Dip and hat pattern: a phonological contrast of German?

Ambrazaitis, Gilbert; Niebuhr, Oliver

2008

Dip and hat pattern: a phonological contrast of German?

Gilbert Ambrazaitis[1] *& Oliver Niebuhr*[2]

[1]Linguistics and Phonetics, Centre for Languages and Literature, Lund University, Sweden
[2]Laboratoire Parole et Langage, Université de Provence, Aix-en-Provence, France
[1]Gilbert.Ambrazaitis@ling.lu.se, [2]Oliver.Niebuhr@lpl-aix.fr

Abstract

Is the high plateau in a 'hat pattern' a *phonetic* artefact, or does it reflect a *phonological* feature? Can it contrast with a low plateau, i.e., a 'dip pattern'? The presented perception experiment supports the phonological point of view, since it shows that the dip/hat contrast can disambiguate German *oder*-constructions, which are interpretable as 'alternative' or 'yes/no-questions'. This specific function may be derived from a more general substance–function relation: While a hat pattern has a 'bracketing function', a dip signals detachment.

1. Introduction

German has a frequently occurring intonation pattern involving two pitch accents – a rise and a fall – concatenated by a high plateau. Several terms have been used to label this pattern, e.g., 'bridge accent' ('Brückenakzent', [14]), 'hat pattern' ([4], inspired by the IPO model for Dutch [1]), 'I-contour' ('I-Kontur', [8]), or simply 'rise-fall contour' [6]. Despite the wide consensus about its *existence*, there is little consensus about the *characteristics* of this 'hat pattern'. The list of open questions includes functional, distributional, phonetic and phonological aspects. This study deals with the nature of the concatenation pattern itself, i.e., the tonal contour connecting the two accents:

- Is the high plateau in a **'hat pattern'** a *phonetic* consequence of the pitch accents involved, in part due to time restrictions, or does it contrast with a low plateau, i.e., a **'dip pattern'**, independently of the pitch accents, thus reflecting a *phonological* feature?

The Kiel Intonation Model (KIM) [4] assumes *phonologically* different concatenation patterns, which may be dipped or non-dipped. The KIM distinguishes between peak and valley accents. A hat, then, is defined as a non-dipped sequence of two peak accents. The KIM postulates (in its original version) three semantically distinct peak accent categories differing phonetically mainly in the timing of the pitch contour (early, medial, late peak). With two logical exceptions (early in the first, late in the second position), all combinations of peak accents may occur in a hat, yielding four basic hat pattern types: medial/early, medial/medial, late/early, and late/medial.

By contrast, within the autosegmental-metrical (AM) framework of intonational phonology (e.g., [12], [7]), the concatenation contour is usually treated as the result of a *phonetic* interpolation process (exceptionally, the high plateau in a hat pattern has been characterized as a high non-accentual floating tone in [14]). In terms of GToBI [2], for example, a distinction between a dip and a hat pattern would have to be explained as a phonetic effect of phonological phenomena such as pitch accent composition or phrasing: The sequence H* T- H* (two pitch accents with 'intermediate phrase' boundary in between) would probably surface as a hat pattern for T = H, but as a dip for T = L; likewise, the pitch accent composition L*+H H+L* would probably surface as a hat, while H* L+H* would exhibit a dip, and even H* H* may be realized with dipped, or sagging F_0 transition, if the time interval between the accents is sufficiently large, cf. also Pierrehumbert's [12] original account of English: H*+H H* (high plateau); H* L+H* (pronounced dip); H* H* (slighter dip, sagging transition).

Of course, a possible time dependence of a dip pattern would favour a 'phonetically-driven' concatenation hypothesis. However, in a comparison of hat and dip patterns in spontaneous speech, no salient difference in mean duration was found, and the standard deviations were large for both samples [11]. Furthermore, by means of systematic resyntheses of dip and hat patterns in all possible pitch accent contexts provided by the KIM, it has been shown that a dip/hat contrast can (acoustically and perceptually) be maintained even when the two accents are placed on short vowels in syllables immediately following each other; this result was subsequently supported by spontaneous speech data [9]. Thus, the production of a hat as opposed to a dip is not necessarily an artefact of time constraints, but may be semantically or pragmatically motivated.

