & Jacobsen, 2008). This kind of lateralization difference was not observed between the speech and nonspeech conditions in the present study. Regarding the processing of duration in the speech condition, it was hypothesized that the categorical assignment of vowel duration would affect the amplitude of the *M50* to the second syllable. Unlike atypical vowel durations at the category boundary, phonologically correct vowel durations should have a gating effect on the perception of the subsequent syllable. In line with the P50-suppression paradigm, gating should result in decreased *M50* amplitudes (Cadenhead et al., 2000; Light et al., 2001; Mathiak et al., 2011). The P50-suppression effect describes the decrease of neural activity – gating – when known stimuli are processed, in contrast to the processing of unknown stimuli (Cadenhead et al., 2000; Light et al., 2001; Lijffijt et al., 2009; Mathiak et al., 2011). Although such a result was expected only for the speech condition, it was found in both conditions (Figure II. 5a&b). However, the conditions differed in two points, to which we now turn our attention.

Amplitude increase and decrease of the M50 to the second syllable as a function of duration

The amplitude of the M50 to the second syllable showed an increasing and decreasing i.e. squared pattern, as a function of vowel duration in both conditions. Two explanations are possible for this finding.

1) Unlike in the speech condition, the *M50* to the second syllable in the nonspeech condition did not show a distinct *M50* peak at the left edge of the continuum (Figure II. 4, stimuli 1-3). *M50* amplitudes probably decreased since they were cancelled by the negative *M100* to the first syllable (Figure II. 5b). Further, while the amplitudes of the *M100* to second syllable stayed level in the speech condition (Figure II. 5a), they increased in strength in the nonspeech condition with increasing tone duration (Figure II. 5b). This increase possibly affected the amplitude of the *M50*. Taking this into account, the squared pattern of the *M50* to

the second syllable might be a secondary effect of overlapping responses to the two syllables in the nonspeech condition whereas in the speech condition the results could be regarded as a P50-suppression effect. Prototypical durations in category short may have gated the second syllable, decreasing *M50* amplitudes; untypical instances did not match the phonological top-down expectancy resulting in enhanced amplitudes.

2) Since so far P50-suppression effects have been shown only for very short stimuli and not for longer stimuli, a different explanation appears to be more likely. The present analysis found alpha band activity to significantly affect *M50* amplitudes (Figure II. 6c-e). Negative alpha slopes, calculated for the time points of the peak of the *M50* in response to the first and second syllable (see Table II. 1), were associated with an enhancement, positive alpha slopes with an attenuation of the *M50* amplitudes. Amplitude changes of evoked activity due to alpha activity and its slope have been already shown by Barry et al. (2004); and Fellinger et al. (2011). This effect was stronger in the second than in the first syllable, which could be explained by the assumption that the first syllable imposed an entrainment upon the alpha phase. The attenuation/enhancement effect could have been responsible for the bell-shaped pattern of *M50* amplitudes in response to the second syllable as a function of duration (Figure II. 5a) because the durational difference between short and long vowels of ca. 100ms approximately matches the period duration of alpha activity. The *M50* peak coincided with different alpha slopes depending on the varying vowel duration. In line with Barry et al. (2004), the *M50* was enhanced when it coincided with a negative alpha slope (A in Figure II. 6a); it was attenuated when it coincided with a positive alpha slope (B in Figure II. 6a). Since brain oscillatory dynamics generally entrain to rhythmic auditory stimulation (Will et al., 2007) and are flexible with regard to the “beats-per-minute” of the stimulation (Cummins, 2009; Nozaradan et al., 2011; Zion Golumbic, Poeppel & Schroeder, 2012), such neural oscillations would not have a fixed frequency but be adjustable to temporal parameters like speech tempo (Schröger, 2005; Luo & Poeppel, 2007, Hertrich et al., 2012). Possible triggers could be the onset responses to each syllable (Hertrich et al., 2012) and/or the amplitude modulations of the speech signal (Abrams et al., 2008; Peelle, Gross & Davis, 2012).

It was found that *M50* amplitudes as a function of vowel duration (Figure II. 5a) correlated significantly with the identification response times from Tomaschek et al. (2011) (Figure II. 1d). This correlation supports a close relationship between the *M50* deflection to the second syllable and the categorization process with respect to vowel duration (see Lijffijt et al. (2009) who found a positive correlation between sensory gating and response times). Considering the present paradigm, however, a definitive explanation for the squared patterns in both conditions cannot be provided.

The lack of M100 to the first and M50 to the second syllable for short durations in the nonspeech condition

M100 deflection to the first and *M50* to the second syllable became fully distinct only for *tone* durations longer than 79ms (see Figure II. 4, stimuli 1-4). This finding could be the result of the temporal structure of the nonspeech stimulus. Adding together the durations of the noise, the tone and the occlusion, a tone duration of 79ms corresponds to the second syllable's onset time of 211ms after stimulus onset. 211ms just exceeded the full long temporal window of integration of 200ms for acoustic signals (Boemio et al., 2005; Poeppel, 2003). It is possible that, for short durations, the onset of the second nonspeech syllable fell into the first long temporal window of integration. Since all events within the integration window are treated to a certain extent as a unitary stream (Yabe, Koyama, Kakigi, Gunji, Tervaniemi, Sato & Kaneko, 2001a; Yabe et al., 2001b), the auditory system may not have registered the onset of the second syllable as an independent event, which would explain why *M100* and *M50* deflections were not visible. When the onset of the second part fell into a new temporal window, the deflections became distinct (see Figure II. 4, stimuli 5 – 12). In contrast to the nonspeech condition, *M100* to the first and *M50* to the second syllable were distinct for all speech stimuli, even for the shortest vowel durations, indicating that the system registered the onset of the second syllable. There are two possible explanations for this pattern, which probably interact with each other:

1) In the nonspeech condition, the onset of the second syllable fell into another temporal window of integration than the first syllable. This would imply that the size of the temporal windows of integration was different between speech and nonspeech condition, in line with Nusbaum & Henly (1992) who proposed adaptive integration windows modified by the lexical status of the stimulus. Yabe, Tervaniemi, Reinikainen & Näätänen (1997); Yabe, Tervaniemi, Sinkkonen, Huotilainen, Ilmoniemi & Näätänen (1998) showed that the minimal size of the integration window was 150ms. Luo et al. (2007) and Peelle et al. (2012) also argued that theta oscillations (4-8Hz) function as the integration mechanism. The specific size of the integration window could depend on the frequency of the oscillations which are adjusted to temporal parameters like speech tempo (see Abrams et al. (2008); Ahissar et al. (2001); Deng & Srinivasan (2010); and Hertrich et al. (2012); Luo et al. (2007) for speech tempo effects). The amount of temporal windows of integration probably represents the

perceptual counterpart of phonological timing units, as already proposed by Poeppel (2003). Phonological theory provides rules of how many timing units have to be occupied by a single phonetic segment (vowel/consonant) in order to be categorized as short or long. The perceptual system translates the timing units into the amount of integrated temporal windows (see below).

2) Yabe et al. (2001b) have shown that the segregation mechanism takes precedence over temporal integration when a signal containing large spectral changes is processed. Regarding the present study, this would not imply that the integration windows were different between the conditions, but rather that the effects observed were a result of the hierarchy of segregation and integration. In the speech condition, segregation took precedence over integration. Due to the drastic spectral change the temporal integration was terminated in the speech condition after the onset of the second speech syllable. Hence, the second syllable elicited a distinct *M50* response. In the nonspeech condition, integration took precedence over segregation; the onset of the second syllable was integrated into the ongoing temporal window. This is probably why it did not elicit a distinct response.

The additional DP / DN deflection: Phonotactic processing and phonological considerations: The neural mechanisms reflected by DP

DP/*DN* occurred between *M100* to the first and the *M50* to the second syllable when vowel duration exceeded the category boundary. Three explanations are presented considering the neural mechanisms reflected by the additional deflection.

1) *DP* has a mean latency of 191ms measured from the onset of the /a/ vowel (Table II. 1), fitting into the time window of the *P200/M00*. Its latency correlated with vowel duration. In line with Alain et al. (1997) who showed a positive correlation between the amplitude of the electroencephalographic *P200* and stimulus duration, *DP* might be interpreted as *P200m*. This might indicate the processing of the speech stimulus at the level of the sensory memory (Menning et al., 2000; Tremblay, Shahin, Picton & Ross, 2009). 2) Regarding the different temporal structure of the stimuli and the effects of neural entrainment on the latency of the deflections (see Barry et al. (2004)), *DP* might have been masked by the *M50* to the second syllable for short vowel durations and became visible for longer durations. It seemed to emerge for vowel durations just “before” the category boundary (Figure II. 4, Stim. 5-6). Other studies have shown that *DP* coincides with the time window of phonological processing in speech signals (Connolly et al., 1994; Kujala et al., 2004; Newman et al., 2003). Thus, its occurences in the speech condition in contrast to the nonspeech condition might have

mirrored a process which compared the representation of measured duration with predictions derived from the phonological lexicon.

3) *P200/M200* has been observed in speech as well as in nonspeech conditions (Edmonds et al., 2010; Eulitz et al., 1995; Houde, Nagarajan, Sekihara & Merzenich, 2002; Kuriki et al., 2006; Tiitinen et al., 1999), which is why the present lack of an *M200* deflection in the nonspeech condition is surprising. This leads to the hypothesis that *DP* could not be an *M200* component that was masked or cancelled by other deflections. Instead, this complex could be seen as a magnetic field component reflecting the categorical status of additional vowel duration.

This view is supported by the statistical assessment of the deflection which showed that *DP* emerged abruptly near the category boundary across the durational continuum (Figure II. 5d&e). It is possible that this deflection mirrors the "registration" of additional duration by specialized neurons, which are relevant for phonological processing. Such category sensitive neurons have been shown by Leon et al. (2003) by means of intracellular recordings in trained macaque monkeys which had to categorize varying visual signal durations to as 'short' or 'long'. They found that cells in the posterior parietal cortex fire depending on the 'category' of the presented duration.

The identification response times from Tomaschek et al. (2011) also speak in favor of this interpretation: The information that vowel duration was long seemed to occur already during the perception of the vowel suggesting that the processing of categorically long vowels depends on a critical duration threshold in terms of time slots. Otherwise one could not explain the finding that, in the identification test, response times were faster in category long than in category short when measured from the offset of the stimulus.

The implications of DP for the phonological vowel system

The present study investigated whether the theoretical phonological representation of vowel duration would be mirrored in the neural correlates of German native speakers. Taking the present findings into account, some phonological implications can be made. Phonological formalisms (Becker, 1998; Zec, 2007) represent vowel duration with a short vowel being connected to only a single syllabic slot (Figure II. 1b), a long vowel to two syllabic slots (Figure II. 1a). Thus, the occurrence of *DP* in words with long vowels can be seen to represent two aspects: 1) Luo et al. (2007) found theta-oscillations during speech perception and have argued that these represent the temporal integration mechanism at the syllabic level

(Poeppel, 2003). Correspondingly, a peak in the theta-band was found for the present evoked responses (Figure II. 6b), imposing that the theta-cycles could as well be understood in terms of perceptual time slots. As long vowels are covered by two theta-cycles, *DP* represents the integration of the vowel by two time slots instead of by one time slot. 2) In phonological terms, *DP* represents the additional intra-syllabic constituent of long vowels in contrast to short vowels (Figure II. 1a&b), which occurred when the second slot was opened during auditory processing. The present results therefore support the theory that vowel duration in German is perceived categorically, i.e. vowel duration is attributed to either the short or the long category (Tomaschek et al., 2011).

One could argue that *DP*, reflecting additional neural activity, characterizes the long vowel as "marked" in contrast to the short vowel. A marked phoneme/segment needs 'more' feature specifications than an unmarked phoneme/segment which would result in 'more' neural activity. This was found by Eulitz et al. (2007) who showed prolonged beta-band desynchronization when marked in contrast to unmarked vowels were processed (round vs. non-round vowels). However, the current phonological theory of German does not distinguish between short and long vowels in stressed syllables with respect to prosodic prominence. Since in German the minimal nucleus of a stressed syllable consists of two constituents which need to be occupied (Becker, 1998; Hall, 1992), a long vowel which occupies two constituents is equivalent to a 'short vowel + consonant' sequence (Figure II. 1a&b). But in words with a single intervocalic consonant, this results in ambisyllabicity. Accordingly, it follows that German vowel duration is not used as a marker for syllable stress, because in certain circumstances short stressed vowels can be shorter than unstressed vowels (Heike, 1970; Hirata, 2004). Moreover, since in general stress-timed languages use vowel duration as a salient stress marker, stressed syllables with short vowels can be considered as "marked" or more complex in some respect. Finally, *DP* might be interpreted in terms of syllable weight. However, current German phonology regards CVC and CVV syllables to have equal weight (Féry, 1998; Knaus & Domahs, 2009).

At the feature level, [-long] or [+long] specify whether the vowel occupies one or two syllabic constituents, respectively. The phonological specification of phonetic duration as [-/+long] is necessary not only to differentiate short from long vowels, but also, at least as a pre-processing step, to differentiate [+high] from [-high] vowels: The short lax [ʊ] and [ɪ] and the long tense [oː] and [eː] overlap in their spectral distribution to such an extent that the primary cue to differentiate between them seems to be duration (Bennett, 1968; Hoffmann, 2011;

Weiss, 1974). Because of this, phonological categorization within the German vowel system requires a durational specification, even for the correct assignment of phonological vowel quality with regard to spectral characteristics of the acoustic signal. *DP's* emergence after the category boundary. This indicates the setting of the feature [+long]. By contrast, [-long] is set when the onset of the second syllable occurs before the setting of [+long]. Since stressed syllables require two slots to be filled in the case of short vowel duration, the following consonant must be taken to fill the second slot in the case of short vowels. It thus follows that in terms of syllabic stress, the processing of short vowels could be considered to be more complex as compared to long vowels which represent a kind of "default" pattern for stressed syllables. However, the present study did not find any signs of additional effort related to the ambisyllabicity of the second consonant of the stimuli in the case of short vowels. This is consistent with findings, in speech production where ambisyllabic and non-ambisyllabic consonants are produced similarly in German (Hertrich & Ackermann, 1997) even though they differ in orthography.

The present study concentrated on the perception of linearly changing duration in a speech and nonspeech condition by speakers of German, a language which uses vowel length in a contrastive way. This paradigm could easily be extended in future studies to speakers of languages without contrastive duration (e.g. Polish, Russian) to compare directly how an auditory system not acquainted with phonological length copes with the phonologically induced changes in phonetic vowel duration.

Conclusion

Study II investigated the influence of vowel duration in structurally matched speech and nonspeech signals on event-related magnetic fields (EMF) and how a phonological category boundary would be reflected. In general, each syllable in the speech and nonspeech signal produced an *M50*/*M100* deflection. The various effects of different vowel durations gave rise to the following conclusions:

1) In contrast to speech, responses to short nonspeech stimuli did not show distinct M50/M100 responses to the two syllables, indicating that speech induced a perceptual mode with more flexible temporal integration windows.

2) In the speech condition, an additional magnetic-field component (duration sensitive positivity '*DP*') was found in the case of phonologically "long" vowels, which might represent a neural correlate of an additional phonological constituent in the syllable structure.

3) Behavioral response times in the identification task correlated with the *M50* amplitude in response to the second syllable, confirming the interpretation of the *M50* as an aspect of sensory gating.

Taken together, the temporal structure of speech and its phonological relevance seems to be directly reflected in the time course of evoked MEG activity. Accordingly, the present study contributes to our understanding of the neural processes underlying phonological processing.

Study III: Perceptual characteristics of the length-quality interaction

Introduction

The German vowel system

The German vowel system shows a complex structure based on an interaction between vowel duration and vowel centralization. In contrast to the /a/ - /a:/ distinction, where the phonetic difference is more or less restricted to vowel duration (Hoffmann, 2011; Wiese, 1988), quantitatively contrasting mid and high vowels exhibit an interaction between vowel duration and spectral quality (Hall, 1992). In the vowel triangle, short vowels are more centralized, long vowels are more peripheral (Figure III. 2a). The present study investigated the phonological status of these acoustic cues in German by assessing their psychoacoustic characteristics. In terms of Massaro et al. (1983) the question arises which cue will "... result in [...] discrete changes along a perceptual dimension" and therefore can be regarded as a lexicalized phonological feature.

In phonological terms, the short-long distinction is called *length* contrast, the peripheral-central distinction *quality* contrast, reflected by [+/- long] and [+/- tense] features, respectively. Concerning the distribution of these cues, the length contrast is limited to stressed syllables whereas in unstressed syllables it is largely non-existent. De-stressed instances of non-low vowels may still exhibit some differences in centralization. This distribution gave rise to the assumption that the differences, irrespective of the length aspect, refer to a lexical specification of *quality* (Moulton, 1962b; Reis, 1974). Thus, the primary lexicalized feature would be [+/- tense]. However, phonotactic restrictions due to *length* (Becker, 1998; Hall, 1992), the allophonic distribution of (physically) short [+tense] and [-tense] vowels in unstressed syllables (Ramers, 1988), as well as the exclusively durational /a/-/a:/ contrast (Hoffmann, 2011) suggest that *length* [+/- long] seems to be the lexicalized feature. Thus, in order to avoid a phonological over-specification, length-associated differences in vowel quality should be considered as secondary modifications that do not rely on lexical phonological specification. As such, tenseness should be regarded as allophonic (Becker, 1998).

Previous studies on the primary acoustic cue

So far, no consensus was found in the literature which of the two acoustic cues in the German vowel system – duration and centralization – is used primarily and can thus be seen as the lexicalized feature. The findings of Bennett (1968), Weiss (1974) and Sendlmeier (1981) speak in favor of vowel quality. In a series of identification tests, using disyllabic real words, they have shown that non-low vowels seem to be categorized mainly on the basis of quality. Nevertheless, vowel length cannot play a completely subordinate role, since the distinction between /a/-/a:/ is performed on the basis of duration. As such, Heike (1970) has found that the location of the category boundary was supported by a test which could be interpreted as a goodness rating. This finding was reproduced by Study I where it has also been shown that the perception of the /a/-/a:/ contrast matches the criteria of categorical perception: it has a sharp boundary between category short and long which is concomitant with a increase in discrimination sensitivity (Liberman et al., 1957). The change in phonetic quality due to shortening affects the position of *high short lax* vowels to such an extent that they overlap strongly in formant frequencies with *mid long tense* vowels (Hoffmann, 2011). Although the distinction between these vowels is not based on a primary phonological distinction, since it involves two features – vowel height and vowel length –, the *mid long tense* vowel /oː/ and the *high short lax* vowel /ʊ/ are differentiated mainly on the basis of duration (Bennett, 1968; Heike, 1970; Weiss, 1974).

Psychoacoustic characteristics of a phonological contrast

It has been shown that the existence of a phonological contrast affects the psychoacoustic characteristics along the acoustic continuum when it ranges from one end of the contrast to the other. The juxtaposed phonological categories decrease the discrimination performance within the category, but increase it next to the category boundary. This interaction between the category boundary and the discrimination ability has been described as Categorical Perception (Goldstone et al., 2009; Liberman et al., 1957). Ideally, no discrimination is possible within the category, only across-boundary items can be discriminated. However, this traditional concept of categorical perception has been questioned in several studies since within-categorical discrimination could have been shown, even for stop consonants which were argued to be perceived only in a categorical way (Carney et al., 1977; Hanson, 1977). Furthermore, it has been shown that the traditional comparison of identification and discrimination performances cannot reveal how speech is categorized, since discrimination results are influenced by the test person's "subjective criteria" (Ylinen et al., 2005) during the discrimination task (Gerrits et al., 2004; Lively & Pisoni, 1997; Lotto et al., 1998; Massaro et al., 1983; Schouten et al., 2003b). However, in line with Ylinen et al. (2005), the analysis of the discrimination performance along the continuum, especially at the category boundary, can be used to show whether a phonological category exists in the first place. As Ylinen et al. (2005) stated it appropriately: “[...] discrimination performance at the category boundary [should be improved], if the categories are accessed”. In this sense, Ylinen et al. (2005) have shown that the phonological status of vowel duration in Finnish produces an increase in discrimination performance at the category boundary in native speakers of Finnish, but not in native speakers of Russian (which do not have phonological vowel length). Hence, unlike non-phonological contrasts, phonological contrasts will increase the discrimination performance at the category boundary (Flege et al., 1994; Miyawaki et al., 1975; Pisoni et al., 1974b; Stevens et al., 1978; Xu et al., 2006; Ylinen et al., 2005).

