Prosodic Encoding of Sentence Mode and Focus in L2

Johannes Schliesser

Dissertationsschrift zur Erlangung des akademischen Grades

Dr. phil.

Eingereicht durch Johannes Schliesser,

geboren am 19.06.974 in München,

bei der Philologischen Fakultät der Unversität Leipzig

Contents

1 Introduction

The fact that second language prosody is often different from the target language intonation is obvious for many people in contact with language learners. So, this issue has encouraged a rather long tradition of investigations of specific "errors" in second language speech (e.g. Gårding, 1981; Grosser, 1982; Lepetit, 1989; Lepetit & Martin, 1990; Mairs, 1989; Neufeld, 1987; Pürschel, 1975; Willems, 1982) and the role of prosody in foreign accent (e.g Anderson-Hsieh, Johnson, & Koehler, 1992; Els & Bot, 1987; Jilka, 2000; Jilka & Möhler, 1999; Major, 2001).

However, one must state that sound research on this topic is often still hampered by two fundamental problems. The first is that especially earlier work was heavily afflicted by a lack of methods of prosody description / transcription and the accessibility of technological tools for speech analysis. A part of the work up to the 80s of the last century is simply a description of the authors intuitions about the differences between native and foreign speech contours. This problem is certainly not completely overcome nowadays[1], but with the upcoming of autosegmental models of prosody transcription and a spreading of powerful software and hardware, widely accepted tools for a replicable analysis exist.

The other point that is still constraining research is a lack of detailed and well founded theoretical models of prosody processing for both, perception and production. No wonder: fundamental assumptions like the categoriality of prosodic contours and whether they carry a specific meaning are still under discussion (e.g. Baumann, Becker, Grice, & Mücke, 2007; Kohler, 1987, 2004). Even the most detailed models of speech production in the native language (Levelt, 1989) do not allow for precise predictions of the speaker's output. Models of second language learning, and more specifically, models of the acqusition of second language phonology (e.g. Flege, Munro, & MacKay, 1995; Major, 2001) are strongly bound to the assumption of meaningful categories and effects of (miss-)categorization: borrowing from them may be helpful, but must be undertaken carefully and considering the diverging caracteristics of prosody.

[1] For instance with regard to a considerable inter-rater variance within approved transcription schemes, or frequent artifacts in analytic software

This doctoral thesis is not meant to solve these problems. However, the hypotheses and discussions presented will be linked to these issues.

1.1 The Aim of the Study

This thesis investigates the production of sentence mode and contrastive focus in English sentences spoken by advanced German language learners. It focuses on two main questions. The first is **whether there is a systematic pattern of second language prosody ?** And second, if there is, **how can it be accounted for in terms of processing ?**.

From these general purposes, several more specific aims can be derived. First, research on a possible systematic prosodic pattern of L2 encoding of sentence mode and contrast entails a detailed description of the produced utterances. More precisely, utterances for which a context unambiguously induces the communicative functions of the sentence modes statement and (echo-)question and a specific position of a corrective contrast are analyzed in the light of (categorical) tonal contours as perceived by an expert listener (the author), as well as six acoustic parameters that have been shown to be affected by the above functions: syllable duration, the maximum, the mean, and the range of the fundamental frequency, the relative position of the pitch maximum within the syllable, and intensity. These correlates capture the suprasegmental and sentence-level properties of speech.

Second, in opposition to many studies investigating first or second language prosody, the full range of variation within and between speakers will be reported and included in the analysis. Instead of isolating THE one appropriate contour, the full range of perceived tones will be reported and the quest is not for an ideal or specific correlate, but rather for function (and position) specific preferences. A similar approach is taken for the acoustic parameters which will be analyzed with the help of appropriate statistical tools.

A third focus of this study is that the L2 utterances will be cautiously compared to similar utterances of native speakers. This is necessary, as the large majority of the published studies on L2 prosody explain learner specific patterns with transfer from the native language. However, a detailed study of native speaker behavior in similar utterances is barely undertaken. Two experiments that analyze the prosodic correlates of the encoding of sentence mode and contrast by native speakers of the target and the native language of the learners will precede the analysis of the L2 patterns in order to assess detailed predictions in the case of transfer from the native language of the learners (L1) or perfect acquisition of the target language patterns.

This issue introduces a fourth main issue of the study, namely that other possible explanations of L2 prosody should be considered. With very rare exceptions, L2 patterns that transfer could be accounted for by transfer had to "remain unexplained"(Archibald, 1998, p.269). However, there are good reasons to include a set of 'universals' in the set of possible explanations of the source of the retrieved patterns. The idea that sentence mode and focus are expressed prosodically in a similar way in many languages has a long tradition, and are supported by recent results of experiments on prosody perception. The "Biological Codes" (Gussenhoven, 2002, 2004), though imprecise in detail, provide a reasonable base for predictions of possible L2 patterns of prosody production following universal principles. Of course, a discussion of the alternatives of transfer or recourse to universals touches the question of the initial state of prosody acquisition, more precisely whether a learner first applies his L1 knowledge to the planned L2 speech production or initially relies on properties that are possibly shared between him and the listener. It has to be seen in how far the speaker groups involved in the study provide sufficient support for one or the other hypothesis about the source of L2 patterns.

And finally, as no testable model of second language prosody is available, the findings shall be compared to native language prosody production. There are a number of models or modellings available: the most deeply elaborated is Levelt's 'Speaking' model (Levelt, 1989). Furthermore there is Fujisaki's model of speech generation (Fujisaki, 1980 Fujisaki & Hirose, 1982) which is wide-spread in research on synthetic speech generation. Also, the more recent PENTA model (Xu, 2005) will be discussed in the light of the data.

1.1.1 Structure of the Presentation

Following short definitions of the communicative intentions that will be prosodically encoded by the speakers in the next section, chapter 2 will present the expected intonational correlates for sentence mode and corrective contrast for the native and the target language of the learners, namely German and English. It includes a summary of the expected differences between the languages (2.3) which are a necessary prerequisite to derive predictions on the basis of L2 production models based on the Contrastive Analysis Hypothesis (2.4). Furthermore, models and possible predictions from 'universal' aspects of prosody processing are presented in 2.6. Chapter 3 presents previous findings related to supralexical prosody in a second language. Next, the implementation of prosody production in models of speech or prosody production are introduced (chapter 4). More specifically, the key concepts of Levelt's detailed 'Speaking' Model, Fujisaki's technically influential proposals of prosody generation, and the more recent PENTA model by Xu are presented. Then the experiment to assess the prosodic correlates of sentence mode and contrast encoding in English by native speakers is reported (chapter 5), followed by the similar experiment for German spoken by native speakers (chapter 6). The key results for the two languages are compared in chapter 7. Now, the experiment investigating the encodings of German native speakers in their L2 English is reported (chapter 8). The results are discussed in chapter 9 with a comparison between the learner language and the utterances of the two native speaker groups. Also, the two possible sources of the L2 behavior, transfer and a recourse on universals are weighted against each other. The last chapter 11 evaluates the drawbacks of the results on models of prosody production.

1.2 Definitions of Sentence Mode and Contrast

As mentioned above, this study examines prosodic correlates of sentence mode and contrast for L2 speakers. These terms need to be defined.

1.2.1 Sentence Mode

For the current study, critical sentences have been elicited in form of a statement and a corresponding echo-question. The notions of statements and questions, and the conditions which allow a classification of either two have a long history of discussion in semantics, pragmatics and philosophy of language. As these arguments have no direct relevance to the research question, the section is restricted to the absolute necessities for a sound explanation of the choice of the stimuli in the current study.

Statements

What makes an utterance a statement ? Bartels (1997) distinguishes three approaches for a classification depending on the perspective on the different parts of an utterance: the message (proposition) itself, or the speaker's or the hearer's attitudes. For the first view, one can conclude that a statement must be propositional, in that it represents "a complete proposition which is put in correspondence with external reality. This usually mean that the assertion can be assigned truth conditions" (Bartels, 1997, p. 59). The second perspective includes the speaker's attitudes towards the proposition, namely "in the idealized and not uncommon case the fact that the speaker believes the proposition he puts forward to be true" (Bartels, 1997, p. 64). A more listener-orientated or interactive approach assigns assertiveness for sentences that are uttered to change the addressee's mutual beliefs, or, as summarized in Bartels's definition of assertiveness, which includes the speaker, the addressee and the proposition: "a speaker expresses an instruction to the addressee to reduce his context set b all those possible words incompatible with the speaker's commitment to that proposition." (Bartels, 1997, p.90). For the current study these explanations are sufficient. But note that the intention of changing the speaker's mutual beliefs is even literally true for corrections: and all statements in the following empirical study are explicit correcting the hearer's belief. Note also, that the terms of statements, declaratives, and assertions are used synonymously in the current study.

Questions and Echo-questions

A number of researchers (e.g. Lyons, 1977; Bartels, 1999) distinguish 'questions' as a semantic/pragmatic category, and 'interrogatives' as a syntactic category, specified by a distinct auxiliary position, or the presence of *wh-* words (cf. Iwata, 2003, p.186).

Generally, a minimal consensus about 'questionhood' or interrogatively is that questions are "utterances that convey perceived lack of information – simply put, speaker uncertainity – regarding a relevant aspect of the propositional content" (Bartels, 1997, p. 14). It is discussed in how far the possible implication to the hearer to give the answer to the question is part of a linguistic definition of questionhood. While for Lyons (1977, p. 754) "the indication that the addressee is expected to give an answer is not part of the question itself", there are strong proposals like in Lewis (1969, p. 186) – "Questions are imperatives"– and weaker ones like in (Searle, 1969, p. 67), who considers questions as "an attempt to elicit this information" (cf. discussion in Bartels, 1997).

[2]Researchers (including Bartels, 1997 and Haan, 2002) commonly distinguish nine 'core' types of questions: yes-no questions (with inversion of the finite verb and unbiased towards the answer), wh-questions (including a wh-word), alternative questions (coordinating more or less explicitly alternatives with the OR operator), tag-questions (merely seek confirmation but not new information), declarative questions (non lexical or syntactic interrogativity marker, sometimes subsumed as special yes-no questions, sometimes assumed to be more biased towards one answer than the other), echo questions (repeating (parts of) the preceding utterance, seeking for confirmation and expressing surprise or improper understanding), elliptic questions (comparable to wh-questions), rhetorical questions (to be regarded functionally

[2]This paragraph is a summary of the survey of Haan, 2002, p. 12 - 18

as a statement despite of sharing lexical and syntactic properties with other question types (cf. Haan, 2002, p. 18)), and embedded questions (equally functionally non-questions).

Bolinger (Bolinger, 1957, 1989) distinguishes two functionally different categories: questions can be 'reclamatory', in case they are uttered if the speaker wants to get information about something he thinks that the hearer may know. 'Reflex' questions include the first function, but also express affects of the utterer, like surprise, incredulity, etc.

In the current study, only echo-questions are used to analyze intonatory correlates of questionhood and their interaction with focus. There are two main reasons for that procedure: First, echo-questions can have exactly the same surface form in terms of an exact repetition of the lexical content and syntactic structure of the corresponding statement, and thus allow for an analysis comparing the prosodic properties of the two sentence mode without introducing artifacts from these sources. The second reason is that echo-questions that are related to a narrow-focus statement take the same focus domain. Both reasons shall be discussed in some more detail below. In my opinion these arguments outweigh the general concern that echo-questions are "less prototypical" (Haan, 2002, p. 21) than other question types.

Before introducing major assumptions about echo-questions, it is evident but noteworthy that echo-questions, unlike other questions, are uttered AFTER the statement they are related to. They "repeat as a whole or in part what has been said by another speaker" (Quirk, Greenbaum, Leech, Svartvik, & Crystal, 1985, p. 835). Parker and Pickeral (1985) distinguish two types structural echo-questions. If the constituents remain in the same position as in the related statement, they are Type I echo-questions. These can be an exact copy of the statement, or one constituent can be replaced by a wh-word. If there is an inversion of a constituent, they are Type II echo-questions. Echo-question do not add content to the previous statement, and they have to be different from it in either lexical, structural or intonational way. According to Parker and Pickeral (1985) echo-questions without wh-word cannot end in a falling contour. In Bolinger's functional point of view they have two key functions: confirmation and rectification. "If the former, the reprise will be *yes-no* (echo or reflex); if the latter, it will be *wh-* (reclamatory)." (Bolinger, 1987, p. 264). For Bolinger, the lexical or syntactic structure is no criterion: The wh- questions can be structurally closer or further away from the stimulus. It goes from substituting one constituent by a wh-word up to substituting all constituents, it can be a reduction up to a simple "What ?" or 'wide-focus reclamatory' questions like "What did you say ?" (cf. Bolinger, 1987, p. 263). While structurally, there is no inherent difference to 'regular' questions, echo-questions are special as they are rarely uttered by pure curiosity, to be understood as a quest to fill a lack of information. They mostly express "surprise or incredulous response, especially if the intonational range is up somewhat"(Bolinger, 1987, p. 266). In these cases a falling or low end is possible.

The detailed analysis of intonatory correlates of different kinds of questions by Bartels (1997) in American English, Bartels links echo-questions obligatorily to a final rise. However, their phrase accent can be high in case the hearer wants a confirmation of something that he might not have understood properly an therefore asks for a repetition of the previous statement – or low, in case the utterance mainly functions as a reaction to express attitudes of an already evoked content. However, the main criterion to distinguish echo-questions from other echoing utterances is that echo-questions "invite for a repetition of the original utterance"Bartels (1997, p. 282). Structurally and functionally similar echo-exclamations do not ask for a repetition and are uttered – irrespectively of the sentence-mode they are echoing – with a falling or low end. However, although Bartels states a "sharp tonal distinction" (p. 282) between rising echo-questions and falling echo-exclamations, they "overlap greatly in distribution" (p. 465).

The echo-questions that will be used are lexically and structurally exact copies of the previous statements. An assertative sentence like in (1)

(1) Peter flew to Paris.

can be echoed by the sentence like in (2)

(2) Peter flew to Paris ?

and it will be followed by a statement confirming the content of the utterance, or – more precisely the focused part of it (more on that below).

The answer is obligatorily "yes" or "no". So, the echo-questions in the sense they are applied here are declarative yes-no questions. However, it is obvious that echo-questions that aren't uttered with the unique intention of asking for a repetition of the the previous sentence – for instance to make sure that one has understood the preceding statement correctly – are asymmetric in what they expect as an answer: namely a confirmation of the previously said, most shortly by 'yes'. The point what echo-questions exactly ask about should be elaborated in some more detail, as it is relevant for one important property for the current study: the focus.

The qualities of focus in echo-question is a special issue in Iwata's analysis (Iwata, 2003) of the semantics / pragmatics of this kind of questions. The core tool for his analysis, which could also be used as a predictor for the form of echo-questions, is the 'focus of metarepresentation'. The term 'meta-representation' comes from Relevance Theory (Sperber & Wilson, 1986). For our case it is sufficient to say that is a representation that resembles an utterance in terms of form or in content, but it can also represent a thought or intention attributed to someone (cf. Iwata, 2003, p. 189). The key notion is the ressemblance: this is how echo-questions that are reduced, or which are comprehensible only by a shared knowledge can be treated. For echo-questions it is important that the metarepresentation is attributed to the speaker of the original utterance. An echo-question can be paraphrased typically by a sentence "I ask whether you said ...", which is a paraphrase that "might be taken to reflect a request for confirmation" (Haan, 2002, p. 17), but also expresses unexpectedness of the main part of the previous utterance.

The focus of metarepresentation is the "portion of an EQ [echo-question, the author] that has given rise to uncertainity" (Iwata, 2003, p. 210) and it "is contrastively stressed" (ibid.). The focus of metarepresentation "remains unchanged from the original even when other parts of the metarepresentation are reformulated" or omitted (cf. Iwata, 2003, p. 211). Hence follows: what is clearly focused in the original utterance is very probably the portion of the meta-representation that is asked to be confirmed, and thus the focus of metarepresentation of the echo-question, too.

They differ to their preceding assertative sentences only through intonation, but not by the lexical materials or the sequence of the constituents. Thus, in a comparison of prosodic correlates, confounds by micro-prosody[3] or different words are avoided.

1.2.2 Contrastive Focus

Definitions of focus in the literature would deserve (and fill) a dissertation on their own. Here, only a precise description of the kind of contrast used in the focus-eliciting dialogues will be presented. These are "backwards related corrections with contrastive focus" (Steube, 2001). They consist "of at least two sentences, the corrigendum and the corrigens. [...] A corrigens interrupts the sequential development of a text and offers the contrastively marked entity as a replacement for the corresponding incorrect element in the corrigendum."(Steube, 2001, p. 215). In simpler words, the speaker of the corrigendum thinks / knows that for the hearer an entity α is true in the context. But he negates that α is true and replaces it by an

[3]Prosodic differences due to correlates to segmental features.

equivalent entity β. In the current study, I will refer to "backwards related corrections with contrastive focus" as focus, contrast, correction, corrective contrast, and contrastive focus synonymously during my description of the experiments. For the literature review, other kinds of focus may be relevant, they will be explicitly defined.

There are several reasons to use corrective contrast for the assessment of second language prosody. First, it is applicable to all syllables (possibly even to a number of phonetic segments) in the utterances. The position of the (prosodically encoded) corrections is not determined by the structure or the word class of the targets. At least for sentences with default, simple word order (SVO) the location of the corrigens is not limited by the syntactic structure. Furthermore, corrective contrast is the initiation of focus that is suspected to receive the highest level of prosodic prominence and outranks other focus assignment rules or preferences, and word-accent needs: it is at the highest level of a focus hierarchy (Molnár, 2006). Furthermore, the occurrence of acoustic correlates of contrast seems to be quite reliable[4]. As it is unknown what exactly the second language speakers will do, it is desirable that they do something.

For a more concrete picture, let us reconsider the sentence in example (1), "Peter flew to Paris". Corrective contrast is possible on every syllable in every position. Most commonly, the subject (3) or the nominal constituent in the prepositional phrase (4) can be corrected.

(3) Context : b. *I met Peter at the airport, yesterday. I think he wanted to go to Rome.*

- a. He didn't fly to Rome. Peter flew to PAris.
- b. Peter flew to PAris ? Last time I met him, he said he hates the Frenchmen.

(4) Context: b. *Which of your brothers flew to Paris last week. Was it John ?*

- a. It wasn't John. PEter flew to Paris.
- b. PEter flew to Paris ? Last time I met him, he said he hates the Frenchmen.

But also the verb can be collectively contrasted.

(5) Context: b. *Peter went to Paris last week. Did he take the train ?*

- a. He doesn't like the Eurostar. Peter FLEW to Paris.
- b. Peter FLEW to Paris ? He is always so much afraid of flying...

[4] Dietrich (1990) claims that for his stimuli, acoustic correlates were found in all cases (cf. Dietrich, 1990, p. 421)

A strong accent on verbs, without a context inducing correction, can also be interpreted as the so-called "verum-focus" (Höhle, 1982). This means that not only the action, but the whole proposition of the sentence is corrected or confirmed. But intuitively, there are no acoustic differences between the realizations of corrective contrast on verbs and a verum focus.

To sum up, we have seen that corrective contrast can be assigned to every syllable of a sentence, and its function remains to replace a contextually given element by another element of the same category in the listeners world. Corrective contrast is highest ranked in the focus hierarchy and should result in strongest acoustic correlates of prominence. Furthermore, corrective contrast can also be induced in echo-questions. As far as these are concerned, we have seen that echo-questions in a narrow sense are an exact copy of the lexical content, word order and information structure of a preceding statement. They are yes-no questions, special in that they presumably do not only ask for confirmation, but also express some sort of surprise, incredulity. And they are a reasonable medium for corrective contrast.

2 Language Specific Prosody and Perspectives on L2 Intonation

This chapter will first summarize a number of relevant studies that investigated the correlates of sentence mode and focus for the languages of English and German. Then the key issues of the Contrastive Analysis Hypothesis will be mentioned, together with an influential model of foreign language speech processing, Flege's Speech Learning Model. Finally, proposals of universality in prosodic encoding will be discussed.

2.1 Correlates of Focus and Sentence Mode Encoding in English

First, correlates of focus will be discussed. Among the many investigations in the literature there are two studies by Cooper and colleagues which are especially relevant to the experiments presented later.

Cooper, Eady, and Mueller (1985) reported two experiments which investigate the effects of contrastive focus on duration and fundamental frequency at different positions in declarative sentences. They found that duration increases about 30-40 % if the emphasized word is at the beginning or in the middle of a sentence, and it is 10-15 % longer if it is at the end of the sentence. With regard to mean fundamental frequency, they found that there is "deaccentuation": In sentence initial position, the contrasted word does not show higher pitch than the same word in uncontrasted condition, but the rest of the sentence is lower than in the other conditions. In both sentence medial positions (third to sixth syllable in the sentence), the mean fundamental frequency is locally higher if the word is contrasted than if it is not, and there is deaccentuation afterwards. Contrast in final position induces higher pitch, too. A second experiment with longer sentences replicated these results and gave further evidence for the retroactive hypothesis, i. e. that an upcoming contrastive accent is not necessarily prominent through a higher mean pitch on it, but by the strong fall, or deaccentuation, afterwards.

There are two main points that one can criticize in their study. First, the recorded sentences are not really coherent in several aspects. For example, the first content word was either the first or the second word in the sentence, so the emphasized syllable was either utterance initial, or preceded by an unaccented determiner. Second, the analyzed words had different numbers of syllables (1, 2, 3 and one 4-syllabic words were used). Third, the position of the accented syllable of polysyllabic words was not identical for all words, which themselves were not matched in length. Furthermore, the whole prosodic dimension of intensity is ignored, and the mean fundamental frequency is the only pitch related parameter. It may be reasonable for the incoherent stimuli, but we do not get information about pitch range, nor the direction of pitch movement, nor the absolute maximum of the fundamental frequency.

These points were investigated in more detail in a recent study by Xu (2005). They confirmed deaccentuation – in terms of a comparable lower peak of the fundamental frequency – of all postfocus words, affecting also unstressed syllables in the focused words. Effects of prefocal relative pitch lowering were found for some speakers but not all. However, there was an increased pitch maximum in all positions of the declarative utterances, including the initial one (cf. p. 173 Xu, 2005). Under specific conditions (longer syllable, not word- or sentence-final) the peak of the fundamental frequency was shifted towards the end of the syllable.

Eady and Cooper (1986) extended Cooper et al. (1985) to the acoustic correlates of (information) focus in declarative sentences compared to lexically and structurally identical yes-no questions. There were three focus conditions elicited by questions: neutral focus (What is happening ?) and two narrow focus conditions asking for the first or the last content word in the target sentences. The question mode was induced by the command "Ask ...". Accuracy in terms of perceptibility of adequate focus position and sentence mode of the produced sentences was blind rated: the 6 speakers chosen for further acoustic analysis produced an average of 94 % correct sentences: Correct is to be understood in a sense that the perceived main accent of a sentence coincided with the position of the contextually induced focus position.

Eady and Cooper (1986) analyzed duration and fundamental frequency of all six conditions. For duration they found that the focused word was significantly longer irrespective of position and of sentence mode than the same word in unfocused condition. The mean fundamental frequency was not significantly higher in the sentence initial position, but in the other three key word positions. Sentence initial narrow focus was encoded by deaccentuation of the following constituents. In statement final position, mean F0 is significantly higher if the word is focused than if it is not focused or in neutral focus condition. Questions seem to use an overall higher register than statements irrespective of focus condition (cf. also Lieberman, 1967), with the exception of the first content word. Effects of question initial focus do not appear on the initial content word, but possibly they are the reason for a significantly higher fundamental frequency of the second and third critical word in that condition. But there are no focus induced effects on pitch at the end of questions.

The word-internal alignment of the pitch peak (or "proportional peak placement" in their terms) showed a significant effect for sentence mode on the final word, with later peaks - and thus rising contours - for questions.

Eady and Cooper's study can be criticized very similarly to the previous by Cooper et al. (1985). They ignored intensity and the linguistic status of the key words was not carefully controlled. Further, now the positions of focus were limited to the edges of the sentences, there were no sentence-medial foci. Nonetheless, their investigation provides a rather straightforward analysis of acoustic correlates of sentence mode and focus in English.

Yet, there is no reason why they neglected intensity. Beckman (1986) clearly found significant intensity differences for accented vs. non-accented syllables in American English for word accent minimal pairs. The important contribution of intensity to the perception of prominence is demonstrated in a recent corpus-based account by Kochanski, Grabe, Coleman, and Rosner (2005). They measured the effects of a number of acoustic parameters on perceived prominence in a corpus of British English. Their essential finding was that prominent syllables are mainly longer and louder than non-prominent syllables. For a good prediction of prominence, these two parameters are sufficient. Fundamental frequency also has greater range and higher peaks, but these parameters provide little value to a predictive computational model. Unfortunately, Kochanski et al. (2005) did not account for different positions in a sentence and for sentence mode.

As for questions, the clear preference of a rising or high end of questions in English as found by Eady and Cooper (1986) and other researchers, is corroborated in various studies. In a survey of question intonation in British English dialects, Grabe (2004) found a varying percentage of rising / falling contours depending on the kind of question (wh-question, inversion-question and declarative question[1]) and the dialect of the speaker. For example, London English uses falling contours in 96 % of the declarative sentences in the corpus, and it assigns 56 % of the wh-questions with a low boundary tone, but only in 5.6 % of the declarative questions. English by speakers from Bradford shows a high preference for

[1]These are questions that have declarative syntax and morphology, but question function; they are often called intonation question.

falling tones: all declarative sentences and 83 % of the wh-questions and more than 20 % of the declarative questions have a low boundary tone (falling or low end). In Newcastle and Belfast, one finds a notable proportion of rising patterns for declarative sentences (17 % and 83 % respectively). Also, Syrdal and Jilka (2004) show that falling contours in questions are frequently produced and perfectly acceptable for American English.

They confirm the detailed proposals of intonational correlates for different kinds of questions from a more semantic-pragmatic perspective by Bartels (1997, 1999). She states that falling contours are common and perfect especially for syntactically determined yes-no questions, and wh-questions. I already referred to her work in the section of prosodic correlates to syntactically ambiguous echo-questions (see section 1.2.1). In her point of view, questions that lack a syntactic cue to questionhood have to end in a rising contour, otherwise they express attitudes of the speaker but do not ask for information or confirmation. However, as she assumes an questionhood attribute that is expressed by either syntactic properties or intonation, the argumentation is cyclic. If a morphologically or syntactically non-determined sentence is uttered with a falling contour, it lacks questionhood and is therefore treated as exclamation. Furthermore, she did not test her predictions empirically but relies exclusively on her intuition.

Nonetheless, it is clear that the final contour for questions in (American) English is not necessarily rising, even for yes-no questions, as already Fries (1964) found out: in his analysis of American English Yes-No-question intonation in a corpus of TV quiz shows, he found that more than 60 % had a falling tone. He even concludes that “at least in American English there is no question intonation as such” (cf. in Cruttenden, 1997, p. 36).

To sum up: although the studies are not really comparable, for English, the clear distinction of falling contours or low boundary tones for statements and rising contours or high boundary tones must remain open. Surely, there is a strong tendency for falling contours in statements, but a weaker tendency of rising contours for questions, even yes-no and intonation questions. In the case of echo-questions, falling contours with an overall higher register may be quite frequent, as an expression of “surprise” is pronounced together with the question of confirmation (yes-no). As for focus, one can expect higher fundamental frequency peaks on the accented word, together with longer durations in statements. If focus is in initial position, the fundamental frequency might not be locally higher, but the following syllables should be deaccented. The only correlate of focus in questions was a longer duration for focused words in initial position. Furthermore, one must emphasize the increased intensity for accented (and probably focused) words, which unfortunately is ignored in many studies.

2.2 Correlates of Focus and Sentence Mode Encoding in German

In terms of acoustic correlates, sentence mode was investigated by various measurements by Batliner (1989b). In German, utterances (one-word or more words) which are syntactically declarative have to have a rising contour to be perceived as questions (Batliner, 1989a, 1989c). Indeed, although intuitively possible, falling contours for yes-no questions are not common[2], and do not play a role in Batliner and colleagues' default intonation contours.

Focus is encoded by fundamental frequency and duration. Braun (2004) presented detailed acoustic analyses of contrastive (but not corrective) themes[3] in German. She analyzed target sentences in which the contrastive focus condition was established by a context and recorded 10 speakers from two different regions of Germany. From a number of pitch related parameters, she found that the height of the peak and the range of the F0-rise were significantly higher for contrasted compared to non-contrasted targets. A logistic regression revealed that the range of the rise was most important (Braun, 2004, p.74 ff.). The absolute duration as well as the duration ratio of the target constituent in the utterance were significantly longer if they were contrasted. The slope of the rise did not differ, but the rise was longer, thus the peak later in the target, if it was contrasted (cf. Braun, 2004, p.77ff). In case the target was contrasted, the pitch peak was found in the unstressed syllable following the most accented one. So there is - as reported also for English (K. E. A. Silverman & Pierrehumbert, 1990), Korean (Jong, 1994), and Greek (Arvaniti, Ladd, & Mennen, 1998) - a probable peak dislocation from the accented syllable to the right.