On the basis of spontaneous speech data [11], a set of functional distinctions between hat and dip patterns has been suggested: (a) A hat signals 'known, self-evident information', as opposed to 'unknown, non-self-evident information' signalled by a dip. (b) A hat is 'neutral' and 'not anticipating the result' ('ergebnisoffen'), a dip is 'insistent' ('nachdrücklich'), or 'challenging' ('auffordernd'). For example in alternative questions (e.g., *"Bei IHnen oder bei MIR?"* (pitch-accented syllables are capitalized) "At your place or at my place?"), a hat provides a neutral presentation of the two alternatives, whereas a dip stresses the contrast between them. (c) By *acoustically* 'embracing' two pitch accents, a hat signals a *semantic* connection between the accented words. That is, a hat has a 'bracketing' function, a dip signals detachment.

While we tend to agree with the suggestions under (a) and (c), we propose that, in question constructions with *oder* ("or"), the hat/dip contrast may signal a more fundamental functional distinction than the one mentioned under (b). As Wunderlich [14] has pointed out, an *oder*-construction such as *"Willst Du KAFfee oder TEE?"* ("Would you like some coffee or tea?") is ambiguous, since it may not only function as an 'alternative question' ("Which one (of the two alternatives) would you like: coffee or tea?"), but also as a 'yes/no-question' ("Would you like some hot beverage (for example, coffee or tea)?"). Observe that both readings may initialize a dialogue, i.e., the intended interpretation of an *oder*-construction does not depend on the preceding linguistic context. Wunderlich [14] suggests that the intonation contour can disambiguate an *oder*-construction: While a hat pattern signals an alternative question, a final rise is used in a yes/no-question. Another possibility to signal a yes/no-question, we hypothesize, is to use a low-dipped concatenation pattern. However, the most suggestive reading of an *oder*-construction (even if intonation is missing as in written text) is an alternative question, the yes/no-reading being more specific, or contextually restricted. That is, the hat pattern is most likely not a *necessary* cue to the alternative-question reading, but, we hypothesize, it is a sufficient cue: using a hat pattern when intending a yes/no-question would result in miscommunication, and responding to a 'hatted' *oder*-construction by "yes" or "no" would be pragmatically odd. This implies that at least the absence of a hat pattern – achieved by, e.g., a dip pattern – is a necessary prerequisite for the yes/no-question reading.

The hypothesis can be summarized as follows: The concatenation contour connecting two pitch accents may be employed for communicative-functional purposes. One such purpose is the disambiguation of *oder*-constructions. In that, a hat pattern will trigger the reading as alternative question, while a dip pattern *can* turn the *oder*-construction into a yes/no-question. This hypothesis was tested in a perception experiment. A confirmative result will support the idea that the concatenation contour is not a *phonetic*, but a *phonological* feature.

2. Method

2.1. Experimental design

A perception experiment involving an indirect identification task was conducted. Listeners were presented F_0-manipulated versions of the ambiguous *oder*-construction *"Wollen wir dann den SAMStag oder SONNtag nehmen?"* ("Shall we take Saturday or Sunday?") as auditive stimuli, all based on a single natural utterance, spoken by a 28-year-old male speaker of Standard Northern German (SNG). The stimuli were presented in a constant dialogue context *"JA, das WOCHenende passt mir GUT."* ("Yes, the weekend would be excellent."), spoken by a 28-year-old female speaker of SNG. A pause of 500 ms was inserted between the question (stimulus) and the answer (context). The listeners' task was to judge whether or not the answer semantically matched with the question. That is, a *'matching'* judgement is to be interpreted as the identification of a *yes/no-*, and a *'non-matching'* judgement as the identification of an *alternative* question (cf. 1). Thus, the identification was indirect.

This design could, of course, introduce a bias towards the yes/no-question. However, recall that the interpretation of an *oder*-construction is *inherently* biased (cf. 1): If an answer suitable for an alternative question (e.g., "Let's take Saturday.") were offered as a context, the question (stimulus) would probably always 'match' with the answer (context), irrespective of the stimulus intonation. Our experimental design is a means to cope with this inherent bias of *oder*-constructions. Furthermore, a direct identification task was disregarded, because we believe that the cognitive load of such a task would be too high for the given research question, since it would imply at least three steps: first, imagining a suitable situational context; second, evaluating the intonation contour with respect to that context; third, transferring the result of this evaluation into (labels for) linguistic categories (yes/no- or alternative question). That is, we need a contextualization of the stimuli, which we achieve by our indirect identification design.