The present study

The present study is seen as a contribution to the discussion about which acoustic cue is used primarily to distinguish centralized short from peripheral long vowels in the stressed syllable. For this, an identification test to find the category boundaries as well as spectral and temporal discrimination tasks to find the discrimination performance were performed. A fine-graded two-dimensional continuum was used, covering F1 frequencies of peripheral and centralized /u/ as well as ranging from short duration to long duration. Since long instances of the short [ʊ] are perceived as [oː] (Bennett, 1968; Heike, 1970; Weiss, 1974), additionally to the phonological features which distinguish between short and long vowels ([+/- long] and [+/- tense]), a third feature is needed to distinguish between /u/ and /o/ in terms of vowel height (i.e. [+/- high]) (Hall, 1992). Table III. 1 presents the specification of all four tested vowels.

		Length	
Height	**Tenseness**	**[- long]**	**[+ long]**
[+ high]	**[+ tense]**		/u:/
	[- tense]	/ʊ/	
[- high]	**[+ tense]**		/o:/
	[- tense]	/ɔ/	

Table III. 1: Phonological feature representation of the short-long (length), peripheral-central (tenseness) and high-mid (height) vowel contrast.

The present reasoning starts with a fully specified model, where each vowel differentiated by [+/- high], [+/- tense] and [+/- long] should be represented by a phonetic prototype (Diesch et al., 1999; Iverson et al., 1993; Kuhl, 1991; Kuhl, 2004). The phonological length contrast is associated to a phonetic durational distinction, while the phonological distinction in terms of [high] and the one in terms of [tense] are relevant to changes of the first formant F1. The prototypes are expected to be distributed in a two-dimensional vowel space like in Figure III. 1a. This distribution is based on the following assumptions: i) The phonological length distinction [+/- long] leads to a duration distinction that is comparable to the [ʊ,uː] contrast and the [ɔ,oː] contrast; ii) as known from productional studies (Hoffmann, 2011), a [+high] prototype is systematically lower in F1 than a [-high] prototype and iii) a [-tense] prototype is systematically higher in F1 than its [+tense] counterpart. The line in Figure III. 1a represents the expected category boundary between u-vowels and o-vowels, i.e. the distinction in terms of [+/-high].

Concerning the course of the category boundaries between short central and long peripheral vowels, there are three hypothetical constellations, depending on the perceptual relevance of duration and of the first formant. i) If quality was the only cue taken into account to distinguish [ʊ] from [uː] and [ɔ] from [oː], then the vowel space should be structured as shown in Figure III. 1b. ii) Different results are expected in the inverse case, when duration is the primary cue as shown in Figure III. 1c. iii) Finally, if both temporal and spectral cues are relevant, a division of the vowel space is expected as shown in Figure III. 1d. Concerning the discrimination performance, if the short-central vs. long-peripheral contrast is based primarily on *length*, the perception of duration should be affected by duration categories and, thus, exhibit a maximum in temporal discrimination at the short-long-boundary. Such a result was already shown for the /a/-/a:/ contrast in study I. If the /ʊ/-/uː/ contrast is based primarily on qualitative differences, spectral discrimination should exhibit a maximum at the central-peripheral boundary.

Methods

Subjects

20 native speakers of German were recruited and 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.

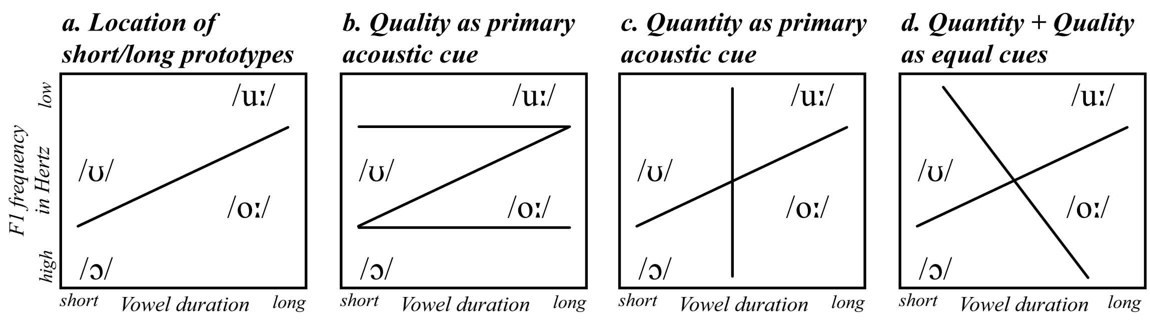

Figure III. 1: *Hypothetical division of the two-dimensional vowel space with the respective primary cue.*

Stimuli

The stimulus material for each test, instances of the trochaic nonsense words /gubə/, /gu:bə/, /gobə/, /go:bə/, was drawn from a two-dimensional spectro-temporal continuum between /u/ – /o/ and short – long. Vowel quality between /u/ and /o/ was changed in the F1 dimension (Hose, Langner & Scheich, 1983); vowel length was changed in the duration dimension. 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 in German (Féry, 1996). Thirdly, 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; see Gottfried et al. (1990) and Hirata (2004) for speech tempo effects. Although there is a statistical tendency in German to 'avoid' voiced plosives after short vowels (Lenerz, 2000), the consonant /b/ was used in order to increase the phonological distance from existing lexical words such as, e.g. /guːtə/ or /goːtə/. The voiced plosive affects the perception of the preceding vowel insofar as it slightly changes the location of the category boundary toward longer durations (see Lehtonen (1970, P. 80) for effects of consonant voicing on vowels.)

In order to exclude any accidental variations of natural speech, these items were synthesized by concatenating noise segments with vowel segments. The principal validity of synthesized stimuli regarding phonetic/phonological research questions has been shown in several studies (Ainsworth, 1972; Blomert et al., 2004; Bunton et al., 2009; Sams et al., 1990). Stimuli were synthesized using the formant synthesizer of Hertrich et al. (1999); 2007) which computes formants as amplitude- and phase-modulated sinusoids on the basis of the F0 frequency. Aperiodic speech segments, which represent fricatives or stop consonant bursts, are generated by adding a random parameter into the formants' phase progression. Formant values and the temporal aspects of the stimuli were approximately derived from a real recording of the first author, analyzed with PRAAT (Boersma et al., 2009). Segment durations and the vowel formant frequencies of the endpoints of the continuum can be seen in Figure III. 2b.

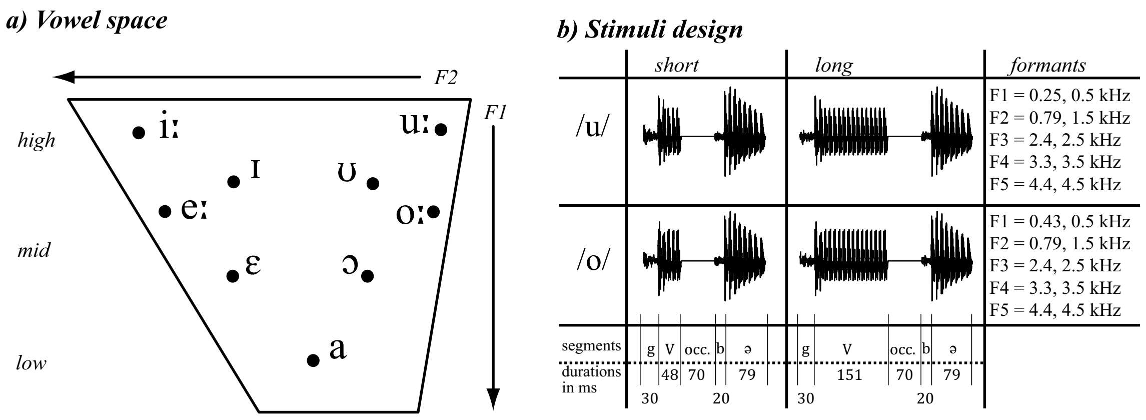

Figure III. 2: ***a)*** *Schematics of the locations short (V) and long (V:) vowels in the vowel space. Short vowels are lower and more central than their long instances.* ***b)*** *Stimulus design indicating segment durations and formant frequencies of the respective vowel (V) and the schwa (first and second number). The endpoints of the tested spectro-temporal continuum are presented.*

Formant transitions were implemented in the stimuli according to Strange (1987) and Strange et al. (1998) within the initial 20 ms of each vowel. Importantly, no dynamics with regard to articulation such as shorter consonants (Gay, 1978) and formant changes for shorter stimuli (Lindblom, 1963) have been implemented to prevent additional duration cues. The only 'dynamic' change that was implemented was some minimal intonation in order to increase naturalness. This comprised a pitch accent on the first (stressed) syllable, followed by a terminal fall, which represents a frequent pattern for single-word utterances in German (Grabe, 1998). F0 of the /u,o/ vowels started with a rise from 100 to 110 Hz. During the final schwa, F0 declined from 105 to 80 Hz. Regarding the minimal F0 increase from 100 to 110 Hz in the first syllable no critical effects on the category boundary were assumed since it has been shown that F0 changes in the vowel do not affect the perception of vowel length in German and the location of the category boundary (Lehnert-LeHouillier, 2007), unlike changes in the spectral energy of the formants (Lehnert-LeHouillier, 2010). The rising pattern on the stressed vowel was preferred to a rising-falling pattern in order to avoid any interaction of perceived intonational variability with vowel length. For the adaptive discrimination test, durations were available up to 536 ms and frequencies up to 730 Hz in case of poor discrimination performance.

Procedures

Every subject participated in three sessions in a sound proof chamber at the Department of Neurology of the University of Tübingen. In the first session, two temporal discrimination tests and one spectral discrimination test were performed. In the second session, a second spectral discrimination test and the identification test were performed. In the last session, the goodness rating was performed. Subjects had to answer via keyboard. Before each session, subjects performed a short training in the presence of the first author. Instructions were given twice: Once orally by the first author with the possibility to ask questions in case of misunderstanding and a second time via headphones before each test. During the tests, no repeated listening option was available.

Stimuli were presented via Sennheiser HD 201 headphones at an intensity agreeable for the subject. The entire test was controlled by a MATLAB (version R2009a, Mathworks) procedure running on a laptop (acer Extensa) including both stimulus presentation and acquisition of behavioral data. Technical response time of the keyboard was ca. 50 ms (SE = 0. 6 ms) and was not subtracted from the measured response times. The analysis of the behavioral data was also performed with MATLAB, and the statistical calculations were done with the CRAN R. version 2. 15, statistic package {R_Development_Core_Team, 2010} using the lme4 package (Baayen, 2008).

Experiments

Identification test

The identification test covered a two-dimensional continuum (10 steps in the F1 frequency and 12 steps in the duration dimension) with vowel duration ranging from 48 ms to 151 ms and F1 ranging from 250 Hz to 430 Hz. Each item was categorized 10 times in a forced four-choice identification test to /gube/, /gu:be/, /gobe/, /go:be/. Pisoni et al. (1974b) showed that behavioral response time (RT) is shorter when subjects have to categorize categorical items as compared to ambiguous items located at the boundary, irrespective of their absolute acoustic distance. This is why RT represents an additional criterion for examining the space of auditory categories and was used, too.

The aim of the analysis was to get the most precise individual category boundaries along the spectro-temporal continuum. To these ends, for each spectral configuration the individual answer probability "*vowel was perceived as long*" was fitted by the psychometric logistic function (Lapid et al., 2008; Wichmann et al., 2001) as a function of duration, and for each vowel duration the answer probability "*vowel was perceived as /o/*" was fitted by the psychometric logistic function as a function of F1 by MATLAB's least squares estimation using the model in (1).

$$(1) \quad \Psi = \frac{1}{1 + e^{-(x-\alpha)/\beta}}$$

where 'Ψ' represents the probability of "*vowel was perceived as longer*", 'x' represents the vowel duration, 'α' the curve's midpoint (P = 50%) in the x-axis and 'β' the slope of the curve at that midpoint. Hence, parameter 'α' provided the location of the category boundary. See study I for an examplary fit.

In order to reduce interindividual variance, the individual RTs were normalized by subtracting individual means and adding the group mean. RT longer than three standard deviations from the group mean in each duration bin were excluded from calculations concerning the response times (6% of all trials). For the exact extraction of a numeric RT-maximum, RTs were fitted by a Gaussian function modelling the rising-falling shape of the data curves (see Figure III. 3b). The maximum value of the Gaussian fit was picked and interpreted as the RT maximum.

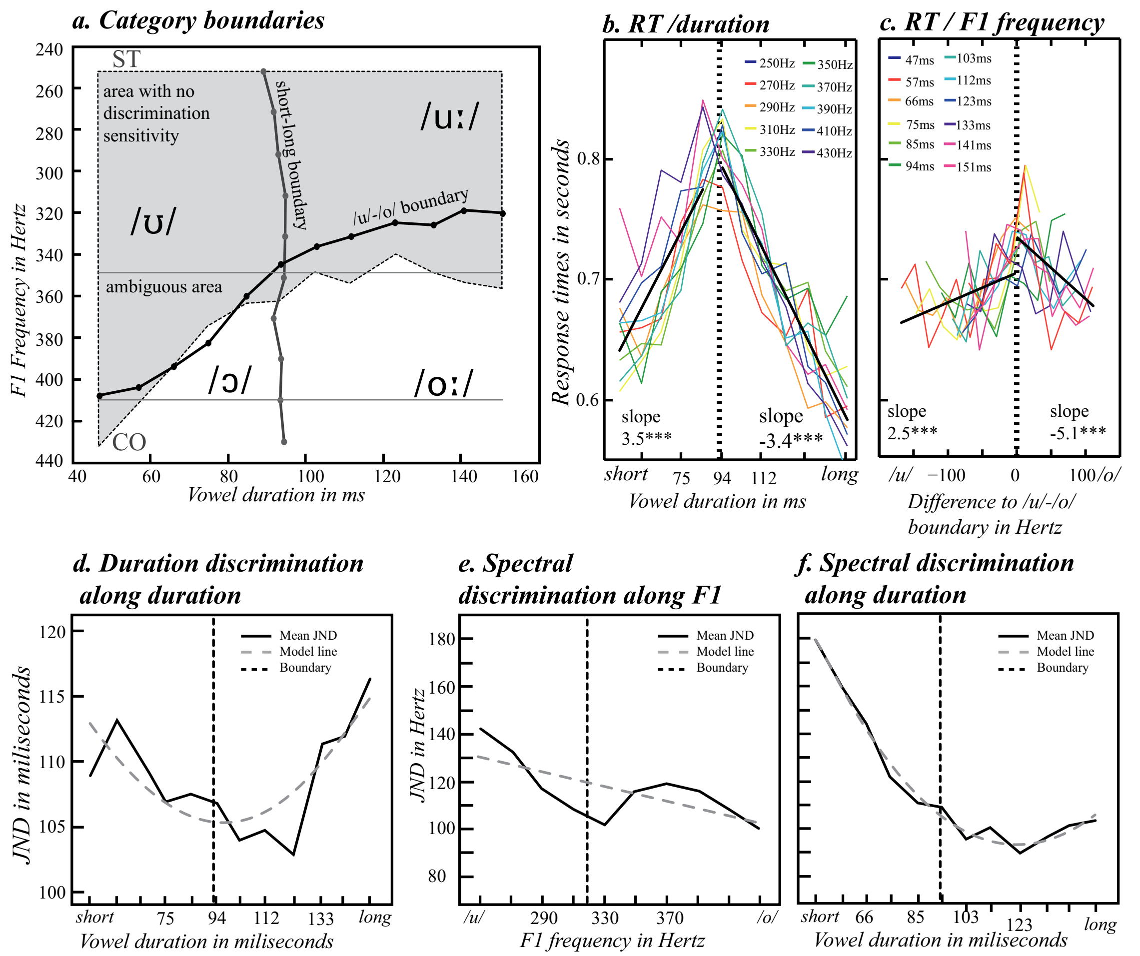

Figure III. 3: ***a)*** *Identification results: The vertical dotted grey line represents the short-long boundary; the black dotted line the /u/-/o/ boundary. Frequencies from the ambiguous area are perceived either as /ʊ/ or /oː/, depending on duration. The area is marked between the grey horizontal lines. The grey shaded area indicates the results from the discrimination test (d), where no discrimination between adjacent items is possible.* ***b-c)*** *Response times (RT). The black bold line indicates linear model fits. The asterisks mark the goodness of the fit. The vertical dashed line indicates the location of the category boundary.*

d-f) *Discrimination results in a A-X sequence. The black solid line indicates mean results; the grey dashed line shows the model fit. The vertical dotted line represents the category boundary.* ***d)*** *F1 frequency was kept constant at 250Hz in both, standard and comparison.* ***e)*** *F1 frequency in standard was kept constant at 250Hz.* ***f)*** *Vowel duration was kept constant at 250Hz in both, standard and comparison.*

Discrimination tests

The adaptive discrimination tests covered a) the temporal /u/-/u:/ continuum (12 steps) with a fixed F1 frequency of 250 Hz and b) the F1 /u:/-/o:/ continuum (10 steps) with a fixed vowel duration of 151 ms. Each continuum was submitted to a i) temporal and ii) spectral discrimination test.

i) The temporal discrimination test asked whether in the A-X word sequence the /o,u/ vowel in X was longer than in A. Subjects were instructed to push the button “yes” when they perceived the target vowel in X as longer than in A and "no" when they perceived the target vowel as equally long as or shorter than in A. Subjects were particularly instructed to pay attention to the actual durations rather than just to the phonological durational categories short and long. To these ends, they were given examples with clear within- and across-categorical duration changes. The set consisted of 12 A stimuli with an F1 frequency of 250 Hz and with durations ranging from 47 to 151 ms. The set was played in a pseudo-randomized order. A staircase procedure was performed for each of the A (Lapid et al., 2008). The procedure was as follows: After one positive answer, the duration of X was decreased by one pitch period (~10 ms), i.e. duration was made more similar. To increase the probability that subjects will perceive a duration difference after one negative answer, the duration of X was increased by three pitch periods after one negative answer. Possible X duration ranged from 47 ms to 536 ms. The test started with a difference of ~60 ms, a duration found as an optimal starting point in study I. The test ended when each A stimuli was played 30 times.

ii) The spectral discrimination test asked whether in the sequence "A –X" the /o,u/ vowel in X sounded exactly the same like in A. The set consisted of 10 A stimuli with a duration of 151 ms and with F1 frequencies ranging from 250 Hz to 430 Hz. The set was played in a pseudo-randomized order. Subjects were instructed to push the button "yes" when they perceived the target vowel in X as equal as in A and "no" when they perceived the target vowel as different from the vowel in A. They were given examples where they perceived clear within-and across-categorical duration changes. A staircase procedure was performed for each of the A (Lapid et al., 2008). The procedure was as follows: After one positive answer, the F1 frequency of X was increased by 10 Hz making it subtly different from A.

To increase the probability that subjects will perceive the items as equal after one negative answer, the F1 frequency of X was decreased by 30 Hz. Possible X duration ranged from 250 Hz to 730 Hz. The test started with a difference of ~40 Hz, a distance found as an optimal starting point in a preliminary study. The test ended when each A stimuli was played 30 times.

The aim of this analysis was to find the just-noticeable-differences (JND) along the continuum which serve as an indicator of the discrimination sensitivity. This was done by fitting the answer probabilities for each A item by the psychometric logistic function in three bins in each subject:

a. "*vowel duration was perceived as longer*" as a function of "X – vowel duration".

b. "*vowel quality was NOT perceived as equal*" as a function of "X – F1 frequency".

c. "*vowel quality was NOT perceived as equal*" as a function of "X – vowel duration".

Due to the adaptive procedure, the actual data concentrated around the perceptual threshold, and the margins of the scale were underrepresented. Therefore, in order to fit the psychometric logistic curve more smoothly two 0% values at the right and two 100% values at the left end of the abscissa were added.

A probability of 75% was chosen as threshold for the just-noticeable-difference (JND), mirroring the ¼ to ¾ distribution of yes-to-no answers in the staircase test (Lapid et al.) and was calculated by (2).

$$(2)\quad \mathrm{T} = \alpha + (-\beta * \log(\frac{1}{P} - 1))$$

where 'T' represents the duration and frequency threshold, i.e. the point when a difference between X and A was perceived, and 'P' the probability of 75%. 'α' and 'β' were the results from the fitting procedure with the psychometric logistic function.