A number of insights in the correlates of different kinds of focus and inter-speaker variation comes also from Baumann and colleagues. Baumann, Grice, and Steindamm (2006) analyzed the productions of six German native speakers of one sentence which contained broad, three narrow(er), and contrastive focus as answer to focus-inducing questions. They analyzed tonal correlates in a GToBI transcription as well as the acoustic parameters duration, the pitch maximum, pitch range, and the relative position of the peak. The main findings are that the narrower a focus is, the less frequent is a downstep on the nuclear phrase accent (the object of the sentences). In sentences with contrastive focus, no downstep was found at all, which means that the peak height of the beginning of the sentences was maintained. The acoustic parameters were, all in all, increasing with the reduction of the focus domain: the most clear differences to broad focus were found with contrastive focus. A similar investigation

[2]Yet, falling questions are common if they are information questions, for instance wh-questions.
[3]Words that already are given by previous context.

was published by Baumann et al. (2007): This time the focus conditions were reduced to broad, narrow, and contrastive focus, but more target sentences were recorded, again as an answer to focus-inducing questions. The target word was the object of verb-final sentences. A comparison of the GToBI annotated tonal correlates of the targets confirms the preference of a downstepped high nuclear tone for broad focus, and an up-stepped or "unmodified"[4] tone. The focus type also induces different pitch maxima on the target; the highest values were found for contrastive focus. This paper also presents effects of focus on the production of the accented vowels, and states more frequent hyperarticulations for narrow/contrastive focus than for broad focus. Both papers also indicate inter-speaker differences for the choice of the pattern which encodes narrow focus.

A study very much comparable to the work of Kochanski et al. (2005) for British English was undertaken by Mixdorff (2002). He investigated the acoustic correlates of perceptually prominent syllables in a corpus. Fundamental frequency excursion and duration had a significant impact, whereas intensity was no reliable parameter for perceptually prominent syllables. The reliability of duration decreased at the end of the sentences, as phrase final lengthening confunded accent-induced lengthening.

Also Elsner (2000) investigated the acoustic correlates of perceived accent in an annotated German corpus. Fundamental frequency excursions and durational parameters were fairly good predictors of perceived accent. Intensity was no reliable parameter. This is also coherent with the findings of Dietrich (1990).

To sum up, sentence mode is expressed by the final contour of the utterances. A falling or low tone indicates a statement, and a rising or high final tone is used to encode questionhood, independently of the type of questions. Narrow or contrastive focus is correlated with an increased length and pitch maximum of the syllables, but intensity does not provide consistenc cues.

2.3 Expected Differences between English and German

There are numerous studies that compare intonational features of different languages, starting with Delattre (1965).

Very detailed comparisons between English and German can be found e.g. in Grabe (1998). For contrast especially, she regarded pitch peak alignment: in English, the fundamental

[4]Unmodified with respect to the height of a prenuclear accent tone.

frequency peaks during the accented syllable, while in German the peak is aligned with the right edge of the syllable or even later[5]. Another central point is a distinction between German as a truncating language and English as a compressing language. This means, that in German, in case of phrase-final accented words of statements (falling contour), an accentuation contains a lower pitch range the shorter the accented vowel is. It truncates the accentuation. English instead, increases the slope of the rise in case of a short accented vowel, and therefore compresses the full pitch movement to a shorter time interval.

Trim (1988), in his contrastive investigations of German and English, assumes that final rises in German are always used for questions and other non-final utterances, whereas they can also occur at the end of English sentences as markers of contradiction and contrast in terms of a fall-rise.

To summarize, we expect almost uniquely final rising contours for German questions, but a notable proportion of falling questions in English, possibly together with an overall higher register. Acoustic correlates of questions should be a high and late peak on the sentence-final syllable in German, but a falling contour and a relatively low peak in this position in English. The English questions will possibly show an overall higher level of fundamental frequency.

2.4 The Contrastive Analysis Hypothesis and the Acquisition of Phonology

The Contrastive Analysis Hypothesis has emerged in the 1950s and is still very influential in second language research and applied linguistics. For many studies it is the explicit or implicit "philosophy" behind the actually treated research questions. It refers to Weinreich's hypothesis about languages in contact (Weinreich, 1953) and was elaborated e.g. by Lado (1957, 1964) and Thomason and Kaufman (1988). The principal idea is that the second language speaker is the smallest instance of intensive contact between languages, namely the native language and the target language. The main assumptions of the original (or "strong" (Wardhaugh, 1970)) version are as follows:

1. All L2-speakers' linguistic performance can be explained by the relationship of the native and the target language.
2. Therefore, these two languages have to be compared in detail for similarities and differences.

[5] At least if the accent is at the end of a phrase.

3. Because the main process is transfer (or interference). It means that features of one language are transferred to the other. Transfer can be positive if the languages are share certain features. Or it can be negative if unshared features are transferred, which results in "typical" errors for learners of a certain target language with a certain linguistic background .

The language of the learner uses is often called "interlanguage". It can be defined as a "system intermediate between the mother tongue and the target language" (Corder, 1994, p. 23). The second language speaker thus uses features of the target language as well as missplaced features of the native language.

This strong assumption could not be held for long. The main caveats were that first, not all learners of one native language committed the same errors when learning a specific second language. And second, it could not account for gradual differences in difficulty (for a more detailed discussion, see e.g. Major (2001)). These gradual differences come to light especially for the learning of the phonologic/phonetic characteristics of a target language.

One of the most influential models concerned with the learning of speech characteristics is the Speech Learning Model (SLM), developed by Flege and colleagues (e.g. Flege, 2003; Flege & Hillenbrand, 1987; Flege et al., 1995; Flege, 1997; McAlister, Flege, & Piske, 1999). This model is very much concerned with the relationship of the L1 and the target language of a learner as "bilinguals cannot fully separate their L1 and L2 phonetic subsystems." (Flege, 2003, p. 326). More specifically the two categorically organized systems influence each other. With this core assumption, some key properties of the learning process and specialities of foreign language speech can be derived. The first is that sounds of the target language that are similar to sounds in a speaker's native language are more difficult to learn than sounds that are clearly distinct to the phonological categories of the native language[6] "[...] the SLM predicts that the greater is the perceived phonetic dissimilarity of an L2 speech sound from the closest L1 sound, the more likely it is that a new category will be created for the L2 sound." (Flege, 2003, p. 328). The relative ease of the acquisition of "new" compared to "similar" (Flege & Hillenbrand, 1987; Flege, 1997) sounds is caused by the fundamental processes of categorical processing in phonology. If a sound of L2 is subsumed to be a variance of a sound in L1, a creation of a new phonetic representation is suppressed, as the L1 sound seems to be appropriate for the acquired sound[7]. The sound system can be reorganized in the time course of acquisition, when the learner has acquired more fine-grained representations of the phonological inventory of the target language.

[6] See also C. T. Best and Strange (1992).
[7] See also the Perceptual Assimilation Model (PAM) by Best and colleagues e.g. in C. T. Best (1994); C. T. . Best (1995); C. T. Best, McRoberts, and Goodell (2001).

More controversially discussed is Flege's assumption, that the phonetic properties of a second language can be learned by learners of all ages, although the perception and correct production of a "similar" sound are less likely when age increases. This is contrary to the so-called Critical Period Hypothesis, which essentially says that the acquisition of new sounds in a native-like way is impossible if the learner has lost child-like perceptual and productive plasticity (Johnson & Newport, 1989)[8].

But the key for all predictions based on the SLM or comparative models is the categorization and re-categorization process for phonemes. The properties of the phonetic categories of the native language of a learner interfere with the acquisition of the categories, or category boundaries of the second language. To summarize the section about the contrastive analysis perspective and the related popular speech learning model, one may cite Corder, 1994, p. 23:

> "the mother tongue is the starting point for the acquisition of the second language, which then proceeds by a series of restructurations of the mother tongue or a sequence of approximate systems progressively more similar to the target language. [...] the acquisition of the pronunciation of a second language is indeed largely a matter of progressively restructuring the mother tongue in the direction of the target language."

The assumption, that the native language is the starting point of second language acquisition has been challenged, mostly by empirically grounded studies which assume universals in second language production.

2.5 Universals in Second Language Speech Production?

One of the challenging findings in the field of speech production is the appearance of phenomena that are neither present in the native nor in the target language. For example, Broselow, Chen, and Wang (1998) found instances of final devoicing in L2 English by Chinese native speakers. There is no final devoicing in English and there are no closed syllables in Chinese, so the second language pattern is neither derivable from the acquisition of the target language system nor from L1 interferences. Broselow et al. (1998), in an Optimality Theory approach, explained that pattern as the "emergence of the unmarked", which means

[8] A discussion of the age effects on phonetic learning would be far beyond the interest of this study, as all the speakers examined later have started L2 learning as adults. The interested reader can be referred to an overview of evidence supporting or corrupting a negative correlation of Age of Acquisition and L2 pronunciation performance e.g. by DeKeyser and Larson-Hall (2005).

that the foreign language learners do not apply the native constraint ranking to L2, but they have not yet acquired a target language like reranking of the constraints. The result is, that they rely on a constraint ranking directly issued from universal grammar, a kind of phonological grammar that has not been overformed by language specific conventions. This is not the place to discuss these far going assumptions issued from the theoretical fundamentals of their study, but it has to be emphasized that L2 patterns may not always be accounted for by properties of the native or the target language, respectively.

This point of view is defended and illustrated with a large number of empirical evidence in syntax, concluded in Klein and Perdue (1997). They describe properties of the *Basic Variety*. This idea issued from substantial research on word order patterns in immigrant speech with various native language - target language relations, conducted with beginners in an un-supervised learning environment. The key features of the Basic Variety can be described as a

> "relatively stable system [...] which
>
> - seemed to be determined by the interaction of a small number of organizational principles,
> - was largely (though not totally) independent of the specifics of source and target language organization,
> - was simple, versatile and highly efficient for most communicative purposes" (Klein & Perdue, 1997, p. 303)

Most of the organizational principles are based on semantics and information structure, for instance the "control asymmetry between referents" (Klein & Perdue, 1997, p. 314f.) and focus. Klein and Perdue (1997) did not examine the intonational properties of their data. Instead, they point out that focus marking in the Basic Variety mainly relies on word order and not on intonation (Klein & Perdue, 1997, p. 317). However, there are three points which make their work relevant for the current study:

First, they reject the classical "target deviation perspective [...] The 'learner variety' is not perceived and studied in terms of what it is, but in terms of what it is not." (Klein & Perdue, 1997, p. 306). Investigations of second language speech are thus often only focused on potential difficulties that arise from differences between the native and the target language. But one will probably overlook features that are typical for second language speech, independent of whether such qualities differ between the two languages. So, this view is clearly opposing the Contrastive Analysis Hypothesis.

Second, they studied unsupervised learning. Prosodic patterns are not (yet) integrated into the curricula of foreign languages classes, neither in school nor in university. And none of the learners reported explicit instructions of how to produce appropriate intonational patterns. Thus, the learning of prosody can be regarded as (largely) unsupervised.

And third, the word order principles that Klein and colleagues found are based on semantics and information structure. Of course, focus is one of the main categories of information structure. And the Basic Variety "represents a particularly natural and transparent interplay between function and form in human language." (Klein & Perdue, 1997, p. 304). The transparency between form and function is an essential point in intonation research, which will be presented in the following section.

2.6 Prosody as a Universal in Human Communication

One of the fundamental concepts of modern linguistics is the arbitrariness of the sign and the meaning. And just this point is challenged by some evidence and theories in the case of prosody. In the following paragraphs I will give a short overview of the central points of the discussion.

The assumption that prosody is rather a prelinguistic than a language specific way of information transmission is defended in the work by Dwight L. Bolinger. For him prosody is defined as "all uses of fundamental pitch that reflect inner states." (Bolinger, 1989, p. 3) Its central function is to express the speaker's affects. This leads to strong universalist statements, like: "The fact is that human speakers everywhere do essentially the same things with fundamental pitch." (Bolinger, 1978, p. 515).

The encoding of focus and sentence mode can be interpreted as an expression of affects. The intonatory accent occurring with a focused element, for him is "at the service of emphasis" (Bolinger, 1989, p. 3) and "not determined by syntactic structure but by semantic and emotional high-lighting" (Bolinger, 1972, p. 644). Accent is gradual – "A bigger thing produces a bigger feeling" (Bolinger, 1989, p. 22) – and by no means grammatical in terms of distinct categories. Syntax may interact as certain positions or structures may be emphasized more likely than others, but there is no obligatory syntax-accent relation[9].

[9]This view is contrary to many studies, who e.g. discuss the position of a so-called sentence accent, or neutral focus position (e.g. Féry, 1993; Jacobs, 1988; Höhle, 1982; Kiss, 1998; . Pierrehumbert, 1980, and many others).

Similarly, the distinction of questions vs. statements on an intonatory base is not a grammatical one, but a distinction between "curiosity and confidence". The actual sentence type, and its pragmatic content is not necessarily determined by its intonation. "No intonation is an infallible clue to any sentence type: any intonation that can occur with a statement, a command or an exclamation can also occur with a question." (Bolinger, 1989, p. 98) But he is ready to prudentially allow for a global tendency for questions:

> "From all available descriptions, it appears that the one distinctive trait is HIGHER PITCH, either throughout or, especially, toward the end. Though the rising terminal seems to be most widely used (and even where less used, may come through as 'more surprised', hence 'more curious'), it is by no means universal." (Bolinger, 1989, p.103f.)

To summarize Bolinger's point of view, one may state that first his connection of intonation to affects allows for rather strong assumptions about the function of intonational contours throughout all human languages, as affects can be assumed to be universal. On the other hand he emphasizes that strong predictions of concrete intonational forms corresponding to the universal functions are not possible. There might be tendencies, but even the position of the sentence accent is only predictable "if you are a mind-reader" (Bolinger, 1972); and presumed global tendencies for sentence mode encoding are not universal in the sense of a global rule.

The most important problem of this point of view is that it hardly results in testable predictions.

However, there are stronger hypotheses about a universal form-meaning relation in prosody. Ohala (Ohala, 1983, 1984), has proposed a controversial position regarding the use of fundamental frequency among species. In an ethological perspective he observed that there is a tendency that fundamental frequency is negatively correlated with power. The lower the fundamental frequency is, the higher is power and vice verso. The reason of this correlation is that lower voices are connected to larger vocal tracts, thus larger species, and thus, the dominating ones. And he applied this to the use of fundamental frequency in human communication. For example, polite utterances are spoken in a higher register than commands of an authority. And, relevant to the current study, a statement is more powerful than a question, because the one who utters a statement signals the power of knowledge of what he is talking about. On the other hand, the one who utters a question is less powerful, as the speaker is dependent on the one who should answer.

This point of view has been taken up by Gussenhoven in his Biological Codes Theory (Gussenhoven, 2002; Gussenhoven, 2004). He distinguishes three Universal Codes: The Frequency Code, the Effort Code, and the Production Code. The Freqency code is an extension of Ohala's assumptions on the meaningful use of fundamental frequency. He cites a number of perception experiments that support the theory that higher pitch is interpreted as friendliness, politeness, uncertainity (e.g. Biemans, 2000; A. Chen, Rietveld, & Gussenhoven, 2001; Uldall, 1964; Rietveld, Haan, Heijmans, & Gussenhoven, 2002), especially towards the end of the utterances. Uncertainity vs. certainity is highly related to the intonation of questions vs. statements. For example, Gussenhoven and Chen (2000) report a perception experiment in which native Dutch, Chinese, and Hungarian listeners judged CVCVCV sequences varying in peak height and peak alignment for being statements or questions. The participants were instructed that the sequences were extracted from an unknown language. Irrespective of the default cases of their native language background (Dutch using high end pitch, Chinese an overall higher register, Hungarian a marked accent distribution), higher peaks and higher ends were preferably interpreted as questions.

Nonetheless, there are many languages which do not associate higher or rising pitch with questions, like Swedish, or also some dialects of English. Gussenhoven's theory can handle these exceptions by assuming "grammaticalization". Prosodic form-function relation are grammaticalized. Most frequently the follow the predictions of the Frequency Code. But, if some "unnatural" contour is used for a specific form/function relation (e.g. a falling contour in questions), a speaker community can agree on a deviation from the "natural" constraints.

The Effort Code summarizes effects of forcefulness on certain elements of an utterance to signal their importance. Thus, the Effort Code is correlated with focus, but also with affective surprise. More effort leads to higher articulation precision and a higher pitch excursion. This is what is generally found with focus. But pitch height per se is never a decisive criterion for 'emphasis', only its relative height compared to the speakers register and the actual context of the syllable are strong cues. Gussenhoven has "no examples of "unnatural" grammatical focus expression." (Gussenhoven, 2002, p. 51) Unfortunately, Gussenhoven does not include intensity in the Effort Code. Its meaning would be straightforward: more effort is correlated with higher intensity.

The third Biological Code is what Gussenhoven calls the Production Code. It is very similar to Lieberman's breath groups (Lieberman, 1967), and thus accounts for the usual declination thoughout an utterance. In the beginning, there is more subglottal air pressure than at the end which favors intensity and pitch decreasing over the time between two breaths / breaks.

Gussenhoven presents both, evidence for shared interpretations of prosodic correlates of linguistic (and paralinguistic) functions, but also deviances in details due to language specific “grammaticalizations”, especially for the Frequency Code. However, he does not include intensity as a relevant parameter for a higher effort on emphasized syllables. And, up to now, evidence for language specific and language independent use of the Biological Codes is tested empirically only for perception.

To conclude this section, one must state that an assumption of strong, detailed and testable features that are similar for all languages under all circumstances is not found in the literature. But there is Bolinger’s view of a dominant prelinguistic function of prosody: it is encoding affects which are presumably similar among all humans, and there are tendencies across languages that some form - function relations correlate. But the predictive value of his view is very limited. More valuable for a derivation of testable hypotheses are Gussenhoven’s Biological Codes. They connect prosodic patterns to the psycho-physiological correlations of speech production. The presumed universal form-function relations can be specified and deviated by language specific grammaticalizations. All in all, the Biological Code provides rather strong predictions for a “universal” encoding of sentence mode and focus, while it still is capable to handle language specific conventions.

3 Second Language Prosody: Issues and Findings

The number of studies investigating second language prosody is rather limited. Moreover, most of the existing publications are barely comparable: there is a wide range of annotation systems or description models, and we find different subject populations from various languages, heterogeneous in number and level of language acquisition. Relevant work can be roughly can be classified as an assessment of the influence of prosody in foreign language speech, as a presentation of previous findings of focus, contrast, or emphasis encoding in L2, and as a summary of the processing sentence mode encoding by learners of a second language.

3.1 Prosody and Foreign Language Accent

One part of previous studies is motivated by the fact[1] that prosody plays an important role for the perceived foreign accent of L2 speakers.

For example, Els and Bot (1987) try to quantify the effects of intonational deviances of a perceived foreign accent and the traceability of the L2 speakers native language. They compared the performance in native language identification for Dutch sentences spoken by native speakers of Dutch, French, English, and Turkish by Dutch language teachers. The participants heard three versions of each sentence: original, monotized (flattened pitch signal) and low-pass filtered (poor segmental information). Correct foreign language background identification reached 68 % of the sentences in the original version, 43 % of the sentences with monotized fundamental frequency, and 20 % of the low-passed sentences. The differences were significant at each level. So, prosody has an effect for the classification of foreign speech, but it is less strong than filtering out segmental information.

[1] Or as Els and Bot (1987, p. 147) put it: "the empirically substantiated observation that hardly any foreign language learners of over eleven or twelve years of age manage to acquire such proficiency in pronunciation in the foreign language that they are consistently taken for native speakers."

A concurrent ranking of the factors influencing foreign speakers performance was found by Anderson-Hsieh et al. (1992). They investigated the correlation of ratings of the pronunciation of second language speech to segmental errors, syllable structure errors, and a rating of prosodic native-likeliness. Short sentences of sixty participants from 11 language groups were examined. All examined parameters had a significant influence on the pronunciation judgements. Prosody showed the highest correlation to the goodness of pronunciation, compared to both other factors. Also, prosody showed a significant correlation for all subgroups[2], but not the number of errors in segments and syllable structure.

A more recent study by Jilka (2000) is concerned with the acoustic details of English produced by proficient German native speakers and German produced by proficient (American) English native speakers. He investigated both, the production and the perception of foreign language speech and compared the learners to native speakers. He used different speaking situations (read sequences, free speech and repetitions of heard prompts) to assess the main intonational characteristics of the foreign languages, and was interested in accent placement and the phonetic realizations of certain prosodic patterns. He used the analysis by synthesis principle, which means that foreign language utterances are synthesized, certain specific qualities, like the height or the steepness of the fundamental frequency is changed, and if the synthesized versions sound more similar to the target-language, the changed deviation is a specific pattern of foreign language speech.

Jilka distinguishes four groups of deviations of L2 prosody: Category type or placement errors, category transfer, interferences in the specific acoustic realizations, and differences in the overall characteristics of the prosodic pattern (cf. Jilka, 2000, p. 83 ff.). The first are, for instance, a higher number of accents [3] or a wrong placement of a tonal category. They can be motivated neither by native language nor by target language patterns. Possible cases of categorical transfer are, according to Jilka rather difficult to identify, as many function-form mappings are similar, or realized with floating category boundaries. But, for example, in a sentence repetition task, the German learners of English did not realize the original fall-late rise for a short question, but used the German type of rising nuclear accent. As an example for phonetic transfer, German speakers of English have difficulties to start a phrase in the upper half of their pitch range, which is frequent for native American English if an accented syllable is close to the sentence initial position, but not in German. And

[2]The "subgroups" did not necessarily contain languages that were historically related, but they were regionally distinct (e.g. East Asia languages, Indian Subcontinent languages). Certainly, one has to raise serious and reasonable doubt against this grouping. But the main point of their study is that prosody continues to be correlated to the overall ratings, but segmental errors or deviances in syllable structure do not, irrespectively of the linguistic heterogeneity of the language "subgroups".

[3]Too many phrasal accents are also reported in Grosser (1997), Ueyama and Jun (1998) and Archibald (1998).

overall, English native speakers transfer the high rate of rather steep pitch movements to German. Manipulations of these factors in synthetisized speech clearly showed their influence in better foreign accent rating. On the other hand, segmental information still seems to be more present in the ratings of foreignness.

To conclude, one can say that prosody contributes significantly to the perception of foreign accent, and that it contains information that indicate that the speaker is non-native of the language, some may be specific transfer from the learner's linguistic background.

3.2 Prosodic Encoding of Phrasal Stress in a Second Language

While there are some studies dealing with the assignment of word-level stress in a second language (e.g. Gut, 2003; Juffs, 1990; Mairs, 1989; Shah, 2004; Ueyama, 2000), the number of investigations of prosodic stress encoding on a phrasal level is more limited.

Gårding (1981) investigated phrasal intonation of Swedish by a French native speaker and French by a Swedish native speaker. She made laryngograph recordings of two assertive sentences and compared the syllable and accent distribution of the speakers in their native and the corresponding second language. She sums up her impressions as "it is quite easy to hear the difference between the Swedish and the French speaker. And it is next to impossible to hear the difference between the two speakers' native and foreign performances. The reason is, of course, that the prosodic transfer has been almost total." (Gårding, 1981, p.158)
For example, French spoken by the Swedish native speaker contained too many accents which were combined with lengthening as is the case in Swedish, but not in French. The French native speaker speaking Swedish had no phrase internal accents and no length variation. According to Gårding, prosodic features and, above all, features that are combined in one language (length and pitch accent in Swedish) and independent in another (in French) are "extremely persistent" in second language speech and "very difficult to learn" (cf. Gårding, 1981, p.162).
One can certainly object the lack of a larger database (only two sentences), and the absence of statistical examinations. Furthermore, other parameters, like intensity, are not investigated at all.

The latter is examined in a study by Kelm (1987). He compared Spanish utterances spoken by (Mexican) Spanish native speakers, (American) English native speakers, and (American) English learners of (Mexican) Spanish who have lived in Mexico for an average of one year.

He used two structured free-speech tasks (activities and picture comparison) that elicited contrasts. He compared the fundamental frequencies of contrasted syllables to the “normal tone” - which he aimed to find for the first syllable of declarative sentences. The same syllables were used to compute an intensity ratio between contrasted syllables and the “normal intensity”, the mean intensity of the first syllable.

These two differences, contrasted vs. normal fundamental frequency and contrasted vs. normal intensity were examined between the three groups. His results showed that pitch and intensity are used to a significantly larger extent by the native English speakers, than by the English learners of Spanish, than by the Spanish native speakers. The the differences between normal and contrasted syllables in intensity were significant only between the English and the Spanish native speakers, while the English learners of Spanish did not differ from either native group. Kelm concluded that “advanced non-native Spanish speakers continued to use pitch and intensity to express contrastive stress even when speaking Spanish.” (Kelm, 1987, p. 633)

This strong statement for a complete transfer is probably too simplistic. For example, the use of intensity is weaker for the Spanish spoken by English native speakers. Furthermore, relating all accents to the first syllable of utterances (his “normal” tone and intensity) can be tricky. In this position, deaccented syllables (like determiners) are likely to occur, different sentence modes and different utterance lengths which may influence pitch and energy assignment, as well as the preparation effects of upcoming accents that increase the onset, at least for English (cf. e.g. Cooper et al., 1985). Moreover, Kelm (1987) does not consider other explanations except for transfer of L1 at all.

Intonational correlates of several discourse functions are reported in Wennerstrom (1994), among them contrasted or new information against given information and phrase boundaries. She investigated the speech of Spanish, Thai, and Japanese learners of English, and (American) English native speakers. Recordings of a read text and two free speech tasks eliciting new and contrasted information by ten speakers of each group were examined. Her results showed that none of the foreign language learners used prosodic correlates to the same amount as the English native speakers. Worst performers, also with regard to the inappropriate use of falling boundaries between closely connected sentences, were the Thai native speakers. Wennerstrom does not attempt to relate all patterns to transfer from L1, but she claims that intonation in Thai as a tone language does not cover discourse functions.

Just the opposite interpretation of the role of an L1 tone language was explicated in Y. Chen, Robb, Gilbert, and Lerman (2001). There, the acoustic correlates of sentence stress of proficient native Mandarin learners of English were investigated and compared to utterances by native speakers of (American) English. They studied fundamental frequency, duration, and intensity in the positions of subject, verb, and object in the sentence "I bought the cat there." Mandarin is generally considered as a syllable-timed language without strong stress patterns and a relatively constant syllable duration within a sentence (cf. Santen & Shih, 2000)[4]. The results showed that Mandarin females use significantly higher pitch excursions to mark for stressed vowels in English than American female speakers on all positions. For male speakers, this effect was not significant. As for duration Y. Chen et al. (2001) report a significantly smaller difference between an accented target word compared to a corresponding non-accented target word for the female and male L2 speaker group. There was no significant difference between the groups in the intensity of stressed and unstressed vowels. These effects are mainly accounted for by transfer: Lexical tone interferences should explain the higher F0 excursions and the less extensive usage of duration and intensity should result from a less pronounced deaccentuation (less low values for unstressed vowels), caused by an interplay of Mandarin more balanced syllable / sentence organization.

A very similar study using the same target sentence and elicitation method was presented in Yeou (2004b). He investigated the acoustic correlates of contrastive focus of 10 (5 male, 5 female) Moroccan native speakers speaking English compared to the utterances of 40 (20 male, 20 female) American English native speakers. He found that the Moroccan speakers used a higher F0 and duration differences for accented constituents compared to the same constituents in unaccented conditions than the American speakers. This accounts also for duration: Moroccan English has longer durations on the accented constituent than American English. But the American speakers used intensity more effectively than the Moroccan speakers. Unfortunately, the paper did not contain any statistical confirmation of the results. Yeou (2004b) mainly explains this pattern by transfer from L1. Longer duration and higher F0 excursion were also found for medial focus in Moroccan Arabic (Yeou, 2004a). The comparably less increasing of intensity was accounted for by a possible compensatory effect of the higher pitch and duration excursion for accented words in Moroccan English.

Certainly, the studies by Y. Chen et al. (2001) and Yeou (2004b) can reclaim the plus of a larger number of speakers involved. On the other hand, none of them presented a similar dataset from the native language of the learners. But as the only source of L2 prosodic features regarded is L1 transfer, a substantial source of comparable L1 intonation would be necessary. And of course, one single sentence is very probably not representative for patterns of L2 speech. Finally, eliciting sentence stress by visual emphasizing of the target

[4] Nonetheless, Mandarin Chinese speakers seem to use higher F0 and longer duration, as well as higher intensity for stressed syllables (Shen, 1993; Xu, 1999)

words (uppercase italics) is quite crude: from a linguistic perspective , stress should be related to communicative functions.