2.2. Material and stimuli

The two original utterances (cf. 2.1) were recorded digitally (at 44.1 kHz and 16 bit) in the anechoic chamber at the Humanities

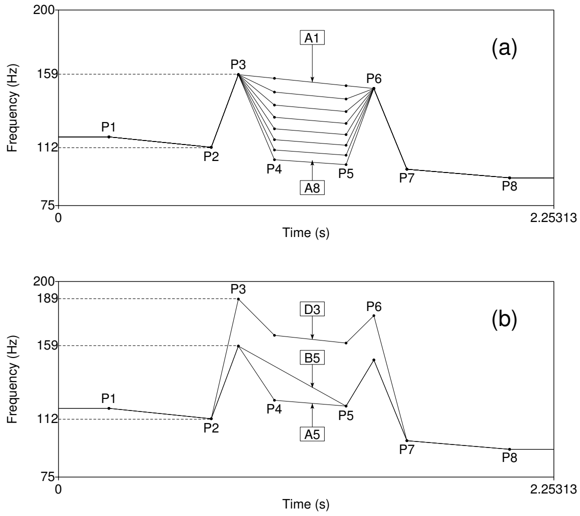

Figure 1: *(a) Stylized F_0-contours of the A-series, defined by 8 pitch points (P1-P8). (b) Comparison of the three series A, B, and D: stylized F_0-contours of stimuli A5, B5, and D3.*

Lab, Lund University, Sweden, using a *Neumann U87 Ai* microphone. Both recordings were subsequently amplified with individual (manually selected) factors in order to guarantee that they were matching in loudness when played as a dialogue. The context utterance (answer) was not manipulated any further. The software Audacity (`http://audacity.sourceforge.net/`) was used both for the recording and the amplification.

The question utterance was spoken with two medial peaks in terms of the KIM and a dip concatenation, realized as a medium low plateau. The peak maxima were reached shortly after the onset of the post-vocalic nasal of the stressed syllables ([m] or [n] in [ˈzamstak] or [ˈzɔntak], respectively). On the initial *("Wollen wir dann den")* and the final *("-tag nehmen")* part of the utterance, the contour was flat and low.

The original F_0 contour was replaced by a stylized F_0 contour, using the PSOLA manipulation of Praat (`www.praat.org`). The stylized contour was defined by 8 pitch points, reconstructing the general shape of the original contour. In particular, the original timing was used concerning onset, maximum, and offset of the two peak accents. The first peak maximum was set near the original one, at 159 Hz, i.e., 8 semitones (st) above 100 Hz; the second peak was set to 150 Hz, i.e., 1 st below the first one, in order to model a slight declination, typical for a hat pattern [1], [4]. Likewise, the utterance-initial and -final plateaux, together with the dip plateau between the peak accents, were modelled as a low, slightly declined F_0 baseline. Finally, the utterance-final word ([neːmː]) had a rather long original duration of approx. 370 ms. It was shortened to approx. 310 ms.

This F_0- and duration-manipulated version (displayed as A8 in Figure 1.a) served as the basis for the stimulus generation. Three stimulus series (A,B,D) were created, containing 8 (A,B), or 11 (D) stimuli, respectively. For the A-series (cf. Figure 1.a), the low concatenation plateau was raised in 7 steps of 1 st, yielding a continuum of 8 stimuli (A1–A8), containing a hat pattern (A1) and seven degrees of a dip pattern (A2–A8).

Based on the A-series, a B-series was created by deleting pitch point 4 for each dip level (1–8), resulting in a 'pointed dip', lacking the low plateau, and exhibiting a slower fall of the first peak accent as compared to the A-series. Finally, the D-series differs from A in that the two peaks have a maximum pitch 3 st higher. However, as in A, the step size between the stimuli was set to 1 st, and the lowest plateau was as low as in A. That is, the D-series contains 11 stimuli in total, from D1 (hat pattern) to D11 (low plateau corresponding to A8). Figure 1.b compares A, B, and D by displaying one example from each series. A further series (F) comprising 8 stimuli was included in the experiment, but is excluded from the current paper.