In order to eliminate inter-subject variance, individual JNDs in each set were normalized for subject variability by subtracting the subject means as a function of vowel duration or F1 frequency from all threshold values and adding the group set average. Since the JND values were not normally distributed (Kolmogorov-Smirnov-Test: $p<0.001$), median instead of mean values were used. In order to find the general trend of the answer patterns, the median values in each set were fitted by a linear and square model with the JND values as a function of vowel duration and F1 frequency, respectively. Both models were compared with each other by means of a likelihood ratio test (Baayen, 2008). The more complex squared model was rejected when it did not differ significantly from the linear model which was the case for the spectral discrimination as a function of F1 (i.e. bin (c)).

Results

Results in the identification test

Regarding the hypothetical effects of the duration and spectral cue on the division of the two-dimensional vowel space in Figure III. 1b-d, the present results (Figure III. 3a) of the identification test resemble most that in Figure III. 1c. It was hypothesized that this pattern would emerge when duration would be used as the primary acoustic cue. The short-long-boundary runs vertically at a mean duration of ~ 93. 5 ms (95% confidence interval = 92. 3 – 94. 8ms). The category boundaries are differently reflected by the response times (RT). RT as a function of vowel duration for each underlying F1 frequency are shown in Figure III. 3b, showing a maximum at the short-long boundary. Individual RT maxima correlated significantly with individual short-long-boundaries (R = 0. 35, t = 4. 7, p < 0. 001). The same holds true for RT as a function of F1 frequency. Since putative RT maxima at the /u/-/o/-boundaries should wander along the F1 frequency continuum (see Figure III. 3a), RT values in Figure III. 3c were centered to the /u/-/o/-boundary for each vowel duration. Individual RT maxima correlated significantly with individual /u/-/o/ category boundaries (R = 0. 44, t = 5. 4, p < 0. 001) after excluding all maxima at the very edges of the F1 continuum (i.e. 250 Hz & 430 Hz). Regarding the slopes of the RT toward the boundaries, slopes along the duration continuum are steeper than along the F1 frequency continuum (slopes are shown in Figure III. 3b&c).

The locations of the category boundaries were analyzed with a mixed-effect model (Baayen, 2008), using F1 frequency and vowel duration as fixed effects for the category boundary between short and long as well as between /u/ and /o/, respectively. Subjects were included as random effects and the respective fixed effects as random slopes in order to account for repeated measures (Schielzeth et al., 2009).

The mixed-effect analysis fitting the short-long-boundaries as a function of F1 frequency did not yield any significance (model slope = 0. 01, t = 1. 4). This could be seen as a negative test that the model slope is shallow, which is why F1 frequencies did not affect the location of the short-long boundary. The mixed-effects analysis fitting the course of the /u/-/o/-boundaries as a function of vowel duration yielded significance (model slope = -0. 96, t = -17. 5). Thus, vowel duration affected the location of the /u/-/o/ boundary. This primary result speaks in favor of duration as the primary distinctive cue in contrast to quality.

Nevertheless, the picture is more complex, when one looks at the /u/-/o/-boundary courses in category short and long, and at the short-long-boundary courses between /ʊ/-/uː/ and between /ɔ/-/oː/. This was analyzed by a mixed-effects model including an interaction between category (see below) and slope as fixed effects and as random slopes, accounting for repeated measures.

i) The slope of the /u/-/o/ boundary differs between category short and category long. Regarding the categorization to short or long, all /u/-/o/ boundaries belonging to vowel durations shorter than 100 ms were regarded as category short. A mixed-effects analysis found that the slope was steeper in category short (slope = -1. 8 Hz/ms, t = -9. 4) than in category long (slope = -0. 3 Hz/ms, t = 6. 5). This was supported by a significant interaction between category and slope steepness ($\Delta_{Category\times Slope}$ = -1. 51199, t = -7. 11).

ii) F1 frequency seems to affect the short-long boundary in category /u/, but not in category /o/. Regarding the categorization to boundary between /ʊ/-/uː/ or /ɔ/-/oː/, all spectral boundaries belonging to F1 frequencies smaller than 340 Hz were regarded as belonging to the /ʊ/-/uː/ boundary. A mixed-effects analysis found that the short-long boundary was not affected by frequency between the /ɔ/-/oː/ boundary (slope = 0. 008 ms/Hz, t = 0. 46) ; but it was significantly affected by frequency between the /ʊ/-/uː/ boundary (slope = 0. 07 ms/Hz, t =5. 8). The difference between the slopes was significant ($\Delta_{Category\times Slope}$ = 0. 06, t = 2. 5). Importantly, the short-long boundary differed by 5. 5 ms (CI = 7. 8 – 3. 3, Student's T-Test, t = 5. 2, p < 0. 001) between F1 frequency = 250 Hz and 310 Hz.

Results in the discrimination tests

The course of the category boundaries as shown in Figure III. 3a is a necessary proof to show which is the primary acoustic cue. It is not sufficient, though, in order to conclude which might be the primary phonological feature. For this, the effects of hypothetical prototypes on the discrimination performances, which are distributed in the two-dimensional vowel space, might provide the necessary insight. It was hypothesized that the existence of a phonological category would affect the perception of the linearly changing continuum by showing a maximum in discrimination performance at the category boundary in contrast to the discrimination performance within the categories. The discrimination performance of spectral and durational changes is shown by means of just-noticeable-differences (JND), a value that indicates the perceptual distance which is needed to perceive a difference between two subsequent items (Lapid et al., 2008). The lower the JND, the higher the discrimination performance. Spectral and temporal JNDs along the short-long and F1 continuum are presented in Figure III. 3d-f. The statistical significance of the present results was assessed by fitting linear models to the JND courses. The parameters for each model in the sets a – c are presented in Table III. 2. Significant values are marked with asterisks. The temporal discrimination along the duration continuum shows a temporal JND minimum near the short-long boundary. Such a result matches the hypothetical effect of duration categories changing the perception. The duration JND is best represented by a positive square model (Figure III.

3d) with all parameters being significant (Table III. 2a). The minimum of the square curve is located at 96. 3 ms in the near proximity of the short-long-boundary of 93. 3 ms.

The picture is somehow different considering the spectral discrimination performances. Along the /u/-/o/ continuum, the spectral JND decreases with increasing F1 frequencies (Figure III. 3e), best represented by a negative linear model (Table III. 2b). Next to the /u/-/o/ boundary, the spectral JND deviates 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 /u/-/o/ categories. However, only the JND for 310 Hz differs 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), this becomes not significant.

	fit significance	**Intercept**	**Linear parameter**	**Square parameter**	**curve minimum**
set a)	p = 0. 01	134. 3ms ***	- 0. 6JND/ms **	0. 003**	96. 3 ms
set b)	p = 0. 02	169. 0Hz ***	-0. 15JND/Hz *	----	---
set c)	p < 0. 001	322. 1Hz ***	-3. 7 JND/ms ***	0. 02***	122. 0 ms

*Table III. 2: Parameters of the linear and squared model fittings to discrimination results as shown in Fig. 3d-f. Asterisks indicate significance (p: * = 0. 01, ** = 0. 001, *** < 0. 001).*

The primary goal of the present study was to investigate whether there would be ‘central and peripheral vowel categories’ which reflect the quality contrast. If the peripheral-central contrast was represented by categories, the spectral discrimination performance within the /u/ category would be increased at the hypothetical peripheral-central vowel boundary as it was shown for the category boundary between /u:/ and /o:/. The spectral JNDs along the short-long continuum (Figure III. 3f) show that this cannot be the case. Spectral JNDs increase steadily with decreasing vowel durations. Their course is best represented by a positive square model (Table III. 2c), showing a rightward shift with respect to the short-long-boundary with a minimum located at ~122. 0 ms. Unlike at the /u/-/o/ boundary, no local discrimination maximum at a putative peripheral-central vowel boundary is present within the area which is reserved for /u/. This is shown by means of the grey shaded area in Figure III. 3a. The upper edge of the area represents the location of the A stimuli, the lower edge represents the location of the X stimuli in the continuum, Hence, the spectral JNDs along the short-long continuum are so large that the compared item in the standard-comparison sequence has to be located at the /u/-/o/ boundary or cross it, in order to perceive a difference between A and X.

Discussion

The psychoacoustic characteristics of the duration and quality

The aim of the present study was to investigate the perceptual relevance of phonetic vowel quality and vowel duration for the phonological German tenseness/length contrast between 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 fine graded two-dimensional duration and F1 continuum. Considering the identification results, the location of the boundaries between category short and long, as well as between the /u/ and /o/ category, was differently affected depending on the categories which were subdivided by the respective boundary (Figure III. 3a). Regarding the /u/-/o/ boundary, its course as a function of duration was affected significantly in both length categories. It was steeper in category short than in category long. Regarding the short-long boundary as a function of F1 frequency, its course between the categories /ʊ/ and /uː/ was affected significantly by F1 frequency, but not between /ɔ/ and /oː/. The present results partially reproduced the findings of Lehnert-LeHouillier (2010), who have shown that the short-long category in German native speakers was affected by vowel quality. Taking into account the hypothesis regarding the effects of each cue on the course of the category boundaries between the short central and long peripheral vowels (Figure III. 1b-d), the present results represent a middle way between Figure III. 1c and 2d. Since vowel quality affected the short-long boundary only minimally, vowel duration, however, affected the /u/-/o/ boundary strongly, it is concluded that these results speak in favor of duration as the stronger cue for distinguishing short lax from long tense vowels. Vowel quality, at least in the domain of F1, was not used as the main cue distinguishing short from long vowels, since vowel quality was stronger affected by duration than vice versa. This was also reflected in the response times (RT) which increased strongly at the short-long boundary, but not at the /u/-/o/ boundary (Figure III. 3b-c). The differences between the category short and long regarding the slope of the /u/-/o/ boundary courses indicate that in category long, /uː/ and /oː/ were discriminated mainly on the basis of F1 frequency; in category short, /ʊ/ and /ɔ/ were discriminated using both cues, F1 frequency and duration (Escudero & Boersma, 2004). This is probably the reason why the short-long boundary was not affected by F1 frequency: It competed with the /ʊ/-/ɔ/ boundary. Between

/ʊ/ and /uː/, the boundary could be moved toward the shorter category since no competing category was present.

The present study reproduced the findings of Bennett (1968) and Weiss (1974). However, the present interpretation differs from the latter. Bennett (1968) and Weiss (1974) argued that /ʊ/ was mainly differentiated from /uː/ by spectral information. Their results, nevertheless, may have been biased by the lexical items provided for identification and the range of possible answers.

In the literature, it has been shown that the presence of a phonological category reduces the discrimination performance in the category but increases it toward the boundary (Flege et al., 1994; Massaro et al., 1983; Miyawaki et al., 1975; Pisoni et al., 1974b; Stevens et al., 1978; Xu et al., 2006; Ylinen et al., 2005). Taking this into account, it was hypothesized that a category boundary would be reflected in a decrease in just-noticeable-differences (JND), i.e. an increase in discrimination sensitivity. Regarding the temporal discrimination along the short-long continuum (Figure III. 3d), the present result matched to a certain account the hypothesized pattern, supporting the existence of duration categories and thus duration as an important cue. The present results regarding the temporal discrimination performance mirror the findings of study I, where a discrimination peak was found at the short-long boundary for the /a/-/a:/ contrast, which they interpreted as an indicator that vowel duration is perceived in a categorical way.

The results of the spectral discrimination along the /u/-/o/ continuum did not seem to match the hypothesized effect. Spectral JNDs decreased with increasing F1 as shown by the linear model line (Figure III. 3e). On the one hand, one might interpret these results as a strong ‘Perceptual Magnet Effect’ (PME) for /u/ in contrast to /o/ (Kuhl, 1991; Kuhl, 2004). Unlike Categorical Perception (Goldstone et al., 2009; Liberman et al., 1957), the PME posits that discrimination is still possible within the category but decreases toward the category prototype. In the present results, discrimination performance decreased almost linearly toward the /u/ category. On the other hand, it has been shown that spectral JND *decreases* with *increasing* frequency along a continuum between 20 Hz and 1000 Hz not only in speech (Mermelstein, 1977) but also in non-speech stimuli, which do not have a prototype (Akin & Belgin, 2005; Dai & Micheyl, 2011; Moore, 1973; Wier, Jesteadt & Green, 1977). In this sense, the present JND decrease seems to be a natural pattern. It seems as though the presence of /u/-/o/ categories in the F1 continuum is rather reflected by a local JND decrease at the /u/-/o/ boundary (Figure III. 3e). Indirectly these findings mirror the assumption that the boundary between vowel categories is less sharp as compared to consonants and is not necessarily associated with a discrimination peak (Cebrian, 2006; Diesch et al., 1999; Fry, Abramson, Eimas & Liberman, 1962).

From the articulatory point of view, the short lax – long tense contrast is reflected by a change in duration and centralization (Hoffmann, 2011). The present identification results have shown that vowel quality was a weaker cue to differentiate short lax from long tense vowels than vowel length. This finding is supported by the spectral just-noticeable-differences along the duration continuum, which were tested in a standard-comparison (A-X) sequence. The spectral discrimination performance increased strongly with shorter vowel duration (Figure III. 3d). This increase was so strong that the F1 frequency in the X had to be located at the category boundary between /u/ and /o/ (Figure III. 3a, grey shaded area). Thus, in the vowel space reserved for category /u/, no differences between [u] and [ʊ] in disyllabic words could be perceived. No spectral area was reserved for peripheral or central /u/. Rather, the entire F1 continuum from 250 Hz until the /u/-/o/ boundary was attributed to only one /u/ category (Figure III. 3a).

The present discrimination results seem to indicate that length is not only a primary cue, but that quality is no cue at all. Although this finding would support all proponents of vowel length as the primary cue (Becker, 1998; Hall, 1992; Ramers, 1988), the present 'either /u/ or /o/ result' with regard to the spectral discrimination performance might be an artifact. First, it was found that the short-long boundary, though minimally, was affected by F1 frequency. Taking this into account, it seems that, although at the discrimination level no difference was made between /ʊ/ and /u/, lowering the vowel and changing its quality serves as a secondary, supporting cue for identification. Taking this into account, the tenseness distinction can be regarded as allophonic.

Mechanisms underlying centralization

It has been shown that when a specific vowel is shortened due to increased speaking rate (Flege, 1988; Gay, 1978), it is produced more central than its longer, i.e. slowly produced instance (Lindblom, 1963; Moon & Lindblom, 1994). In this sense, the horizontal centralization of short, in contrast to categorical long vowels, is a natural articulatory mechanism. The lowering of categorically short vowels, however, is not a natural articulatory mechanism, since shorter vowels tend to be higher due to increased speaking rate, i.e. they undershoot their target position (Lindblom (1963). See also Flege (1988); and Gay (1978) who have shown that subjects increased peak velocity in order to avoid undershoot).

An explanation for this counter-natural behavior of categorically short vowels can be provided when one sees the process of articulation and perception as a reciprocal process in order to guarantee the transmission of information (see Summers, Pisoni, Bernacki, Pedlow & Stokes (1988) for articulatory adaptation in noise). E.g. Gussenhoven (2007) showed that lower vowels (e.g. /a/) which are equal in duration to higher vowels (e.g. /e/) are perceived *shorter*. He reasoned that this is due to the listener's compensation for the articulatory lengthening of low vowels in contrast to high vowels. Meister et al. (2011) replicated these findings by showing that the short-long boundary occurred at a shorter duration in high vowels than in low vowels. In this sense, the lowering of categorically short vowels in contrast to categorically long vowels enhances the perception as short and as such contributes to the categorical perception of vowel length. This would account for the present finding that the short-long boundary is located at a later point for higher F1 frequencies than for lower F1 frequencies. Such an enhancement is probably used to differentiate between short and long vowels when speaking rate is increased so strongly that acoustic durations of the two categories overlap (Hirata, 2004).

Another explanation might account for the finding that shorter vowels transfer less vocalic information than longer vowels (Wallace et al., 2009) since the auditory system has less time to extract it (Obleser et al., 2004; Roberts, Flagg & Gage, 2004b; Scharinger et al., 2011). By lowering the vowel, F1 frequency is increased (Figure III. 2a) and so is the spectral information density per time unit. Taking this into account, lowering the vowel might be seen as a compensatory mechanism to guarantee the correct extraction of spectral information which is needed to attribute the vowel to a category (Maeda, 1990; Summers et al., 1988).

Implications for phonology

If one accepts the notion that speech perception and speech production are reciprocal (see Hickok (2001) for a review of neurological findings concerning this matter), one is allowed to infer from the present psychoacoustic characteristics a phonological system needed for production. Regarding the question, which phonological feature is primary and lexicalized (Becker, 1998; Hall, 1992; Moulton, 1962b; Ramers, 1988; Reis, 1974), in light of the present results the answer would be *length*, i.e. the [+/- long] feature: Duration functions as a distinctive cue between /ʊ/ and /oː/ as well as between /ʊ/ and /uː/. Though not fully significant, /u/-/o/ categories seem to affect the psychoacoustic characteristics, a finding that can be interpreted as the perceptual reflect of the phonological [+/- high] feature. These two features seem to be sufficient to guarantee a *perceptional* distinction between the [-long, +high] /ʊ/ and the [+long, -high] /oː/ as well as between [-long, +high] /ʊ/ and [+long, +high] /uː/. In the following, a vowel perception model is presented. For this, one can assume a rather basic representation of the spectro-temporal vowel space (Figure III. 4a) whose processing is rule-driven. Due to statistical learning of the first language (Kuhl, Conboy, Coffey-Corina, Padden & Nelson, 2008; Ramscar, Yarlett, Dye, Denny & Thorpe, 2010), the system *knows* where the category boundaries are located between /u/ and /o/ (Escudero et al., 2009), as well as between short and long (study I to study III) and, most importantly, it *knows* how to adjust these for different speaking rates (Hirata, 2004; Hirata & Tsukada, 2004; Lindblom, 1963).

During the first step of processing, the incoming vowel duration is measured (Figure III. 4b). Depending on whether vowel duration is shorter or longer than the durational category boundary (CB), the length feature [-/+ long] is set (study II). Depending on the length feature, the location of the [+/- high] boundary is different. During step two of processing, the extracted spectral information is set in relation to the [+/- high] boundaries, attributing the [+/- high] feature. The resulting vowel is thus a combination of the sequential information processing and feature setting. So far, the extraction of intracategorical quality and the setting of a tenseness feature was omitted. This was due to the fact that the perception of short and long vowels may be based on merely the features [+/-long] and [+/-high]. Nevertheless, since quality is seen as a secondary, supporting cue to vowel duration (Gussenhoven, 2007), it should be included into the present model. In light of the present results and those of Hirata (2004), the tenseness feature becomes necessary when the two duration categories cannot be differentiated correctly. This happens when vowel durations of category short and long overlap in high speaking rate. The [+/-tense] feature is set during the extraction of spectral information based on the vowel's relative position to both, short and long F1 boundaries. When the spectral information is located between these two, the feature is set to [-tense], resulting in the perception of a "short" vowel. When the processed spectral information is located above both boundaries, the feature is set to [+tense], resulting in the perception of a "long" vowel.

The extraction of tenseness seems to be contradictory to the present results on discrimination performance (Figure III. 3a&d), since they posit that no discrimination was possible between central and peripheral /u/. One possible explanation of this finding is that subjects were unable to discriminate between different instances within the /u/ space since the extracted spectral information has already been transformed into a phonological representation, apply to either [+high] or [-high] (see Pisoni (1973a) for the decay of acoustic information). A putative tenseness feature was not present. This interpretation is supported by the finding that only discrimination between [+high] and [-high] was possible (Figure III. 3a&d). The inability to transform central and peripheral instances to phonological features speaks in favor of the allophonic status of phonological tenseness.