A simple and strong effect of transfer is also challenged by the results of Grosser (1982) and Grosser (1997). He presents a longitudinal study with young (10-12 years old) learners of English from Austria. They were tested at several points from the very beginning of acquisition (formal instruction of 5 weeks) to two years after the beginning of instruction. In the beginning the learners stressed too many syllables, more than one would assume for the native and the target language. And in one detail, the peak alignment of accented syllables, the learners develop from some, possibly “default” peak position, to the peak position of their native language – and not to the one of the target language.

The peak alignment of accents is also the major issue in Mennen (2004). She compares the utterances of Dutch native speakers in their second language Greek to their utterances in Dutch, as well as native Dutch and Greek control groups. She found bi-directional interferences in the accented syllables in the utterances of four out of five L2 speakers. The F0 maximum was somewhere in between the native and the target peak position. One speaker was able to use specific peak alignment patterns for both, his L2 (Greek) and his L1 (Dutch). Mennen (2004) made bi-directional interferences responsible for the pattern of the first subjects, while the latter apparently has acquired the language-specific distribution of the peak positions in accented syllables.

Gut (2000) reported a related experiment, assessing the acoustic correlates of emphasis in native English, native German, and utterances of German learners of English. The participants read lists containing five similar words, and were instructed to emphasize an underlined word in three different (medial) positions. The main findings was that emphasized syllables in native English differ from non-emphasized in a higher fundamental frequency and less consistent and less strongly different intensity. In native German utterances, fundamental frequency excursions were used less consistently for emphasis encoding, but higher intensity is used to a greater extent in the analyzed syllable. As for German English, Gut found that the observed patterns reflect encoding preferences in German. If a German learner of English uses the fundamental frequency for signalling emphasis in his native language, he will also use it when speaking English: if he or she prefers intensity in German, it will also be highly correlated with emphasis in his second language.

There are some pitfalls in that study: first, no statistical analysis is presentend, not even the mean values over all speakers are given – the data only shows single speaker examples. This would be helpful, especially as the language specific role of fundamental frequency and intensity for accent encoding are in contrast to findings in larger corpora (for English:

Kochanski et al., 2005; for German: Wagner, 2002). Second, vowel-intrinsic correlates of fundamental frequency and intensity (cf. Beckman, 1986, p. 126 ff. and 141 ff. and literature therein) are neglected, although the analyzed vowels of German and English are different. And third, as the control group for native German and the learners of English were the same people, a strong effect of speaker-specific preferences is not surprising. Additionally, bi-directional interferences as proposed by Mennen (2004), could also be considered.

3.3 Prosodic Encoding of Sentence Mode in L2

First, one of the earlier works will be presented. Backman (1977) compared the prosodic properties of utterances of a (Mexican) Spanish native speaker beginning to learn English, and an advanced learner. She measured some aspects of fundamental frequency movement, and used expert hearers' judgements and correction propositions for an assessment of learner specific intonation deviation. She found that the range of the final fall in statements was very similar to the Spanish pattern for the beginner, and intermediate between Spanish and English for the advanced learner. For yes-no questions, a stronger global elevation of the overall pitch level was found for English, but by far not as strong as for Spanish. The beginner still used almost the Spanish elevation, while there was a stronger approximation to the lower level for the advanced learner. On the other hand, while the native questions in both languages contained a final rise (higher in English than in Spanish), the beginner used a weak fall, and the advanced learner a rise weaker than the native and the target language. So the learners used a form-function relation that – in the Beckman's point of view – is not occurring in neither the native nor the target language of the learners, and therefore cannot be explained by transfer. She continues her argumentation by referring to a paper examining Dutch learners of English (Jenner, 1976) which states exclusively falling contours, too. This, and other facts like reduced pitch range were similar between the Spanish native and the Dutch native learners. So she raised the question of "universal problems in acquiring a second intonational system" (Backman, 1977, p. 34). There is no need to note the lack of representativeness of Backman's study, as she analyzed the utterances of only one speaker per group. And after all, she ignored that – at least in context – a fall is a perfectly acceptable contour in syntactically determined yes-no questions, as pointed out for instance by Bartels (1997); Fries (1964); Grabe (2004).

Willems (1982) found – in an analysis of contours in wh-questions – the opposing tendency. The English native speakers used a steep final fall in 94 % of all cases. The Dutch learners instead used a rising contour in more than half of the questions, and in the cases of a fall, the range was only half as much. (cf. Willems, 1982, p. 105). For yes-no questions, he found that Dutch learners of English used a higher pitch excursion for than an English native control group. He concludes : " No doubt under the influence of the usual rise in the intonational intonation system of their native language, they created the impression that they emphatically wanted to stress the interrogative character of yes-no questions by exaggerating the final rise."(Willems, 1982, p. 107).

A comparable result is presented in the recent study of the prosodic properties of English utterances by German native learners in Jilka (2000). The pitch range of the final rise in the L2 questions was higher than in the same questions uttered by native English speakers.

Shen (1990) explored the perception and the production of French statement vs. question intonation by Chinese speakers. Chinese question intonation for unmarked questions is a higher beginning of the sentence with a high or falling tune at the end of the phrase, often correlated to a higher pitch register, overall. French question intonation, following Shen, is a phrase final rise. She was interested in how Chinese native speakers acquire the function of final pitch movement which is not distinctive in the native language. First, a perception study showed that Chinese native speakers who had no knowledge of French were able to categorize French utterances in being falling or rising with the same accuracy than French native listeners. Second, the utterances of six proficient Chinese native learners of French were compared to native speech. The L2 laryngograph recordings were judged by French native listeners of being nativelike in a wide range of 19 to 80 % of the utterances per individuum. More interestingly, the utterances were judged as statements or questions, and there was no difference in correct categorizations between the learners and the native control group. Thus, the sentence mode encoding by means of a final rise for questions and a final fall for statements was seemingly sucessfully acquired.

But a phonetic difference between the native and the learned question marking rise was encountered: the Chinese L2 speakers either produced a higher rise than the French control group, or the introduced a pre-final fall. (cf. Shen, 1990, p. 130). Shen recurs on "hypercorrection" or "articulatory habits" (ibd.) as an explanation, as well as possible effects of a "phonological filter"[5], which would favor an exaggerated rise in the learner's utterances.

More recently A. Chen and Mennen (2008) investigated the tonal correlates of questions of four untutored learners of English whose native language was Italian, more specifically several Italian dialects. They analyzed the phrase final contours (falls / rises), as well as

[5]According to Rossi (1971) pitch height differences of less than 20 Hz are not perceivable in normal conversation. Therefore, the learners increased the range strongly to make sure that they would be understood.

the height of the pitch maximum and the alignment of the pitch peak within the accented syllable with respect to the native and target language contours and the question types of *wh*-questions and declarative questions. They showed that overall, *wh*-question tend to end in a fall, while the declarative questions are rather produced with a final rise. However, the preferred final contours for the declarative questions differ considerably between the speakers. While two learners show a clear preferences for a rise at the end of the utterances, one speaker is indifferent in the usage of falls and rises, and the forth one clearly prefers falls over rises. The target language (London English) clearly prefers a final rise in declarative questions. A. Chen and Mennen (2008) prudentially argue that transfer from the different dialects of the learners may be a reason for the different preferences of the final contour. Peak height and the alignment of the pitch maximum, two other potential cues to question intonation, are not used consistently. The authors argue that this might be caused by the "relative importance of the cues in the target language."

3.4 Summary

Several points should be summarized from the above section. First, literature on L2 prosody is very scarce. There are certainly less than 20 studies concerning prosody beyond the word (accent) level published. Second, the number of native and target languages is very limited. With some exceptions, most of the time English is one of them. Studies on the nature of prosodic patterns in a second language thus must be influenced by the characteristics of English. Third, the studies considering L2 prosody are barely comparable as they use different methods of description. Fourth, the studies that present detailed analyses of L2 speech are mostly based on a very small number of learners (five or less!), or do not submit the reported patterns to statistical analysis. Fifth, none of the studies investigate systematically the influence of other factors than transfer. Patterns which are not explainable by native language interference remain "unexplained" or unexplored.

And finally, and this is not surprising considering the lack of reliable data and systematic exploration of possible mechanisms, there is no connection to models of speech (prosody) production. They will be considered in the next section.

4 Models of Prosody Production

In this chapter, an overview of how three different models of prosody production would handle the encoding of contrast and sentence mode is given.

4.1 The Production of Intonation in Levelt's (1989) Model

Still, there is no model of speech production available that can compete in completeness and rigor with Levelt's model of speech production presented 20 years ago in "Speaking" (1989). Levelt divides the speech production activities into three subsystems: The Conceptualizer, the Formulator, and the Articulator. The encoding of sentence mode (mood in Levelt's terms) and focus is extensively treated for all of them.

4.1.1 Planning Sentence Mode and Focus

The planning of the message is undertaken in the Conceptualizer. Levelt assumes an incremental two-stage process with macroplanning and microplanning that results in preverbal messages which are the input for the Formulator. During macroplanning, the speaker transposes a subset of his communicative intentions in a sequence of speech acts [SA]. These are "specified for intended mood (declarative, interrogative, imperative) and content." (Levelt, 1989, p. 109). In principle, macroplanning is a process that is not language specific[1].

Microplanning is the process that deals mainly with information structure, thus prominence, focus, and topicalization, and language specific requirements of the input to the Formulator. Here, the elements of the message receive their markers of the information structure. The resulting preverbal message contains an abstract representation of the information structure,

[1]But see his discussion about the language specificity of semantic structures (Levelt, 1989, p. 105f.).

conserves the previously established sentence mode markers and now is adapted to language specific requirements.

4.1.2 Encoding of Sentence Mode and Focus

Once the preverbal message enters the Formulator, the processes of Grammatical and Phonological Encoding take place subsequently. The link between them is the surface structure. It includes a marker for the sentence mode, but Levelt is very vague in describing its mechanisms and qualities[2]. The case of focus is more explicit and clear in Levelt's model. Every intonational phrase has to have at least one focused constituent. On the surface structure, it receives the focus diacritic f. If the constituent spans more than one lemma and there is no other information in the preverbal message, the nuclear stress rule (cf. Selkirk, 1984) applies, and the diacritic symbol is passed down to the rightmost lemma under the focus node. So, the surface structure contains all information of prominence assignment and - in a yet unknown manner - information about the sentence mode.

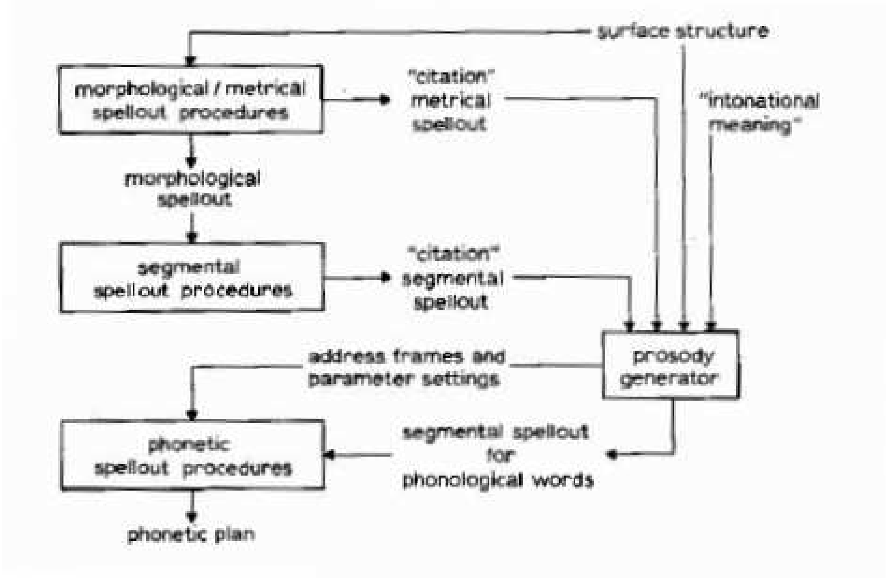
4.1.3 Intonational Plans and the Prosody Generator

The encoding of intonation takes place in the Prosody Generator. Its input are the segmental and metrical spellout, the intonational meaning, and the surface structure which carries the information we are interested in here, sentence mode and focus. The output of the Prosody Generator is all information needed for the phonetic spellout procedures. It thus integrates all information available. Levelt assumes meaningful, functional tones that are assigned incrementally to syllables. These tones are modified by prenuclear tunes. He lists seven tones (but expected more) that can be chosen for the syllable sequence from the nucleus to the end of an intonational phrase. Among them, there is the "high fall", which implies "seriousness" and a "matter-of-fact-way" and is "probably the most common tone for declaratives" (Levelt, 1989, p. 312). The other tone which one could expect for the elicited echo-questions is the "high-rise", with a nucleus starting slightly up from the previous tune and then rising.

The tunes and tones are moderated by a *key* for every intonational phrase in longer utterances. One of the three keys (low, mid, high) delimits the peak extention in a phrase, and thus serves as a cue for backgrounding and foregrounding information. The other modulation

[2] " .. how mood and modality are indicated in surface structure is not well known; we suppose that such indicators are recognized by the Phonological Encoder, which generates the appropriate pattern of intonation (in particular the characteristic boundary tones)."(Levelt, 1989, p. 180)

Figure 4.1: Phonetic encoding and the Prosody Generator, from Levelt, 1989, p. 366



of the tones is the *register*, the "pitch level of the baseline" Levelt (1989, p. 316). A higher register transports a higher level of emotions, but also weakness, helplessness. Storage of and access to these tones is not clear in Levelt's model: "The speaker surely has no intonational lexicon with the whole ready-made contour templates. The closest thing to that may be something like a tone lexicon, a relatively small set [...] of canonical tonal contours. [...] In the process of phonological encoding, they have to be projected onto the stretch that extends from the nucleus to the end of the intonational phrase" (Levelt, 1989, p. 316). Levelt, like most of his contemporal colleagues, suspects potential universality:

> "The meanings of these major nuclear tones are, in part, conventional, and specific to English (or even to certain dialects of English). Still, there are aspects to these meanings that are probably more universal. Tones with a final fall express completeness, finality. Tones with a final rise express non-finality, openness."(Levelt, 1989, p. 315).

4.1.4 The Generation of Intonation

Levelt distinguishes three different subprocesses in the Prosody Generator. Two are global (over the whole sentence), and one is local. The first is the declination of the fundamental frequency contour over the whole (declarative) utterance, which is found more due to "physiological factors " than to the "speaker's phonetic plan" (Levelt, 1989, p. 399). The next process is the setting of key and register. The "key" information is accessed every time a pitch accent is made, and determines the amount of pitch excursion for the speaker in the actual context[3]. The register is the relative height of the fundamental frequency in the speaker's range. There are two possible processing ways: Either, the speaker has a global opposition of foregrounded and backgrounded information or a new setting for each pitch excursion, "depending on the accessibility or the contrastiveness of the particular lexical item." Levelt (1989, p. 400f)

As for the generation of tones, first the nuclear pitch movement has to be set. The nucleus can be higher or lower than the previous syllables. Then the *boundary tone* to the end of the intonational phrase is generated. The generation of the tones is strictly incremental. Different meaningful tones can be assigned to a sequence of syllables or coincide on one. But they cannot override or extinct each other (cf. Levelt, 1989, p. 401f.).

How should the Prosody Generator know which tone to select? Part of the information is available from the unfolding surface structure, among them sentence mood and focus. But that information can be overruled by spontaneous changes of the sequence of utterance, e.g. deciding to add another clause. A planned falling tone will thus be supplemented by a continuation rise. Information from the 'Intonational Meaning' like attitudes and emotions enter directly into the Prosody Generator. But the details of the Prosody Generator process are left open: " The Prosody Generator will, in some yet unknown manner, integrate these various sources of activation in making the final choice of tone". (Levelt, 1989, p. 403).

To conclude, Levelt's assumptions of prosody production mainly follow Bolinger's functional perspective, in that tones that have certain - sometimes language specific - functions and universal tendencies. It is important that he links prosodic planning to the essential communication functions and the other processes of formulating and phonological encoding. Yet, for predictions of second language prosody the model is restricted, as the access to the "set of canonical contours" is not described.

4.2 De Bot's Model of Second Language Speech Production

De Bot (1992) presents an adaption of the "Speaking" model for bilingual speech production. Although concentrated mainly on lexical access and resulting problems - like Code Switching

[3]One may think of it as a factor that each pitch accent is multiplied with.

- de Bot's bilingual production model checks all components in Levelt's model whether they are language specific or are not.

As for the Conceptualizer, de Bot argues, like Levelt, that the macroplanning component is not language specific itself, but it selects the language to be produced. Microplanning is language specific and adds language choice information to the preverbal message.

The language marker in the preverbal message activates principally language specific Formulators. This principal separation of Formulators for each language a speaker has learned is modified by several factors: first, if the languages are closely related, one can assume that the speakers will "for the most part use the same procedural and lexical knowledge" (de Bot, 1992, p. 9) which is (obviously) impossible in the case of very distant languages.

The input into the Prosody Generator summarized above "is largely language specific" (de Bot, 1992, p. 17), e.g. the meaning / form relation, the information of the surface structure, the segmental structure, and the intonation contours (tones). "If there were separate systems for the above-mentioned components for each language then it would be reasonable to assume that these systems would not influence each other; " (ibid.). But, on the other hand, various instances of transfer from L1 to L2, the determination of foreign accent by prosodic patterns, and the fact that prosody is hardly learned by late bilinguals convinces de Bot that "the existence of two systems is very improbable. [...] There is only one Articulator for bilingual speakers which has an extensive set of sounds and pitch patterns from both languages to work with. The extent to which these sounds and patterns are more or less perfect models depends on the frequency and quality of contact with the L2." [ibid.].

To summarize these proposals: de Bot assumes one Prosody Generator which is language independent. This allows to account for the difficulties in the acquisition of prosodic patterns in a second language. The properties of the L2 prosody should be derivable from either the native or the target language of the learner, or be gradually intermediate between them. This point of view can be subsumed as a perspective of the Contrastive Analysis Hypothesis. De Bot does not discuss the issue of how to deal with potential universal tendencies in prosodic form-function relations.

4.3 Fujisaki's Proposals of Speech Generation

Fujisaki's model of pitch contour analysis and generation (Fujisaki, Hirose, & Sugito, 1980a; Fujisaki & Hirose, 1982, 1984) is very influential in computer assisted speech analysis and synthesis. It has successfully been applied to a number of languages (e.g. English and

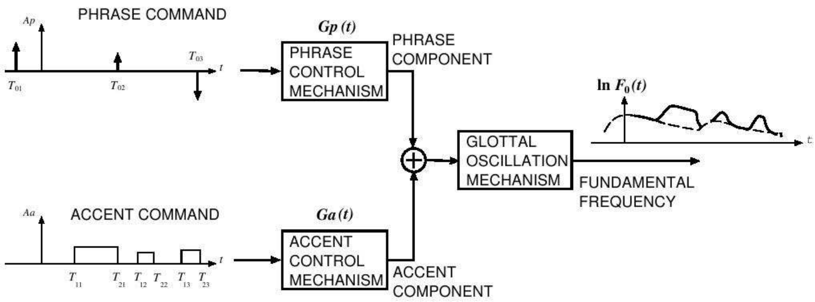


Figure 4.2: Scheme of Fujisaki's Model of prosody production, from Mixdorff, 2002, p. 55

Japanese in Fujisaki, 1980; German in Mixdorff, 1998 and Möbius, Pätzold, & Hess, 1993). It's core is a two-component additive function modelling fundamental frequency contours. The first component is called phrase component which starts at a base level, then rises rather quickly to a first peak from which a falling contour continues to the end. The second component is the accent component which means time-limited pulses of pitch height. Both components are logarithmically added and smoothed to the resulting F0 contour. The appealing innovation of this model is that all parameters (base values, height of the peak and length of the phrase component, as well as height and duration of the accent components) may be computed by the analysis of corpora and then be applied to generate (more or less) naturally sounding speech (e.g. Mixdorff, 2002).

Language specific parameters of the phrase component and accent components to several linguistic functions including word accent and perceived prominence (focus), can be analyzed from annotated corpora, and subsequently fed into speech generation models.

The advantage of the Fujisaki model is that it can provide rather accurate standard values for fundamental frequency patterns of a language while using only two components. Even the rising question intonation can be modelled by an additional accent component at the end of the utterance. It's disadvantage is that the values of the phrase and the accent component have to be computed for a lot of communicative situations if one wants to capture at least the main prosodic patterns of a language in use. They are static, and describe pitch behavior only for situations previously annotated by other means (syntax, perceived prominence). It is therefore more applicable to standardized speech production, e.g. in schedule announcements at railway stations. However, also the experimental stimuli used in the experiments presented later encode only two functions, namely corrective contrast and sentence mode. Now, in case one would clearly find different accent components for the native language of the learners (German) and the target language of the learners (English), a straightforward application

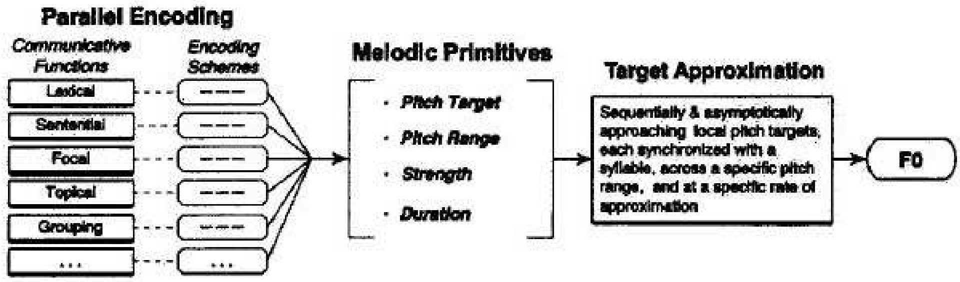


Fig. 11. A schematic sketch of the PENTA model. See text for explanations. The unnamed block at the bottom left indicates communicative functions yet to be identified.

Figure 4.3: Xu's PENTA model of speech melody processing. (Xu, 2005, p. 243)

of the Contrastive Analysis Hypothesis would result in a prediction that English spoken by German native speakers would use the native or the target language properties, or an intermediate accent component size. So, although Fujisaki's model is not psycholinguistic in a sense to be explicit for the encoding process, it allows for the most precise predictions of all prosody generation models.

4.4 The PENTA Model of Prosody Production

Finally, a recent model of prosody production by Xu (2005) will be presented. This model jointly considers the generation of lexical tone and suprasegmental intonation[4]. Instead of autonomous phonological structures - like models referring to meaningful "tones", and invariant acoustic correlates like in Fujisaki's model, he describes three main mechanisms: "a) articulatory implementation, b) language specific target assignment, and (c) parallel encoding of multiple communicative functions. " (cf. Xu, 2005, p. 222)

The "Parallel Encoding and Target Approximation (PENTA) model" starts with a parallel encoding of communicative functions on the lexical, sentential, focal, topical, grouping, or other level. They activate function specific encoding schemes, which are symbolic or numeric form-function correlations. The encoding schemes specify melodic primitives, including "the local pitch target, pitch range, articulatory strength, and duration." (Xu, 2005, p. 243). Then, the Target Approximation follows, an incremental process that aligns the melodic

[4]He subsumes lexical tone occurring in tone languages and suprasegmental intonation by "speech melody".

primitives to syllables and the previously produced F_0 patterns. This process results in the production of the actual fundamental frequency pattern.

Three points of the PENTA model should be emphasized: First, it explicitly assumes a "parallel transmission of *multiple* communicative functions" (Xu, 2005, p. 246), which all activate correlating encoding schemes. The encoding schemes themselves "are either universal or language specific" [ibid.]. Furthermore, the new parameter of pitch strength determines an actual weight of the single encoding schemes in the context of the other, simultaneously transmitted functions, and defines the likeliness in how far this pattern will come to light in the actual fundamental frequency.

Xu's model is elegant in a way that it can account for both: tones (as a function of targets) and graduality (as a function of target strength). Thus it can handle prototypical or mean encodings of communicative functions in a controlled context, as well as gradual deviations. Unfortunately, Xu does not cover intensity. And also, he does not give predictions on the behavior of second language speakers, his encoding schemes remain vaguely language-specific or universal.

4.5 Summary

I have presented three models of prosody production. Levelt convinces by the insights into the depth of prosodic processing in language production and the careful disentangling of the information needed to produce the actually appropriate prosody. But for the central process, the access and the storage of prosodic patterns, his model is forced to remain very vague due to a lack of empirical data. This problem cannot be solved by the application of his model to L2 speech by de Bot. Fujisaki's model on the other hand, reveals very precise predictions of certain prosodic functions and be very detailed for language specific patterns. But it completely lacks a relation to mental processing and therefore one has to apply the principle of the Contrastive Analysis Hypothesis – that L2 speech is derivable from properties of the native and the target language – to get precise predictions about the second language prosody patterns. Third, Xu's model tries to join the insights of models with abstract, phonological targets, and gradual deviations of these due to cumulative encoding of a number of communicative functions and the "strength" of the target. But also his model does not give hints of probable patterns for L2 speakers: the encoding schemes are either language-specific or universal.

5 English Native Speakers Speaking English

This chapter will present the data of a control group of native speakers of English. The prosodic correlates of sentence mode encoding (statements and questions) and of corrective contrast in three different position will be analyzed. The analysis will include an overview of perceived contours and a detailed assessment of acoustic correlates. The utterances will represent a plausible target for the L2 speech by the learners. As discussed in more detail in section 2.1, one can expect a longer duration in all positions, and a higher fundamental frequency in all but the initial position (cf. Eady & Cooper, 1986) for contrasted constituents. With Kochanski et al. (2005) one can assume a higher intensity for these syllables, too. Statements should be encoded by an overall declination of the fundamental frequency, falling to the bottom of the speakers' range. Questions, instead are supposed to either rise their F0 towards the end (Bartels, 1997; Bolinger, 1989; Eady & Cooper, 1986; Grabe, 1998) or contain a final fall, often associated with an overall higher pitch register (Bolinger, 1987; Grabe, 1998).

5.1 Eliciting Correlates of Sentence Mode and Corrective Contrast

5.1.1 Subjects

Sixteen female native speakers of British English took part in the experiment. Their age ranged from 20 to 29, the mean age was 24.6 years. Eleven of them originated from (Greater) Manchester, one was raised in London, the other four came from different locations in Northern and Central England (Preston, Leeds, Birmingham). The choice of British English, and more specifically the variety spoken in England was motivated by the assumption that most of the German second language learners – who will be presented in the third experiment –

were exchange students in England. So, in case that the acquisition of prosody was successful, they would have acquired the variety they were immersed into. Therefore, observed deviances to the target language by the L2 speakers would be hardly attributed to the learning of a different variety of English. Furthermore, the the constraint in the local origin should increase homogeneity of the control group, as the prosodic encoding, especially of questions, seems to vary from region to region in the British Isles (cf. Grabe, 2004). All subjects were paid for their participation.

5.1.2 Materials and Design

The elicitation of the investigated utterances was induced by short read dialogues (4-5 turns). This is seen as a compromise between explicit instructions to emphasize a certain constituent (e.g. Liu & Xu, 2005), and a corpus of uncontrolled speech. A latter would always need a pragmatic annotation, or a perceptual intonation analysis which at least encodes prominent vs. non-prominent syllables (Fant & Kruckenberg, 1989; Kochanski et al., 2005; Wagner, 2002). But as the goal was to assess the encoding process of specific communicative functions, dialogues that clearly induced the desired sentence mode and contrast position seemed to be more appropriate.

Twenty-four critical sentences were used. They contained only monosyllabic words, mainly to eliminate effects of a different word accent placement between the two control groups (English and German), for example the fact the default main accent on bi-syllabic verbs is on the second syllable in English, while it is on the first syllable in German. So, the sentences were very short SVO sentences, containing three syllables. All nominal constituents (subjects and objects) were proper names and monosyllabic transitive verbs were used in present and simple past.

Corrective contrast was elicited in all three positions: a) in sentence initial position, the subject and first syllable of the sentences; b) in sentence medial position, the verb second syllable of the sentences, and c) in sentence final position, the object and last syllable of the sentences. Thus, differences in prosodic correlates due to the syntactic function and position in the utterance cannot be disentangled. Nor can one give evidence to the acoustic qualities of non-content-words or unaccented syllables in polysyllabic words, as there were none. These restrictions were accepted in favor of a maximum of comparability between the target language and the native language of the second language learners.

In addition to the three critical contrast conditions one filler condition was inserted. It contained the same sentences, but had maximal focus projection, and all new information.

In assertative mode it was the answer to some question like “Do you have news from home?” or “What has happend in the mountains? ” and a following echo-question, mostly connected by “Pardon” or similar, as a key of asking for a repetition or confirmation.