2.3. Subjects and procedure

The 35 stimuli (A, B, D, F) were rated three times each, resulting in 105 stimulus presentations, which were played in randomized order. In each stimulus presentation, the dialogue consisting of stimulus (question) and context (answer) was played once, followed by a pause of 4 s for the rating. Twenty native speakers of German (15 female aged 20–39 ($\bar{x} = 24{,}3$); 5 male aged 21–26 ($\bar{x} = 22{,}6$)), with no reported hearing impairments, took part in the perception experiment. They heard the stimuli via loudspeakers in a sound-treated experimental studio at the IPdS Kiel, Germany, and were asked to mark their judgement on a sheet of paper. The subjects received instructions in written form, including a careful description of the pragmatic contrast under discussion. To this end, an example of an *oder*-construction was presented, and its interpretation in two different situations was discussed.

2.4. Inferential statistics

We are approaching our research question by formulating and testing three sub-hypotheses, derived from the general hypothesis (cf. 1) and an additional assumption (cf. below): (i) Different concatenation patterns will evoke different 'matching' values in per cent, with high values (>>50%) for the deepest dip, and low values (<<50%) for the hat pattern. (ii) A dip pattern with a *low plateau* (A-series) will render higher 'matching' values than a *pointed* dip (B-series). (iii) Two dip patterns of equal depth (in st, measured from the peak maximum) but different peak heights (A vs. D) will evoke different 'matching' values, with the dip higher in the speaker's tonal range (D) yielding lower values than the dip lower in the speaker's range (A).

An additional assumption underlying (ii) and (iii) is that the functional contrast under discussion is best realized (and perceived) with the 'extreme' realizations of a hat or a dip pattern, respectively. More specifically, for a dip pattern we assume that a certain low pitch level (probably near the speaker's F_0 baseline) has to be reached in order to signal a yes/no-question. The power of the dip pattern to signal that function, then, will be greater when the low level is reached earlier (cf. ii). Likewise, it is not sufficient to fall a certain number of semitones, if the critical low level is not reached through this fall (cf. iii).

A repeated-measures ANOVA with two factors was calculated: (1) dip depth (8 levels; stimuli D9-11 were excluded), in order to test hypothesis (i), and (2) series type (3 levels), in order to test hypotheses (ii) and (iii). The dependent variable was the sum of 'matching' responses per stimulus and subject, which may take values from 0 to 3 (since there were three repetitions).

3. Results

The 'matching' ratings in per cent across all subjects are plotted in Figure 2. For each series there is a transition from 'not matching' (<20% matching) for the hat pattern (stimuli A1, B1, D1-2) to 'matching' (>80%) for the low-dipped concatenation

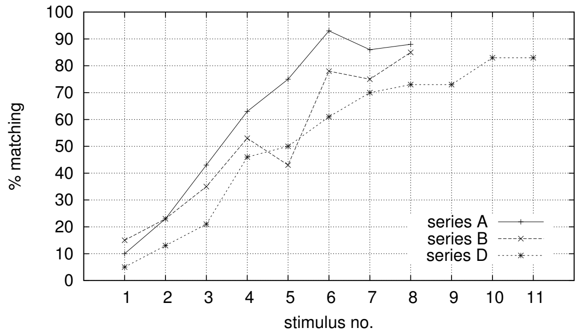

Figure 2: *'Matching' responses in % across all 20 listeners and 3 repetitions.*

(A6-8, B8, D10-11). According to the ANOVA (*df* adjusted according to Greenhouse-Geisser), the effect of dip depth is significant ($F = 42.907$; $df = 3.079$; $p<.001$).