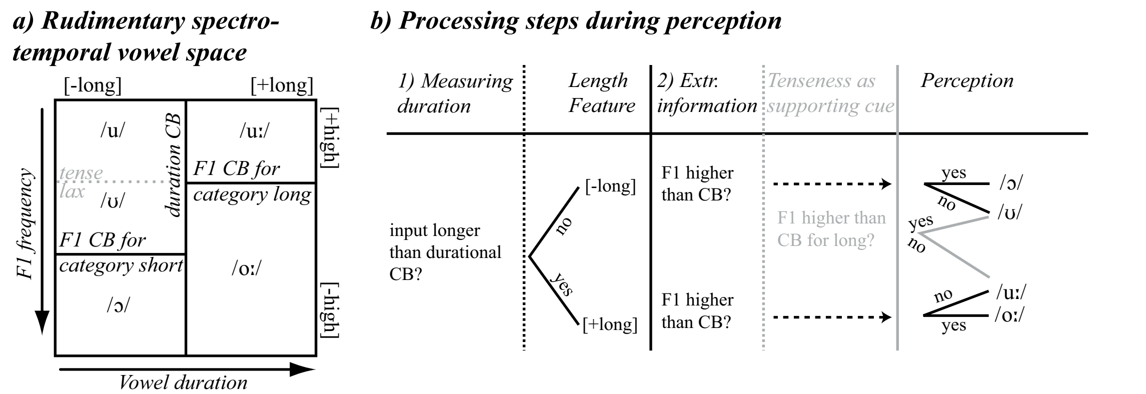

Figure III. 4: ***a)*** *Basic spectro-temporal vowel space used to model the perception of short and long vowels. The category boundaries (CB) between the vowel categories and between the duration categories are based on statistical language learning.* ***b)*** *The steps during the processing of vowel quality and duration. 1) Based on the duration measurement, different CBs are assumed for the vowel categorization. 2) Depending on the categorical length the spectral information vowel is differently categorized. In case, duration cannot be measured properly, tenseness functions as a supporting cue.*

The question remains whether a [+/-tense] feature is necessary for production, i.e. for an underlying phonological system. The present paper assumes that the [+/- tense] feature is needed in order i) to guarantee the centralization of the vowel during articulation and ii) to reflect the secondary, supporting status of the quality cue. The proposal follows the phonological literature (Hall, 1992; Ramers, 1988) in assuming that the [+/-tense] feature is filled in during the phonological derivation. Its phonetic equivalent is centralization which is used as a secondary, supporting cue. [+/- tense] can be derived by means of two rules:

a) V → [+tense] / [+long], else [-tense]

b) V → [+tense] / __$, else [-tense]

Rule (a) is based on Hall (1992); Ramers (1988) who argue that vowel length is lexicalized. According to rule (a), a long vowel is attributed the feature [+tense] and is therefore centralized at the surface. Rule (b), where $ indicates the syllable boundary, is based on Becker (1998) who argues that tenseness is a phonetic reflect of syllable structure. Open syllables become tense, closed syllables become lax. In the end, the articulatory result is equal to rule (a). However, the choice between the two rules is left open.

Conclusion

Study III analyzed the interaction between vowel duration and vowel quality in German using a two-dimensional /u/-/o/ and short-long continuum. It was found that duration is used as a primary cue to distinguish between /ʊ/ and /uː/ as well as between /ʊ/ and /oː/. /ʊ/ and /oː/ are differentiated by duration due to the decrease of spectral discrimination sensitivity as vowel duration shortens. Furthermore, although vowel quality is not necessary for the distinction between /ʊ/ and /uː/, it serves as a supporting cue. Thus, tenseness is allophonic. Mapping these perceptual results onto a phonological representation, vowel duration is presumably lexicalized as a length feature. Vowel quality seems to be assigned at a later stage, derived by phonological rules, in order to guarantee the central-peripheral articulation.

Study IV: The sensitivity to acoustic and phonologic information: A priming study

Introduction

Vowel identity reflected by the M100/N100

In recent years a large amount of evidence accumulated that different aspects of speech are reflected in the electrophysiological and neuromagnetic correlates of the brain as early as 100ms after the signal onset. Specifically, it has been shown that the amplitude of the electrophysiological N100 and its neuromagnetic M100, i.e. the negative reaction 100 ms after stimulus onset, is larger for periodic, i.e. vowel like signals, than for aperiodic, i.e. fricative like segments of speech (Alku, Sivonen, Palomäki & Tiitinen, 2001; Hertrich et al., 2000). Furthermore, although not statistically significant, the amplitude tends to be larger for low vowels than high vowels, reflecting higher neural activity when the first formant is higher (Alku et al., 2001; Roberts et al., 2004a). The latency of the M100 is sensitive to frequency, too. When tested with non-speech tone stimuli, latency correlates negatively and linearly with tone frequency till 1kHz (Roberts et al., 2004a; Roberts et al., 1996). However, the latency of the M100 to vowels increasing in F1 frequency form a plateau according to the category (Roberts et al., 2004a). Importantly, this mechanism is independent of pitch, showing that the M100 reflects vowel rather than mere acoustic information (Poeppel, Phillips, Yellin, Rowley, Roberts & Marantz, 1997; Poeppel et al., 1996). Roberts et al. (2004a) argued that the latency plateaus are a result of distinct tonotopic cohorts corresponding to native language vowel categories. This is supported by Obleser et al. (2004) and Scharinger et al. (2011) who have shown that the generator of the M100 is distinct for each vowel category and, importantly, feature.

The German vowel system

Taking this into account, the M100 suits perfectly to investigate one of the core questions regarding the German vowel system. The German vowel system shows a complex structure based on an interaction between vowel duration and vowel centralization (Hall, 1992). Short vowels are more centralized and have a higher first formant (F1) frequency, long vowels are more peripheral and have a lower F1 frequency (Figure IV. 1a). These two phonetic cues, duration and centralization, are differentiated in phonology by two features: The length feature [+/-long] reflects the long-short distinction; the tenseness feature [+/-tense] reflects the peripheral-central distinction (Hall, 1992; Ramers, 1988).

Regarding the distribution of these phonetic cues, the duration distinction is limited to stressed syllables whereas it is largely non-existent in unstressed syllables. Nevertheless, de-stressed instances of vowels may still exhibit some differences in centralization. This gave rise to the assumption that these differences, irrespective of the length aspect, refer to a representation of *quality* in the mental lexicon. Thus, the lexicalized feature would be [+/-tense] (Moulton, 1962a; Reis, 1974). Importantly, only *distinctive* and *non-predictable* information is ‘lexically specified’; information which is non-distinctive and can be predicted by means of phonological rules remains ‘underspecified’ (Eulitz et al., 2004; Keating, 1998; Lahiri et al., 2002).

There are some shortcomings which ‘lexicalized quality’ cannot explain. i) German shows phonotactic restrictions due to vowel duration (Becker, 1998; Hall, 1992). ii) short [+tense] and [-tense] vowels in unstressed positions are distributed in an allophonic way to open and closed syllables, respectively. This allophony speeks rather for a subsyllabic specification of vowel quality (Becker, 1998; Ramers, 1988). iii) It is a general consensus that the /a/-/a:/ contrast is exclusively durational since the quantitatively contrasting instances overlap to a high degree in the F1/F2 vowel space (Hoffmann, 2011; Sendlmeier et al., 2006). These points rather speak for *length* [+/- long] as the lexicalized feature. Thus, in order to avoid a phonological over-specification, length-associated differences in vowel quality, i.e. centralization of short vowels, should be considered as secondary modifications that do not rely on lexical phonological specification.

However, no consensus can be found in the literature which of the two acoustic cues in the German vowel system – duration and centralization – is used primarily in perception and can thus be seen as the lexicalized feature. The findings of Bennett (1968), Weiss (1974); (1978) and Sendlmeier (1981) seem to speak in favor of vowel quality. By means of behavioral identification tests, they found that subjects relied more on the quality cue when they had to attribute vowels to category long or short in a forced choice test. However, mainly speakers from the north of Germany participated in their study. They furthermore used lexical items that might have caused a bias. Vowel length cannot play a completely subordinate role, since the distinction between /a/-/a:/ as well as /ɛ/-/ɛ:/ is performed on the basis of duration (Hoffmann, 2011). This is supported by study I where non-linear changes in the discrimination performance along the short-long continuum has been found. It was argued that these changes were due to lexicalized duration categories. Also, study II found an additional neural correlate that emerged for categorically long vowels, only. Shortening, furthermore, affects the relative position of centralized vowels to such an extent that *high short lax* vowels overlap to a large extent in formant frequencies with *mid long tense* vowels (Hoffmann, 2011; Sendlmeier et al., 2006) (see schematics in Figure IV. 1a). Hence, the perception depends on phonological length when it comes to the distinction between *short* [-tense, +high] /ʊ/ and *long* [+tense, - high] /oː/ (see Bennett (1968); Heike (1970); Weiss (1974) and results from study III). Having assessed the structure of the spectro-tempral continuum between short and long as well as between /u/ and /o/ and having tested the discrimination ability for spectral changes along the short-long continuum, it was argued in study III that the [+/- tense] feature cannot be lexicalized. Figure IV. 1b shows the results of study III regarding the identification of the spectro-temporal continuum between /u/ and /o/. A hypothetical [+/- tense] boundary is included. In that test, the ability to perceive spectral spectral changes decreased in short vowels to such an extent that vowels were perceived either as /u/ or as /o/. Within the /u/ vowel space, no distinction was made between alleged /u^{+tense}/ and /ʊ$^{-tense}$ /. Nevertheless, it was found that the category boundary between /ʊ/ and /uː/ was, though minimally, affected by F1 frequency. It was concluded that the lax-tense distinction was perceivable at the phonetic level, but it was not transferred into a phonological representation. This is why the tenseness distinction is allophonic, supporting the categorical perception of vowel duration (Gussenhoven, 2007). The present study investigated whether the change from [u^{+tense}] to [ʊ$^{-tense}$] might be reflected in the neuromagnetic correlates. As such, it investigated further the discriminative sensitivity between different spectral information in the /u/ vowel space. This was performed by means of a priming experiment with a probe-target sequence, in which the target did or did not match in phonetic and phonologic information with the probe.

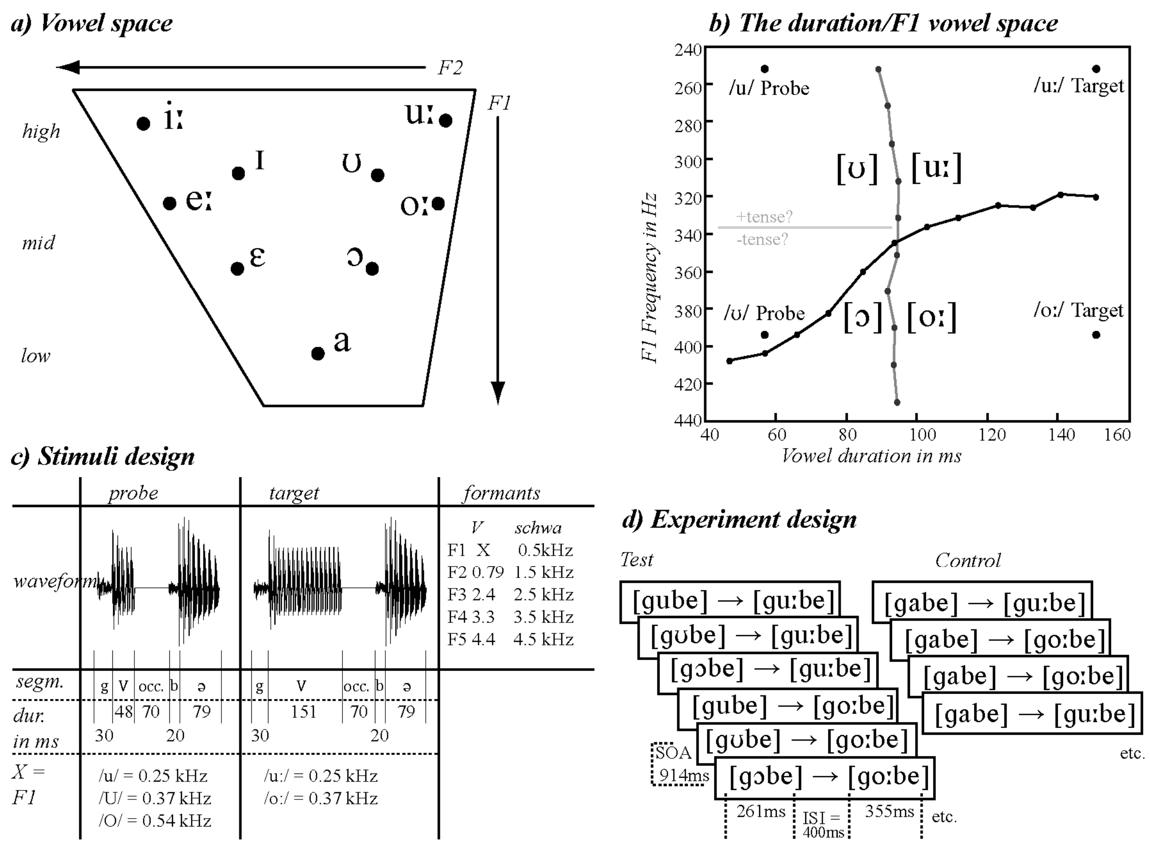

Figure IV. 1: a) Schematics of the German vowel space. Short vowels are more central and lower than long vowels. b) Results from Study III showing the structure of the spectro-temporal continuum between /u/ and /o/ and short and long. The location of the first two probes and the targets are indicated. The hypothetical lax-tense boundary is indicated. c) Waveforms of the synthesized stimuli. The F1 frequency (X) of the vowel in question and segment durations are presented. d) Experiment design for the priming condition and the control condition.

The priming paradigm

Generally, studies concerned with phonological processing used a magnitude of different real words and non-words (Connolly et al., 1994; Desroches, Newman & Joanisse, 2008; Newman & Connolly, 2009; Newman et al., 2003). These words overlapped to a certain amount either at the beginning (Fat – Fan, Fat - Fun) or at the end (Fat – Bat, Fat - but). Using paradigms similar to the oddball design, where a phonological expectancy was violated, the time course of phonological processing and its relation to lexical access was investigated. A neural component called phonological mismatch negativity (PMN) was identified, peaking ~300ms after the phonological violation. However, none of these studies concentrated on one single segment in the word, not to mention one feature like [+/-high]. Furthermore, an oddball design needs a many-to-one ratio in order to produce a mismatch response (Näätänen et al., 2007). The more combinations one wants to test, the longer the duration of the experiment. In order to reduce experiment duration, but maintain the possibility of comparing several constellations, the present study used a probe-target paradigm.

Priming studies make use of the fact that the an information overlap between two subsequent words, be it semantic or phonological, affects the processing of the second word and as such production times (Slowiaczek et al., 2000) and response times (RT) in lexical decision (Dumay et al., 2001; Fear et al., 1995). The effects of a probe on the processing of the target are generally assessed in behavioral studies. A facilitating effect is found when the RT to the target decrease in comparison to a control baseline; an inhibitory effect is reflected by a RT increase to the target (Zwitserlood, 1996). Unfortunately, a phonetic overlap between probe and target, be it word initial or word final, can result in both, inhibition and facilitation (Goldinger, 1998).

In neurological studies, priming attenuates cortical activity in lexical auditory priming (N400: Brown et al. (1993); Radeau et al. (1998), fMRI: Bergerbest et al. (2004); in visual repetition priming Gruber et al. (2004); Gruber et al. (2005a)). To a certain degree, the P50-suppression – the reduction of the P50 response to the second signal in a sequence of two identical signals – can be seen as priming (Cadenhead et al., 2000; Light et al., 2001; Lijffijt et al., 2009; Mathiak et al., 2011). The reduction is explained by the brain’s capacity to differentiate between relevant and irrelevant incoming information and as such adapt its sensitivity to auditory input (Boutros et al., 1999). Lijffijt et al. (2009) have shown that this ‘gating’ mechanism is not restricted to the P50 but also affects the N100 and P200 and correlates with behavioral discrimination performance. Lijffijt et al. (2009) relate the N100/P200 gating to a mechanism protecting higher cognitive functions from uncontrolled input.

The present study

The present study had two questions. The broad question was whether a single segment in a probe can prime another single element in the target and affect neural correlates as early as the M100. The answer to this question is a narrower question: What is the neural response to frequency change in a short vowel. Taking these two questions together, it was of interest how much phonetic and phonologic information can be extracted from a short vowel within a disyllabic word by means of a priming experiment. Since the M100 seems to reflect vowel information (Alku et al., 2001; Obleser et al., 2004; Scharinger et al., 2011) and since it has been shown that the processing of a target can be affected by information contained in the probe (N400: Brown et al. (1993); Radeau et al. (1998); fMRI: Bergerbest et al. (2004)) it was hypothesized that the target’s M100 latency and amplitude could be manipulated by the probe’s F1 frequency and whether phonetic and phonologic information matched or did not match.

Five structurally identical disyllabic nonwords (/g_bə/) were used as stimuli. They were drawn from the two-dimensional vowel continuum from study III which is presented in Figure IV. 1b. The continuum covered F1 frequencies from /u/ to /o/ and vowel durations from short to long and was categorized to /gubə/, /gu:bə/, /gobə/, /go:bə/. Probes contained the short vowels /u$^{+\text{tense}}$/, /ʊ$^{-\text{tense}}$/ and /ɔ$^{-\text{tense}}$/. Targets contained the long vowels /uː$^{+\text{tense}}$/ and /oː$^{+\text{tense}}$/. The vowels can be described by the phonological features [+/-long], [+/-tense] and [+/-high] (Hall, 1992) as shown in Table IV. 1a. Probes and targets were chosen so that they did or did not match regarding their F1 frequency (250Hz, 370Hz, 540Hz), their height feature [+/- high] and their supposed tenseness feature [+/- tense] (Table IV. 2b-c). Although it is know that the optimal vowel duration to extract the vowel identity is 150 ms and less information is conveyed with shorter vowels (Wallace et al., 2009), the vowel in the probe had to be short so that there would be a constellation matching in F1 frequency but not in the height feature ([ʊ$^{-\text{tense}}$]→[oː]).

a. Feature Representation

		- long	+ long
+ high	+ tense		/u:/
	- tense	/ʊ/	
- high	+ tense		/o:/
	- tense	/ɔ/	

b. Frequency Match

	/uː/ 250Hz	/oː/ 370Hz
/u^{250Hz}/	+	–
/ʊ^{370Hz}/	–	+
/ɔ^{540Hz}/	–	–

c. Height Feature Match

	/uː/ [+high]	/oː/ [-high]
/u$^{+\text{high}}$ /	+	–
/ʊ$^{+\text{high}}$/	+	–
/ɔ$^{-\text{high}}$/	–	+

d. Tenseness Feature Match

	/uː/ [+tense]	/oː/ [+tense]
/u$^{+\text{tense}}$/	+	+
/ʊ$^{-\text{tense}}$/	–	–
/ɔ$^{-\text{tense}}$/	–	–

*Table IV. 1: **a)** Phonological feature representation of the short-long, peripheral-central and high-mid vowel contrast. **b-d)** The tables illustrate the information matching between primes (rows) and targets (columns). ‘+’ stands for a match.*

Comparing the matches it is obvious that the match in the secondary tenseness feature (Table IV. 1d) is confounded by a match in the primary height feature (Table IV. 1c) and in the frequency match (Table IV. 1b). This is why no direct effects of a tenseness match vs. no-match were expected. Rather, it was hypothesized that a match in phonetic frequency and/or phonological height feature would affect the M100 amplitude and/or M100 latency. However, since in the priming literature both facilitating and inhibitory effects have been shown (Behavioral: Fear et al. (1995), neural: Bonte et al. (2004)), no hypothesis in this regard could be stated. Regarding the discussion about which is the lexicalized feature, the reasoning was indirect. In the behavioral tests, study III found contradictory results: On the one hand, their participants were not able to discriminate between short lax and short tense instances within the category /u/. On the other hand, the location of the short-long boundary (~90ms) was located significantly at a later point in the continuum for high than for low F1 frequencies, indicating some sensitivity to frequency change at average durations. The question arose whether the minimal sensitivity would remain in very short vowels. Hence, should the /u/ and /ʊ/ probes, containing a short vowel, affect the target's M100 amplitude in different ways, the auditory system would be sensitive enough to extract phonetic information, which is necessary to discriminate between different instances of short /u/. Such a finding would support study III's conclusion that tenseness is used in an allophonic way.

Methods

Subjects

20 native speakers of German were recruited and paid for their participation (10 males, 10 females. Mean age: 26. 4 years, SD = 8. 2 years). All subjects were speakers of Standard German, coming mainly from the south of Germany. All had normal hearing as confirmed by an audiometric screening test and were right handed as confirmed by a German version of the Oldfield (1971) test. Written consent was obtained from all subjects in line with the guidelines of the Ethics Committee of the University of Tübingen.

Stimuli

The stimulus material consisted of the trochaic nonsense words /gubə/, /gʊbə/, /gɔbə/, /guːbə/, /goːbə/. Stimuli were taken from a two-dimensional vowel space (Figure IV. 1a) extending from /u/ to /o/ and from short to long and which has been categorized to /gubə/, /gu:bə/, /gobə/, /go:bə/ in study III. 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 in German (Féry, 1996). Thirdly, 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 (Hirata, 2004); see Gottfried et al. (1990) for speech tempo effects. Although there is a statistical tendency in German to 'avoid' voiced plosives after short vowels (Lenerz, 2000), the consonant /b/ was used in order to increase the phonological distance from existing lexical words such as, e.g. /guːtə/ or /goːtə/. The voiced plosive affects the perception of the preceding vowel insofar as it slightly changes the location of the category boundary toward longer durations (see Lehtonen (1970, P. 80) for effects of consonant voicing on vowels.)