Sentence mode, characterizing a statement or assertative sentence and a question or interrogative sentence, was elicited by context. The assertative sentences were clear statements, ensuring the truth of their content as opposed to the previously established context. The questions were echo-questions in the strict sense: they entirely repeated the previous statement without adding lexical or syntactic question markers. Their position in the dialogues (see the scheme below) and the preceding “What ?” clearly indicated that they had to be understood as echo-questions. Furthermore, there were indications of the sentence mode by punctuation: The corrective statements were accompanied by exclamation marks, the echo-questions were presented with a final question mark.

So, the twenty-four sentences were distributed over 96 dialogues, either forcing most unambiguously a corrective contrast (CF) on one of the constituents (S, V, or O) or introducing a neutral focus (filler condition), both spoken as statements by one speaker, and as questions by the other speaker of the dialogue. This leads to a 3 (critical focus positions) x 2 (sentence modes) design.

All dialogues were checked by a native speaker, a linguist who was familiar with information structure research for a most unambiguous elicitation of corrective contrast in the intended position.

For the recordings, the dialogues were divided into four lists following Latin Square procedure, so that each speaker spoke every sentence in one of the eight conditions, one dialogue containing the same critical sentence in assertative (for one speaker) and question mode (for his dialogue partner). All in all, each sentence in its specific sentence mode and focus condition was recorded twice, and the same sentence mode and focus conditions occurred three times within the utterances of one speaker.

A schematic formulation of the principles of the dialogues and an example are given in Table 5.1 and Table 5.2.

.

Table 5.1: Semantic scheme of the dialogues

Speaker	Proposition
A	utters the proposition P_i. P_i contains an agent S_i, an action V_i and a patient O_i. So $P_i(V_i, S_i, O_i)$ is true for speaker A, and B is overtly instructed about A's assumptions of a certain content. But A is not sure about one of the three parts of P_i, so he puts into question either S_i or V_i, or O_i.
B	negates the truth of the constituent put into question by A. Then B utters the first target sentence, in which he introduces a (contextually new) replacement for the part of P_i put into question. Thus, there is a new proposition P_k, which corrects exactly the part of Pi that has been put into question, while the other constituents remain valid (given). Formally, if the corrected part is the subject, P_k is $P_k(V_i, S_k, O_i)$, if the corrected part is the verb, P_k is $P_k(V_k, S_i, O_i)$, and if the corrected part is the object, P_k is $P_k(V_i, S_i, O_k)$.
A	Then Speaker A takes the turn again. First, there is an expression of surprise, like "What ?" or "Really ?". Then A utters the second target sentence, the echo-question Q_k which is identical in lexicon, word order and information structure to P_k. So, Q_k is $Q_k(V_i, S_k, O_i)$, $Q_k(V_k, S_i, O_i)$, and Q_k is $P_k(V_i, S_i, O_k)$, for echo-questioned subject, verb, and object, respectively.
B	Some tail sentence(s) follow which continue to motivate the intended contrast.
(A)	Sometimes, another reply (tail) was used to finish the story

5.1.3 Procedure:

The lists of dialogues were printed on paper and subjects were invited to read them in pairs. The recordings took place in a sound-proof booth at the department of linguistics at the University of Manchester. The distance between the two speakers and the microphone was more or less equivalent.

Every dialogue partner could read the whole dialogue, also including their partner's parts. They were instructed to fully inspect the dialogues first and then read their parts as freely as possible, to imagining to be "an actor in a radio play". They were encouraged to repeat their sentences if they had the impression that the reading was not appropriate. The experimenter was outside the booth listening. He could give signs to repeat the dialogues if they were not spoken correctly, especially if there were (lexical) reading errors in the critical sentences. The speech was recorded with an AKG C414 high quality microphone via an external sound-card and amplifier (Mobile USB-Pro) and recorded directly to a PC with 44100 Hz sampling rate and 16 bit encoding.

Table 5.2: Examples of the contexts to elicit corrective contrastive focus in English for the target sentence *Fred kissed Sue*. Corrected constituents in the table in capitals

neutral focus: filler condition A: Tell me, what happend later that night? B: Well... Fred kissed Sue. A: Really ? Fred kissed Sue ? B: Oh yes, and they're really in love now.	*Subject corrected* : CF S A: Who was the guy who kissed Sue last night ? Wasn't that Harry? B: Oh no, that wasn't Harry! FRED kissed Sue ! A: What ? FRED kissed Sue ? I thought he didn't like her at all. B: Well, apparently he changed his mind about her.
Verb corrected : CF V A: I heard that Fred met Sue in the disco last night. Did you see them talking to each other ? B: Yes, but they were not only talking. Fred KISSED Sue ! A: What ? Fred KISSED Sue ? I thought that he was Mary's boyfriend... B: That's what I thought too. But they kissed each other right on the dance floor...	*Object corrected* : CF O A: Who was the girl that Fred kissed last night ? Was that Marion? B: Oh no, That was not Marion. Fred kissed SUE ! A: What ? Fred kissed SUE ? He never liked her. B: Well, I don't know. I think he was drunk.

Preprocessing

The critical sentences were extracted from the dialogues and word boundaries were tagged. As the sentences and words were very short and easy, no perceivable pauses occurred [1].

Then a three-layer TextGrid was created for every sentence using the PRAAT software package (Boersma, 2001; Boersma & Weenink, 2007). The first layer contained information about the contextually induced corrective contrast. On the second layer, the perceived sentence accent was marked (with the letter “A”). On the third layer, a basic description of the perceived tones was given for every word.

This system is a simplified version of the ToBI annotation system, originally established for (American) English intonation (. Pierrehumbert, 1980; K. Silverman et al., 1992), but also adapted to German (GToBI: (Grice & Benzmüller, 1995; Reyelt, Grice, Benzmüller, Mayer, & Batliner, 1996; Grice, Baumann, & Benzmüller, 2005). My annotation scheme contains only two tone levels, low and high, and combinations thereof. There are two practical arguments against a full application of the ToBI annotation: first, an encoding of phrase boundary tones is not necessary due to the simple structure of the sentences: There is only one relevant boundary, which is the end of the sentences. Second, due to the monosyllabic constituents it would be difficult do distinguish boundary and accent tones, and they would overlap. The monosyllabic constituents would also make it difficult to distinguish finer grained differences for instance between an early accented rise L*H and a late accented rise LH*. The rather rough method applied in this study will be compensated by the analysis of the acoustic correlates later.

In short, the inventory to a sufficient description of the perceived contours contained only six entities: a rise (LH), a fall (HL), a high (H) tone or a low (L) tone without perceivable pitch movements, and a rise-fall (LHL) and a fall-rise (HLH), but no finer distinction (e.g. strong or weak fall, mid-level tones) or functional assignment (boundary tones etc.).

Figure 5.1 shows an example; for a better understanding, the first layer contains a transcription of the sentence instead of the intended contrast marker (K).

All acoustic analysis was done with the PRAAT software package, too (Boersma, 2001; Boersma & Weenink, 2007). As PRAAT is very sensitive to both micro-prosody and creaky voice, the assigned fundamental frequencies had to be controlled for plausibility. Occurring octave jumps (mostly in the neighborhood of obstruents) were corrected. Random pitch

[1] In the very rare case of some milliseconds of silence between two words, the beginning of the second word was marked as the border between them.

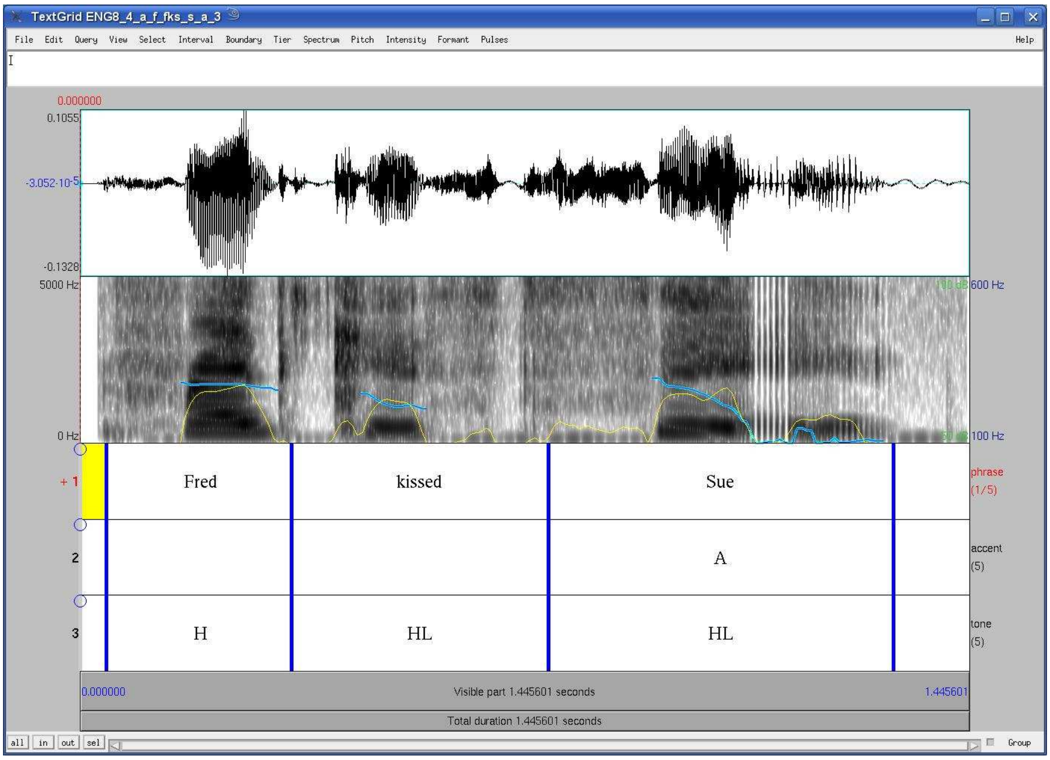


Figure 5.1: The target sentence "Fred kissed Sue" spoken by speaker 8a in assertative mode with corrective contrast on the object

assignment due to creaky voice (mostly at the end of the sentences) was deleted. Then the perceptual and acoustic data were collected using automated scripts.

5.2 Perceptive Analysis

All perceptual analysis was undertaken by the author, a native speaker of German. Although language-specific deviances induced by the different native language of the annotator may not be ruled out completely, the overall correctness of the analysis can be assumed for several reasons: First, the detection of the main sentence accent is one of the issues of highest inter-rater reliability of the perceptual prosody research (Gut & Beyerl, 2004). Second, the simple sentence structure and the low number of three syllables per sentence exclude a large number of possibilities of the AM-approach. Third, the descriptive inventory is simplified: Thus, there is no distinction between nearby categories that are highly vulnerable to rater-specific preferences, (e.g. L*H with the accent on the lower part of the rise, and LH* with the – later – accent towards the end of the rise, or distinguishing low vs. mid fall). And finally, the relative large number of sentences (48 per condition) should reveal valid tendencies, subsuming cases that are more ambiguous than others[2].

5.2.1 Perceived Accent

In a first step the positions of the contextual contrast and the perceived main sentence accent are compared. In Table 5.3 the percentage of "correct" stress assignments per condition is given. "Correct" in this case is to be understood as the percentage of target sentences in which the perceived main phrase accent corresponds to the intended corrective contrast position.

Table 5.3: Percentage of sentences with the perceived main accent in the position of the intended contrastive focus by sentence mode and focus position (CF = contrastive focus; on subjects (CF S), verbs (CF V), or objects (CF O)

	CF S	CF V	CF O	mean
Statement	89.6	77.1	93.8	86.8
Question	68.8	50.0	97.9	72.2
mean	79.2	63.6	95.9	79.5

As shown in Table 5.3, overall four out of five critical utterances received the main accent on the contrastive focus position. The number of correct accent placements differs by sentence mode and by contrast position. Contextual contrast reaches about 80 % of perceived main accent on the subject. Seventy-seven % of the declarative but only 50 % of the interrogative sentences received the main accent in the verb position, if it was contrasted. Most successfully, the object received the main phrase accent if it was contrasted. Almost all of the intended contrasts were perceived in that position in both sentence modes.

Statements attained a higher rate of correctly placed accents than questions, which have a strong preference for stressing the final syllable/constituent.

Discussion The fact that only 80% of the utterances contained the main sentence accent on the contextually focussed constituent is surprising at first sight. The more, as the corrective contrasts are said to induce the strongest perceivalbe prominence marking available (Molnár, 2006). On the other hand, also the simple one-sentence contexts of Cooper et al. (1985) and

[2]One can speculate what would have happened if there were more than one perceptual annotation, possibly with annotators of different native languages. Remembering the difficulties with some of the sentences and results of an investigation of inter-rater variance (e.g. Gut & Beyerl, 2004), I would predict a certain amount of variance. However, considering the facts that the syntactic structure as well as the information structure properties are very restricted, the tone inventory simplified, and the comparably large amount of data, I think it is reasonable to assume that the overall results would be very similar.

Eady and Cooper (1986) result in five to ten percent inconsistent accent placements. An unknown proportion of "unusable" accent placements – for word accents – is also reported in Beckman (1986, p. 148). So, the fact THAT there are missplaced accents is already reported (although only scarce).

There are three reasons to explain the comparably high percentage of failures: First, in this experiment corrective contrast was induced by a comparably complex context. The speakers had to understand the introducing sentences, and then oppose the critical sentence against the previous context spoken by their dialogue partners. This is more complex than reacting on one focus-inducing context question as in Cooper et al., 1985. The second reason may have been the instruction to "act" the dialogues: It led to a certain involvement of the speakers who were naive actors. The (desired) expressiveness may have overrun a straightforward assignment of properties of the information structure. One indication for that involvement is the distribution of accents in the neutral focus position. Here, the main phrasal stress is always expected on the last constituent / syllable. But the object of the sentences received the main stress only in 75 % of all utterances, but also in 24 % in the subject position, and even 1 % on the verbs. The third reason is that one possibly has to assume that – at least in speech in context – focus and even not corrective contrast is not expressed by a perceivable accent. This point will be discussed in more detail below.

If we consider the percentage of correct stress assignment for each position separately, we find the following:

In the subject position, the number of "correct" accent placements is average overall. But while the contrast in statements is perceived in about 90 % of the cases, it is found only in 69 % of the questions. The less constant accentuation of subjects in question may be explained by a lower overall rate of prosodic correlates of contrast in questions: this issue will be dealt with in extension in the discussion of the acoustic correlates in section 2.1.

Contrastive accent on verbs seems to be dispreferred. Although intuitively possible, no study has systematically explored the prosodic effects of a corrective contrast on verbs of an utterance. Indeed, already constructing the contexts that elicit corrective contrast unambiguously on the verbs was rather difficult. Furthermore, sentences "out of the blue" with a strong accent on the finite verb are assumed to induce the so-called "verum focus" (Höhle, 1992), which would contrast not the verb as a single constituent, but the whole proposition. On the other hand, error rates of 25 % in statements and 50 % in questions show that a contrastive focus on verbs indeed HAS effects on the perceived main phrasal stress. It is far from a global stress assignment on the last constituent (which would imply no perceived main accent on the verbs) and still different from chance (which implies about one third of all phrasal accents on the verb).

The high number of correctly assigned main phrasal stress on objects is not surprising as the object would also receive the 'default accent' (Chomsky & Halle, 1968) in the case that no context is given. This accounts for both, statements and questions.

All in all, three points can be made. First, the contextually induced contrast caused strong effects on the perceived main phrase accent. Contrastive focus overrides in most cases the tendency of the main sentence accent on the last content word. Second, there is no one-to-one connection between contextually induced corrective contrastive focus and perceived main accent. Although they are highly correlated, and corrective contrast has effects on the position of the main sentence accent in all positions[3], corrective contrast induced by a context does not necessarily result in a perceivable strong accent. Therefore, one has to distinguish accent-focus from information-structure focus. And finally, and this point is crucial for all of the analyses following: The functional, contrast-inducing load was comparable for all sentences. So, one should to assume that speakers intended to encode focus in the desired positions. And, at least from a perspective speech PRODUCTION, there is no way to say why some contrasted syllables were perceptively accented while others were not. So one has to make a choice: One possibility is to restrict the analyses of correlates of corrective contrast to sentences that bear the sentence accent in the position of the focus, thus treat the sentences in which focus and accent position do not co-occur as errors. This choice is made by most studies in the literature – their analysis of prosodic correlates of focus is therefore merely an analysis of correlates of accented parts of utterances (under specific focus conditions). I think that, when investigating the prosody production processes, it is more appropriate to the goal to INCLUDE the data in which the position of the intended contrast and the perceived accent differ. It means to accept that focus OFTEN determines the position of the perceived sentence accent in (almost) naturally contexts, but NOT ALWAYS. Thus, the analyses reported below will treat the prosodic correlates of contextual corrective contrast as it is, and not correlates of accentedness.

However, parts of the analyses have been repeated for focus-accent matching utterances only for reasons of comparison. They will be mentioned time given. Generally, restricting the data to these sentences partly strengthens the issues found in the total set of utterances but it does not change the tendencies, and it does not increase the number of significant

[3] As for the phrase final object position, one can argue that the percentage of perceived phrase accents for object contrasted sentences is higher than in sentences with wide focus, and thus a focus projection on the object of the sentences.

differences between contrasted and non-contrasted constituents in the analyses of acoustical data reported later.

5.2.2 Distribution of Tones

In this section, the perceived tonal contours are analyzed. The distribution of the basic tones of low (L), high (H), rising (LH), falling (HL), and risefall (LHL) and fallrise (HLH) was investigated for each constituent, respectively syllable, in the utterances. The assignment of these tonal contours was partially difficult, as the units of annotation were unisyllabic and, and sometimes only had little – and sometimes even no[4] – sonorant parts.

As the theoretical and methodological approaches are very different, it is difficult to establish concrete expectations for the unisyllabic constituents in the correctively contrasted condition and the same constituents in sentences in which the focus was elicited in other positions. Jackendoff (1972) assumed a 'B-accent', a fall rise contour on contrastive topics, which would correspond to the subjects in the critical sentences. Following . Pierrehumbert (1980) one would predict a complex tune of either a fall-rise (H*LH%) or a rise-fall with a high boundary (L+H*LH%). Later, a rise (L+H*) was predicted for contrast and especially corrective contrast (J. Pierrehumbert & Hirschberg, 1990, p. 296f.) in earlier positions, while phrase final corrective contrast would still receive the LH% boundary tone. Steedman (2000) also prefers a correlation of a more simple rise (L+H*) for contrast encoding. Bolinger (1961, p. 87) instead, along with Pike (1945), argues that "nothing [in the sense of a pitch accent] is uniquely contrastive."

Perceived Tones in Sentence Initial Position

We start with an overview of the perceived contours in the sentence initial position, the subjects of the critical utterances. The results are presented in Table 5.4. Remember, that if the subject is contrasted (CF S - condition) it is a new proper noun correcting a previously established agens. In the other two conditions, the subject-noun is already explicitly established in the context, and thus is "given", in terms of information structure, in the utterances.

Five different perceptual contours have been observed. Independent of sentence mode and contrast condition, the most frequently perceived contour is a rise (LH) which accounts for 48 (statem. CF V) to 69 (statem. CF S) percent of all sentence initial syllables. Further

[4] e.g. in case of a reduced /ksd/ for 'kissed'

Table 5.4: Distribution of perceived tones on subjects for English by native speakers in percentage per condition

		H	HL	L	LH	LHL
Statem	CF S	25.0	4.2		68.8	2.1
	CF V	27.1	18.8	6.3	47.9	
	CF O	31.3	2.1		62.5	4.2
mean (Statem)		27.8	8.3	2.1	59.7	2.1
Quest	CF S	35.4	8.3		52.1	4.2
	CF V	39.6	2.1	2.1	56.3	
	CF O	35.4	4.2		60.4	
mean (Quest)		36.8	4.9	0.7	56.3	1.4

25 (Statem CF S) - 40 % (Quest CF V) of the perceived contours on the subject of the target phrases are high (level) tones (H) without perceivable pitch movement. They are more frequent for questions (36.8 %) than for statements (27.8 %). A low level tone (L) occurs only in case of contrat on verbs, which also increases the percentage of falling tunes (HL) in statements. A complex rise-fall (LHL) is rare, and appears only in questions only in case the subjects is contrasted. As for the different contrast conditions, there seems to be nothing special with the contrasted (CF S) conditions. They do not behave differently from the other sentences. The only conspicuity is the rather high number of falling (HL) or low (tones), and a lower number of rising contours on subjects of statements with an upcoming contrasted verb (Statem CF V). This pattern can be explained by a preparation effect for the dominant pattern of a contrasted verb, discussed below.

Sentence mode may already be encoded by a preferred usage of a high level tone (H) instead of a rising tone (LH) for questions.

Perceived Tones in Sentence Medial Position

The perceived tonal patterns in the verb position are presented in Table 5.5.

On the verbs in the target sentences we find clear differences between the conditions. While declarative sentences only have 6 % of high tones, 64 % of the verbs in questions are uttered with a high level tone (H). Also contrast on verbs has a clear effect: for statements we find the highest percentage of all high tones (H) and the overall highest percentage of rising (LH) contours in the CF V condition (contrast on the verbs), but the least falling (LH) and low (L) tunes. In questions, the correlation of contrast and questions is less strong: only

Table 5.5: Distribution of perceived tones on verbs for English by native speakers in percentage per condition

		H	HL	L	LH	LHL
Statem	CF S	6.3	50.0	41.7	2.1	
	CF V	10.4	27.1	6.3	47.9	8.3
	CF O	2.1	62.5	35.4		
mean (Statem)		6.3	46.5	27.8	16.7	2.8
Quest	CF S	47.9	33.3	16.7	2.1	
	CF V	64.6	16.7	2.1	16.7	
	CF O	79.2	14.6	2.1	4.2	
mean (Quest)		63.9	21.5	6.9	7.6	

the high number of rising contours (17 %) for contrasted verbs compared to non-contrasted verbs (2 % and 4 %) indicates contrast specific tendencies in that position. Otherwise, one discovers a falling or low tone dominance if the (preceding) subjects are contrasted and the almost unique usage of a high level tone if the upcoming object is contrasted. This leads back to the higher percentage of low or falling contours on the previous subjects if the verb is contrasted.

Also, the effects of contrast have to be seen in connection with the comparatively poor performance in 'correct' accent placement, as discussed in section 5.2.1. If one takes into account the contrasted verbs that actually bear the main sentence accent only, the effects for the perceived tones increase considerably: 62 % of the accented verbs in statements, and 29 % of accented verbs in questions are perceived as rise (LH).

To sum up, in sentence medial position, on the verbs of the utterances, we find clear tonal correlates for contrast and sentence mode. Contrast induces a higher proportion of rising instead of falling contours. In this position, accent and tone are strongly correlated. As for sentence mode, the most common contour for questions is a high level tone, but a fall for statements. Furthermore, preparation effects of an upcoming contrast in the object position can be stated.

Perceived Tones Sentence Final Position

The tones on the objects of the critical sentences are presentend in Table 5.6. The object position in the target sentences is especially interesting as it is also the place of the default accent. If the main sentence accent is induced in other positions than the object, one may predict tonal effects of deaccentuation (Cooper et al., 1985; Eady & Cooper, 1986). Furthermore, it is likely that it is the place of a (rising) boundary tone that indicates question mode (. Pierrehumbert, 1980; J. Pierrehumbert & Hirschberg, 1990).

Table 5.6: Distribution of perceived tones on objects for English by native speakers in percentages per condition

		H	HL	HLH	L	LH	LHL
Statem	CF S	4.2	22.9		47.9	22.9	2.1
	CF V	2.1	35.4		45.8	12.5	4.2
	CF O	8.3	37.5		12.5	18.8	22.9
mean (Statem)		4.9	31.9		35.4	18.1	9.7
Quest	CF S		45.8	6.3	12.5	22.9	12.5
	CF V		62.5	6.3	6.3	16.7	8.3
	CF O		66.7	16.7		12.5	4.2
mean (Quest)			58.3	9.7	6.3	17.4	8.3
mean (Overall)		2.4	45.1	4.9	20.8	17.7	9.0

Most striking is the comparably high percentage of the complex LHL-tone in statements with contrasted objects. More than 20 % of the sentences with a contextually contrasted object had a rise fall (often correlated with long duration) on the last syllable. Also, the percentage of a final high (H) level tone (8.3 %) is twice and four times higher if the object is contrasted than in the other two critical conditions in declaratives (8.3 % for CF O compared to 4.2 % for CF S and 2.1 % for CF V).

On the other hand the percentage of simple low syllables (L) is less than one third for the CF O condition compared to the other two conditions in statements (12.5 % for CF O compared to 47.9 % for CF S and 45.8 % for CF V). In questions, a low level tone (L) does not occur at all if the object is contrasted, but shows up if the subject (12.5 %) or the verb (6.3 %) is contrasted. On the other hand, the complex HLH-contour is almost three times more frequent for contrasted final syllables in the target questions than for interrogatives that have other positions of contrast (16.3 % for CF O, compared to 6.3 % for CF S and CF V).

The question-statement encoding should also have its main locus at the end of the sentences. However, that distinction is not clear at first sight. While statements clearly prefer a low or falling tone, questions do not seem to opt preferably for rising or high tones. If one sums up the percentages of high ending and low ending contours (H, LH, HLH as high, and L, HL, LHL as low), one gets the data summarized in Table 5.7.

Table 5.7: Low/Falling vs. High/Rising tones on OBJECTS by Sentence Mode (summed percentages

	Assert. S.	Questions
low/falling	77	72.9
high/rising	23	27.7

As Table 5.7 shows, 77 % of the declaratives and 73 % of the questions end low or falling.

Looking at the unique tunes instead, the differences between questions and statements are clear: There is no HLH contour on objects in statements, while there is no single H tone in questions. The percentage of low level (L) tones is more than one third for the statements (35.4 %), but only 6.3 % of the questions. But the falling end tone is the most frequent for questions (58.3 %), while it occours only in 31.9 % of the the statements.

Alltogether, for the perceived tones in the object position, we find evidence for a different distribution of the tones for contrasted vs. non-contrasted objects, like the higher proportion of complex LHL tones. But we cannot differentiate statements from questions by a simplified boundary tone assumption of a low or falling tone for the first and a high or rising tone for the latter. However, there are differences in the distribution of the unique contours on the final syllable.

5.2.3 Summary and Discussion of the Perceived Contours

To conclude the analysis of the perceived tones as correlates of sentence mode and contrast, I want to emphasize tree points.

First, in each of the three positions more than one contour is perceivable. So intonational grammars that assume more or less obligatory tunes for a certain communicative function – whether specified for syntactic position or not – are too simplistic. With one exception[5] less than 70 % of the syllables bear the same tone. That fact cannot be ignored in models of prosody production (see chapter refspp).

Second, correlates of focus are only partly perceivable as distinct contours on the contrasted syllables. There is no particular preference for certain tones for contrasted vs. non-contrasted subjects; and in the object position, we see that about one quarter (22 %) of the sentence

[5]Verbs in questions preceding a contrasted object (CF O - condition) yield a high level tone in 79 % of the cases.

final syllables in statements carry the complex rise-fall (LHL) as contrast indicator. Only in the sentence medial position, on the verbs of the examined sentences, one can assess a preference of a rise instead of a fall if the verb is contrasted, and that a perceived main phrase accent is strongly correlated with that change of the slope of the pitch movement.

And third, the high rate of falling contours for questions must be discussed. In the experiment, the proportion of falling and rising contours at the end of the utterances does not differ between statements and questions. About three quarters of the questions and end in a falling or low tone. This is surprising at first sight, but yet plausible: On the one hand, one cannot completely exclude that a slightly present rising boundary tone for questions is just dominated or overwritten by the fall: the last constituent is monosyllabic, and thus, there is only one possible place for the accent and the boundary tone. But, on the other hand, the origin of most of the speakers (Manchester) is also the region in which one of the highest rates of falling question intonation is found: In the nearest region in Grabe's regional variety investigation (Bradford), 83 % of the wh-questions, 17 % of yes-no questions, and 22 % of the declarative questions ended in a low boundary tone[6] (cf. Grabe, 2004). Fries (1964) already found that there is no difference between statements and questions in the final contours in an investigation of a corpus of (American) English. The dialogues are by far more natural ways of inducing sentence mode than e.g. the instruction of "Ask ..." (Eady & Cooper, 1986), or any introspection on how to produce a question.

From her introspective account, Bartels (1997) would threat the echo-questions ending in a fall as echo-exclamations, and not as questions. As she argues that the state of the echo-utterance whether it is a question or an exclamation entirely depends on the final tune, it is impossible to be falsified. However, the questions indeed DO ask for confirmation, more precisely for a confirmation of the focus-metarepresentation (Iwata, 2003), which contains the constituent under corrective contrast in the preceding statements. The follow-up answers in the dialogue explicitly confirm the entity put into question. So, the echo-questions are ask for something and elicit an answer. However, one must take into consideration that the echo-questions are elicited in a context that induces surprise: All components of the question have already been given in the preceding statement which is repeated identically, so they do not ask for new information. Furthermore, the questions are introduced with a preceding "What?", which also indicates surprise.