The highest 'matching' ratings (around 90%) are obtained for the A-series. Furthermore, the transition from 'non-matching' to 'matching' happens earlier in the stimulus continuum for the A-series than for B/D. While for A a dip depth of 3 st (A4) is already rated as 'matching' in the majority of the cases (>60%), 5 st are needed in the D-series to obtain a comparable result (D6). In the B-series, the 50%-mark is slightly exceeded with a dip of 3 st (B4), but a direct comparison with the other two series is difficult, due to the deviating result for stimulus B5, which may be an artefact caused by the randomization used. At large, the stimuli 3-8 of the B-series have received fewer 'matching' ratings than the corresponding A-stimuli, but more than the D-stimuli. However, for the deepest dip, the pointed target (B8) matched as well as the low plateau (A8). According to the ANOVA, the effect of series type is significant ($F = 14.993$; $df = 1.664$; $p<.001$). In order to test for significant differences between the single series types, we performed pairwise comparisons with Bonferroni correction. According to these post-hoc tests, the ratings for the A-series differ significantly from those of both B ($p<.01$) and D ($p<.001$), but the differences between B and D are not significant ($p>.05$).

4. Discussion

The results lend strong support for our hypothesis. In particular, (i) since we interpret a 'non-matching' rating as an indirect identification of an alternative question, we conclude that the subjects identified stimuli with a hat pattern or a slight dip as alternative, and stimuli with low dips as yes/no-questions. (ii) Except for the deepest dip, stimuli with a low plateau (A) are more frequently identified as yes/no-questions than stimuli with a pointed (and late) low target (B). (iii) In stimuli with higher peaks (D), deeper dips were necessary in order to be judged as yes/no-questions than in stimuli with low peaks (A). In summary, only a rather extreme realization of the concatenation pattern can guarantee a high likelihood (>80%) of signalling the corresponding functional category: That is, a factual high plateau is needed for the functional category 'hat pattern' (even a slight dip of 2 st yields rather uncertain ratings for the A- and the B-series); and a 'functional dip pattern' seems to have a low pitch target near the speaker's baseline (A6-8; B8; D10-D11).

The fact that neither hat nor dip pattern were identified as 100% matching/non-matching does not substantially weaken our conclusion, because, first, naive speakers of German are usually not aware of the ambiguity of *oder*-constructions until it is explained and exemplified to them. Second, there are certainly several ways of signalling the intended interpretation of an *oder*-construction (cf. [14], sec. 1). Third, German speakers are usually not aware of their intonational contrasts in general. Fourth, other dimensions than F_0, such as duration and energy, may contribute to the percept of intonational categories [10], but have been neglected in this experiment. Taking into account these circumstances, the results are very clear.

But why does a hat pattern, or a dip, respectively, in *oder*-constructions lead to an interpretation as alternative question, or yes/no-question, respectively? Our preliminary account takes up the idea of a possible substance–function relation: While a hat pattern has a 'bracketing function', a dip signals detachment, cf. (c) in 1, [11]. In our case, then, a hat pattern signals that the object of the test sentence *("Samstag oder Sonntag")* forms a single closed unit. This restricts the scope of the conjunction *oder*, and thus the addressee's choice, to exactly the alternatives mentioned within this unit (either Saturday, or Sunday). The dip pattern does not mark such a closed unit, and hence, does not restrict the scope of *oder*. On the contrary, a dip pattern connects the accented lexical items acoustically and semantically in a way that would easily allow the addition of further accented items. Consequently, a potentially open list of alternatives is indicated by the dip pattern, inviting the addressee to choose not only from the examples offered by the enquirer. Possible answers include these examples (Saturday, Sunday), but also the complete rejection of the enquirer's suggestion ("not at the weekend"), or its general acceptance without deciding for any details yet ("yes, the weekend should work"), as well as any other suggestion ("why not on Tuesday"). This account explains why an *oder*-construction with hat pattern *must* be interpreted as an alternative question, while a dip *can* turn the *oder*-construction into a yes/no-question. It thus also explains why an *oder*-construction with dip still should be interpretable as an alternative question, as we suggest in 2.1.

5. Conclusions and future research

Our goal was to experimentally test the hypothesis that a certain *phonetic* contrast *can* be employed to signal a certain *functional* contrast. Thus, the aim of this study was neither to survey all possible expressions of that functional contrast, nor to examine all functions that might be associated with that phonetic contrast. The disambiguation of an *oder*-construction may indeed be a rather rare case, but it served well to illustrate – and to experimentally test – the hat/dip distinction.