In order to exclude any accidental variations of natural speech, these items were synthesized by concatenating noise segments with vowel segments. The principal validity of synthesized stimuli regarding phonetic/phonological research questions has been shown in several studies (Ainsworth, 1972; Blomert et al., 2004; Bunton et al., 2009; Sams et al., 1990). Stimuli were synthesized using the formant synthesizer of Hertrich et al. (1999); 2007) which computes formants as amplitude- and phase-modulated sinusoids on the basis of the F0 frequency. Aperiodic speech segments, which represent fricatives or stop consonant bursts, are generated by adding a random parameter into the formants' phase progression. Formant values and the temporal aspects of the stimuli were approximately derived from a real recording of the first author, analyzed with PRAAT (Boersma et al., 2009).

Formant transitions were implemented in the stimuli according to Strange (1987) and Strange et al. (1998) within the initial 20 ms of each vowel. In order to increase naturalness, some minimal intonation was implemented in the test stimuli comprising a pitch accent on the first (stressed) syllable, followed by a terminal fall, which represents a frequent pattern for single-word utterances in German (Grabe, 1998). No significant effects were assumed to occur due to the minimal intonation since i) it was equal in all vowel durations and ii) since it has been shown that F0 changes in the vowel do not affect the location of the category boundary (Lehnert-LeHouillier, 2007), unlike changes in the spectral energy of the formants(Lehnert-LeHouillier, 2010). F0 of the /u,o/ vowels started with a rise from 100 to 110 Hz. During the final schwa, F0 declined from 105 to 80 Hz. The rising pattern on the stressed vowel was preferred to a rising-falling pattern in order to avoid any interaction of perceived intonational variability with vowel length. Figure IV. 1c displays the durations of the segments and the vowel formant frequencies of the probes and targets. The matching constellations between probe and target can be found in Table IV. 1b-d. For the control – target condition, the same consonantal environment was used with the short vowel /a/ in the first syllable (F1 = 850 Hz, F2 = 1340 Hz, F3 = 2600 Hz, F4 = 4100 Hz, F5 = 5400 Hz). The probe and target words, thus, equaled in all segments but the vowel in the first syllable.

Data acquisition

Subjects performed an identification test in a sound proof chamber at the Department of Neurology of the University of Tübingen. Stimulus material consisted of probe-target sequences in the first run and control-target sequences in the second run. Each probe-target / control-target sequence ([u]→[uː],[ʊ]→[uː], etc.) was played 10 times. See Figure IV. 1d for the experiment design. Subjects were informed to ignore the probe and to identify the vowel in the target either as [uː] or as [oː] by pressing a respective button with the left hand and the right hand, respectively. Test and control condition were performed in two separate blocks since it was found in preliminary tests that the control with /a/ induced a mismatch. Response times were measured from the offset of the target.

In the MEG session, subjects were comfortably seated in upright position and had to listen to the same probe – target and control – target sequences as in the behavioral test in a passive listening task. An entire session was subdivided into eight runs comprising 102 stimuli each. Run 1 and 5 consisted of the control-target sequences, the other runs of the probe-target sequences. In preliminary tests it was found that the /gabə/ control produced as mismatch negativity when it was pooled with the other probes. This is why the control condition was specifically played in a separate run. The ISI of 400ms and the SOA of 914ms were fixed for all trials. Although the fixed durations increased the probability of stimulus onset expectancy, they were chosen in order reduce the timing variance between probe and target and possible effects of entrainment on the M100 deflection (see Fellinger et al. (2011) for effects of the phase of underlying neural oscillations on the P50). Within one run, each probe-target sequence was played 17 times in a pseudo-randomized order. Each control-target sequence was played 56 times in each run. The high number of the same control-target sequences in one run produces a habituation effects, which attenuates the deflections (Mutschler, Wieckhorst, Speck, Schulze-Bonhage, Hennig, Seifritz & Ball, 2010). This was taken into account, since it was assumed that target intrinsic differences would nevertheless emerge in the M100. The control-target sequence was used to guarantee the same temporal structure in a trial as in the probe-target sequence in order to omit structure dependent effects on the magnetic fields. Subjects had to look at a fixation cross attached at a distance of ~50cm from their eyes. MEG recordings were performed using a 275-channel whole-head system (VSM MedTech Ltd., Coquitlam, Canada, SQUID gradiometer sensors) in a sound-attenuated, magnetically shielded chamber. Three coils generating magnetic fields were attached at three fiducial points (nasion, preauricular points on each side) of the subjects' head. These coils recorded continuously the head position in relation to the MEG sensor array. Only recordings with a maximum movement below 0. 5 cm during the recording were further analyzed. Stimuli were presented via insert earphones and plastic air tubes (Ear tone 3A system) at an comforTable IV. intensity for the subjects (ca. 70 dB). Stimulus presentation and synchronization of stimuli with MEG data was controlled by a custom-made computer program running on a DOS machine (yes, DOS machine!). MEG data were recorded at a sampling rate of 585. 93Hz in epochs of 1800ms duration including a 100-ms pre-stimulus baseline. The peculiar sampling rate was chosen, since the master clock of the MEG electronics used to work with a frequency also used by a local radio station leading to interferences and artifacts.

Analysis of the behavioral data

Behavioral response times (RT) were normalized for individual effects by subtracting the individual mean and adding the grand mean. Only response times of the first and last stimulus in the continuum, i.e. those which were used as targets in the MEG experiment, were analyzed. In order to delete the target inherent effects, the RT in control condition was subtracted from the RT in the test condition (i.e. Ψ = test – control). Importantly, negative/positive values cannot be interpreted as priming/inhibition since the control condition was tested in separate runs.

In order to model the sigmoid shape of the identification results and to extract the individual category boundary, each individual probability bin "*vowel was perceived as [oː]*" depending on the probe was fitted by an psychometric logistic function by MATLAB's least squares estimation (Lapid et al., 2008; Wichmann et al., 2001). The model in (1) was used:

$$(1)\ \Pi = \frac{1}{1+e^{-(x-\alpha)/\beta}}$$

'Π' represents the probability of "*vowel was perceived as longer*", 'x' represents the vowel duration, 'α' the curve's midpoint (Π = 50%, i.e. the category boundary) in the x-axis and 'β' the slope of the curve at that midpoint.

Preprocessing of the MEG data

For dipole fitting, the MEG data were averaged over all trials and conditions (i.e. probes) for each subject after pre-stimulus baseline correction (-100 ms – 0 ms) ; eye-blink rejection (threshold criteria; frontal sensors > 2 nAmp). In order to exclude effects of the control condition on the dipole location, fitting was performed on the individual grand averages of the test condition. Neuromagnetic responses were offline band pass filtered (4–40 Hz) ; the low pass frequency was so chosen that the sustained negativity after the first stimulus would vanish. Two subjects were excluded from further analysis due to a high eye blink rate (more than 1/3 of all trials). Two dipoles were placed in the left and right hemisphere of each subject. Since the M100 deflection is not generated by a single but by multiple sources (EEG: Alain et al. (1997), MEG: Pantev et al. (1989b); Scharinger et al. (2011)), and since no hypothesis could be stated which cortical area might have been affected by the probe, a large temporal window was chosen for the fit (50 ms to 200 ms after the onset of the second stimulus). Also, the time window was chosen so that the dipoles could be fitted consistently in all subjects and gave a good representation of the total activity within the auditory cortex (Hertrich et al., 2000; Mathiak et al., 1999). Dipoles were fitted with the CTF software 'DipoleFit' which used a spherical standard head model including all gradiometer sensors. The courses of the dipole moments were determined by means of subspace projection, providing the partial amount of overall activity for each sample point that can be accounted for by a given dipole structure (Hertrich et al., 2000). The sign of dipole orientations was adapted to obtain a negative *M100* moment (Mathiak et al., 1999). The bilateral dipoles accounted in average for 85% of the variance. They were located in the upper temporal lobe in the region of the central auditory system. Figure IV. 2a shows the group mean dipole locations placed in one random subject. Figure IV. 2b shows the neuromagnetic response for the grand average across the conditions (i.e. probes) and the typical field topography of the M50 and M100.

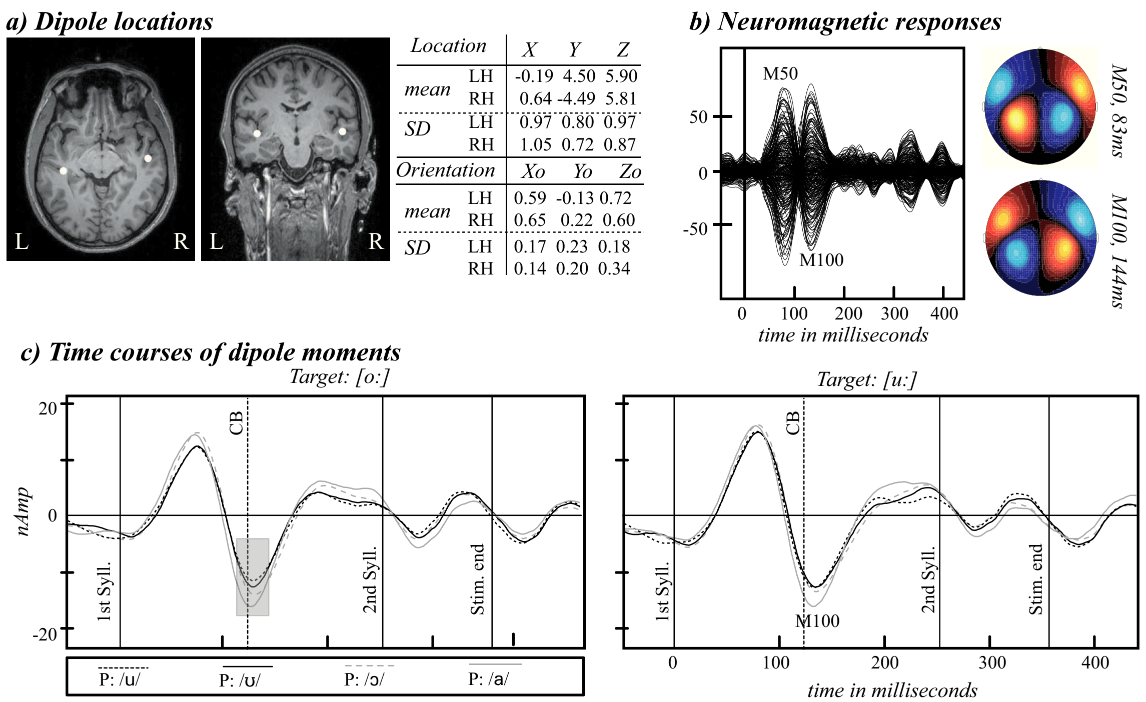

Figure IV. 2: ***a)*** *Group mean dipole locations in the left and right hemisphere plotted in one exemplary subject, including their location and orientation in the CTF source space.* ***b)*** *Grand average neuromagnetic responses to the target and the typical topographies of the M50 and M100.* ***c)*** *Time courses of dipole moments for the two targets affected by the probes (P). Left panel: target contained /u:/, right panel: target contained /o:/. First and second syllable onsets are indicated. The grey shaded square shows the analyzed mean vowel amplitude.*

Analysis of the MEG data

Using individual dipole locations, mean M100 amplitudes were calculated for each priming condition (*u1→uu*, *u2→ oo* etc.) in the time window between 110 ms and 150ms for each subject (grey shaded area in Figure IV. 2c). In order to exclude target inherent effects on the M100, the amplitude in control condition was subtracted from the amplitude in the test condition (i.e. Ψ = test – control). Ψ values will mimic the M100 pattern with respect to zero insofar as positive Ψ values reflect a attenuation of the M100 amplitude in comparison to the control condition and negative Ψ values reflect an enhancement of the M100 amplitude in comparison to the control condition (i.e. Ψ > 0: attenuation; Ψ < 0: enhancement). Importantly, negative/positive values cannot be interpreted as priming/inhibition since the control condition was recorded in separate runs. The effect strength is negatively correlated with the amplitude attenuation and mirrors the M100 behavior: Higher Ψ indicate a stronger *attenuation* and lower Ψ indicate a stronger *enhancement* of the M100 amplitudes in comparison to the control condition.

The effect values Ψ were analyzed with CRAN R, version 2. 15. 0 (R_Development_Core_Team, 2010) using mixed-effects models provided by the lme4 package (Baayen, 2008). Mixed-effects models include random effects and fixed effects as well as random slopes and are able to handle empty cells. All models used for the present study included SUBJECTS as random effects. The presented models were the result of model comparison between a less complex and a more complex model (Baayen, 2008). As Scharinger et al. (2011) stated: “[The] rationale of these model comparisons is to find the model that provides the best fit for the observed data without over- or underfitting. Model comparisons are made on the basis of the Akaike Information Criterion and the Bayes Information Criterion, essentially reflecting the entropy of the model. This entropy should be minimal, and models with a significantly lower Akaike Information Criterion or Bayes Information Criterion are evaluated using likelihood (L) ratios that are associated with a p value. ” Hence, more complex models, which did not yield $p > 0.05$ and thus did not achieve a significantly better fit than the less complex model, were rejected and will not be presented in the further analysis. In contrast to the ANOVA analysis, mixed-effects models do not provide any F-values nor p-values, but t-values. These are associated to differences between factor levels or slopes for linear regressions as a function of covariates (Baayen, 2008). Generally, a difference/slope (Δ) can be seen as larger than zero and thus significant when $t > 2$. The present analysis used various fixed effects/random slopes: PROBES ([u],[ʊ],[ɔ]) ; TARGETS ([uː], [oː]) ; HEMISPHERE (LH, RH) and F1FREQUENCY (250Hz, 370Hz, 540Hz) as a linear covariate. Fixed effects were included as random slopes in order to account for repeated measures (Schielzeth et al., 2009) when model comparison yielded a better fit.

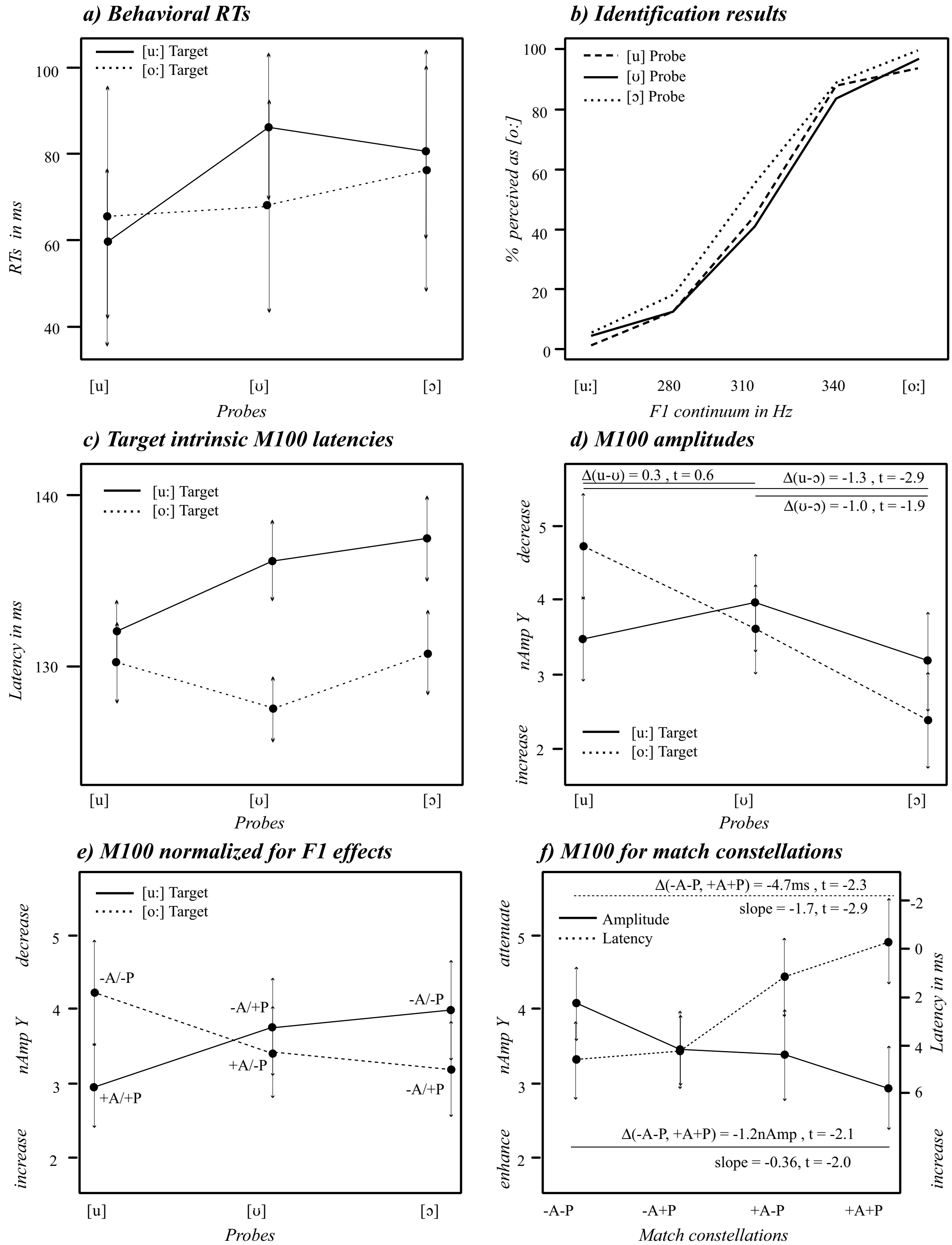

Figure IV. 3: *Results of the behavioral and neural tests. Whiskers indicate the confidence interval. When possible, the results of the mixed-effects analysis are presented. Δ indicates the difference between two factors.*

Behavioral test: ***a)*** *Response times to the categorization of the target as /u/ or as /o/ affected by the probes.* ***b)*** *Identification results of the continuum between the two /u/ and /o/ targets. No significant effect of the probes was found.*

Results

Behavioral response times (RT) and identification results.

Figure IV. 4a shows behavioral RTs. A mixed-effects analysis with an interaction between TARGETS and PROBES did not find any significant main effects of either factor. The identification test showed that subjects perceived the stimulus at the left edge of the continuum as [uː] and the stimulus at the right edge as [oː] (Figure IV. 3a). These stimuli served as targets in the MEG experiment. A mixed-effects analysis did not find any significant effects of the PROBES on the location of the category boundaries, nor on the slope of the psychometric function. Mean category boundary between [uː] and [oː] was located at 310 Hz (SD = 0. 3).

M100 Latencies

M100 latencies were found by looking for the position of the minimal amplitude in the time window mentioned above expanded by 10 ms to the left and right. A mixed-effect model with TARGET (/uː/, /oː/) as the main factor found that the M100 was significantly faster for the /oː/ target than for the /uː/ target ($\Delta_{([\text{u:}],[\text{o:}])}$ = 5. 6ms, t = 3. 6). After subtracting the TARGET intrinsic effects (test – control), no significant effect due to PROBE was found. Model comparison showed that a model with PROBE as fixed factor did not achieve a significantly better fit than a model with F1FREQUENCY (χ^2 = 1. 8, p = 0. 86). Rerunning the analysis with F1FREQUENCY and TARGET as fixed factors found additionally a marginal main effect of F1FREQUENCY (Δ_{Slope} = 0. 01, t = 1. 74).

In order to analyze whether within-categorical changes affected the M100 latencies in the target ([u] vs. [ʊ] in Figure IV. 3c), an additional mixed-effects analysis was performed, excluding M100 latencies affected by the [ɔ] probe. No main effects were found, but the interaction between the two probes and the two targets yielded significance ($\Delta_{slopes\ (/u/,/o/)}$ = 6. 9, t = 2. 1).

An analysis of M100 latencies with regard to the match/no-match constellations shows a general trend to decrease the latency with the amount of matching information (Figure IV. 3f). The additional mixed-effect analysis with MATCHCONSTELLATION (see Figure IV. 3f) as fixed effect and HEMISPHERE as random slope found that both were significant, i.e. the difference between a full match and no match, as well as the decreasing trend.