This leads to the argumentation in Bolinger (1987, p. 266) who considers falling contours in morphosyntactically unmarked questions as perfect in that case. Concentrated on the

[6]Grabe (2004) did not consider echo-questions. But the Bradford region shows one of the highest rates of falling contours in questions compared to the others in the study.

aspect of surprise, the phrase-final falls can also be explained by an implementation of the Frequency Code: if falling contours express confidence and authority, the speaker shows his (confident) disbelief to the content he has heard.

However, it would be shortcoming to reduce the echo-questions to statement-like expressions of surprise. The contexts (follow-up answers that confirm the entity in question) and another aspect which did not show up in the analysis of tones so far clearly distinguish the echo-questions from statements: the higher pitch register. This, and other aspects in the acoustic signal are treated in the following sections.

5.3 Acoustic Correlates

In this section, the acoustic correlates of sentence mode and corrective contrast as measured in the acoustic signal[7] are presented. Six dependent variables have been analyzed for each constituent. An overview and a definition is given in Table 5.8.

5.3.1 Overview of the F0 Contours

Before we go into details of the acoustic correlates of sentence mode and corrective contrast for each constituent, we shall get an overview of the fundamental frequency patterns of the whole sentence. Figure 5.2 shows the mean fundamental frequency for every quarter of each constituent. Each constituent was divided into four parts of equal length, and the mean frequency for that period was computed. Thus, the three-constituent sentences were divided into twelve parts. As the whole utterances had a duration of about one second overall,

[7] All acoustic correlates were analyzed with PRAAT software package version 4.6.34.
Pitch settings :
Time step : 0.001 s
Pitch minimum : 100 Hz
Pitch maximum : 600 Hz
Single values : Hertz linear
Minimum and Maximum: Hertz Parabolic
Intensity settings : Default
Pitch minimum: 100 Hz
Time window : 3.2 / 100 = 0.032 s
mean energy subtracted
In preliminary tests these settings had shown the least number of artifacts. The values for the fundamental frequency affected by the pitch minimum and the pitch maximum. The comparably wide range used in this study limited the number of wrong or missed values. It is known that different pitch settings influence the outcome in PRAAT. However, the analyses of the encoding of sentence mode and contrast, as well as the comparison of the different speaker groups presented later, are not affected, because the settings were kept constant.

Table 5.8: Acoustic parameters in the analysis

Duration ratio	$\frac{Duration\ of\ the\ constituent}{Duration\ of\ the\ whole\ sentence}$
Pitch maximum	$Pitch\ maximum\ of\ the\ constituent$
Mean pitch	$Mean\ F0\ value\ of\ the\ constituent$
Pitch range	$Pitch\ max.\ -\ Pitch\ min.\ of\ the\ constituent$
Alignment ratio of the Pitch maximum	$\frac{(Time\ of\ pitch\ max. - Onset\ of\ the\ constituent)}{Length\ of\ the\ constituent}$
Mean intensity ratio	$\frac{Mean\ intensity\ of\ the\ constituent}{Mean\ intensity\ of\ the\ whole\ sentence}$

the the picture represents the mean pitch in time windows of about 80 ms and gives a rather fine grained contour of the utterances per condition, normalized for unequal absolute durations of the different constituents in different conditions. Statements and questions are represented by the black and the grey lines, respectively, and show the mean fundamental frequency of four time windows per constituent (Su1-Su4 = subjects, Ve1-Ve4 = verbs, Oj1-Ob4 = Objects) in three different contrast conditions (CF_S = contrast on subjects, CF_V = contrast on verbs, CF_O = contrast on objects).

First, the question – statement distinction will be considered. Questions clearly use a higher register than statements. There is an overlap only in the third quarter of the objects, and only if the object in statements is under contrastive focus. Otherwise, fundamental frequency is higher from the beginning until the end of the sentences. Also, the declination of the sentences in both sentence modes is visible. Statements start at about 240 Hz, have a maximum of about 260 Hz during the first constituent and end with around 200 Hz (210 Hz in case of contrast). Questions start with about 290 Hz, go up to 320 Hz during the first constituent and fall to an average of about 220 Hz in the end.

Next, the graph already shows some correlates of contrastive focus. In statements, the contours show the highest relative values for the contrasted constituents compared to the same sentences with contrast in a different position. A contrasted subject is about 10 Hz higher during the third and fourth quarter of it's duration if it is contrasted (condition S CF S) than a non-contrasted; a contrasted verb is about 60 Hz higher at is maximum than its non-contrasted counterparts (condition S CF V); in the object position, contrast adds about 40 Hz at the most to the mean fundamental frequency (condition S CF 0).

The picture is different for questions: A clear correlation of the mean fundamental frequency

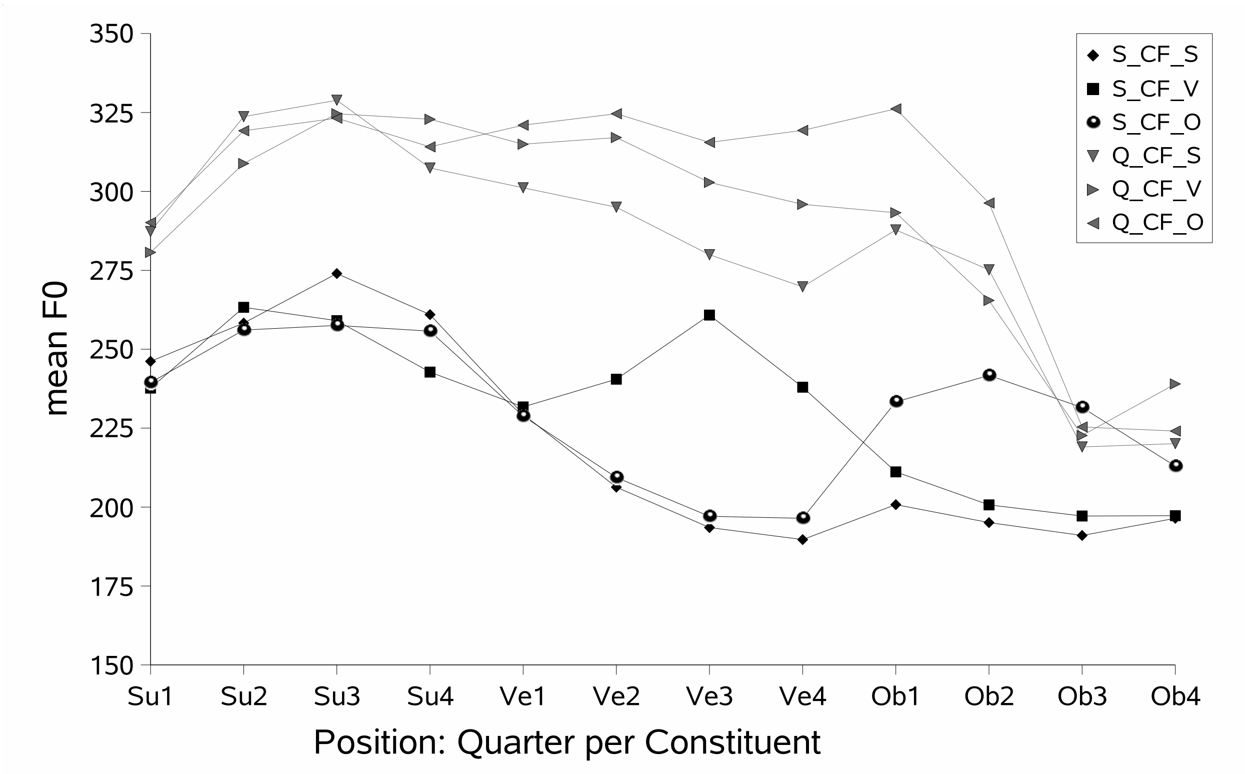


Figure 5.2: Time normalized fundamental frequency contours for English by native speakers

and the position of the contrast cannot be found. One can see a relatively steep declination for the subject–contrasted questions, but no especially high values on the subject itself. Verb–contrasted questions never show specific excursions due to contrast. Only object–contrasted questions retain a higher fundamental frequency until the second half of the objects.

These visual impressions shall now be investigated statistically, including other acoustic correlates like duration, intensity and pitch peak position.

5.3.2 Duration Ratio

Duration ratio, that is the duration of the constituent divided by the duration of the whole sentence, was computed separately for every constituent. Note that if each constituent had the same length, the mean duration ratio would be 33.33 for each. Computing the ratio normalizes the duration for speech rate artifacts. As each of the speakers only uttered one sentence in one condition, and the complete set of all conditions was distributed over six (including the filler condition even over eight) speakers, this normalization is necessary to exclude inter–speaker random effects. The data are summarized in Table 5.9 and graphically presented in Figure 5.3

Table 5.9: mean duration ratio for English by native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}$	Statem	Quest	$\overline{X}$
CF S	31.99	30.61	31.30	29.59	29.66	29.63	38.40	39.73	39.06
CF V	27.56	27.98	27.77	36.20	31.70	33.95	36.25	40.33	38.29
CF O	27.11	28.45	27.78	29.66	28.90	29.28	43.22	42.67	42.94
$\overline{X}_M$	28.89	29.01	28.95	31.82	30.09	30.95	39.29	40.91	40.10

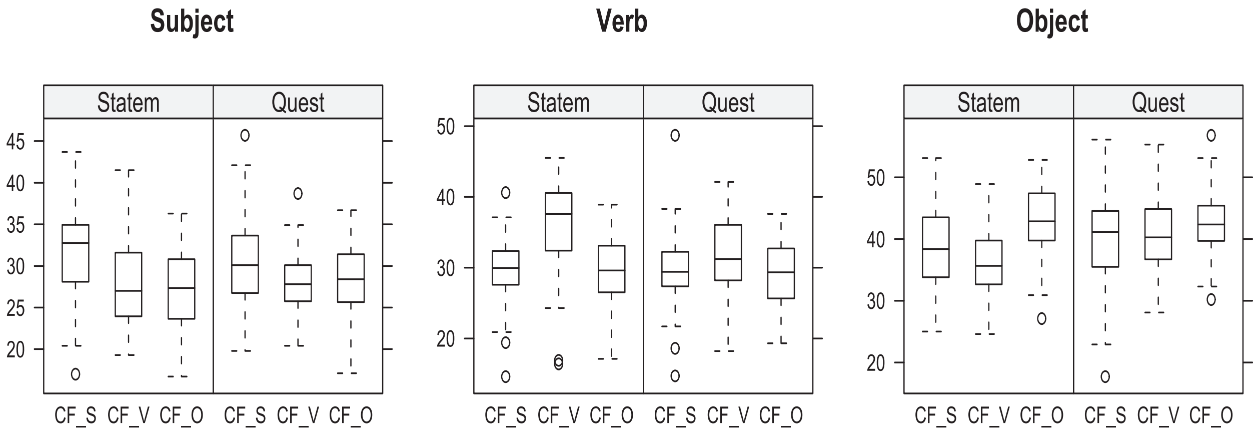


Figure 5.3: Mean duration ratio for English by native speakers for each constituent by sentence mode and contrast position.

Although the subject, verb, and object all consisted of one syllable only, we can clearly see a final lengthening effect. The sentence–final syllables are about 33 % longer than the other two: subjects and verbs occupy about 30 % of the sentence, the objects about 40 %.

As for corrective contrast, we see that the relative duration of the focused constituent is always longer than the same syllable in both other conditions. As duration ratio normalizes to the length of the sentences, differences in the time for questions and statements cannot be assessed as a whole, but only in the relative duration of the constituents in the utterances.

These findings are approved by statistics. A repeated measures ANOVA was run for each constituent in a subject-based and an item-based aggregation. It contained two factors ((Sentence Mode and Contrast) with two (statement vs. question) and three levels (Contrast on subjects, verbs, and objects), respectively. In addition, Bonferroni-corrected pairwise comparisons were run to identify significant differences between the three levels of Contrast for each Sentence Mode separately.

Sentence initially, in the subject–position, there is a significant main effect for Contrast with $F1(2,30) = 10.72$; $p < .001$; $F2(2,46) = 17.80$, $p < .001$. The main effect for Sentence Mode is not significant: $F1(1,15) = 0.03$; p = ns.; $F2(1,23) = 0.065$, p = ns. The interaction between Contrast* Mode only approaches significance: $F1(2,30) = 2.77$, $p = 0.079$; $F2(2,46) = 3.11$, $p = 0.054$. The post-hoc tests confirm that contrasted words are longer than the uncontrasted words in statements as well as in questions.

In sentence medial position, we find a significant main effect for Contrast ($F1(2,30) = 38.32$, $p < .001$; $F2(2,46) = 23.32$, $p < .001$) and an (almost) significant main effect for Sentence Mode ($F1(1,15) = 4.32$, $p = .055$; $F2(1,23) = 10.28$, $p < .01$). The interaction Contrast*Sentence Mode is also (almost) significant ($F1(2,30) = 3.10$, $p = .06$; $F2(2,46) = 7.78$, $p < .001$) . The post-hoc tests confirm that contrasted verbs are significantly longer than the verbs in the non-contrasted statements. This lengthening effect is not approved in questions. Furthermore, the effect for Sentence Mode can be traced back to significantly longer contrasted verbs in statements than in questions, which reflect the focus-effects of contrast in statements only.

In the object position, we find a significant main effect for Contrast($F1(2,30) = 14.55$, $p < .001$; $F2(2,46) = 15.90$, $p < .001$) and (almost) significant for Sentence Mode ($F1(1,15) = 3.81$, $p = 0.07$; $F2(1,23) = 5.65$, $p < .05$). The interaction Contrast*Sentence Mode ($F1(2,30) = 3.24$, $p < .05$; $F2(2,46) = 6.23$, $p < .01$) is also significant. The post-hoc comparisons confirm that contrasted objects in statemens are longer than the non-contrasted objects, but no such effect is consistently found in questions. As for sentence mode, the pairwise comparisons of identical focus position confirm a longer duration for questions than statements in the verb-contrasted condition.

To conclude, it can be confirmed that contrast induces syllable lengthening, which is consistent with the findings in the literature. We also see that this lengthening effect is obligatory in all positions only for statements, but not for questions: in the sentence medial and sentence final position, the significant interaction of the ANOVA, and the pairwise comparisons show that contrasted syllables are not significantly longer than uncontrasted if they occur in questions.

5.3.3 Pitch Maximum

The pitch maximum is the highest excursion of the fundamental frequency during the syllable. Literature predicts a higher pitch maximum for contrasted – and thus accented – syllables in general. The difference between contrasted and uncontrasted syllables may be weaker in the sentence initial position (Cooper et al., 1985). During statements, the maximum of the fundamental frequency should mirror an overall declination of the fundamental frequency, thus should decrease from constituent to constituent. Questions are expected to have a higher fundamental frequency, especially towards the end of the sentence.

The results of an analysis of the pitch maximum for each of the mono-syllabic constituents are presented in Table 5.10 and Figure 5.4

Table 5.10: Pitch maximum for English by native speakers; mean by Sentence mode and Contrast conditions

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	298.0	357.1	327.6	241.5	320.5	281.0	218.1	294.2	307.1
CF V	278.7	348.1	313.4	279.8	353.1	316.4	232.6	381.1	306.8
CF O	273.4	350.2	311.8	238.6	348.1	293.4	282.8	401.7	342.3
$\overline{X}_M$	283.4	351.8	317.6	253.3	340.6	296.9	246.2	392.9	319.7

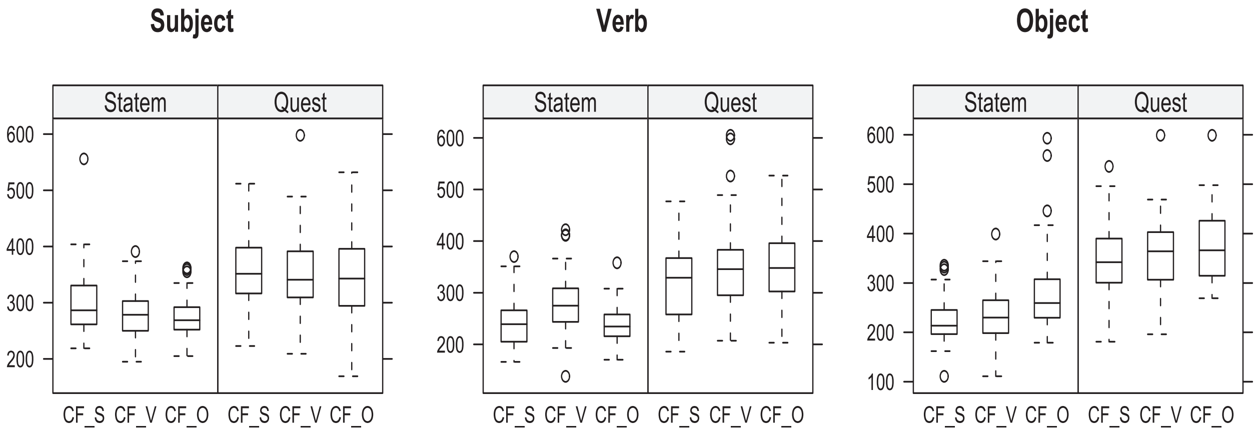


Figure 5.4: Pitch maximum for English by native speakers for each constituent by sentence mode and contrast position

We see that questions have a higher pitch maximum than the statements, irrespective of their position in the utterances. We also see that for verbs and objects in statements, the focused conditions are clearly higher than the two unfocused conditions. This effect is weaker in the subject position, and yet weaker for questions in general.

These points are confirmed in the statistical analyses. In sentence initial position, there is no significant main effect for Contrast(F1(2,30) = 3.163, p = ns; F2(2,46) = 1.775, p = ns.), but significant for Sentence Mode: F1(1,15) = 54.47, p < .001; F2(1,23) = 107.61, p < .001. The interaction Contrast*Sentence Mode is not significant either: F1(2,30)= 1.039, p = ns.; F2(2,46) = 0.59, p = ns. In short, the pitch maximum of the subjects of the sentences statistically differs only in terms of a higher register for questions than for statements, but there are no (local) effects of focus.

In the verbal position, the maximum of the fundamental frequency clearly shows a significant main effect for Contrast: $F1(2,30) = 14.\ 28$, $p < .001$; $F2(2,46) = 6.36$, $p < .01$, and for Mode: $F1(1,15) = 98.99$, $p < .001$; $F2(1,23) = 319.79$, $p < .001$. The interaction Contrast*Mode is also significant: $F1(2,30) = 3.89$, $p < .05$; $F2(2,46) = 5.14$, $p < .05$. The Post-Hoc analyses (Bonferroni corrected paired comparisons) show that contrasted verbs (CF V) have a significantly higher pitch maximum than the two non-contrasted conditions in statements. In the subset of questions, contrasted verbs are not higher than non-contrasted.

In the object position, there is a significant main effect for Contrast($F1(2,30) = 9.14$, $p < .01$; $F2(2,46) = 10.86$, $p < .001$), as well as for Mode ($F1(1,15) = 100.69$, $p < .001$; $F2(1,23) = 268.20$, $p < .001$). The interaction Contrast*Mode is not significant ($F1(2,30) = 1.36$, p = ns.; $F2(2,46) = 1.73$, p = ns). This analysis would suggest that Contrast induces effects of the same size and direction for both sentence modes. However, the Post-Hoc Bonferroni adjusted pairwise comparisons confirm a significantly higher pitch peak for the contrated words compared to the non-contrasted only within statements, but not (or only marginally (see Table 5.15) for questions.

In conclusion, one can state first, that effects of Contrast on the maximum of the fundamental frequency in English do not occur in the sentence initial position, and they are found systematically only for statements in sentence medial and sentence final position. Questions are clearly encoded by a higher pitch compared to statements, independently of the position, which means that the speakers use a higher register not only at the end of the sentences but from the beginning on.

5.3.4 Mean Pitch

The mean pitch per constituent and the pitch maximum presented above are obviously highly correlated. While the latter is more sensitive to steep movements that can appear earlier or later in the syllable, the mean pitch shows the overall height of the voiced parts of the constituent. The data are presented in Table 5.11, and the corresponding boxplot indicating the main characteristics of the distribution is shown in Figure 5.5.

The results are very much comparable to those of the pitch maximum analyzed above. We see a higher pitch register for questions than for statements. Contrastive focus results in a

Table 5.11: Mean pitch for English by native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	259.1	316.0	287.5	205.9	290.1	248.0	181.3	290.0	236.2
CF V	253.1	310.9	282.0	242.1	312.2	277.2	189.8	285.1	237.4
CF O	252.4	315.9	284.1	208.0	320.6	264.3	226.3	275.3	250.8
$\overline{X}_M$	254.8	314.3	284.6	218.7	307.6	263.1	209.9	273.2	241.5

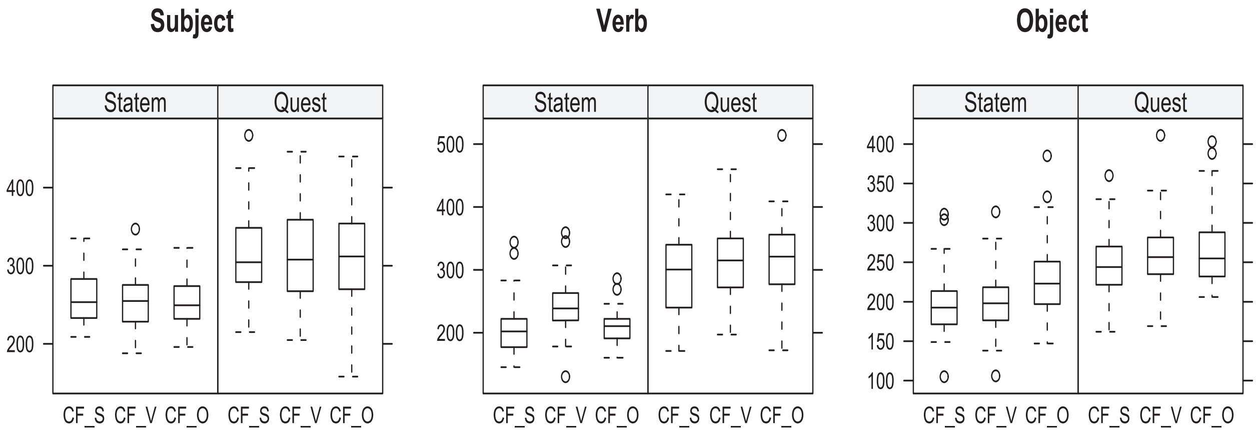


Figure 5.5: Mean pitch for English by native speakers for each constituent by sentence mode and contrast position

clearly higher mean pitch in sentence medial and sentence final position of statements, but not sentence initially and not in questions. This is confirmed by the statistical analyses.

With regard to the subjects of the sentences, the main effect for Contrast is not significant (F1(2,30) = .07, p = ns.; F2(2,46) = 0.30, p = ns.), but questions are significantly higher than statements (F1(1,15) = 63.21, p < .001; F2(1,23) = 143.91, p < .001). The interaction Contrast*Mode is not significant (F1(2,30) = 0.30, p = ns.; F2(2,46) = 0.17, p = ns).

For verbs, we find a significant main effect for Contrast (F1(2,30) = 10.543, p < .001; F2(2,46) = 5.25, p < .01) and for Mode (F1(1,15) = 89.79, p < .001; F2(1,23) = 690.04, p < .001), as well as a significant interaction Contrast*Mode (F1(2,30) = 7.07, p < .01; F2(2,46) = 6.51, p < .01). Indeed, as the Post-Hoc tests show, only contrasted verbs in statements are higher than non-contrasted, but no significant differences are found in questions.

In the sentence final position, there is a significant main effect for Contrast (F1(2,30) = 10.34, p < .001; F2(2,46) = 10.67, p < .001), and for Mode (F1(1,15) = 42.08, p < .001; F2(1,23) = 173.01, p < .001), but no significant interaction Contrast*Mode (F1(2,30) = 2.56, p = ns.; F2(2,46) = 1.45, p = ns). As for the pitch maximum, this would indicate similar effects for Contrat for both sentence mode. However, the Post-Hoc pairwise comparisons again confirm a higher mean fundamental frequency for contrasted objects in statements only.

5.3.5 Pitch Range

The pitch range is the difference between pitch maximun and pitch minimum of the syllable, irrespective of where they occur: Thus, it does not differentiate between a rising and falling range. The literature suggests a higher pitch range associated with an accented syllable. It also indicates a strong movement of the fundamental frequency to encode the interrogative sentencemode, especially at the end of the utterances. Table 5.12 shows the data obtained for each of the three unisyllabic constituents, and their corresponding distributions are presented in Figure 5.6.

Table 5.12: Mean pitch range for English by native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	78.44	92.62	85.53	58.04	61.08	59.56	57.02	176.27	117.27
CF V	60.31	83.33	71.82	77.21	83.88	80.54	74.45	160.25	117.35
CF O	48.71	71.10	59.91	51.38	58.46	54.92	115.54	205.14	160.34
$\overline{X}_M$	62.49	82.35	72.42	62.21	67.81	65.01	82.30	186.36	134.47

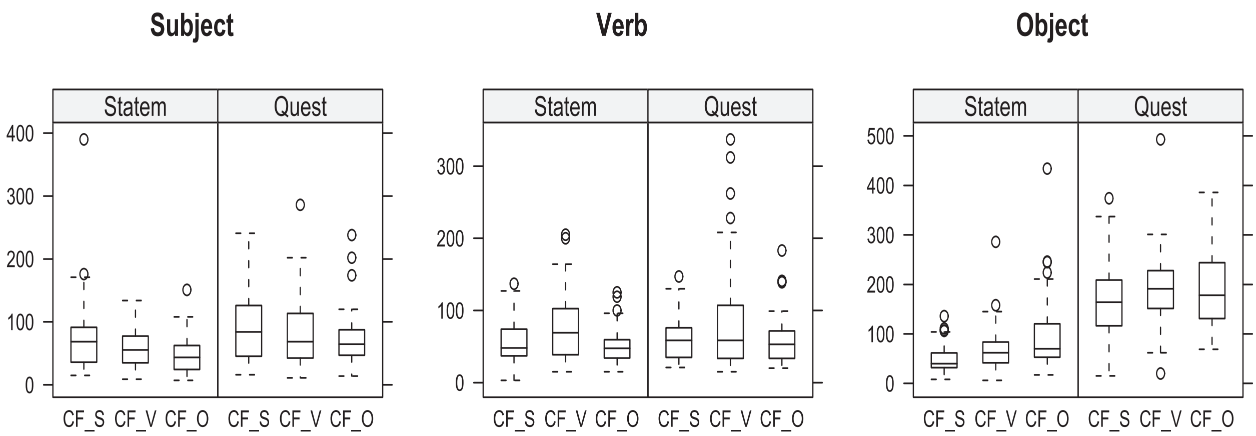


Figure 5.6: Pitch range for English by native speakers for each constituent by sentence mode and contrast position

We see that the pitch range is higher for questions compared to statements irrespectively of the position, but it is almost two times higher in the objects position of questions. So, there is a strong pitch movement at the end of the questions. We also see that a contrasted constituent always has the highest pitch range, irrespectively whether it occurs in statements or in questions. Numerically the contrast–induced differences are a little weaker in questions than in statements, but they occur in both sentence modes.

In the subject position, we find a significant main effect for Contrast ($F1(2,30) = 6.42$, $p < .01$; $F2(2,46) = 6.97$, $p < .01$) and for Sentence Mode ($F1(1,15) = 13.60$, $p < .01$; $F2(1,23) = 9.81$, $p < .01$). The interaction Contrast*Sentence Mode is not significant ($F1(2,30) = 0.32$, p = ns.; $F2(2,46) = 0.37$, p = ns). However, the Post-Hoc analysis with Bonferroni-adjusted pairwise comparisons shows that there are no significant differences in the pitch range of questions, and that the pitch range of the subjects in statements is higher only compared to subjects in utterances with a corrected object, but not to subjects in verb-correcting utterances.

In the verbal position, the pitch range shows a significant main effect only for Contrast ($F1(2,30) = 13.55$, $p < .001$; $F2(2,46) = 12.18$, $p < .001$), but no differences between questions and statements ($F1(1,15) = 1.104$, p = ns.; $F2(1,23) = 0.79$, p = ns).The interaction Contrast*Mode is not significant, either ($F1(2,30) = 0.07$, p = ns.; $F2(2,46) = 0.10$, p = ns). However, the Bonferroni adjusted pairwise comparisons confirm higher pitch range only in statements (against both other conditions in the item-based analysis, and against statements with a contrasted object only in the subject-based analysis), but no significant differences induced by contrastive focus in questions.