Without referring to any empirical evidence, it has recently been stated that *'the intonation contour between the two accents in hat contours is neither grammatically relevant, nor perceived by the listeners'* ([13], our translation). The fact that listeners can easily discriminate hat and dip patterns, however, has been established in a series of perception experiments with discrimination tasks [3]. Yet, perceptual discriminability does not imply a communicative, and hence not a phonological function. Our function-oriented experiment, however, has supported the idea that the concatenation contour is a *phonological*, rather than a *phonetic* feature of German intonation.

In order to show that the concatenation pattern may signal a functional contrast, independently of the pitch accents involved, a constant pitch accent frame according to the KIM has been constructed: the KIM would analyse our stimuli as a sequence of two medial peaks with *different concatenation patterns*. Within current AM models, however, our stimuli would probably be analysed as sequences of *different pitch accents*: e.g., L+H* L+H* for the dip vs. L+H* H+L* for the hat.

These AM representations imply that the hat/dip contrast cannot be implemented independently of the pitch accent composition. However, all four peak accent compositions according to the KIM (medial/early; medial/medial; late/early; late/medial) have been attested both as hat and as dip patterns in the spontaneous speech corpora used by [11], [5].

Yet, in order to further strengthen the evidence for the independence, and hence, the phonological status of the concatenation pattern, it should be tested whether the essential semantic components of both the pitch accents and the concatenation pattern are conserved in the global interpretation of the utterance. For that, further experiments should be performed with stimuli similar to those of this study, but with different pitch accent compositions. By means of suitable contextualizations, subjects will then have to judge both the semantic effect of the concatenation, as in this experiment, and the semantic effect caused by changes in pitch accent composition.

6. References

[1] Cohen, A.; 't Hart, J., 1967. On the anatomy of intonation. *Lingua*, 19, 177-192.

[2] Grice, M.; Baumann, S.; Benzmüller, R., 2005. German intonation in Autosegmental-Metrical phonology. In *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.-A. Jun (ed.). Oxford University Press, 55-83.

[3] Hertrich, I., 1991. Perceptual categorization in terminal intonation patterns of two-accent utterances of German. *AIPUK*, 25, 243-294.

[4] Kohler, K.J., 1991. A model of German intonation. *AIPUK*, 25, 295-360.

[5] Kohler, K.J.; Niebuhr, O., 2007. The phonetics of emphasis. *16th ICPhS*, Saarbrücken, 2145-2148.

[6] Krifka, M., 1998. Scope inversion under the rise-fall contour in German. *Linguistic Inquiry*, 29/1, 75-112.

[7] Ladd, D.R., 1996. *Intonational Phonology.* Cambridge University Press.

[8] Molnár, V.; Rosengren, I., 1997. Zu Jacobs' Explikation der I-Topikalisierung. *Ling. Berichte*, 169, 211-247.

[9] Niebuhr, O, 2005. Sequenzen phonologischer Intonationsgipfel. Theoretische Möglichkeiten und empirische Realisierung im segmentellen Kontext. *AIPUK*, 35a, 55-95.

[10] Niebuhr, O., 2007. Categorical perception in intonation: a matter of signal dynamics? *Interspeech 2007*, Antwerp, 109-112.

[11] Peters, B.; Kohler, K.J.; Wesener, T., 2005. Melodische Satzakzentmuster in prosodischen Phrasen deutscher Spontansprache. *AIPUK*, 35a, 7-54.

[12] Pierrehumbert, J.B., 1980. *The phonology and phonetics of English intonation.* MIT.

[13] Sudhoff, S.; Steube, A.; Hogrefe, K., 2004. Syntaktische Positionen bewegungsbeschränkter I-Topiks im Deutschen. *LAB*, 81, 147-158.

[14] Wunderlich, D., 1988. Der Ton macht die Melodie – Zur Phonologie der Intonation des Deutschen. In *Intonationsforschungen*, H. Altmannn (ed.). Niemeyer, 1-40.

AIPUK = **A**rbeitsberichte **IP**dS **U**niversität **K**iel
LAB = **L**inguistische **A**rbeits**b**erichte Universität Leipzig