Amplitudes

Figure IV. 2c shows the time courses of the dipole moments for the two targets. Since no differences between the hemispheres were found (see analysis below), left and right hemisphere were collapsed for better clarity. An initial mixed-effect analysis with PROBES, TARGETS and HEMISPHERE as fixed effects showed that TARGET and HEMISPHERE had no significant effect on the M100 amplitude. Significant effects of PROBE on the M100 amplitude are shown in Figure IV. 3d. Looking at Figure IV. 3d where the x-axis was arranged according to the F1 frequency of the probe, it seems as though the higher the F1 frequency, the stronger the M100 are enhanced (i.e. the lower is Ψ). A comparison between two models with PROBE and F1FREQUENCY as fixed factors, respectively, showed that the model with PROBE did not achieve a better fit than the less complex model with F1FREQUENCY ($\chi^2 = 6.1$, $p = 0.3$). The analysis was repeated with a F1FREQUENCY*TARGET with F1FREQUENCY centered to zero to calculate to correct intercept. It showed that F1FREQUENCY affected M100 amplitudes significantly (slope = -0. 008 nAmp/ms, $t = -3.7$) and that it showed a significant interaction with TARGET ($\Delta_{(/u:/,\ /o:/)} = 0.007$, $t = 2.3$). Hence, the linear effect of F1FREQUENCY was different depending on the target. This was probably a result of the amplitude enhancement for the target [uː] in comparison to [oː] when probed with [u] (Student's T-test: $\Delta_{(/u:/,\ /o:/)} = -1.3$, $t = -1.6$, $p = 0.1$), and the amplitude attenuation when the targets were probed with [ɔ] (Student's T-test: $\Delta_{(/u:/,\ /o:/)} = 0.8$, $t = 1.0$, $p = 0.3$). In order to evaluate the effect of the different probes without the effect of linear increasing frequency, M100 amplitudes were normalized for the F1 frequency effect (Figure IV. 3e) by subtracting the mean amplitude in each probe bin and adding the grand average. Normalization resulted in *increasing* M100 amplitudes for the target /oː/, but *decreasing* M100 amplitudes for the target /uː/ when F1 frequency were enhanced. This orthogonal pattern was supported by a significant TARGET*F1FREQUENCY interaction (slope $\Delta_{(/u:/,\ /o:/)} = -0.007$ nAmp/ms, $t = -2.2$). Looking at the confidence intervals in Figure IV. 3e it was obvious that post hoc tests which analyzed the differences between the M100 amplitudes due to the probes for each target would not yield any significance. This is why the interpretation and discussion of the present results will rely on the linear trends due to the probes as shown by the significant TARGET*F1FREQUENCY interaction.

In order to analyze whether within-categorical changes affected the M100 amplitudes in the target ([u] vs. [ʊ] in Figure IV. 3d), an additional mixed-effects analysis was performed, excluding M100 amplitudes affected by the [ɔ] probe. Unlike the latencies, the amplitudes missed a significant interaction between the two probes and the two targets ($\Delta_{slopes\ (/u/,/o/)}$ = 1. 6, t = 1. 6), as well as main effects of PROBE ($\Delta_{([u],[u])}$ = -1. 1, t = -1. 6) and TARGET ($\Delta_{([u:],[o:])}$ = -1. 3, t = -1. 9).

An analysis of M100 amplitudes with regard to the match/no-match constellations shows a general trend to enhance the amplitude with the amount of matching information (Figure IV. 3f). An additional mixed-effect analysis with MATCHCONSTELLATION (see Figure) as fixed effect and HEMISPHERE as random slope found that both, the difference between a full match and no match, as well as the increasing trend were significant. A mixed-effects analysis with amplitudes as responses and M100 latencies as a covariate, including covariate random slopes for each subject, showed a significant negative correlation between normalized M100 amplitudes and latencies (slope = 1. 4 nAmp/, t = 3. 6). I.e. the later M100 peaked, the weaker it was.

Discussion

The aim of the present study was twofold. The broader aim was to investigate whether a single segment in a word could prime the processing of a single segment in another word. The narrower aim was to investigated the sensitivity to within and across-categorical vowel changes. This was done by using disyllabic nonwords (/g_bə/) in a probe-target sequence by means of a behavioral identification task and a passive listening task in the MEG. The short vowel in the probe's first syllable did or did not match in acoustic (F1 frequency) or phonologic information (height feature) with the long vowel in the target (see Table IV. 1b-c for matching constellations). While no significant results could be obtained in the behavioral test, the match/no-match constellations affected significantly the M100 response to the target regarding amplitudes and latencies. A match constellation enhanced the M100 amplitude and attenuated latencies in contrast to a no-match constellation. In the following, the results are discussed.

Priming by a single speech segment

In the behavioral test, subjects had to identify the vowel in the target in probe-target sequences as [uː] or [oː], ignoring the probe. In such a design, response times (RT) to the target are generally inhibited and inhibition correlates positively with the overlap between probe and target (Mayr & Buchner, 2007; Tipper, 2001). However, no significant effects of the probes neither on the response times (Figure IV. 3a), nor on the location of the category boundary (Figure IV. 3b) could be found. One might argue that the negative result in the present study was due to the large amount of speech material between the vowels in question. In their production priming task, Radeau et al. (1998) found priming effects of an phonological overlap only in monosyllabic words, but not in disyllabic words. They argued that the information of the first syllable was deactivated during the processing of the second. On the other hand, Fear et al. (1995) found in their lexical decision task that an overlap in the first syllable inhibited, in the second syllable primed response times. Importantly, studies investigating phonological priming effects, be it in production or perception, concentrated on the amount of overlap between probe and target using large sets of different words. So far, no study investigated the priming effects of a single phoneme and its phonologic and phonetic information. Thus, the present finding cannot be compared directly with previous studies.

In the MEG experiment, the effects of the probes were analyzed by means of the latency and the amplitude of the neuromagnetic response occurring ~100 ms (M100) after the vowel onset in the time courses of dipole moments (Figure IV. 2). The M100 was chosen since it has been identified to reflect vocalic category information with regard to its latency (Roberts et al., 2004a) and generator location (Obleser et al., 2004; Scharinger et al., 2011). Effects of acoustic or phonologic information match/mismatch were assumed to manifest in this time window.

In line with previous studies (Roberts et al., 2004a; Roberts, Ferrari, Stufflebeam & Poeppel, 2000), the M100 peaked earlier in the target containing [oː$^{370\text{Hz}}$] than in the one containing [uː$^{250\text{Hz}}$] irrespective of the prime (Figure IV. 3c). The difference can be attributed to the higher information density in higher F1 frequencies which need a shorter integration in contrast to lower frequencies (Alain et al., 1997). M100 peak latency was slightly affected by the F1 frequency in the probe's vowel: The higher the F1 frequency was, the later peaked the M100 in the target (Figure IV. 3c).
After the subtraction of target intrinsic latency differences, the latency correlated negatively with the information overlap: The more information matched between probe and target, the faster M100 peaked (Figure IV. 3e).
As hypothesized, the target's M100 could be manipulated by the acoustic and phonologic information contained in the probe's vowel. Primarily, the target's M100 amplitude correlated significantly with the F1 frequency in the probe's vowel (Figure IV. 3d). Enhancement effects in probe-target constellations have been shown in recent studies using frequency modulated tone sweeps and band pass filtered noise (Heinemann, Rahm, Kaiser, Gaese & Altmann, 2010; Soeta, Shimokura & Nakagawa, 2008). To the knowledge of the author, however, no

study so far has reported a correlation like the present one. An attempt to explain these findings would be very vague, which is why the discussion concentrates on the M100 target amplitudes which were normalized for F1 frequency effects. Like the M100 latencies, the resulting amplitudes reflected the acoustic and phonological match / no-match constellations between the probe and the target (Figure IV. 3e&f). The authors are fully aware that the effects shown in Figure IV. 3e&f were relatively small and would not yield significance when tested pair wise. Although it has been shown that the integration window for vowels can be as small as 30 ms (Yrttiaho, Tiitinen, Alku, Miettinen & May, 2010), short vowels convey significantly less information than longer vowels (Wallace et al., 2009). Thus, one reason for the small size of the effects is that the auditory system did not have the time to extract sufficient information from the short vowel in the probe.

Generally, priming studies report that the neural responses to the target are attenuated (Bergerbest et al., 2004; Brown et al., 1993; Gruber et al., 2004; Gruber et al., 2005a; McDonald, Thesen, Carlson, Blumberg, M.Girard, Trongnetrpunya, Sherfey, Devinsky, Kuzniecky, Dolye, Cash, Leonard, Jr., M.Dale & Halgren, 2010; Radeau et al., 1998). The present findings rather reflect those of Bonte et al. (2004), who found that the encephalographic N100 amplitude was enhanced by an overlap between probe and target when stimuli were non-words, in contrast to real words. However, these studies did not concentrate on one single segment not to mention on one feature.

In the present study, M100 amplitudes were enhanced by a full match and attenuated by full no-match between probe and target. Constellations with only one match, be it in F1 frequency or in the height feature resulted in average amplitudes (Figure IV. 3f). This amplitude pattern correlated negatively with M100 latencies. In summary, the present study reports contradictory results with respect to priming: Facilitation regarding peak latencies and inhibition regarding processing activity. Grill-Spector, Henson & Martin (2006) propose two models to account for faster processing of the target. The *sharpening model* states that less neurons become activated by the target to specify the target’s information. In contrast to the probe, the ‘read out’ (sic!) of a smaller number of neurons responding to the target is faster. The *facilitation model* states that the information is gathered faster during the processing of the target, since it was already present due to the probe, also resulting in earlier responses. However, both models predict weaker neural responses to the target, which was not the case in the present study.

There might be another, though very speculative explanation regarding the present M100 amplitudes when one equates their enhancement to an inhibitory process. Inhibition regarding response times has been shown in behavioral priming experiments when subjects were instructed to ignore the probe and answer only to the target (Mayr et al., 2007; Tipper, 2001). Slower response times are seen as the result of an active suppression process of the information from the ignored stimulus. When probe and target are equal regarding their information, the suppression of the information has to be terminated and the information has to be re-established for processing. This, of course, takes time, resulting in slower response times. In the present behavioral task, subjects were instructed to ignore the probe and answer to the target. Although there was no significant inhibitory effect in the behavioral task, subjects may have transferred the 'ignore-answer' strategy to the passive listening task in the MEG. Like in the negative priming paradigms, this resulted in an inhibition of the processing, thus, enhancement of the M100 amplitudes.

Regarding the choice of experimental paradigm, the present study decided to use a priming instead of an oddball design for two reasons. i) In the oddball design, effects are studied by means of the mismatch negativity (MMN) emitted by a infrequent deviant stimulus in a row of frequent standard stimuli (Pakarinen, Lovio, Huotilainen, Alku, Näätänen & Kujala, 2009). Taking into account the number of combinations in the present study ~3500 trials would be needed with a many-to-one ratio of 80-20%. The present study needed 816 trials. ii) The lack of the mismatch negativity, that starts to peak ~100ms after the mismatch made it possible to study effects on the early deflections like the M100.

Implications of the results for the phonology of German vowel length

In a behavioral discrimination test, study III has shown that within-categorical discrimination ([u] vs. [ʊ]) was not possible. Subjects differentiated between short vowels on the across-categorical basis (/u/ vs. /o/). However, identification results showed that the auditory system was sensitive enough to F1 frequency changes since though minimally, the location of the short-long boundary was changed by higher F1 frequencies (Figure IV. 1b, compare short-long boundary for F1 = 250Hz, 270Hz, 290Hz). The present results in the MEG experiment support such a sensitivity, since general changes in M100 amplitude have been found even for within-categorical F1 changes (Figure IV. 3c-e). How can this be brought in accordance with the discrimination results in study III?

It was argued in study III that in the discrimination test phonetic within-categorical differences between /u/ and /ʊ/ were leveled out because the disyllabic word caused a high memory load (Pisoni, 1973b). The discrimination was thus performed on the basis of phonological features, resulting in the comparison of across-categorical items. Within-categorical quality differences, however, function as an additional non-phonological cue for the processing of duration as Gussenhoven (2007) has shown that vowel height correlates with the perception of length: Equally long vowels are perceived shorter the lower they are. Taking this into account, lowering the short vowel in contrast to the long vowel, and by these means increasing its F1 frequency (Figure IV. 1a), enhances its perception as short. Although the within-categorical change in quality is not perceived in a phonological way, it serves as an additional cue to the perception of vowel length as a categorical contrast (see results from study I to study III). This is why the auditory system has to be sensitive enough to extract allophonic changes in F1 frequency.

Conclusion

Study IV investigated to what extent an acoustic and phonological information match in a single segment between a probe and a target stimulus affected the processing of the target. Disyllabic non-sense words (/ɡ_bə/) were used, containing a short vowel in the probe and a long vowel in the target, which changed within the category (lax vs. tense quality) or between two categories (/u/ vs. /o/).

Regarding the neural correlates, M100 latencies showed a facilitation effect: A match constellation decreased, a no-match constellation increased peak latencies. On the other hand, M100 amplitudes showed an inhibition effect: A match constellation enhanced, a no-match constellation attenuated M100 amplitudes. The present results showed that i) even single segments in a word can have priming effects and ii) that the auditory system was sensitive enough to extract acoustic and phonological information even in short vowels. This sensitivity is necessary in order to use the tense-lax quality distinction as an additional cue for categorical vowel length.

Study V: Coping with quantity. A priming study on German vowel length.

Introduction

While the phenomenon of priming is widely investigated in terms of semantic categories, the quality and extent of the effect is still unclear in phonetics. Priming can be described in terms of spreading activation or inhibition in a cognitive model that depicts categories within a network (Collins & Loftus, 1975). Since phonemic priming is a matter of categories of form, it presumably differs in some aspects from the classical semantic priming, which is based on meaning categories. There is for example evidence that not only reaction times are influenced by phonemic primes, but also the perception of the target is altered. (Goldinger, 1998, P. 952) defines: „Phonemic priming is observed when perception of a spoken target (e.g., nurse) is modified by prior perception of a related prime (e.g., north), relative to an unrelated prime (e.g., wedge) “. As phonemic contrast we chose the German /a/ - /a:/ contrast. Although vowel perception in general is discussed to be less categorical, for this specific durational contrast there is evidence for a clear categorical perception (Tomaschek et al., 2011). This is crucial, since priming is presumably based on categorical perception. The /a/ - /a:/ contrast is purely durational, allowing us to use speaking rate adaption as a second method to influence phoneme perception. It has been shown that changes in speaking rate lead to shifts in the category boundaries of phonemic contrasts with durational features like voice-onset time and vowel quantity (Miller, 1981)

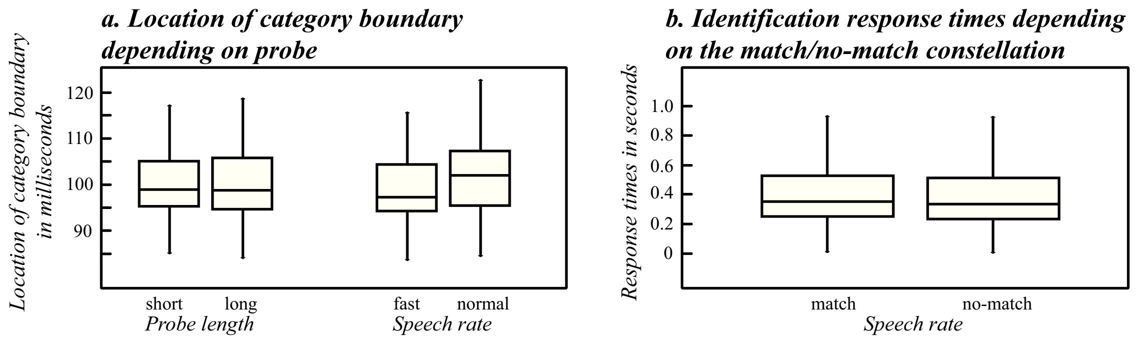

Figure V. 1: Results from Experiment V. Whiskers indicate standard error, boxes the confidence interval and lines in the boxes the group mean.

Methods and Stimuli

The non-word /ta:ke/ was recorded and used to create a ten-step target continuum with vowel durations from ~150ms to ~45ms. The set of primes consisted of 20 bisyllabic german words, ten of them had a long /a:/ as stressed vowel (e.g. /kna:be/), the other half a short /a/ (e.g. /krabe/). Each prime was recorded in two different speaking rates, with a slow rate of 80 and a fast rate of 120 words per minute. As neutral prime a beep of 100ms length was synthesized.

The stimuli were presented in pairs of prime and target, with the prime being either the beep, a long-a word or a short-a word (ISI: 300-350ms). 20 participants performed an identification task for the targets and the reaction time was measured. The stimuli were grouped in two blocks, one of them had only fast spoken primes, one of them only slowly spoken primes. Regarding the above mentioned category activation, prototypical versions of the targets are of primary interest for the analysis of reaction times rather than ambiguous regions in a continuum. We considered the two shortest tokens of the continuum to represent clear members of the category 'short' and the two longest tokens as clear long-vowel representatives. For the analysis a prime-target pair with the same duration (short-short, long-long) was considered as match, opposite pairs (short-long, long-short) as mismatch. Results were analyzed by means of a mixed-effect model using prime duration and speaking rate as main effects and subjects as random effects for category boundaries and prime-target match and speaking rate as main effects for reaction times.

Results

The identification task produced typical s-shaped curves. The category boundary was not significantly affected by short and long primes (Figure V. 1a, $\Delta_{(short,long)} \approx$ 1ms, t = -0. 69). Speaking rate, on the other hand, produced a significant effect (Figure V. 1b, location of the category boundary for fast primes ≈ 99ms, for slow primes ≈ 102ms, t = 2. 61). Reaction times for the whole continuum showed the expected pattern with a peak around the category boundary. Reaction times were slightly faster for normal speaking rate (Δ ≈ 13ms, t=- 1. 6). However, we found no priming effect for matching prime-target pairs (Figure V. 2, Δ ≈ 9ms, t = 0. 796), as well as for the interaction between the main factors (t=0. 325).

Conclusion

Study V found no effects of short and long primes on the location of the category boundary, but an effect of speech tempo. Furthermore, no facilitating nor inhibitory effect was found with respect to the reaction times. This allows two interpretations: i) Phonemic priming is not only rarely investigated but also highly controversial and seems to depend on many factors like length of the ISI, salience of the primed feature and the number of overlapping features. Moreover, evidence suggest an alternative effect, the so-called contrast effect (Lotto et al., 1998), which could be described as anti-priming. It is thus very unclear whether contrast or priming effects could occur in our paradigm if some parameters were changed. ii) The other possible interpretation explains the lack of priming effects with the nature of the durational vowel contrast in German. If this contrast is not perceived as categorical, it cannot be primed, at least following most priming accounts. For a further discussion of this eventuality it will be essential to conduct a discrimination task with our targets and to broadly investigate the general mechanism of phonemic priming.

References

Abrams D, Nicol T, Zecker S, Kraus N. (2008). Right-hemisphere auditory cortex is domant for coding syllable patterns in speech. *The Journal of Neuroscience,* 28:3958-65

Ackermann H, Lutzenberger W, Hertrich I. (1999). Hemispheric lateralization of the neural encoding of temporal speech features: a whole-head magnetencephalography study. *Cognitive Brain Research,* 7:511-8

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *PNAS,* 98:13367-72

Ainsworth WA. (1972). Duration as a Cue in the Recognition of Synthetic Vowels. *Journal of the Acoustical Society of America,* 51:648-51

Akin O, Belgin E. (2005). Hearing characteristics and frequency discrimination ability in musicians and nonmusicians. *Int. Adv. Otol,* 5:195-2ß2

Alain C, Woods D, Covarrubias D. (1997). Activation of duration-sensitive auditory cortical fields in humans. *Electroencephalography and clinical Neurophysiology,* 104:531-9

Alku P, Sivonen P, Palomäki K, Tiitinen H. (2001). The periodic structure of vowel sounds is reflected in human electromagnetic brain responses. *Neuroscience Letters,* 298:25-8

Antoniadis Z, Strube HW. (1984). Untersuchung zur spezifischen Dauer deutscher Vokale ("Study of the specific vowel duration in German Vowels"). *Phonetica,* 41:72-87

Baayen H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Chapter 7 pp. Cambridge: Cambridge University Press.P. 353.