On the objects, pitch range shows a significant main effect for Contrast ($F1(2,30) = 7.77$, $p < .01$; $F2(2,46) = 7.54$, $p < .01$) and a strong effect for Mode ($F1(1,15) = 74.86$, $p < .001$; $F2(1,23) = 209.74$, $p < .001$). The interaction Contrast*Mode is not significant ($F1(2,30) = 0.97$, p = ns.; $F2(2,46) = 1.47$, p = ns.). The Post-Hoc analysis confirms a significant difference between the objects in statements with subject- and object-corrections. Subject-based analysis can confirm this effect for questions, too, but in the item-based analysis this effect is only marginally significant ($p = 0.072$).

To conclude, we can see that the pitch range is clearly increased if a syllable is under contrastive focus in statements irrespectively of the position. It is also strongly higher at the beginning and the end of questions: this results of the gaining of the higher register of echo questions on the sentence initial syllable, and the strong fall at the end of the interrogatives. In sentence medial position, there is no difference in pitch range between statements and questions. However, the statistical analysis of the pitch range also shows contradictory results of the ANOVA and the Post-Hoc comparisons primarily run to discriminate which of the three contrast conditions trigger the main effect for Contrast for each Sentence Mode separately. Of course, they would also solve interactions which are not significant for the parameter of pitch range. In sentence initial and sentence final position, the non-significant interaction Contrast*Sentence Mode suggests a similar behavior of the Contrast-induced pitch range in questions and statements (and vice versae), while the Bonferroni-adjusted comparisons are too conservative to assess significant differences not only in statements, but also in questions.

5.3.6 Alignment of the Pitch Maximum

This acoustic correlate designates the relative position of the peak within the syllable. Later peaks are reported to be perceptually similar to higher peaks (cf. Braun, 2004), and the position of the peaks in accented syllables is language specific (e.g. cf. Mennen, 2004) and an important difference between English and German (cf. Grosser, 1997). The following analysis is based on the computation of the time of the Pitch maximum in one (mono-syllabic) constituent divided through the length of the constituent. This computation normalizes for speech rate, and as all sentences were spoken in all conditions, also for segment-induced pitch peak positions.

A higher Alignment Ratio of the Pitch maximum thus denotes a later peak, a lower value indicates earlier peaks. Furthermore, later peaks can indicate a rising tone (from low to These findings are statistically confirmed. high) and earlier peaks a predominant fall (from high to low). The mean of the alignment ratio of the pitch maximum is presented in Table 5.13. Figure 5.7 shows the data distribution in a boxplot.

Table 5.13: Mean alignment ratio of the pitch maximum for English by native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	0.53	0.54	0.54	0.18	0.21	0.20	0.36	0.42	0.39
CF V	0.41	0.58	0.49	0.46	0.35	0.40	0.28	0.39	0.33
CF O	0.51	0.59	0.55	0.22	0.38	0.30	0.43	0.40	0.42
$\overline{X}_M$	0.48	0.57	0.53	0.29	0.31	0.30	0.36	0.41	0.38

The data suggest that the strongest effect on the position of the peak is found for the encoding of contrastive focus on the verbs. As already seen in the perceptive analysis, in this condition the direction of the contour is changed from the (usual) fall to a rise (see Section 5.2.2).

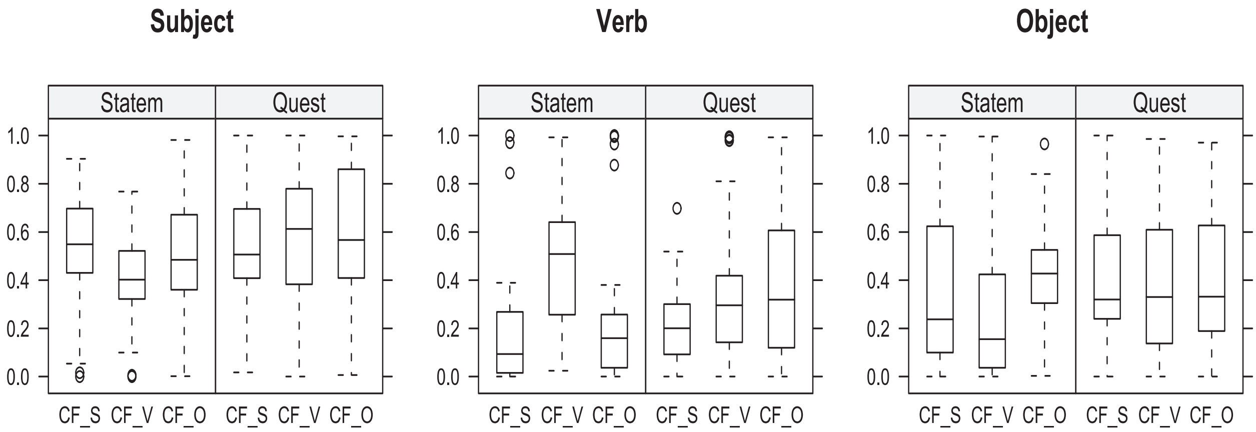


Figure 5.7: Alignment ratio of the pitch maximum for English by native speakers for each constituent by sentence mode and contrast position

The attempt to reach the higher register of the echo-questions does is also mirrored in a later relative peak in the sentence initial position.

The statistical analysis approves this point of view. On the subjects of the sentences, there is no significant main effect for Contrast (F1(2,30) = 1.87, p = ns.; F2(2,46) = 1.68, p = ns.), but for Mode (F1(1,15) = 5.12, p < .05; F2(1,23) = 7.58, p < .05), signalling later peaks for questions. The interaction Contrast*Mode is not significant: F1(2,30) = 2.50, p = ns.); F2(2,46) = 2.53, p = ns.

The observed rise on contrasted verbs is statistically confirmed by a main effect for Contrast (F1(2,30) = 24.03, p < .001; F2(2,46) = 15.94, p < .001). Mode does not induce significant differences (F1(1,15) = 0.92, p = ns.; F2(1,23) = 1.40, p = ns.). The interaction Contrast*Mode is significant (F1(2,30) = 5.97, p < .01; F2(2,46) = 8.19, p < .001). The Bonferroni-adjusted comparisons show that the interaction results from a later peak for contrasted verbs compared to verbs of both other contrast positions in statements, but only later than verbs in the subject-contrast condition for questions[8].

The ANOVA for the alignment ratio of the pitch maximum on objects shows no significant main effect for Contrast (F1(2,30) = 1.88, p = ns.; F2(2,46) = 1.79, p = ns.), nor for Mode: F1(1,15) = 0.84, p = ns.; F2(1,23) = 3.30, p = 0.082, and no interaction Contrast*Mode: F1(2,30) = 2.23, p = ns.; F2(2,46) = 1.49, p = ns.

All in all, the alignment ratio of the pitch maximum seems to be used only marginally to encode sentence mode and corrective contrast in English by native speakers. The later peak

[8]Note that this analysis includes the 50 % of the sentences in which the perceived main accent was not on the verb

for questions in the sentence initial position is expected as the higher register for the echo-questions needs more time to be reached. And very clearly, the skip from a falling to a more frequent rising contour for contrasted verbs is mirrorred in the effects for contrast in that position.

5.3.7 Mean Intensity Ratio

The last acoustical parameter investigated in this study is the mean intensity ratio. Intensity variation is rarely investigated in phonetic and phonological studies on contrast and sentence mode. However, it is presumably highly correlated to perceived prominence in English (cf. Kochanski et al., 2005). Furthermore, intensity is correlated to subglottal air pressure, to which in turn the fundamental frequency is correlated, too. Thus, higher intensity is related to the effort one takes for a syllable - formulated in the Gussenhoven Effort Code.

This study will consider the mean intensity ratio: it is the mean intensity of the constituent divided by the mean intensity of the whole phrase. Thus, the mean intensity ratio normalizes for general differences of the intensity between speakers and as well as the single utterances. Artifacts due to technical reasons, like the distance of the speaker to the microphone, the angle of the speaker to the microphone, and speaker's movements (e.g. head-turning, moving backwards or forwards) are avoided. A mean intensity ratio of 1.0 would mean that the mean intensity of the constituent was exactly the mean of the whole utterance. If it is above 1, the syllable is louder than the average of the whole sentence, if it is below 1, it is quieter than the average of the whole phrase.

As the constituent was a part - and in fact about one third - of the whole target sentence, the ratio differences are very weak, but so is variance. Table 5.14 presents the data and Figure 5.8 shows the distributional properties in a boxplot.

Table 5.14: Mean intensity ratio for English by native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	1.042	1.014	1.028	0.958	0.973	0.966	0.942	0.973	0.958
CF V	1.011	1.013	1.012	1.003	0.990	0.996	0.939	0.975	0.957
CF O	1.021	1.017	1.019	0.950	0.981	0.966	0.993	0.976	0.985
$\overline{X}_M$	1.025	1.015	1.020	0.971	0.981	0.976	0.958	0.975	0.967

Three main points can be made from the data inspection. First, we clearly see effects of the Production Code (Gussenhoven, 2004) or breath-groups (Lieberman, 1967): Intensity

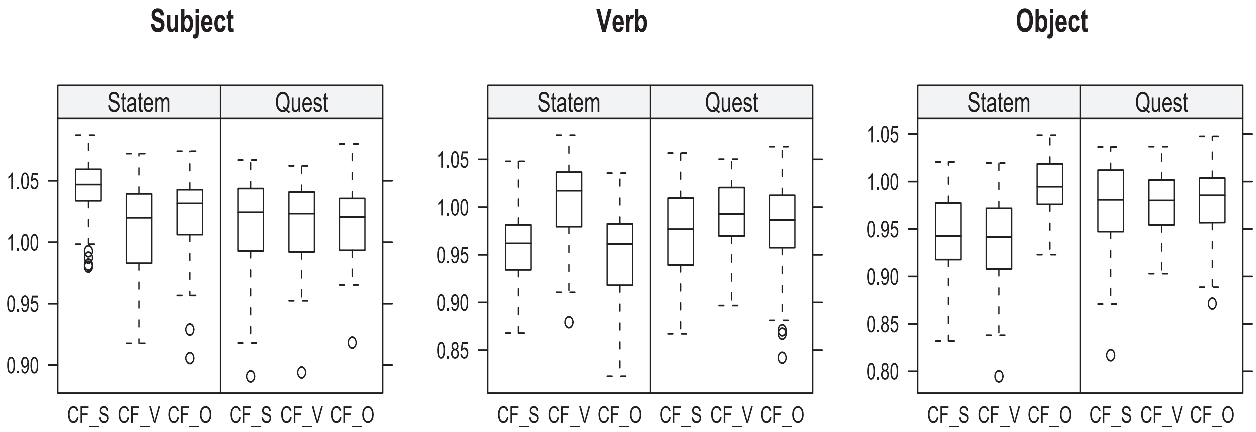


Figure 5.8: Mean intensity ratio for English by native speakers for each constituent by sentence mode and contrast position

decreases over the utterances. This effect is less pronounced in questions than in statements. Statements start with a mean of 1.025, have an intensity ratio of 0.971 sentence-medially and go down to 0.958 in the end. The questions start with a mean intensity ratio of 1.017, decrease to 0.981 in verb position and retain a similar value for the objects: 0.975 over all conditions. So, statements start comparatively louder, approach the question-level in sentence medial position, and they end relatively softer than questions. Focus shows strong local effects, but only for statements: here, the mean intensity ratio of the contrasted constituent is always higher than the same constituent in the non-contrasted conditions. This is not the case for questions.

These points are confirmed by the statistical analysis. In subject position, a significant main effect for Contrast ($F1(2,30) = 3.82$, $p < .05$; $F2(2,46) = 5.64$, $p < .01$) and for Sentence Mode ($F1(1,15) = 5.34$, $p < .05$; $F2(1,23) = 7.93$, $p < .01$) is found. The interaction Contrast*Mode is significant, too: $F1(2,30) = 4.38$, $p < .05$; $F2(2,46) = 7.961$, $p < .001$. The Post-Hoc pairwise comparisons reveal that the contrasted subject in statements is louder than the non-contrasted, but no such effect is found in questions. Differences in Sentence Mode are significant only in subject-contrasted utterances, but not if one compares the intensity ratio of uncontrasted subjects in statements and questions.

In sentence-medial position, we find a significant main effect for Contrast ($F1(2,30) = 11.33$, $p < .001$; $F2(2,46) = 17.82$, $p < .001$). The main effect for Mode is significant only in the item-based analysis: $F1(1,15) = 3.22$, $p = 0.093$; $F2(1,23) = 6.918$, $p < .05$. The interaction Contrast*Mode is also significant: $F1(2,30) = 4.13$, $p < .05$; $F2(2,46) = 9.66$, $p < .001$. The Post-Hoc comparisons reveal that contrasted verbs are louder than non-contrasted ones in statements, but not in questions.

The ANOVA of the Mean Intensity Ratio on the objects of the approves a significant main effect for Contrast ($F1(2,30) = 10.63$, $p < .001$; $F2(2,46) = 13.22$, $p < .001$). Questions generally are louder than statements ($F1(2,30) = 9.49$, $p < .01$; $F2(2,46) = 17.31$, $p < .001$). And the interaction Contrast*Mode is also significant ($F1(2,30) = 8.27$, $p < .001$; $F2(2,46) = 17.64$, $p < .001$). The Bonferroni-corrected comparisons show that the contrasted objects are louder in statements only, but not in questions.

5.3.8 Overview for Contrast

As the large number of ANOVAs[9] and corresponding Post-Hoc comparisons are hard to be reported intelligibly, Table 5.15 shows significant differences of the means of the six acoustic parameters for contrasted vs. the two uncontrasted conditions.

[9] For each experiment, 6 (parameters) x 3 (positions) x 2 (subject-based, item-based aggregations) = 36 within-subject repeated ANOVAs were reported.

Table 5.15: Results of the Bonferroni-adjusted pairwise comparisons between the constituent in contrastive condition and the same constituent in non-contrastive conditions for *English spoken by English native speakers.*

		Subjects		Verbs		Objects	
		F1	F2	F1	F2	F1	F2
Stat.	Dur. Rat.	++	++	++	++	++	++
	Pitch max.	+		++	++	+	++
	mean Pitch			++	++	++	++
	Pitch range	+	+	+	++	+	+
	Peak Align.	+	+	++	++		
	Int. Ratio	++	++	++	++	++	++
Quest.	Dur. Rat.	++	++		+		+
	Pitch max.					+	
	mean Pitch					+	
	Pitch range				+	+	
	Peak Align.			+	+		
	Intensity. Ratio						

Analyses over Subjects (F1) and over Items (F2). "+" denotes significantly different from one other contrast condition, "++" denotes significantly different from both other conditions. All significant differences show a higher value for the dependent variable for the contrasted condition than for the uncontrasted condition(s).

The difference of the prosodic correlates of contrast encoding between statements and questions is striking. Significantly higher values for contrasted elements appear systematically only for statements. The one exception which shows significant higher values due to contrast in questions is the duration ratio in the subject position. Otherwise, the alignment of the pitch maximum is significantly higher for contrasted verbs compared to verbs in the subject-contrast condition, but not compared to verbs in the object-contrast condition. No other parameter shows a significant effect of contrast in both, the subject- and the item-based analysis in questions.

The case is different for statements. The duration ratio is higher for all contrasted elements. The same is true for the mean intensity ratio. So duration and intensity seem to be the most reliable cues for contrast induced effects on syllables, regardless of their position in the (assertative) sentence. The pitch related parameters have to be looked at in a differentiated way:

In the subject position, pitch range is significantly higher if the subject is contrasted compared to the object contrast condition, but not compared to subjects in verb-contrasted statements. The alignment ratio of the pitch maximum is significantly higher only for subjects in the subject-contrast condition compared to the verb-contrasted condition in both, the subject-based and the item-based analysis. So, in the subject position, which always was the very first syllable of the sentences, F0 does not seem to be so much affected by contextual contrast. Pitch related parameters become more reliable in later positions in the sentence.

Most effects of contrastive focus with regard to pitch related parameters are seen in the verb position. Due to the high number of rising contours, the alignment ratio of the pitch maximum is much higher for contrasted verbs than for uncontrasted ones. Also, the mean pitch and the pitch maximum are higher than both other conditions. Pitch range behaves similarly, but here the difference between contrasted verbs and verbs in sentences in which the subject is contrasted does not reach significance ($p = .107$) in the subject-based analysis. In the item analysis, the difference is significant ($p = .047$).

In the object position as the last syllable in the phrases, fundamental frequency seems to be a little less reliably affected by corrective contrastive focus. Only the mean pitch for contrasted objects is significantly higher compared to the objects of both other conditions (contrasted subject and contrasted verb), in both, the subject- and the item-analysis. Pitch range is significantly higher only compared to objects in the subject-contrast condition, but not compared to objects in the verb-contrast condition. Pitch maximum is higher for contrasted objects compared to both non-contrasted objects in the item-based analysis, but the difference between contrasted objects and objects in sentences in which the verbs is contrasted does not reach significance in the subject-based analysis ($p = .131$).

To conclude the findings of the Post-Hoc - tests for Contrast, the three main results are as follows

- Effects of contrast occur almost exclusively[10] in declarative sentences.
- The most reliable parameters seem to be duration and intensity as each analysis confirms significantly higher values for contrasted elements compared to non-contrasted ones, irrespective of their position in the (declarative) sentences and the aggregation of the means.

[10]With the one exception of duration ratio in sentence initial position.

- Pitch related parameters come into play only later during the investigated target statements. The strongest effects are found for the verb. Here, the significant later peak (as shown in the alignment ratio of the pitch maximum) for contrasted verbs compared to uncontrasted ones confirms the difference between falling and rising contours. But also pitch height and range differ significantly between contrasted and uncontrasted versions of this second constituent in the declaratives.

  The effects of contrast on pitch related parameters in the object position are weak: only the mean pitch mean is significantly higher for contrasted objects than both non-contrasted in both aggregations. Pitch maximum on objects fails to reach significance compared to verb-contrasted sentences in the item-based analysis, and the other pitch related parameters, pitch range and alignment ratio of the pitch maximum, do not really convince.

5.3.9 Overview for Sentence Mode

Similar to the effects of contrast, the effects of sentence mode in pairwise comparisons shall be presented in a tabular overview. Table 5.16 shows the results of the pairwise comparisons of statements and questions of the same contrast condition. It is somewhat more complicated, as the difference between the declaratives and interrogatives can be positive (statements show higher values than questions) or negative (statements show lower values than questions) for the different parameters, and also for different contrast conditions.

Table 5.16: Results of the Post-Hoc tests (Bonferroni corrected pairwise comparisons) for Subjects, Verbs, and Objects.

	Subjects		Verbs		Objects	
	F1	F2	F1	F2	F1	F2
Dur. rat.	–	S<Q: O	S>Q: V	S>Q: V	S<Q: V	S<Q: V
Pitch max.	S<Q: S,V,O	S<Q: S,V,O	S<Q: S,V,O	S<Q: S,V,O	S<Q: S,V,O	S<Q: S,V,O
Mean pitch	S<Q: S,V,O	S<Q: S,V,O	S<Q: S,V,O	S<Q: S,V,O	S<Q: S,V,O	S<Q: S,V,O
Pitch range	S<Q: V,O	S<Q: V,O	–	–	S<Q: S,V,O	S<Q: S,V,O
Pitch max. Align.	S<Q:V	S<Q: V	S>Q: V; S<Q: O	S<Q: V	–	–
Intensity	S>Q: S	S>Q: S	S<Q: O	S>Q: V; S<Q: O	S<Q: S,V	S<Q:S,V; S>Q:O

The comparisons are reported for the subject-based and the item-based aggregations as "Subj-A" and "Item-A." respectively. The statement before the colon indicates the direction of the effect: Statements (S) can have higher [>] or lower [<] values than questions (Q). The part after the colon indicates contrast positions [S,V, or O] in which the values significantly comply with the statement before the colon

The seemingly complex pattern of significant differences and higher or lower values for questions can be accounted for in two lines of argument:

The first is the overall higher register for questions which implicates higher rises at the beginning of sentences, and higher pitch up to the end, but with a higher fall on the last syllable. This is shown by the higher values for the parameters related to pitch height: pitch maximum, mean pitch, pitch range.

The second line are the more reliable contrast-induced effects for statements rather than for questions. These result in higher values for statements compared to questions if the investigated constituent is under contrastive focus. This can be studied nicely with for example the mean intensity ratio in the subject and object position, or the alignment ratio of the pitch maximum in the verb position.

5.3.10 The Relative Weight of the Parameters

Now, as the investigation of significant differences in six acoustical parameters is concluded, one of the most interesting questions is the relative importance of the different correlates for contrast. Especially, as English and German presumably differ in the ranking of the parameters, an assessment of the relative weight of the acoustic parameters could be a decisive point for the assessment of possible transfer for second language prosody. Furthermore, the target sentences allow for a comparison of the weight of the parameters in different positions (or grammatical functions): subject, verb, object; or first, second, third syllable ; or sentence initial, medial, final.

From the above analysis of significant differences we know that questions do not show consistent effects of contrast in any position (see Table 5.15). Therefore, an assessment of the weight of the contributions of the parameters is reasonable only for statements.

All six acoustic parameters were included in a generalized linear model as predictors of a binomial function of being contrasted or not. Stepwise backwards, the model was reduced to the parameters contributing significantly to the effect. Table 5.17, 5.18, and 5.19 show the final model for the subject, verb and object position respectively. All presented models show a significant better prediction than an intercept-only model.

Table 5.17: Final model of a stepwise backwards regression in sentence initial, subject position

	Coeff. Est.	Std. Error	z value	Pr($>|z|$)	Sig.
(Intercept)	-39.424931	8.962178	-4.399	1.09e-05	***
Duration ratio	0.164625	0.042117	3.909	9.28e-05	***
Pitch maximum	0.008590	0.004972	1.728	.084056	.
Intensity ratio	30.461809	8.204235	3.713	.000205	***

Residual Deviance Intercept-only model: 183.31; Residual Deviance presented model: 140.79; Deviance: 42.53; P($> \chi$) : .000

In the subject position, as shown in Table 5.17, intensity ratio clearly outweights every other parameter as a predictor of being contrasted or not with a coefficient above 30. The duration ratio has a much lower coefficient, but still has a significant impact. Pitch maximum, as the only fundamental-frequency related parameter surviving the AIC-based reduction of model-components has already a low coefficient, and is only marginally significant (> 0.1)

Table 5.18: final model of a stepwise backwards regression analysis for the verb position

	Coeff. Est.	Std. Error	z value	Pr(> $\|z\|$)	Sig.
(Intercept)	-25.33078	6.03600	-4.197	2.71e-05	***
Duration ratio	0.18394	0.05469	3.364	.000769	***
Pitch maximum	-0.08366	0.03006	-2.783	.005387	**
Mean pitch	0.09252	0.03075	3.009	.002618	**
Pitch range	0.05832	0.01701	3.428	.000607	***
Align. pitch max.	3.65754	1.00599	3.636	.000277	***
Intensity ratio	15.07652	5.89419	2.558	.010532	*

Residual Deviance Intercept-only model: 183.31; Residual Deviance presented model: 90.78; Deviance: 92.53; P($> \chi$): .000

In the sentence medial position, as shown in Table 5.18, all examined parameters are necessary for a good model of the prediction of contrastiveness by prosodic correlates. The highest coefficients are the intensity ratio and the alignment of the pitch maximum. But also all other parameters are significant predictors of a contrasted verb.

Table 5.19: final model of a stepwise backwards regression analysis on statements by English native speakers on the object position of the critical sentences.

	Coeff. Est.	Std. Error	z value	Pr(> $\|z\|$)	Sig.
(Intercept)	-39.80841	7.49727	-5.310	1.10e-07	***
Duration ratio	0.15488	0.04242	3.651	.000261	***
Intensity ratio	33.79549	7.15936	4.720	2.35e-06	***

Residual Deviance Intercept-only model: 183.31; Residual Deviance presented model: 146.54; Deviance: 36.77; P($> \chi$): .000

In the sentence final position, only duration ratio and the intensity ratio remain in the final model and show significant impact on contrast encoding. Again, the highest coefficient is held by the intensity ratio.

To conclude, the two points mentioned above are clearly answered: In subject and object position, only duration and intensity have significant impact on the regression model, which means that they can sufficiently predict (contextual) contrast. In the sentence initial position (the subjects of the critical sentences), duration has the highest z-value (z = 3.9), followed

Table 5.20: Ranking of significant parameters at the last step of a backwise logistic regression model at three essential positions for contextual contrast. English statements spoken by English native speakers

Rank	Subjects	Verbs	Objects
1	Duration	Align. Pitch. max.	Intensity
2	Intensity	Pitch range	Duration
3		Duration	
4		mean Pitch	
5		Pitch max.	
6		Intensity	

by intensity (z = 3.7). In the sentence final position (the objects of the critical sentences), intensity (z = 4.7) outperforms duration (z = 3.7).

The picture is completely different in the verb position (sentence medial): First, all analyzed parameters are needed for a good regression model. Second, highest z-score is obtained by the alignment ratio of the pitch maximum (z = 3.6), which again indicates the higher chance of a rising contour - and thus a late peak - for contrasted verbs. Furthermore, all other pitch related parameters (pitch range, pitch maximum, and mean pitch) have to be included for a valid model and show significant effects. But still, duration and intensity do play a role. For a better overview, the ranking of the significant parameters is shown in Table 5.20

5.4 Summary and Discussion

5.4.1 The Encoding of Sentence Mode

First, the effects of the encoding sentence mode on acoustic parameters will be considered. Foremost, as fundamental frequency is concerned, questions differ from statements in that they use a higher register. This effect is predicted for echo-questions by Bolinger (1987): the echo-questions contextually contain a certain amount of surprise and ask for a confirmation of something that has not been expected. The “surprise” affect is probably further encouraged by the preceding exclamation of “What” or alike. Another indicator for a strong affective component is the higher intensity (if one excludes effects of contrast in statements) of the questions. A higher overall fundamental frequency is also found by Eady and Cooper (1986) for information questions induced by the instruction of “Ask ...”. So, the higher pitch register is in line with predictions from the literature. It is also the prosodic indicator of question mode for the current echo-questions which do not systematically contain an otherwise expected final rise (see the discussion on that point in section 5.2.2).

This discussion leads to the finding that one cannot distinguish statements and (echo-)questions by the contour of the final syllable. The proportion of falling or rising contours is similar for both sentence modes in the English sentences spoken by English native speakers. Not only that three quarters of the questions end in a falling or a low tone, but also one quarter of the statements end high or rising. (Ching, 1982) reports numerous occasions of rising (question) intonation for (functional) statements in dialects of the Southern USA. And as far as the elicited short sentences in the dialogues are concerned, the observation of Fries (1964), that there is no question - or statement phrasal intonation as such, is confirmed.

But what we do find is a relatively longer duration of the objects in questions, which can be accounted for by the larger fundamental frequency downstep (pitch range). Objects in questions are also louder, at least as unaccented objects in statements. Thus, there are acoustic differences between statements and questions in the object position, but not the opposition of the direction of the pitch movement.

To sum up the correlates of sentence mode, both the distribution of tones and the acoustic parameters show that questions and statements cannot be distinguished by the contour of the final constituent. But a higher pitch register as well as a higher intensity and a longer duration of the objects are strong prosodic correlates to prosodically distinguish questions from statements.

5.4.2 The Encoding of Contrastive Focus

Considering the correlates of contrastive focus we have to differentiate between statements and questions on the one hand, and the three investigated positions on the other hand. Consistent with Eady and Cooper (1986), no systematic correlates of corrective contrast were found for questions[11]. Their findings are even approved for corrective contrast which elicits the strongest possible accent (Molnár, 2006), while they “only” had new information focus encoded. As the context strongly maintained corrective contrast for the questions, one has to assume that the encoding of the sentence-mode somehow overrides the prosodic correlates of contrast.

[11] In the summary, the longer duration of contrasted subjects and the higher proportion of rising contours of contrasted verbs can be neglected. None is reaching the salience and statistical reliability compared to the strong effects in the statements.

For statements, the findings of acoustic correlates of contrast are quite clear. The length of the syllables and their intensity is increased if they are contextually contrasted, irrespective of their position. Pitch related parameters are affected more dependent on position: While there are no significant effects for pitch height parameters (pitch maximum and mean pitch) in sentence initial position, they are reliably higher in the sentence medial and sentence final position of statements. These findings are consistend with Eady and Cooper (1986), and Cooper et al. (1985). Pitch range and the alignment of the pitch maximum overall play a less important role in general. But in the sentence medial position the position of the peak is far later for contrasted syllables than uncontrasted, and the alignment ratio of the pitch maximum is the strongest predictor of contextual contrast in this position. The later peak reflects the higher proportion of rising contours for contrasted verbs as revealed by the analysis of tones (see 5.2.2).