Baddeley A. (1992). Working memory. *Science,* 255:556-9

Baddeley A. (2003). Working memory and language: an overview. *Journal of Communication Disorders,* 36:189-208

Barry RJ, Rushby JA, Johnstone SJ, Clarke AR, Croft RJ, Lawrence CA. (2004). Event-related potentials in the auditory oddball as a function of EEG alpha phase at stimulus onset. *Clinical Neurophysiology,* 115:2593-601

Becker T. (1998). *Das Vokalsystem der deutschen Standardsprache ("The Vowel System of the German Standard Language").* 91-138 pp. Frankfurt am Main, Germany: Peter Lang.

Bendixen A, Grimm S, Schröger E. (2005). Human auditory event-related potentials predict duration judgments. *Neuroscience Letters,* 383:284-8

Bennett DC. (1968). Spectral form and duration as cues in the recognition of English and German vowels. *Language and Speech,* 11:65-85

Bergerbest D, Ghahremani DG, Gabrieli JDE. (2004). Neural Correlates of Auditory Repetition Priming: Reduced fMRI Activation in the Auditory Cortex. *Journal of Cognitive Neuroscience,* 16:966-77

Blomert L, Mitterer H. (2004). The fragile nature of the speech-perception deficit in dyslexia: Natural vs. synthetic speech. *Brain and Language,* 83:21-6

Boemio A, From S, Braun A, Poeppel D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *nature neuroscience,* 8:389-95

Boersma P, Weenink D. (2009). *Praat: doing phonetics by computer (Version 5.1.05) [Computer program]*. http://www.fon.hum.uva.nl/praat/, accessed: 01.10.2009

Bohn O-S, Flege JE. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11:303-28

Bonte M, Blomert L. (2004). Developmental changes in ERP correlates of spoken word recognition during early school years: a phonological priming study. *Clinical Neurophysiology,* 115:409-23

Boutros NN, Belger A. (1999). Midlatency evoked potentials attenuation and augmentation reflect different aspects of sensory gating. *Biological Psychiatry,* 45:917-22

Braunschweiler N. (1997). Integrated Cues of Voicing and Vowel Length in German: A Production Study. *Language and Speech,* 40:353-76

Brown C, Hagoort P. (1993). The Processing Nature of the N400: Evidence from Masked Priming. *Journal of Cognitive Neuroscience,* 5:34-44

Bunton K, Story B. (2009). Identification of synthetic vowels based on selected vocal tract area functions. *Journal of the Acoustical Society of America,* 125:19-22

Cadenhead KS, Light GA, Geyer MA, Braff DL. (2000). Sensory Gating Deficits Assessed by the P50 Event-Related Potential in Subjects with Schizotypal Personality Disorder. *Am J Psychiatry,* 157:55-9

Caratini E. (2007). There is no ambisyllabicity (in German). In *14th Conference of the Student Organisation of Linguistics in Europe [ConSOLE XIV] – Decembre 15th-17th 2005. [http://www.sole.leidenuniv.nl/content_docs/ConsoleXIV2005pdfs/published_PDFs/consolе14-caratini.pdf]*, ed. E Schoorlemmer, L Vicente, S Blaho, pp. 37-60. Vitoria-Gasteiz (Spain): Universiteit Leiden

Carney AE, Widin G, Viemeister N. (1977). Noncategorical perception of stop consonants differing in VOT. *J. Acoust. Soc. Am.,* 62:961-70

Cebrian J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics,* 34:372-87

Čeponienė R, Torki M, Alku P, Koyama A, Townsend J. (2008). Event-related potentials reflect spectral differences in speech and non-speech stimuli in children and adults. *Clinical Neurophysiology,* 119:1560-77

Collins AM, Loftus EF. (1975). A spreading-activation theory of semantic processing. *Psychological Review,* 82:407-28

Connolly JF, Phillips NA. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *J. Cognitive Neuroscience,* 6:256-66

Cummins F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics,* 37:16-28

Dai H, Micheyl C. (2011). Psychometric functions for pure-tone frequency discrimination. *Journal of the Acoustical Society of America,* 130:263-72

de Jong K. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics,* 32:493-516

Dehaene-Lambertz G. (1997). Electrophysiological Correlates of categorical phoneme perception in adults. *NeuroReport,* 8:919-24

Dehaene S. (2003). The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences,* 7:145-7

Deng S, Srinivasan R. (2010). Semantic and acoustic analysis of speech by functional networks with distinct time scales. *Brain Research,* 1346:132-44

Desroches AS, Newman RL, Joanisse MF. (2008). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *J. Cognitive Neuroscience,* 21:1893-906

DeWitt I, Rauschecker JP. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences,* 109:E505–E14

Diesch E, Iverson P, Kettermann A, Siebert C. (1999). Measuring the perceptual magnet effect in the perception of /i/ by German listeners. *Psychophysiological Research,* 62:1-19

Dittrich R, Reibisch G. (2006). An acoustic study of /r/-vocalization in word-final position. In *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel 37*, ed. J Harrington, C Mooshammer, F Kleber, pp. 19-26

Dumay N, Benraïss A, Barriol B, Colin C, Radeau M, Besson M. (2001). Behavioral and Electrophysiological Study of Phonological Priming between Bisyllabic Spoken Words. *Journal of Cognitive Neuroscience,* 13:121-43

Edmonds BA, James RE, Utev A, Vestergaard MD, Patterson RD, Krumbholz aK. (2010). Evidence for early specialized processing of speech formant information in anterior and posterior human auditory cortex. *Cognitive Neuroscience,* 32:684-92

Eggermont JJ, Ponton CW. (2002). The Neurophysiology of Auditory Perception: From Single Units to Evoked Potentials. *Audiology and Neurotology,* 7:71-99

Escudero P, Benders T, Lipski SC. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics,* 37:452-65

Escudero P, Boersma P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition,* 26:551-85

Eulitz C, Diesch E, Pantev C, Hampson S, Elbert T. (1995). Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. *The Journal of Neuroscience,* 15:2748-55

Eulitz C, Lahiri A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience,* 16:577-83

Eulitz C, Obleser J. (2007). Perception of acoustically complex phonological features in vowels is reflected in the induced brain-magnetic activity. *Behavioral and Brain Functions,* 3:26

Fear BD, Cutler A, Butterfield S. (1995). The strong/weak syllable distinction in English. *The Journal of the Acoustical Society of America,* 97:1893-904

Feldman NH, Griffiths TL. (2007). A rational account of the perceptual magnet effect. *Proceedings of the 29th Annual Conference of the Cognitive Science Society,*

Fellinger R, Klimesch W, Gruber W, Freunberger R, Doppelmayr M. (2011). Pre-stimulus alpha phase-alignment predicts P1-amplitude. *Brain Research Bulletin,* 85:417-23

Féry C. (1996). German Foot and Word Stress in OT. *Nordlyd,* 24:63 - 96

Féry C. (1998). German word stress in Optimality Theory. *Journal of Comparative Germanic Linguistics,* 2:101-42

Flege JE. (1988). Effects of speaking rate on tongue position and velocity of movement in vowel production. *The Journal of the Acoustical Society of America,* 84:901-16

Flege JE, Munro MJ. (1994). Auditory and catgorical effects on crosslanguage vowel perception. *Journal of the Acoustical Society of America,* 95:3623-41

Flege JE, Munro MJ, MacKay IRA. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication,* 16:1-26

Formisano E, Kim D-S, Di Salle F, van de Moortele P-F, Ugurbil K, Goebel R. (2003). Mirror-Symmetric Tonotopic Maps in Human Primary Auditory Cortex. *Neuron,* 40:859-69

Friedrich CK, Lahiri A, Eulitz C. (2008). Neurophysiological evidence for underspecified lexical representations: Asymmetries with word initial variations. *Journal of Experimental Psychology: Human Perception and Performance,* 34:1545-59

Fry D, Abramson A, Eimas P, Liberman A. (1962). The Identification and Discrimination of Synthetic Vowels. *Language and Speech,* 5:171-89

Gay T. (1978). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America,* 63:223-30

Gerrits E, Schouten M. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics,* 66:363-76

Goldinger S. (1998). Signal detection comparisons of phonemic and phonetic priming: The flexible-bias problem. *Perception & Psychophysics,* 60:952-65

Goldstone RL, Hendrickson AT. (2009). Categorical perception. *Cognitive Science,* 2:69-78

Gottfried TL, Miller JL, Payton PE. (1990). Effect of Speaking Rate on the Perception of Vowels. *Phonetica,* 47:155-72

Grabe E. (1998). Pitch accent realization in English and German. *Journal of Phonetics,* 26:129-43

Greenberg S, Carvey H, Hitchcock L, Chang S. (2003). Temporal properties of spontaneous speech - a syllable.centric perspective. *Journal of Phonetics,* 31:465-85

Grill-Spector K, Henson R, Martin A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences,* 10:14-23

Gruber T, Malinowski P, Müller MM. (2004). Modulation of oscillatory brain activity and evoked potentials in a repetition priming task in the human EEG. *European Journal of Neuroscience,* 19:1073-82

Gruber T, Müller MM. (2005a). Oscillatory Brain Activity Dissociates between Associative Stimulus Content in a Repetition Priming Task in the Human EEG. *Cerebral Cortex,* 15:109-16

Gruber WR, Klimesch W, Sauseng P, Doppelmayr M. (2005b). Alpha Phase Synchronization Predicts P1 and N1 Latency and Amplitude Size. *Cerebral Cortex,* 15:371-7

Gussenhoven C. (2007). A vowel height split explained: Compensatory listening and speaker control. *Laboratory Phonology 9*, pp. 145-72

Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ. (2005). Neuromagnetic Correlates of Streaming in Human Auditory Cortex. *The Journal of Neuroscience,* 25:5382-8

Hakkarainen H. (1995). *Die Phonetik des Deutsche ("The phonetics of German").* München.

Hall TA. (1992). *Syllable structure and syllable-related processes in German.* 27ff & 74ff pp. Tübingen, Germany: Niemeyer.P. VIII, 246 S.

Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics,* 65:413-97

Hanson V. (1977). Within-category discriminations in speech perception. *Perception & Psychophysics,* 21:423 - 30

Heidolph KE. (1981). *Grundzüge einer deutschen Grammatik ["Essential features of a German grammar"].* KE Heidolph. Berlin: Akademie Verlag.

Heike G. (1970). Lautdauer als Merkmal der wahrgenommenen Quantitat, Qualitat und Betonung im Deutschen ("Sound duration as a feature of perceived quantity, quality and accent in German"). In *Proceedings of the Sixth International Congress of Phonetic Sciences Held at Prague, 1967.*, pp. 433-7. Prague, Czech Republic: Acad. Pub. House of the Czech. Acad. of Sciences

Heike G. (1972). Quantitative und qualitative Differenzen von /a(:)/ - Realisationen im Deutschen ("Quantitive and qualitive differences of /a(:)/ - Productions in German"). In *Proc. 7th Intern. Congr. Phonetic Sciences, Montreal, 1971*, pp. 725-9

Heinemann LV, Rahm B, Kaiser J, Gaese BH, Altmann CF. (2010). Repetition Enhancement for Frequency-Modulated but Not Unmodulated Sounds: A Human MEG Study. *PLoS ONE,* 5:e15548

Heinze HJ, Münte TF, Kutas M, Butler SR, Näätänen R, Nuwer MR, Goodin DS. (1999). Cognitive event-related potentials. In *Recommendations fot the practice of clinical neurophysiology: Guidelines of the International Federation of Clinical Physiology*, pp. 91-5: Elsevier Science

Heldner M, Strangert E. (2001). Temporal effects of focus in Swedish. *Journal of Phonetics,* 29:329-61

Hertrich I, Ackermann H. (1997). Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures. *The Journal of the Acoustical Society of America,* 102:523-36

Hertrich I, Ackermann H. (1999). A vowel synthesizer based on formant sinusoids modulated by fundamental frequency. [Download from www.fabian-tomaschek.de]. *Journal of the Acoustical Society of America,* 106:2988-90

Hertrich I, Ackermann H. (2007). Modeling voiceless speech segments by means of an additive procedure based on the computation of formant sinusoids. *Proceedings of the 6th ISCA Workshop on Speech Synthesis*, pp. 178-81. Bonn: University of Bonn

Hertrich I, Dietrich S, Trouvain J, Moos A, Ackermann H. (2012). Magnetic brain activity phase locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. *Psychophysiology,* 49:322-34

Hertrich I, Mathiak K, Lutzenberger W, Ackermann H. (2000). Differential impact of periodic and aperiodic speech-like acoustic signals on magnetic M50/M100 fields. *Neuroreport,* 11:4017-20

Hertrich I, Mathiak K, Lutzenberger W, Ackermann H. (2002). Hemispheric lateralization of the processing of consonant-vowel syllables (formant transitions): effects of stimulus characteristics and attentional demands on evoked magnetic fields. *Neuropsychologia,* 40:1902-17

Hertrich I, Mathiak K, Lutzenberger W, Ackermann H. (2004). Transient and phase-locked evoked magnetic fields in response to periodic acoustic signals. *Neuroreport,* 15:1687-90

Hickok G. (2001). Functional anatomy of speech perception and speech production: psycholinguistic implications. *Journal of Psycholinguistic Research,* 30:225-35

Hickok G, Buchsbaum B, Humphries C, Muftuler T. (2003). Auditory–Motor Interaction Revealed by fMRI: Speech, Music, and Working Memory in Area Spt. *Journal of Cognitive Neuroscience,* 15:673-82

Hickok G, Poeppel D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences,* 4:131-8

Hickok G, Poeppel D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition,* 92:67-99

Hillebrand A, Barnes G. (2002). A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. *Neuroimage,* 16:638-50

Hirata Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics,* 32:565-89

Hirata Y, Tsukada K. (2004). The Effects of speaking rates and vowel length on formant movements in Japanese. *Proceedings of the 2004 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception*, pp. 73-85. Somerville, MA

Hisagi M, Shafer VL, Strange W, Sussman ES. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures. *Brain Research,* 1360:89-105

Hoffmann A. (2011). *A comparison of native and non-native vowels featuring German as a target language: a statistical analysis using corpora*. http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-54494, Tübingen, Germany

Hose B, Langner G, Scheich H. (1983). Linear phoneme boundaries for German synthetic two-formant vowels. *Hearing Research,* 9:13-25

Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. (2002). Modulation of the Auditory Cortex during Speech: An MEG Study. *Journal of Cognitive Neuroscience,* 14:1125-38

Inouchi M, Kubota M, Ferrari P, Robert TPL. (2003). Magnetic mismatch fields elicited by vowel duration and pitch changes in Japanese words in humans: comparison between native- and non-speakers of Japanese. *Neuroscience Letters,* 2003:165-8

Inouchi M, Kubota M, Ohta K, Shirahama Y, Takashima A, Horiguchi T, Matsushima E. (2004). Human auditory evoked mismatch field amplitudes vary as a function of vowel duration in healthy first-language speakers. *Neuroscience Letters,* 366:342-6

Iverson I, Kuhl P. (1993). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *J Acoust Soc Am,* 97:553-62

Jaramillo M, Alku P, Paavilainen P. (1999). An event-related potential (ERP) study of duration changes in speech and non-speech sounds. *NeuroReport,* 10

Jiang J, Chen M, Alwanc A. (2006). On the perception of voicing in syllable-initial plosives in noise. *J Acoust Soc Am,* 119:1092-105

Johnson K, Flemming E, Wright R. (1993). The Hyperspace Effect: Phonetic targets are hyperarticulated. *Language,* 69:505-28

Jusczyk PW, Houston DM, Newsome M. (1999). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology,* 39:159-207

Kasai K, Yamada H, Kamio S, Nakagome K, Iwanami A, Fukuda M, Itoh K, Koshida I, Yumoto M, Iramina K, Kato N, Ueno S. (2001). Brain lateralization for mismatch response to across- and within-category change of vowels. *Cognitive Neuroscience and Neuropsychology,* 12:2467-72

Kaukoranta E, Hari R, Lounasmaa OV. (1987). Responses of the human auditory cortex to vowel onset after fricative consonants. *Experimental Brain Research,* 69:19-23

Keating PA. (1998). Underspecification in phonetics. *Phonology,* 5:275-92

Kenstowicz M. (2004). Generative Phonology. In *Encyclopedia of language & Linguistics*, ed. M Green, CH Reintges. Oxford: Elsevier.

Kirmse U, Ylinen S, Tervaniemi M, Vainio M, Schröger E, Jacobsen T. (2008). Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *International Journal of Psychophysiology,* 67:131-43

Kleber F, John T, Harrington J. (2010). The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics,* 38:185-96

Knaus J, Domahs U. (2009). Experimental Evidence for Optimal and Minimal Metrical Structure of German Word Prosody. *Lingua: International Review of General Linguistics,* 119:1396-413

Koelsch S, Schulze K, Sammler D, Fritz T, Müller K, Gruber O. (2009). Functional architecture of verbal and tonal working memory: An FMRI study. *Human Brain Mapping,* 30:859-73

Kohler K. (1995). *Einführung in die Phonetik des Deutschen ("Introduction to the phonetics of German").* Berlin.

Kohler KJDitpofalp, " Phonetica 36, 332-343. (1979).

Kuhl PK. (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

Kuhl PK. (2004). Early Language Acquisition: Cracking the Speech Code. *Nature reviews neuroscience,* 5:831 -43

Kuhl PK, Conboy BT, Coffey-Corina S, Padden D, Nelson MR-GT. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Phil. Trans. R. Soc. B,* 363:979-1000

Kujala A, Alho K, Service E, Ilmoniemi RJ, Connolly JF. (2004). Activation in the anterior left auditory cortex associated with phonological analysis of speech input: localization of the phonological mismatch negativity response with MEG. *Cognitive Brain Research,* 21:106-13

Kuriki S, Kanda S, Hirata Y. (2006). Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *The Journal of Neuroscience,* 26:4046-53

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,* 12:3777-81

Lahiri A, Reetz H. (2002). Underspecified recognition. In *Laboratory phonology 7*, ed. C Gussenhoven, N Warner, pp. 637-76. Berlin: Mouton

Lapid E, Ulrich R, Rammsayer T. (2008). On estimating the difference limen in duration discrimination tasks: A comparison of the 2AFC and the reminder task. *Perception & Psychophysics,* 70:291-305

Lehnert-LeHouillier H. (2007). The influence of dynamic F0 on the perception of vowel duration: cross-linguistic evidence. *ICPhS XVI*. Saarbrücken

Lehnert-LeHouillier H. (2010). A cross-linguistic investigation of cues to vowel length perception. *Journal of Phonetics,* 38:472-82

Lehtonen J. (1970). *Aspects of Quantity in Standard Finnish.* 85-152 pp. Jyväskylä, Finland: University of Jyväskylä.

Lenerz J. (2000). Zur sogenannten Vokalopposition im Deutschen ("About the so called vowel contrast in German"). *Zeitschrift fur Sprachwissenschaft,* 19:167-209

Leon MI, Shadlen MN. (2003). Representation of Time by Neurons in the Posterior Parietal Cortex of the Macaque. *Neuron,* 38:317-27

Liberman A, Mattingly I. (1985). The motor theory of speech perception revised. *Cognition,* 21:1-36

Liberman AM, Harris KS, Hoffman HS, Griffith BC. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology,* 54:358-68

Light G, Braff D. (2001). Measuring P50 Suppression and Prepulse Inhbition in a Single Recording Session. *Am J Psychiatry,* 158:2066-8

Lijffijt M, Lane SD, Meier SL, Boutros NN, Burroughs S, Steinberg JL, Gerard Moeller F, Swann AC. (2009). P50, N100, and P200 sensory gating: Relationships with behavioral inhibition, attention, and working memory. *Psychophysiology,* 46:1059-68

Lindblom B. (1963). Spectrographic Study of Vowel Reduction. *The Journal of the Acoustical Society of America,* 35:1773-81

Lindblom B. (1968). Temporal organization of syllable production. *Quarterly Progress and Status Report at the Dept. for Speech, Music and Hearing,* 9:1-5

Lindner G. (1976). Urteilsveränderung bei Vokalverkürzung ("Change of judgments as a result of vowel shortening"). *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung,* 29:407-14

Lisker L, Abramson A. (1964). A cross-languange study of voicing in initial stops: acoustical measurements. *Word,* 20:384-422

Lively SE, Pisoni DB. (1997). On prototypes and phonetic categories: A critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception and Performance,* 23:1665-79

Livingston K, Andrews J, Harnad S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 24:732-53

Lotto A, Kluender K, Holt L. (1998). Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America,* 103:3648-55

Lotze S, Gallmann P. (2008). Norm und Variation beim Konjunktiv II. In *Deutsche Grammatik: Regeln, Normen, Sprachgebrauch*, ed. M Konopka, B Strecker, pp. 222-39. Berlin: De Gruyter

Luo H, Poeppel D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron,* 54:1001-10

Macmillan, Creelman. (2004). *Detection Theory: A User's Guide.* Lawrence Erlbaum.