The regression analysis shows that in the initial and the final position of statements, duration and intensity are the best and sufficient predictors for corrective contrast. This is in line with the findings of Kochanski et al. (2005). The case is different in the verb position, for which correlates of fundamental frequency are inevitable for a good model of correlates of contrastive focus. They are probably neglected in the latter study, as verbs are seldomly strongly accented in corpora. But they confirm observations of e.g. Bolinger (1961), Bolinger (1982), Ladd (1996) and K. Silverman et al. (1992)) that focus is likely to be related to a rising (or high) tone.

To conclude, the above analysis is the most comprehensive analysis of prosodic correlates of sentence mode and corrective contrast encoding available. It has been shown that for the same sentences, the echo-questions differ from statements mainly by a higher pitch register, and to a less salient extent by an elongation of the last syllable and more energy in that position. There are no reliable prosodic correlates of contextually induced corrective contrast in questions. For the statements instead, the hypothesis that duration and intensity are most strongly correlated to contextual contrast are supported. Fundamental frequency takes the clue-function over only in the sentence medial position, the verbs in the examined sentences. Here, higher fundamental frequency related parameter values can mostly be attributed to a change of the preferred contour: while verbs had a falling or low tone in most of the unaccented statements, the were rising or high in case the verb was under contrastive focus. This rising tone seems to be the most salient cue for the perception of and accent in that position.

6 German Native Speakers Speaking German

In order to get most comparable data on the prosodic encoding of contrastive focus and sentence mode in the native language of the learners, German utterances of German native speakers are going to be analyzed. Probable prosodic patterns in German have already been summarized elsewhere (see section 2.2), but only a detailed analysis of most comparable utterances can be used as a point of reference to expected interferences (see chapter 3) in the prosodic encodings of the L2 English (chapter 8). The utterances of the German control group should correspond not only in information structure, but also as far as possible in syntactic and lexical properties.

As for English, methods and materials in the literature are very heterogeneous and often inconsistent. However, several predictions can be set. First, we consider the encoding of sentence mode. To my knowledge, there is no previous empirical assessment of prosodic correlates of echo-questions in German. Recurring to the more general yes-no-(declarative) questions, one can expect almost exclusively rising contours. Although intuitively possible in certain circumstances[1], neither experimental assessment (e.g. Batliner, 1989a, 1989c; Oppenrieder, 1989a), nor perceptive intuition based approaches (e.g. Dietrich, 1990) related (yes/no-) questions to a falling contour.

Second, as for the encoding of contrast, research of acoustic correlates is coherent in that the dimension of intensity does not play a significant role to mark prominence in German. Neither experimentally elicited, focused syllables (e.g. Batliner, 1989b; Braun, 2004; Oppenrieder, 1989b), nor accented syllables in corpora (Wagner, 2002) are correlated with an

[1]Normative or instructional books on German pronunciation (e.g. Hirschfeld, Reinke, & Stock, 2007; Kaunzner, 1997) emphasize that wh-questions, questions asking for a decision between alternatives, and "information questions" (*Informationsfragen*) have to be spoken with a final fall. None of these works explicitly treats echo-questions. In Lemke, Graubner, and Lüssing (2006), "inquiries", *Nachfragen* would come closest to echo-questions. They should be intoned with a rising tone.

increased loudness. German contrast encoding seems to rely entirely on duration and fundamental frequency: especially a higher pitch maximum, a higher pitch range, and a later alignment of the pitch maximum are found.

These predictions are now investigated in a production experiment. It mirrors as exactly as possible the materials and procedures of the first experiment: English spoken by English native speakers.

6.1 Eliciting Correlates of Sentence Mode and Corrective Contrast

The goal of the experiment is to asses as precisely as possible the prosodic correlates of the encoding of sentence mode and contrastive focus that would be found in the language learners' speech if they did full transfer.

6.1.1 Subjects

Sixteen subjects were recorded. They were female native speakers of German, originating from Saxony (12), Saxony-Anhalt (1) and Thuringia (3). None of them spoke dialect, but as influences from regional varieties cannot be excluded (cf. Peters, Gilles, Auer, & Selting, 2002), the origin of the speakers was restricted. The speakers were between 20 and 25 years old (mean: 22.4) and were paid for their participation. None of them reported speaking or hearing impairments.

6.1.2 Materials and Design

Materials were designed to be similar as possible to the English materials (see section 5.1.2). Twenty-four three-syllable short SVO sentences were created. With a few exceptions, the proper names in the subject and the object position of the English materials were used again. The verbs were changed to mostly phonologically and/or semantically similar transitive German verbs, demanding accusative (21/24) or dative (3/24) objects. Note that the monosyllabic verbs had to be present tense. Example (1) presents two of the adapted target sentences, one very similar and one rather different to their English counterparts. The full set of materials is collected in Appendix B.

(1) a. Fred küsst Sue.
Fred kisses Sue.
(Eng. materials: Fred kissed Sue)

b. Dyke pflegt Nils.
Dyke takes-care-of Nils.
(Eng. materials : Dough pulled Ned.)

Corrective contrast was induced in the same way as for the English materials. A proposition P containing the entity α (subject, verb, or object) was introduced by the first speaker, who finally asked for the validity of that entity. The second speaker denied the truth of α in P. Then B uttered the critical statement P', with α' replacing α. Follows the next turn by the first speaker starting with a "Was ?" (*What ?*), and repeating P' as an echo-question. Then a 1-2 turn tail sequence followed.

The design was equivalent to Experiment 1: The factor Contrast had three levels: Contrastive focus on the subject (first syllable), the verb (second syllable) and the object (third syllable) of the sentences. All sentences were spoken in the two levels of Sentence Mode: as statements and follow-up (echo-)questions. Additionally, one filler condition, following a context inducing neutral focus (What happend?) was elicited in both sentence modes.

The 24 sentences were distributed over four lists in a Latin square design, so that every speaker uttered each sentence in one condition. Over all subjects, each critical sentence was recorded twice, and each subjects uttered three repetitions of each condition.

6.1.3 Procedure

The dialogues were printed on paper, both partners were able to read both parts of the dialogue. The speakers were instructed to first inspect their parts and then to read them aloud "lively": they should "imagine being actors in a radio play". The recordings took place in a sound proof booth at the University of Leipzig with a Neumann TLM 103 high quality microphone on a pre-amplifier / sound card combination (USB Mobile Pro) and digitized with a 44,100 Hz.

Preprocessing

Preprocessing was analogous to the first experiment with the English native control group. The critical sentences were extracted from the dialogues and word boundaries were tagged. A three-layer TextGrid was created with the PRAAT software package, containing information of the contextually intended position of the correction, the perceived main accent, and the perceived tones as described in section 5.2.2. All perceptual analysis has been undertaken by the author.

The correlates in the acoustic signal were obtained using the same settings of PRAAT as for the English native speakers. The plausibility of the values of the fundamental frequency assigned by PRAAT was controlled. In the case of octave jumps near obstruents the wrong values were corrected. Cases of other implausible values, for instance due to creaky voice, were excluded from the analysis.

6.2 Perceptive Analysis

6.2.1 Perceived Accent

As in Experiment 1, the goodness of the fit between the contextually induced position of the contrastive accent and the perceived main phrase accent was assessed. The percentage of “correct” pitch accent assignment per experimental condition is presentend in Table 6.1.

Table 6.1: Percentage of sentences with perceived main accent on contrastive focus position by sentence mode and focus position for German by native speakers

	CF S	CF V	CF O	mean
Statement	81	79	94	84.7
Question	88	38	85	70.3
mean	84.5	58.5	89.5	77.5

The German native speakers put the main phrase accent on the contextually contrasted constituent in 77.5 % of all critical sentences. The percentage of contrast-corresponding pitch accents is higher for statements (85 %) than for questions (70 %).

This is mainly due to the bad performance for echo-questions in which the verb is contextually corrected. Only in 38 % of all of these utterances, the perceived main accent is in the intended position, which is close to chance (33 %). But the overall stress placement is far from chance, even for questions. In sentence initial position, a contrasted subject receives the main stress in 88 % of the cases, a sentence final stress is found in 85 % of the object-contrast questions. Also, it can not be attributed to the position or syntactic function as the contrasted verbs in statements receive the main sentence accent in 80 % of the cases. Thus, neither the sentence mode nor the position can account for the low number of contrast - accent matches, but only the interaction of both.

There is no comparable data about the error-rate of elicited German corrective contrast available in the literature. Pechmann (1984) and Alter, Mleinek, Rohe, Steube, and Umbach (2001) do not mention rates of "correctly" accented sentences in their experiments: they only state that incoherences of contrastive focus and perceived accent are found.

Overall, the intended vs. perceived main accent ratio seems to be comparable to the data obtained in Experiment 1 for the English native speakers. Only the verb-contrasted echo-questions received a still less coherent accentuation (38% compared to 50% for English native speakers).

6.2.2 Distribution of Tones

Most of the more recent analyses of contrastive intonation in German is based on the ToBI annotation system, especially with it's German correspondence, GToBI (Baumann, Grice, & Benzmüller, 2001; Grice & Benzmüller, 1995; Grice & Baumann, 2002). A rising accent is expected on the accented syllable of a contrasted word (Grice et al., 2005; Steube, 2001) [2].

Nonetheless, there is no study available, that systematically investigated the tonal realization of corrective contrast in different positions of a sentence, including the verb. Now, the data of the constituent-wise perceptive analysis of the perceived contours is presented. It follows the same simplified inventory and assignment regularities as explained in section 5.2.2.

Perceived Tones in Sentence Initial Position

In sentence initial position (Table 6.2), we see that rising (LH) contours are dominant for all contrast conditions and both sentence modes. But their proportion is slightly higher if

[2] Steube (2001) also mentions the so-called root contour LHL as another possible correlate, especially in sentences that contain both, corrigendum and corrigens for corrective contrast.

Table 6.2: Distribution of perceived tones on subjects by sentence mode and contrast position for German by native speakers

		H	HL	L	LH
Statem.	CF S	16.7	2.1		81.3
	CF V	4.2	16.7	6.3	72.9
	CF O	14.6	10.4		75.0
mean (Statem.)		11.8	9.7	2.1	76.4
Quest.	CF S		4.2		95.8
	CF V	10.4	2.1	8.3	79.2
	CF O	10.4	4.2	2.1	83.3
mean (Quest.)		6.9	3.5	3.5	86.1

the syllable is under contrastive focus (CF S). Falling (HL) or low (L) tones are encountered mainly if the (upcoming) verb is contrasted (CF V) in statements.

Considering possible effects of sentence mode on the perceived contours, one can find a higher proportion of LH tones in questions than in statements, mostly on the cost of simple high tones: Especially in the CF S condition, H tones are not found in questions but make almost 17 % of the contours in statements.

But overall, the effects of contrast and sentence mode are rather weak: the dominant rising contour is more convincingly funded by the rising onset of a breath group. Whether the rise or high level tone is higher for subjects under corrective focus has to be tested in the analysis of acoustic correlates.

Perceived Tones in Sentence Medial Position

In the verb position, the second and sentence medial word and syllable of the examined utterances, a rising contour (LH) is expected for contrasted syllables, while they should be falling otherwise. The perceived contours are summarized in Table 6.3.

Considering effects of contrastive, one can state a strongly increased proportion of rising contours: 69% of all statements and 25 % of all questions in the CF V condition carry the LH contour. These proportions are still higher if one considers the perceptively accented verbs only: 87 % of all verbs in "correctly" accented statemtents and 61% of those in questions are spoken with a LH tone. In both other conditions, the fall is the domainant contour. Thus, the perceivable intonation contours seem to be heavily affected by the contrast conditions:

Table 6.3: Distribution of perceived tones on verbs by sentence mode and contrast position for German by native speakers

		H	HL	L	LH
Statem.	CF S	4.2	87.5	4.2	4.2
	CF V	14.6	16.7		68.8
	CF O	8.3	89.6	2.1	
mean(statem.)		9.0	64.6	2.1	24.3
Quest.	CF S	12.5	83.3		4.2
	CF V	10.4	64.6		25.0
	CF O	4.2	85.4	6.3	4.2
mean(quest)		9.0	77.8	2.1	11.1

While for non-accented verbs a fall is dominant, accented verbs are preferably accented with a rising tone.

Effects of sentence mode are directly related to the effects of contrast. The lower proportion of rises and the higher proportion of falls for questions can be attributed to the lower rate of perceptively stressed verbs, hence the contrast-effect.

Perceived Tones in Sentence Final Position

In the utterance - final position, the object of the sentences one can expect some interesting interaction. First, this is where one assumes to see the most salient encoding of sentence mode in German (e.g. Batliner, 1989a, 1989c; Féry, 1993), namely a steep rise of the fundamental frequency in case of questions, and the pitch declination of the declaratives. This declination (fall or low level tone) is opposed to a presumed rise induced by corrective contrast. Table 6.4 gives an overview of the encountered contours.

We find a clear distinction between statements and questions in an opposite proportion of falls and rises: more than 90 % of all questions end with a rising contour, but more than 75 % of the statements end falling or low – if one adds the proportions of the complex rise-fall (LHL) the number even exceeds 95 %.

The rise-fall contour is the only clear correlate of corrective contrast in this position, and it occurs only in statements.

For a better comparison to Experiment 1, a concatenated table of falling or low contours (L, HL, LHL) vs. rising or high contours (H, LH) is presented in Table 6.5.

Table 6.4: Distribution of perceived tones on objects by sentence mode and contrast position for German by native speakers

		H	HL	L	LH	LHL
Statem.	CF S		83.3	4.2		12.5
	CF V		93.8	2.1	2.1	2.1
	CF O	8.3	45.8	2.1	4.2	39.6
mean(Statem.)		2.8	74.3	2.8	2.1	18.1
Quest.	CF S	6.3	2.1		91.7	
	CF V	10.4	4.2		85.4	
	CF O	2.1	2.1		95.8	
mean (quest.)		6.3	2.8		91.0	

Table 6.5: Phrase final low/falling vs. high/rising tones by sentence mode (summed percentages)

	Statements	Questions
low/falling	95.1	2.7
high/rising	4.9	97.3

We see a clear opposite direction of the sentence-final contours for both sentence modes. While statements end almost uniquely falling (95 %), questions are usually encoded by a rising or high tone (97%).

6.2.3 Summary and Discussion

The results of the intonation annotation of the critical utterances are in line with the expectations.

To summarize the findings for the encoding of contrastive focus in German, three points can be made. First, in the sentence initial position there are a slightly higher proportion of rises (LH) and a non-occurrence of low tones (L) that seem to be local correlates of contrastive focus on the subjects of the sentences. Second, in the sentence medial position, corrective contrast is clearly correlated with a local rise instead of a fall in the case the verb is contrasted. This point is clearly visible for the whole declarative dataset including non-accented focused verbs, and even more striking (almost 90 %) if one considers the accented verbs in statements only. But also verbs in questions show that pattern: instead of about 4 % of rises in the non-contrasted conditions, 25 % are found for corrected verbs. The proportion increases to about 60 % if one considers only accent-contrast matching verbs in questions. And third, in the sentence final position correlates of contrastive focus are only found in statements. There, an increased proportion of rising-falling contours (LHL) and the occurence of high level tones (H) at the cost of simply falling contours (HL) can be stated. As up- and downsteps were not annotated, no direkt comparison to Baumann et al. (2006, 2007) can be done: however, his finding of a missing downstep for contrastively focused objects is supported by the almost 10 % of high tones in that position. Properties of the relative pitch height will be analyzed later in the section on the acoustic correlates.

As for the encoding of sentence mode, it is made clear that the statements and the echo-questions only differ in the shape of the final contour[3]: Statements almost uniquely end low (L, HL, or LHL tone), while the questions (almost) exclusively end high (H or LH tones). This is coherent to the predictions in the literature (see section 2.2).

6.3 Acoustic Correlates

The acoustic correlates of sentence mode and corrective contrast are analyzed following the methods and settings of Experiment 1 (see section 2.1). Again, all utterances have been included in the statistical analysis which means that the data set includes sentences in which the perceived main sentence accent and the corrective contrast did not match. The rather high number of items per cell (48) should nonetheless reveal pertinent findings for both factors, using within-subjects ANOVAs and Post-Hoc-Tests to examine the validity of differences.

Six acoustic parameters (duration ratio, pitch maximum, mean pitch, pitch range, alignment ratio of the pitch maximum, and intensity ratio)[4] were examined.

6.3.1 Overview of the F0 Contours

First, an overview of the time course of the fundamental frequency in all six experimental conditions is given. Figure 6.1 shows the mean fundamental frequency for every quarter of each constituent by condition.

First, the question vs. statement distinction shall be considered. In the subject (Su1 - Su4) and the verb position (Ve1 - Ve4) of the sentences, no clear difference between the two

[3]Other aspects yielding small differences between statements and questions in the subject and the verb position can be explained by a more consistend encoding of focus, especially on the verbs in statements.
[4]For the definitions see Table 5.8.

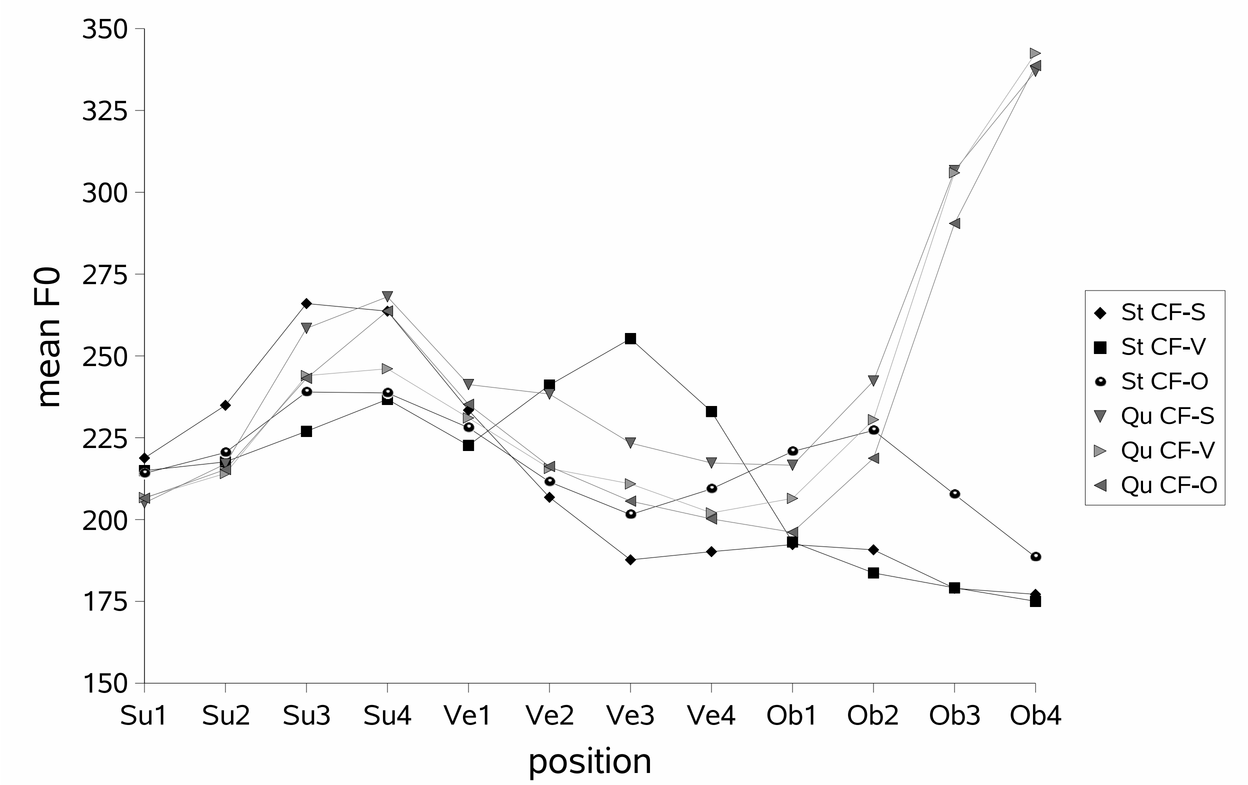


Figure 6.1: Time normalized fundamental frequency contours for German by native speakers

sentence modes (St = statements and Qu = questions) can be made. But in the object position (Ob1 - Ob4) , questions show a steep rise, while statements stay low or fall. So, German encodes sentence mode in the sentence final position only.

Second, the visual inspection of the fundamental frequency contours regarding the correlates of contrast reveals two main points. If we consider the statements only, we see that the contrasted condition always induces the highest value for pitch on that constituent: if the contrastive focus is on the subjects (St CF-S), the line has the highest values on the subjects (Su1 - Su4). For contrasted verbs in statements (St CF-V), the fundamental frequency is higher than for every other condition during the second half of the verb (Ve3-Ve4). In sentence final position, a correcting object (St CF-O) is higher than objects in sentences in which the subject or the verb are accented.

Such clear effects of corrective contrast on the pitch contour do not come to light for questions. Contrasted constituents are only slightly higher than uncontrasted in the sentence medial position, or even lower than the highest in the verb and the object position[5].

[5]However, for the sentence medial position, one might argue that the non-occurrence of a clear contrast-induced fundamental frequency peak is caused by the low number of “correctly” accented verbs. However, if a clear contrast-induced effect on pitch height occurred in that position for the 38 % of accented verbs,

Whether there are significant differences between contrasted and uncontrasted syllables and between statements and questions will be investigated in the following sections, analyzing the six acoustic parameters which are possibly relevant for a prosodic encoding of these two factors.

6.3.2 Duration Ratio

The first parameter which is examined is the duration ratio of the constituents. The duration ratio is the duration of the constituent divided by the duration of the whole sentence. It therefore normalizes possible differences in speech rate between speakers and utterances. If all constituents had the same length, they would amount to 33 (% of the sentence) each.

The data are presented in Table 6.6, the boxplots in Figure 6.2 provide further information about the distribution of the values.

Table 6.6: Mean duration ratio for German by native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	30.04	30.70	30.37	30.31	30.21	30.26	39.66	39.09	39.37
CF V	27.74	28.65	28.19	33.74	32.24	32.99	38.51	39.12	38.82
CF O	27.57	28.62	28.09	29.41	29.68	29.54	43.01	41.69	42.35
$\overline{X}_M$	28.45	29.32	28.89	31.15	30.71	30.93	40.39	39.97	40.18

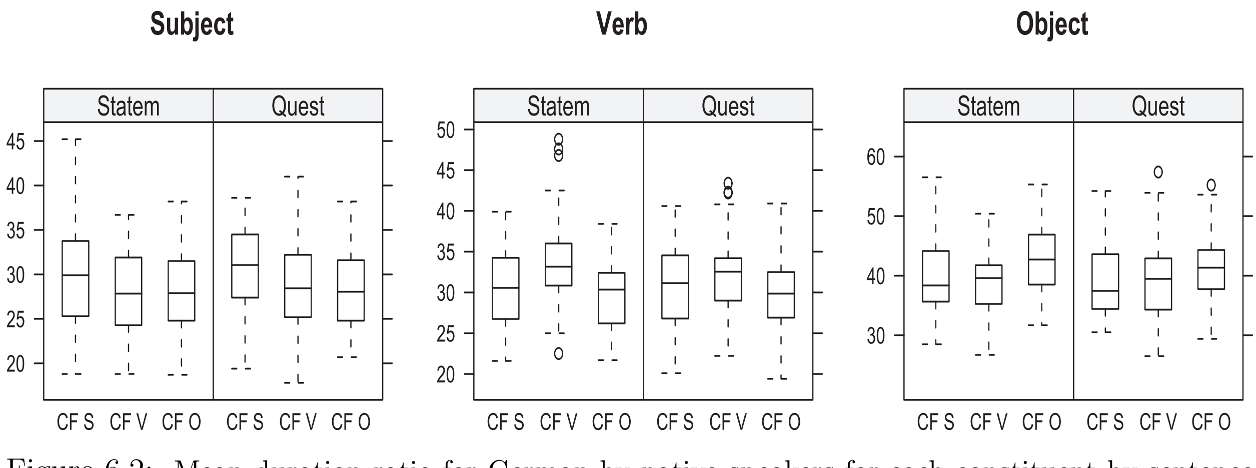


Figure 6.2: Mean duration ratio for German by native speakers for each constituent by sentence mode and contrast position.

it should show up in the means. Furthermore, more than 80 % of the contrasted subjects and objects received the main sentence accent if they were contextually contrasted. One would expect a higher fundamental freqency, but this does not seem to be the case.

With regard to the effects of Contrast on the duration of the constituents, we see that the means of the contrasted syllables are always the longest compared to the uncontrasted conditions. Numerically, this is the case for both, statements and questions, but the differences are smaller for the interrogatives. Differences in the relative duration of the constituents between statements and questions are small, only subjects of questions are about one percent longer than the sentence initial syllables of the declaratives.

In the sentence-initial position, the subjects, this is approved by the within-subjects ANOVAs. The significant main effect for Contrast ($F1(2,30) = 52.84$, $p < .01$; $F2(2,46) = 8.74$, $p < .001$) signals a relatively longer duration. However, the Post-Hoc tests (Bonferroni-adjusted pairwise comparisons) are too conservative to substantiate this finding: there are no significant differences between the contrast conditions neither for the statements, nor for the questions. The main effect for Mode is significant only in the item-based analysis: $F1(1,15) = 1.14$, p = ns.; $F2(1,23) = 5.58$, $p < .05$. Numerically, subjects in questions are longer. There is no significant interaction Contrast*Mode: $F1(2,30) = 0.04$, p = ns; $F2(2,46) = 0.06$, p = ns.

In the verb position, there is a significant main effect for Contrast: $F1(2,30) = 9.04$, $p < .001$; $F2(2,46) = 16.48$, $p < .001$. Here, the Bonferroni-adjusted pairwise comparisons substantiate that the contrasted verbs in statements are significantly longer than verbs in both other contrast conditions, and even a longer duration for contrasted verbs in questions compared to verbs in object-contrasted sentences[6]. There is no difference in lenght of the verbs between statements and questions: $F1(1,15) = 0.41$, p = ns.; $F2(1,23) = 1.55$, p = ns., and the interaction Contrast*Mode is non-significant, too: $F1(2,30) = 1.10$, p = ns.; $F2(2,46) = 1.46$, p = ns.

A similar pattern is observed in the object position: The main effect for Contrast is significant: $F1(2,30) = 8.64$, $p < .001$; $F2(2,46) = 10.34$, $p < .001$. The pairwise comparisons confirm longer contrasted objects than objects in statements with a contrasted subject ($p < 0.01$) and contrasted verbs ($p < 0.05$), but there are no significant differences in questions. The main effect for Sentence Mode is not significant: $F1(1,15) = 0.11$, p = ns.; $F2(1,23) = 0.96$, p = ns., nor is the interaction Contrast*Sentence Mode: $F1(2,30) = 1.59$, p = ns.; $F2(2,46) = 1.06$, p = ns. However, the results of the ANOVA and the Post-Hoc tests are contradictory: the non-significant interaction indicates a similar behavior of contrast in statements and questions, but the Bonferroni-adjusted pairwise comparisons state significantly longer contrasted objects only for statements but not for questions. They must be

[6]Note that there are more than 60 % of contrasted verbs without perceivable sentence accent in the data sample.

regarded as being too conservative to assess a significant difference to the numerically similar effects of contrast in statements.

All in all, we can state that Contrast is encoded by a longer syllable duration in German spoken by native language speakers. However, the effects do reach significance in the Bonferroni-corrected comparisons not in each position and for each sentence mode. Sentence mode encoding does not strongly use duration as a cue.

6.3.3 Pitch Maximum

The first of the four analyzed fundamental-frequency-related parameters is the pitch maximum per constituent. The data are presented in Table 6.7 and visualized with boxplots in Figure 6.3

Table 6.7: Pitch maximum for German by native speakers; mean by sentence mode and contrast conditions for each constituent

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	282.3	285.3	283.8	241.4	265.2	253.3	214.5	356.0	285.2
CF V	248.0	263.3	255.6	279.1	254.2	266.6	214.9	360.3	287.6
CF O	256.0	277.8	266.9	243.8	255.4	249.6	263.1	353.2	308.1
$\overline{X}_M$	262.1	275.5	268.8	254.8	258.2	256.5	230.8	356.5	293.7

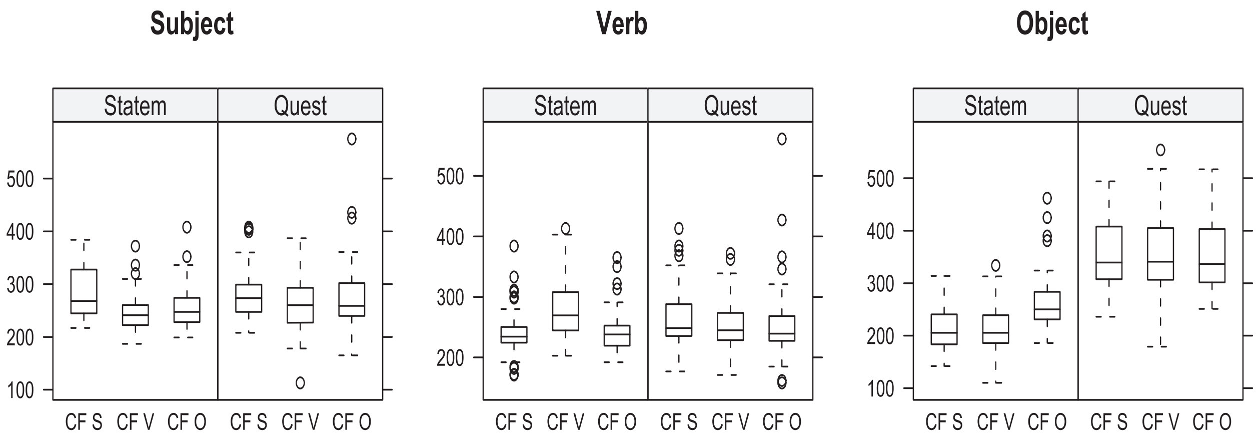


Figure 6.3: Pitch maximum for German by native speakers for each constituent by sentence mode and contrast position.