Maddieson I. (1984). *Patterns of Sounds.* 123-135 pp. Cambridge, New York: Cambridge University Press.

Maeda S. (1990). Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In *Speech production and speech modelling*, ed. WJ Hardcastle, A Marchal, pp. 131-49. Dordrecht: Kluwer Academic Publishers

Massaro DW, Cohen MM. (1983). Categorical or continuous speech perception: A new test. *Speech Communication,* 2:15-35

Mathiak K, Ackermann H, Rapp A, Mathiak KA, Shergill S, Riecker A, Kircher TTJ. (2011). Neuromagnetic oscillations and hemodynamic correlates of P50 suppression in schizophrenia. *Psychiatry Research: Neuroimaging,* 194:95-104

Mathiak K, Hertrich I, Lutzenberger W, Ackermann H. (1999). Preattentive processing of consonant vowel syllables at the level of the supratemporal plane: a whole-head magnetencephalography study. *Cognitive Brain Research,* 8:251-7

Mathiak K, Hertrich I, Lutzenberger W, Ackermann H. (2000). Encoding of temporal speech features (formant transients) during binaural and dichotic stimulus application:: A whole-head magnetencephalography study. *Cognitive Brain Research,* 10:125-31

May P, Tiitinen H. (2010). Mismatch negativity (MNN), the deviance-elicited auditory deflection, explained. *Psychophysiology,* 47:66-122

Mayr S, Buchner A. (2007). Negative priming as a memory phenomenon: A review of 20 years of negative priming research. *Zeitschrift für Psychologie/Journal of Psychology,* 215:35-51

McClelland JL, Elman JL. (1986). The TRACE model of speech perception. *Cognitive Psychology,* 18:1-86

McDonald CR, Thesen T, Carlson C, Blumberg M, M.Girard H, Trongnetrpunya A, Sherfey JS, Devinsky O, Kuzniecky R, Dolye WK, Cash SS, Leonard MK, Jr. DJH, M.Dale A, Halgren E. (2010). Multimodal imaging of repetition priming: Using fMRI, MEG, and intracranial EEG to reveal spatiotemporal profiles of word processing. *Neuroimage,* 53:707-17

McQueen JM. (1998). Segmentation of Continuous Speech Using Phonotactics. *Journal of Memory and Language,* 39:21-46

McQueen JM. (2007). Eight questions about spoken-word recognition. In *The Oxford handbook of psycholinguistics*, ed. MG Gaskell, pp. 37-53. Oxford: Oxford University Press

Mehler J, Dommergues JY, Frauenfelder U, Segui J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior,* 20:298-305

Meister L, Meister E. (2011). Perception of the short vs. long phonological category in Estonian by native and non-native listeners. *Journal of Phonetics,* 39:212-24

Menning H, Imaizumi S, Zwitserlood P, Pantev C. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learning & Memory,* 9:253–67

Menning H, Roberts L, Pantev C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *Auditory and Vestibular Systems,* 11:817-22

Mermelstein P. (1977). Difference limens for formant frequencies of steady-state and consonant-bound vowels. *J Acoust Soc Am,* 63:572-80

Metz DE, Allen K, Kling T, Maisonet S, McCullough R, Schiavetti N, Whitehead RL. (2006). Effect of syllable-initial voicing on vowel duration during simultaneous communication. *Journal of Communication Disorders,* 39:192-9

Meyer E. (1904). Zur vokaldauer im deutschen („About vowel duration in German"). In *Nordiska Studier Tillegnade Adolf Noreen*, pp. 347-56. Uppsala: Norken

Miller JL. (1981). Effects of speaking rate on segmental distinctions. In *Perspectives on the Study of Speech*, ed. PD Eimas, JL Miller. Hillsdale: Erlbaum Associates

Minagawa-Kawai Y, Mori K, Sato Y. (2005). Different Brain Strategies Underlie the Categorical Perception of Foreign and Native Phonemes. *Journal of Cognitive Neuroscience,* 17:1376-85

Miyawaki K, Strange W, Verbrugge R, Liberman A, Jenkins J, Fujimura O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception &Psychophysics,* 18:331-40

Moon S-J, Lindblom B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America,* 96:40-55

Moore BCJ. (1973). Frequency difference limens for short-duration tones. *The Journal of the Acoustical Society of America,* 54:610-9

Moulton WG. (1962a). *The sounds of English and German.* pp. 60 pp. Chicago and London: The University of Chicago Press.

Moulton WG. (1962b). *The sounds of English and German.* Chicago: Univ. of Chicago Pr.P. pp. 60.

Mutschler I, Wieckhorst B, Speck O, Schulze-Bonhage A, Hennig J, Seifritz E, Ball T. (2010). Time Scales of Auditory Habituation in the Amygdala and Cerebral Cortex. *Cerebral Cortex,* 20:2531-9

N'Diaye K, Ragot R, Garnero L, Pouthas V. (2004). What is common to brain activity evoked by the perception of visual and auditory filled durations? A study with MEG and EEG co-recordings. *Cognitive Brain Research,* 21:250-68

Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature,* 385:432-4

Näätänen R, Paavilainen P, Rinne T, Alho K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology,* 118:2544-90

Näätänen R, Paavilainen P, Titinen H, Jiang D, Alho K. (1993). Attention and mismatch negativity. *Psychophysiology,* 30:436-50

Nenonen S, Shestakova A, Huotilainen M, Näätänen R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cognitive Brain Research,* 16:492-5

Newman RL, Connolly JF. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology,* 80:114-21

Newman RL, Connolly JF, Service E, McIvor K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology,* 40:640-7

Nozaradan S, Peretz I, Missal M, Mouraux A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience,* 31:10234 –40

Nusbaum HS, Henly AS. (1992). Constraint Satistfaction, Attention and Speech Perception: Implications for Theories of Word Recognition. In *The auditory processingof speech: from sounds to words*, ed. MEH Schouten, pp. 339 - 48. Berlin, New York: Mouton de Gruyter

Obleser J, Eisner F, Kotz S. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *The Journal of Neuroscience,* 28:8116-24

Obleser J, Lahiri A, Eulitz C. (2004). Magnetic Brain Response Mirrors Extraction of Phonological Features from Spoken Vowels. *Journal of Cognitive Neuroscience,* 16:31-9

Oldfield R. (1971). The assessement and analysis of handedness: The Edinburgh inventory. *Neuropsychologia,* 9:97-113

Paavilainen P, Jiang D, Lavikainen J, Näätänen R. (1993). Stimulus duration and the sensory memory trace: An event-related potential study. *Biological Psychology,* 35:139-52

Pakarinen S, Lovio R, Huotilainen M, Alku P, Näätänen R, Kujala T. (2009). Fast multi-feature paradigm for recording several mismatch negativities (MMNs) to phonetic and acoustic changes in speech sounds. *Biological Psychology,* 82:219-26

Pantev C, Hoke M, Lehnertz K, Lütkenhöner B. (1989a). Neuromagnetic evidence of an amplitopic organization of the human auditory cortex. *Electroencephalography and clinical Neurophysiology,* 72:225-31

Pantev C, Hoke M, Lutkenhoner B, Lehnertz K. (1989b). Tonotopic organization of the auditory cortex: pitch versus frequency representation. *Science,* 246:486-8

Peelle JE, Gross J, Davis MH. (2012). Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cerebral Cortex,*

Phillips C, Pellathy T, Marantz A, Yellin E, Wexler K, Poeppel D, McGinnis M, Roberts T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience,* 126:1038-55

Pisoni D. (1973a). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics,* 13:253-60

Pisoni D. (1973b). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Attention, Perception, & Psychophysics,* 13:253-60

Pisoni D, Lazarus JH. (1974a). Categorical and noncategorical modes of speech perception along the voicing continuum. *J. Acoust. Soc. Am.,* 55:328-33

Pisoni D, Tash J. (1974b). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics,* 15:285-90

Poeppel D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication,* 41:245-55

Poeppel D, Guillemin A, Thompson J, Fritz J, Bavelier D, Braun AR. (2004). Auditory lexical decision, categorical perception, and FM direction discrimination differentially engage left and right auditory cortex. *Neuropsychologia,* 42:183-200

Poeppel D, Idsardi W. (2011). Recognizing words from speech: the perceptionaction-memory loop. In *Lexical Representation - a multidisciplinary approach*, ed. G Gaskell, P Zwitzerlood. Berlin/New York: de Gruyter

Poeppel D, Phillips C, Yellin E, Rowley HA, Roberts TPL, Marantz A. (1997). Processing of vowels in supratemporal auditory cortex. *Neuroscience Letters,* 221:145-8

Poeppel D, Yellin E, Phillips C, Roberts T, Rowley H, Wexler W, Maranth A. (1996). Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Cognitive Brain Research,* 4:231-42

Pompino-Marschall B. (2003). *Einführung in die Phonetik ("Introduction into phonetics").* 324 pp. Berlin / New York: Walter de Gruyter.

R_Development_Core_Team. (2010). R: A Language and Environment for Statistical Computing. *edited by R. F. f. S. Computing (Vienna, Austria),*

Radeau M, BEsson M, Fonteneau E, Castro SL. (1998). Semantic, repetition and rime priming between spoken words: behavioral and electrophysiological evidence. *Biological Psychology,* 48:183-204

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund H-J, Zilles K. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *NeuroImage,* 13:669-83

Ramers K-H. (1988). *Vokalquantität und -qualität im Deutschen ("Vowel quantity and quality in German")*. Niemeyer, Tübingen, Germany. IX, 251 S. pp.

Ramers K-H, Vater H. (1995). *Einführung in die Phonologie ("Introduction to phonology").* Hürth.

Rammsayer T. (2010). Differences in duration discrimination of filled and emtpy auditory intervals as a function of base duration. *Psychophysics,* 72:1591-600

Ramscar M, Yarlett D, Dye M, Denny K, Thorpe K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science,* 34:909–57

Reis M. (1974). *Lauttheorie und Lautgeschichte : Untersuchungen am Beispiel d. Dehnungs- u. Kürzungsvorgänge im Deutschen ("Theory and History of phonems").* Fink, München, Germany. 336 S. pp.

Roberts T, Flagg E, Gage N. (2004a). Vowel categorization induces departure of M100 latency from acoustic prediction. *Cognitive Neuroscience and Neuropsychology,* 15:1679-82

Roberts T, Poeppel D. (1996). Latency of auditory evoked M100 as a function of tone frequency. *Neuroreport,* 7:1138-40

Roberts TPL, Ferrari O, Stufflebeam SM, Poeppel D. (2000). Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. *Journal of Clinical Neurophysiology,* 17:114-29

Roberts TPL, Flagg EJ, Gage NM. (2004b). Vowel categorization induces departure of M100 latency from acoustic prediction. *Neuroreport,* 15:1679-82

Rose MM, Moore BCJ. (2000). Effects of frequency and level on auditory stream segregation. *The Journal of the Acoustical Society of America,* 108:1209-14

Ross B, Borgmann C, Draganova R, Roberts LE, Pantev C. (2000). A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *The Journal of the Acoustical Society of America,* 108:679-91

Rugg M. (1999). Functional neuroimaging in cognitive neuroscience. In *The neuroimaging of language*, ed. CM Brown, P Hagoort, p. 15ff. Oxford: Oxford University Press

Sams M, Aulanko R, Aaltonen O, Näätänen R. (1990). Event-Related Potentials to infrequent Changes in Synthesized Phonetic Stimuli. *Journal of Cognitive Neuroscience,* 2:344-57

Sams M, Kaukoranta E, Hämäläinen M, Näätanen R. (1991). Cortical Activity Elicited by Changes in Auditory Stimuli: Different Sources for the Magnetic N1OOm and Mismatch Responses. *Psychophysiology,* 28:21-9

Scharinger M, Idsardi WJ, Poe S. (2011). A Comprehensive Three-dimensional Cortical Map of Vowel Space. *Journal of Cognitive Neuroscience,* 23:3972-82

Schielzeth H, Forstmeier W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behav Ecol,* 20:416–20

Schouten B, Gerrits E, Hessen Av. (2003a). The end of categorical perception as we know it. *Speech Communication,* 41:71-80

Schouten B, Gerrits E, van Hessen A. (2003b). The end of categorical perception as we know it. *Speech Communication,* 41:71-80

Sendlmeier W, Seebode J. (2006). *Formantkarten des deutschen Vokalsystems ("Formant maps of the German vowel system").* www.kw.tu-berlin.de/fileadmin/a01311100/Formantkarten_des_deutschen_Vokalsystems_01.pdf, accessed: 14.05.2012

Sendlmeier WF. (1981). Der Einfluss von Qualität und Quantität auf die Perzeption betonter Vokale des Deutschen („The Influence of Quality and Quantity on the Perception of stressed German Vowels“). *Phonetica: International Journal of Speech Science,* 38:291-308

Shahin AJ, Roberts LE, Miller LM, McDonald KL, Alain C. (2007). Sensitivity of EEG and MEG to the N1 and P2 auditory evoked responses modulated by spectral complexity of sounds. *Brain Topogr,* 20:55-61

Sheehan K, McArthur G, Bishop D. (2005). Is discrimination training necessary to cause changes in the P2 auditory event-related brain potential to speech sounds? *Cognitive Brain Research,* 25:547-53

Slowiaczek LM, McQueen JM, Soltano EG, Lynch M. (2000). Phonological Representations in Prelexical Speech Processing: Evidence from Form-Based Priming. *Journal of Memory and Language,* 43:530-60

Smits R, Sereno J, Jongman A. (2006). Categorization of Sounds. *Journal of Experimental Psychology: Human Perception and Performance,* 32:733 - 54

Soeta Y, Shimokura R, Nakagawa S. (2008). Effects of center frequency on binaural auditory filter bandwidth in the human brain. *Neuroreport,* 19:1709-13 10.097/WNR.0b013e3283177f99

Stefanics G, Hangya B, Hernádi I, Winkler I, Lakatos P, Ulbert I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *The Journal of Neuroscience,* 30:13578-85

Sternefeld W. (2007). *Syntax - Eine morphologisch motivierte generative Beschreibung des Deutschen. Band 1.* 10-12 pp. Tübingen Stauffenburg Verlag P. 477.

Stevens K, Blumstein S. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America,* 64:1358-68

Strange W. (1987). Information for vowels in formant transitions. *Journal of Memory and Language,* 26:550-7

Strange W, Bohn O-S. (1998). Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies. *Journal of the Acoustical Society of America,* 104:488-504

Strangert E. (2001). Quantity in ten Swedish dialects in Northern Sweden and Österbotten in Finland. *Papers from FONETIK 2001, Working Papers*, pp. 144-7. Lund: Department of Linguistics and Phonetics, Lund University

Strangman G, Boas DA, Sutton JP. (2002). Non-invasive neuroimaging using near-infrared light. *Biological Psychiatry,* 52:679-93

Studdert-Kennedy M, Liberman AM, Harris KS, Cooper FS. (1970). The Motor Theory of Speech Perception: A reply to lane's critical review. *Psychological Review,* 70:234-49

Summers WV, Pisoni DB, Bernacki RH, Pedlow RI, Stokes MA. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America,* 84:917-28

Tiitinen H, Sivonen P, Alku P, Virtanen J, Näätänen R. (1999). Electromagnetic recordings reveal latency differences in speech and tone processing in humans. *Cognitive Brain Research,* 8:355-63

Tipper. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views.

Tomaschek F, Truckenbrodt H, Hertrich I. (2011). German Vowel Quantity: Categorical Perception or Perceptual Magnet Effect? In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, pp. 2002-5. Hong Kong.

Tong Y, Melara RD, Rao A. (2009). P2 enhancement from auditory discrimination training is associated with improved reaction times. *Brain Research,* 1297:80-8

Toscano JC, McMurray B, Dennhardt J, Luck SJ. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science,* 21:1532 - 40

Tremblay KL, Shahin AJ, Picton T, Ross B. (2009). Auditory training alters the physiological detection of stimulus-specific cues in humans. *Clinical Neurophysiology,* 120:128-35

UPSID. (accessed: 25.08.2012). UCLA Phonological Segment Inventory Database. http://web.phonetik.uni-frankfurt.de/upsid.html,

Vater H. (1992). Zum Silben-Nukleus im Deutsche. In *Silbenphonologie des Deutschen*, ed. P Eisenberg, KH Ramers, H Vater, pp. 247 - 84. Tübingen: Gunter Narr Verlag

Vihla M, Eulitz C. (2003). Topography of the auditory evoked potential in humans reflects differences between vowels embedded in pseudo-words. *Neuroscience Letters,* 338:189-92

Wallace AB, Blumstein SE. (2009). Temporal integration in vowel perception. *The Journal of the Acoustical Society of America,* 125:1704-11

Weiss R. (1974). Relationship of Vowel Length and Quality in the Perception of German Vowels. *Linguistics: An Interdisciplinary Journal of the Language Sciences,* 123:59-70

Weiss R. (1978). The German Vowel: Phonetic and Phonemic Considerations. *Hamburger Phonetische Beiträge,* 25:461-75

Weiss R, Dressler WU, Pfeidder OE. (1977). The Phonemic Significance of the Phonetic Factors of Vowel Length and Quality in German. In *Phonologica 1976: Akten der dritten Internationalen Phonologie-Tagung Wien, 1.-4. September 1976*, pp. 271-6. Innsbruck: Inst. fur Sprachwissenschaft der Univ. Innsbruck

Wichmann FA, Hill NJ. (2001). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics,* 63:1314-29

Wier CC, Jesteadt W, Green DM. (1977). Frequency discrimination as a function of frequency and sensation level. *The Journal of the Acoustical Society of America,* 61:178-84

Wiese R. (1988). *Silbische und lexikalische Phonologie - Studien zum Chinesischen und Deutschen ("Syllabic and lexical phonology - Studies about Chinese and German").* 58-113 pp. Tübingen, Germany: Max Niemeyer Verlag.

Wiese R. (2000). *The phonology of German.* Oxford: University Press.

Wiese R. (2011). *Phonetik und Phonologie.* Paderborn: Wilhelm Fink.P. 129.

Will U, Berg E. (2007). Brain wave synchronization and entrainment to periodic acoustic stimuli. *Neuroscience Letters,* 424:55-60

Wilson M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review,* 8:44-57

Wühle A, Mertiens L, Rüter J, Ostwald D, Braun C. (2010). Cortical processing of near-threshold tactile stimuli: An MEG study. *Psychphysiology,* 47:523-43

Xu Y, Gandour JT, Francis AL. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America,* 120:1063-74

Yabe H, Koyama S, Kakigi R, Gunji A, Tervaniemi M, Sato Y, Kaneko S. (2001a). Automatic discriminative sensitivity inside temporal window of sensory memory as a function of time. *Cognitive Brain Research,* 12:39-48

Yabe H, Tervaniemi M, Reinikainen K, Näätänen R. (1997). Temporal window of integration revealed by MMN to sound omission. *Neuroreport,* 8:1971-4

Yabe H, Tervaniemi M, Sinkkonen J, Huotilainen M, Ilmoniemi RJ, Näätänen R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology,* 35:615-9

Yabe H, Winkler I, Czigler I, Koyama S, Kakigi R, Sutoh T, Hiruma T, Kaneko S. (2001b). Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. *Brain Research,* 897:222-7

Ylinen S, Shestakova A, Alku P, Huotilainen M. (2005). The Perception of Phonological Quantity based on Durational Cues by Native Speakers, Second-language Users and Nonspeakers of Finnish. *Language and Speech,* 48:313-38

Ylinen S, Shestakova A, Huotilainen M, Alku P, Näätänen R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Research,* 1072:175-85

Yrttiaho S, Tiitinen H, Alku P, Miettinen I, May PJC. (2010). Temporal integration of vowel periodicity in the auditory cortex. *The Journal of the Acoustical Society of America,* 128:224-34

Zatorre R, Belin P. (2001). Electroencephalography and clinical Neurophysiology. *Cereb. Cortex,* 11:946-53

Zec D. (2007). The syllable. In *Handbook of Phonology*, ed. P De Lacy, pp. 161-94. Cambridge, UK: Cambridge University Press

Zion Golumbic EM, Poeppel D, Schroeder CE. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language,* 122:151–61

Zwitserlood P. (1996). Form Priming. *Language and Cognitive Processes,* 11:589-96