First, with regard to effects of Contrast on the maximum of the fundamental frequency in the three positions of the sentences, we clearly see the excursion for contrasted constituents in statements. For questions, a higher pitch maximum can be found only sentence-initially, if ever. Second, there are minor differences between the two different sentence modes on the first two constituents, but an elevation of more than 100 Hz for questions compared to statements at the end of the sentences. The statistical analysis confirms these points:

On subjects, we find a significant main effect for Contrast: $F1(2,30) = 9.28$, $p < 0.001$; $F2(2,46) = 8.20$, $p < 0.001$. The Post-Hoc comparisons reveal, however, that the pitch maximum is reliably higher only in statements, but not in questions. The main effect for Mode is not significant: $F1(1,15) = 4.03$, p = ns.; $F2(1,23) = 4.09$, p = ns., thus the slightly higher value for subjects in questions is not confirmed. The interaction Contrast*Sentence Mode is not significant: $F1(2,30) = 0.94$, p = ns.; $F2(2,46) = 0.79$, p = ns. Like for the duration ratio analyzed above, the ANOVA would suggest similar effects of Contrast in both sentence modes, but the pairwise comparisons are too conservative to assess the somewhat weaker effects of Contrast in questions than in statements.

In the sentence medial position, the verbs, the main effect for Contrast is not significant at a 5 % level: $F1(2,30) = 3.25$, $p = .053$; $F2(2,46) = 2.57$, $p = .087$, but rather close to it. The adjusted pairwise comparisons show that contrasted verbs have a higher pitch maximum than the other two conditions if they occur in statements, but not in questions. There is no significant main effect for Sentence Mode: $F1(1,15) = 0.14$, p = ns.; $F2(1,23) = 0.44$, p = ns. But interaction Contrast*Sentence Mode is significant: $F1(2,30) = 11.99$, $p < .001$; $F2(2,46) = 6.67$, $p < .01$. As Table 6.7 shows, the pitch maximum is more than 35 Hz higher if the verb is contrasted in statments, but it is 10 and 20 Hz lower than non-contrasted verbs in questions. Together, they weaken the main effect for Contrast, but they show a high interaction.

In the sentence final position, the objects, the ANOVA reveals a significant main effect for Contrast: $F1(2,30) = 7.08$, $p < .01$; $F2(2,46) = 3.49$, $p < .05$. The Bonferroni-adjusted pairwise comparisons support a higher pitch maximum for contrasted objects than for both uncontrasted conditions ($p = 0.004$ against CF S; $p < .011$ against CF V) in statements, but no such differences in questions. The main effect for Sentence Mode is highly significant: $F1(1,15) = 87.60$, $p < .001$; $F2(1,23) = 518.80$, $p < .001$ approving a higher pitch maximum for questions than for statements. The interaction Contrast*Sentence Mode is also significant: $F1(2,30) = 5.49$, $p < .01$; $F2(2,46) = 7.17$, $p < .01$ and supports the finding that the maximum of the fundamental frequency is influenced by Contrast only in statements, but not in questions. The pitch maximum on the objects of the questions may already reach the ceiling of the speakers' usage of pitch for an unambiguous and strong sentence mode encoding in this position.

To conclude, we can certainly state that German is using the pitch maximum to encode contrast in statements. This effect is not found to be robust in questions. As to the encoding for sentence mode, there are no significant differences between statements and echo-questions before the sentence-final position. But there, questions show a the strongly increased pitch maximum, while the statements undergo declination.

6.3.4 Mean Pitch

The mean fundamental frequency per constituent is highly correlated to the pitch maximum presented above. But, as it includes the measures over the whole syllable it is less sensitive to steep pitch excursions, and gives a more salient measure of the perceivable height of the words. The data is given in Table 6.8 and the main characteristics of the distribution can be derived from Figure 6.4.

Table 6.8: Mean pitch for German by native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	245.0	231.6	238.3	203.6	235.1	219.4	184.8	270.1	227.4
CF V	221.2	225.3	223.3	241.4	217.6	229.5	183.0	264.2	223.6
CF O	226.5	228.6	227.6	210.6	214.9	212.8	217.6	252.7	235.2
$\overline{X}_M$	230.9	228.5	229.7	218.5	222.5	220.5	195.1	262.3	228.7

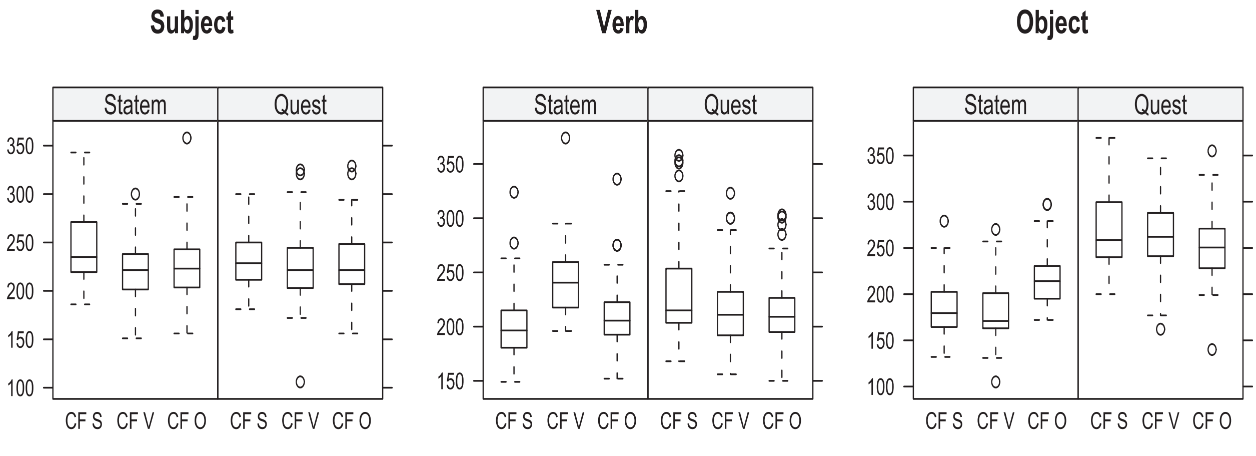


Figure 6.4: Mean pitch for German by native speakers for each constituent by sentence mode and contrast position.

The encoding of Contrast is clearly visible in all positions for the statements, but is not clearly correlated within questions. In the sentence medial and the sentence final position, the values are even lower than other contrast conditions. Sentence mode encoding does not induce strong differences in the sentence initial and medial position, but there is a huge distinctions between the mean pitch of statements and questions at the end.

The within-subjects ANOVAs confirm this view: in the initial position, a significant main effect for Contrast is found: $F1(2,30) = 9.43$, $p < .001$; $F2(2,46) = 4.89$, $p < .05$. The Post-Hoc comparisons reveal that the difference between accented condition and both unaccented conditions in statements is significant in statements, but nor in questions. There is no significant main effect for Sentence Mode: $F1(1,15) = 0.38$, p = ns.; $F2(1,23) = 0.52$, p = ns. The interaction Contrast*Sentence Mode: $F1(2,30) = 3.04$, $p = 0.063$; $F2(2,46) = 1.69$, p = ns. is not significant, either. Again, the occurrence of a significantly higher mean frequency in contrasted sentence-initial syllables of statements but not of questions revealed by the Bonferroni-adjusted t-tests is not reflected by an interaction.

In the medial position, the verbs of the sentences, the main effect of Contrast is also significant: $F1(2,30) = 5.05$, $p < .05$; $F2(2,46) = 3.82$, $p < .05$. The connected pairwise comparisons attribute significant differences between the CF V and CF S ($p < .001$) and CF O ($p < .001$) in statements, but no significant differences in questions. The main effect for Sentence Mode is not significant: $F1(1,15) = 0.31$, p = ns.; $F2(1,23) = 1.37$, p = ns., but the interaction Contrast*Mode: $F1(2,30) = 25.81$, $p < .001$; $F2(2,46) = 12.99$, $p < .001$ is highly significant: it supports findings of the pairwise comparisons that the effects of Contrast on verbs are stronger in statements than in questions.

On the objects, the ANOVA does not show a significant main effect for Contrast[7] $F1(2,30) = 3.14$, p = ns.; $F2(2,46) = 1.85$, p = ns. The Post-Hoc pairwise comparisons find significant differences between contrasted objects and the uncontrasted conditions (CF S: $p < .001$, CF V: $p < .01$) in statements, but not in questions. The main effect for Sentence Mode is highly significant, clearly stating the higher mean fundamental frequency in questions than in statements: $F1(1,15) = 221.87$, $p < .001$; $F2(1,23) = 559.70$, $p < .001$. The interaction Contrast*Sentence Mode is significant: $F1(2,30) = 11.09$, $p < .001$; $F2(2,46) = 12.04$, $p < .001$. Again, this supports the finding of stronger effects of Contrast in questions than in statements.

6.3.5 Pitch Range

The third acoustic parameter investigated in the critical utterances is pitch range, the difference between the pitch maximum and the pitch minimum of the syllable, irrespectively of

[7]But approaching significance in the subject-based analysis: $p = 0.058$.

the time of their occurrence. Thus, there is no differentiation between falls and rises. The data is presented in Table 6.9, and edited in a boxplot in Figure 6.5.

Table 6.9: Mean pitch range for German by native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	70.92	90.54	80.73	62.73	57.33	60.03	49.88	146.5	98.18
CF V	48.15	65.27	56.71	70.67	66.85	68.76	52.19	161.7	106.94
CF O	57.25	81.79	69.52	54.29	67.98	61.14	85.81	169.5	127.65
$\overline{X}_M$	58.77	79.20	68.99	62.56	64.06	63.31	62.62	159.2	110.92

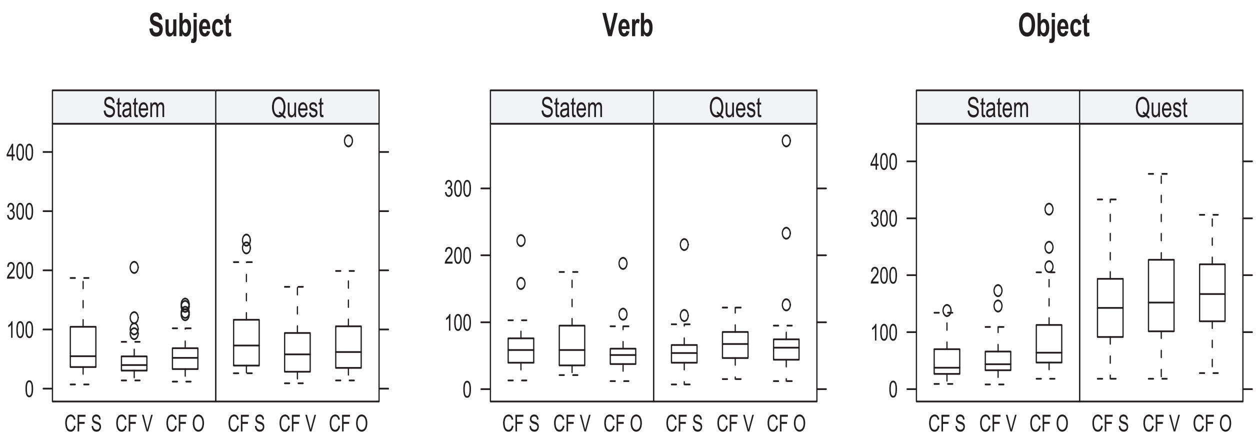


Figure 6.5: Pitch range for German by native speakers for each constituent by sentence mode and contrast position.

The pitch range is clearly influenced by the sentence mode, in that questions have a larger difference between the pitch maximum and the minimum in the subject and the object position of the sentences. In the verb position, the pitch range of declarative and interrogative sentences is almost equal. Contrast induces an increased pitch range in the sentence initial and the sentence final position, but the effects seem to be slightly stronger for statements than for questions. In the verb position, Contrast seems to barely encoded by the pitch range.

These observations are approved by statistics: On the subjects of the sentences, the ANOVA reveals a main effect for Contrast: F1(2,30) = 6.62, $p < .01$; F2(2,46) = 9.12, $p < .001$. The post-hoc Bonferroni-adjusted pairwise comparisons assign significantly higher values for contrasted subjects compared to subjects in verb-contrasted sentences (CF V) only for both statements ($p < .05$) and questions ($p < .05$). The main effect for Sentence Mode is significant, too: F1(1,15) = 11.00, $p < .01$; F2(1,23) = 11.54, $p < .01$ and shows a higher pitch range for questions than for statements. The interaction Contrast*Sentence Mode is not significant: F1(2,30) = 0.16, p = ns.; F2(2,46) = 0.23, p = ns., suggesting similar effects of Contrast for both sentence modes.

In the sentence medial position, no significant effect are found: Neither for Contrast: F1(2,30) = 1.46, p = ns.; F2(2,46) = 1.56, p = ns., nor for Sentence Mode: F1(1,15) = 0.06, p = ns.; F2(1,23) = 0.10, p = ns., and the interaction Contrast*Mode is not significant either: F1(2,30) = 2.40, p = ns.; F2(2,46) = 1.96, p = ns. Thus, the pitch excursion on verbs is not affected by the encoding of contrast or sentence mode.

In phrase final condition, we find a significant main effect for Contrast: F1(2,30) = 8.89, p < .001; F2(2,46) = 5.87, p < .01. The Post-Hoc analysis shows that the differences in pitch range are significant for statements only: both non-contrasting conditions (CF S and CF V) have a smaller pitch range than the object-accenting condition CF O. There is a significant main effect for Sentence Mode: F1(1,15) = 39.94, p < .001; F2(1,23) = 131.78.28, p < .001, showing higher pitch range for statements than for questions. The interaction Contrast*Sentence Mode is not significant: F1(2,30) = 1.00, p = ns.; F2(2,46) = 1.16, p = ns[8].

To summarize, we find that the encoding of Sentence mode affects the pitch range in initial and final position in that questions have a stronger difference between the pitch maximum and the pitch minimum than statements. Contrast induces a higher pitch range on contrasted compared to non-contrasted words too in the edge positions, at least for statements. The pitch range in sentence medial position is not altered to encode echo-questions or contrast on this syllable.

6.3.6 Alignment of the Pitch Maximum

The alignment of the pitch maximum is interesting in three perspectives. First, a relatively late maximum would indicate that the overall pitch contour is a rising one, a relative early one would be found in a falling contour. Second, as the work of Braun (2004) indicates, the usage of peak alignment in relation to pitch height is speaker dependent, most using both of them to indicate focus, but some preferring height, and others preferring a later peak for prominence encoding. The third point is that the alignment of the pitch maximum is different between English and German accents. Thus to assess the second language learners in Experiment 3, a detailed analysis of the native peak location on accented (and non-accented) syllables could be useful.

[8]Again, this is does undermine the findings of the Post-Hoc tests, which confirm significantly higher pitch range only on objects in statements.

The alignment ratio of the pitch peak is computed as the relative time of occurrence of the pitch maximum within the constituent[9]. The maximum value is 1; this would indicate that the peak always occurred at the end of the syllable. The minimum of 0 is found if the peak is in the very beginning of the syllable. The means of the alignment ratio of the pitch maximum are presented in Table 6.10 and a boxplot informs graphically about the data patterns in Figure 6.6.

Table 6.10: Mean alignment of the pitch maximum for German by native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	0.6171	0.7006	0.6589	0.1908	0.2694	0.2301	0.2279	0.7308	0.4794
CF V	0.4945	0.5483	0.5214	0.4605	0.2668	0.3637	0.2350	0.7325	0.4837
CF O	0.6150	0.6355	0.6253	0.2561	0.2362	0.2461	0.3374	0.7201	0.5288
$\overline{X}_M$	0.5756	0.6282	0.6019	0.3025	0.2575	0.2800	0.2668	0.7278	0.4973

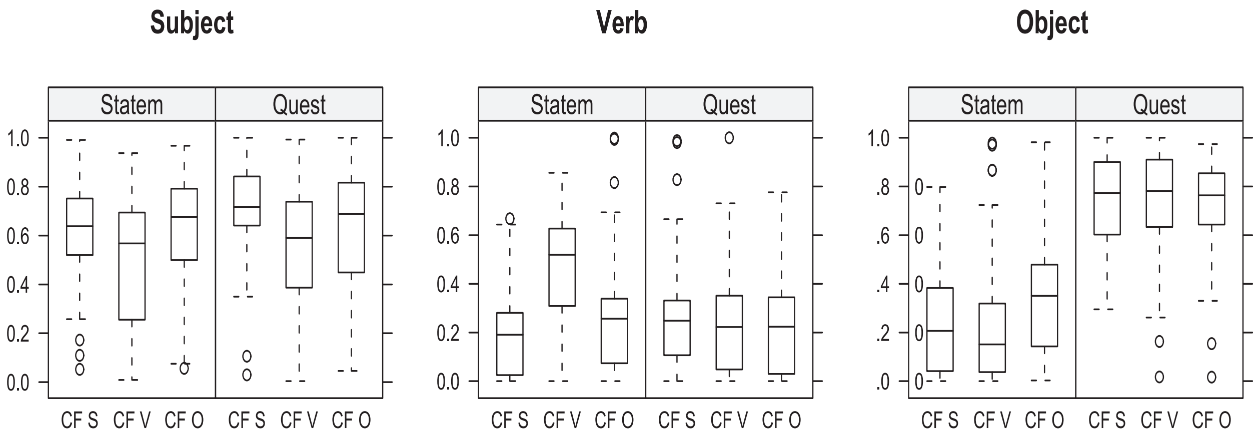


Figure 6.6: Alignment of the pitch maximum for German by native speakers for each constituent by sentence mode and contrast position.

Considering effects of sentence mode first, the means inform us about a slightly later peak in questions than in statements in the subject position, and a strong delay of the peak in the object position. The latter is an acoustic correlate of the rising and falling contours for questions and statements, respectively, at the end of the utterances. Sentence mode encoding does not induce effects in the sentence medial position. The case is different for the encoding of contrast. Here, the effects of contrast are the strongest for verbs in statements, which clearly show a later peak than both uncontrasted equivalents. The delay of the peak is visible but somewhat weaker on the objects in statements. In the subject position, the position of the peak seems to be more importantly influenced by corrective contrast in the

[9]See also Table 5.8.

following syllable, as the CF V condition clearly induces earlier peaks for both, statements and questions. In the other two positions, the verbs and the objects, the position of the peaks does not vary due to contrast in the echo-questions.

Testing these impressions with an appropriate ANOVA on the subjects of the sentences, we find a significant main effect for Contrast: $F1(2,30) = 8.98$, $p < .001$; $F2(2,46) = 8.83$, $p < .001$, and a significant main effect for Sentence Mode: $F1(1,15) = 4.50$, $p = .05$; $F2(1,23) = 4.28$, $p = .05$. The interaction Contrast*Sentence Mode is not significant : $F1(2,30) = 0.37$, p = ns.; $F2(2,46) = 1.03$, p = ns. The Post-Hoc Bonferroni-adjusted comparisons reveal that the peak is later for contrasted subjects than for subjects in verb contrasted sentences for both sentence modes. Thus, contrast does not necessarily show a significant delay of the peak due to a local contrast position, but a preparation effect for the strong peak alignment effect of contrast encoding on the following verbs.

In the sentence medial position, the verbs, there is a significant main effect for Contrast: $F1(2,30) = 11.79$, $p < 0.001$; $F2(2,46) = 10.12$, $p < .001$, but no significant main effect for Mode: $F1(1,15) = 1.43$, p = ns.; $F2(1,23) = 2.90$, p = ns. The interaction Contrast*Mode: $F1(2,30) = 12.72$, $p < .001$; $F2(2,46) = 9.47$, $p < .001$ is significant. The Post-Hoc tests show that the peak is significantly later in contrasted verbs than uncontrasted verbs in both other conditions in statements, but there is no significant difference in questions.

In the object position, the main effect for Contrast is not significant: $F1(2,30) = 1.07$, p = ns.; $F2(2,46) = 2.18$, p = ns. But the peak is significantly later, if the utterance is a question as the significant main effect for Sentence Mode shows: $F1(1,15) = 5.10$, $p < .001$; $F2(1,23) = 163.54$, $p < .001$. The interaction Contrast*Mode is significant in the item-based analysis only: $F1(2,30) = 2.25$, p = ns.; $F2(2,46) = 3.86$, $p < .05$. However, the Bonferroni-adjusted comparisons of the item based aggregation show that the peak is significantly later in contrasted objects compared to objects in subject-contrasted statements ($p < .05$). There are no significant differences due to Contrast in questions.

As a conclusion one can state that the alignment of the pitch peak to encode sentence mode and corrective contrast is used very much depending on the position: It is a strong cue for contrast on the verbs of the target sentences, which reflect the later peak of the dominant rises on contrasted verbs in statements. And in sentence final position, the position of the peak captures the rising contours of questions (late peak) and falling contour of statements (earlier peak).

6.3.7 Intensity Ratio

Acoustic correlates of contrast and sentence mode include the dimension of intensity which is often neglected in research on intonation. The data presented in Table 6.11 and graphically edited in Figure 6.7 show the mean intensity ratio of the constituents: it is the mean intensity of the constituent (subject, verb, object) divided by the mean intensity of the whole utterance. This normalization is necessary as artifacts resulting not only from inter-speaker loudness differences, but also from head turning, moving, and slightly different distances and angles of the speakers to the microphone etc., have to be expected. A value of 1 would mean that the mean intensity of the constituent matches exactly the mean intensity of the whole sentence.

Table 6.11: Mean intensity ratio for German by native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	1.029	1.014	1.021	0.9651	0.9687	0.9669	0.9616	0.9928	0.9772
CF V	1.019	1.012	1.015	0.9898	0.9762	0.9830	0.9594	0.9938	0.9766
CF O	1.022	1.014	1.018	0.9672	0.9666	0.9669	0.9858	0.9939	0.9898
$\overline{X}_C$	1.023	1.013	1.018	0.9740	0.9705	0.9723	0.9689	0.9935	0.9812

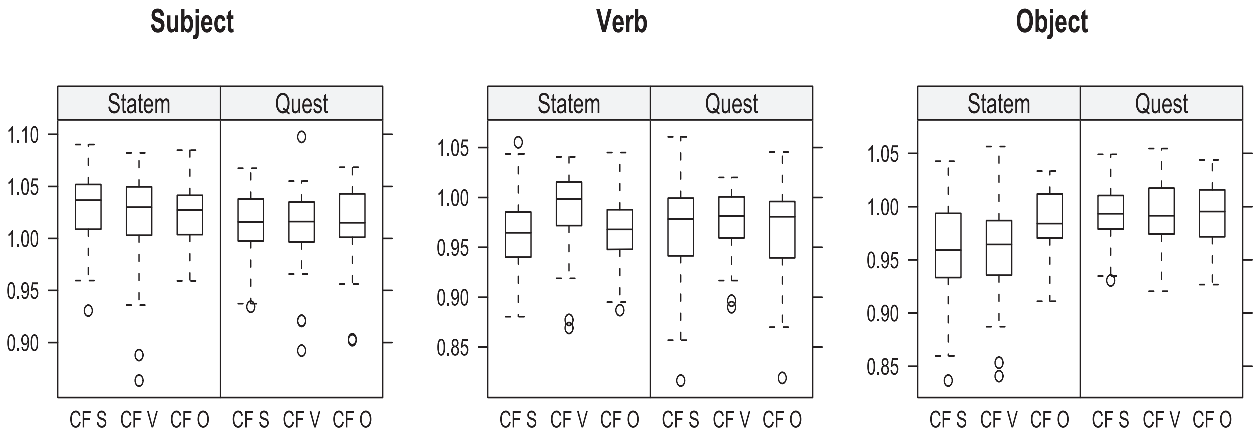


Figure 6.7: Mean intensity ratio for German by native speakers for each constituent by sentence mode and contrast position.

We see that the differences are very small. However, sentence initially the statements are slightly louder than questions, and sentence finally, the questions are louder than the statements, while the sentence modes do not differ in intensity on the verbs. As for contrast, we see that the contrasted constituents are always the loudest, but the differences to the loudest not contrasted counterparts vary between 0.024 (objects in statements) and 0.001 (objects in questions) points in mean intensity ratio.

This is what the within-subjects ANOVA approves for that position: the main effect for Contrast is not significant on subjects: $F1(2,30) = 0.55$, p = ns.; $F2(2,46) = 0.98$, p = ns. But statements are significantly louder as the main effect for Mode shows ($F1(1,15) = 6.67$, $p < .05$; $F2(1,23) = 10.97$, $p < .01$). There is no significant sentence-initial interaction Contrast*Mode : $F1(2,30) = 0.25$, p = ns.; $F2(2,46) = 0.66$, p = ns.

On verbs, we find a significant main effect for Contrast: $F1(2,30) = 4.59$, $p < .05$; $F2(2,46) = 10.11$, $p < .001$, but no significant main effect for Sentence Mode: $F1(1,15) = 0.36$, p = ns.; $F2(1,23) = 1.16$, p = ns., nor a significant interaction Contrast*Sentence Mode: $F1(2,30) = 1.84$, p = ns.; $F2(2,46) = 1.37$, p = ns. The Bonferroni-adjusted comparisons confirm a significant difference between verbs under contrastive focus and non-contrasted verbs in statements, but not in questions. Again, they are probably too conservative to approve significant effects in questions.

In the sentence final position, the ANOVA reveals a significant main effect for Contrast: $F1(2,30) = 3.60$, $p < .05$; $F2(2,46) = 5.10$, $p < .01$, as well as a significant main effect for Mode: $F1(1,15) = 31.87$, $p < .001$; $F2(1,23) = 54.68$, $p < .001$, and (partly) a significant interaction Contrast*Mode: $F1(2,30) = 2.89$, $p = .071$; $F2(2,46) = 6.18$, $p < 0.01$. The Post-Hoc Tests show that the intensity differences are due to objects which are significantly louder if they are contrasted than if not in statements. No significant contrast-induced differences are confirmed for objects in questions.

To sum up effects of contrast and sentence mode encoding on intensity, we can state that questions are uttered louder than statements in sentence final position, but quieter in the sentence initial position. Contrast makes the intensity increase in the sentence-medial and sentence final position of the statements at least, but not in sentence-initial position, statistically stable only in statements.

6.3.8 Overview for Contrast

Finally, the results of the statistical tests on significant differences will be summarized in Table 6.12. It shows significant differences of the Bonferroni adjusted pairwise comparisons between the contextually contrasted constituents and its non-contrasted counterparts in the same position and in the same sentence mode.

Table 6.12: Significant differences in the six acoustic parameters between contrasted and non-contrasted constituents by sentence mode for German by native speakers.

		Subjects		Verbs		Objects	
		F1	F2	F1	F2	F1	F2
Statem.	Duration			++	++	++	++
	Pitch max.	++	++	++	++	++	++
	Mean pitch	++	++	++	++	++	++
	Pitch range		+		+	+	++
	Pitch max. align.		+	++	++		+
	Intensity			++	++	++	++
Quest.	Duration		++	+	+		+
	Pitch max.						
	Mean pitch						
	Pitch range		+				
	Pitch max. align.	+	+				
	Intensity						

Analyses over Subjects (F1) and over Items (F2). '+' means significantly different to one other contrast condition, '++' means significantly different to both other conditions. All significant differences show a higher value for the dependent variable in the contrasted condition than the uncontrasted condition(s).

Considering the results of the pairwise comparisons only, tow aspects become clear: First, contrast is not reliably encoded by any acoustic parameter in questions[10].

Second, the most reliable parameters in statements are the mean and the maximum of the fundamental frequency. Contrasted syllables are always higher than non-contrasted, independently of their position. A longer duration and an increased intensity are not approved statistically in the subject position, but on the verbs and the objects of the sentences. A later peak is induced by contrast only in the verb position, where a (regular) fall becomes a rise in the case of corrective contrast.

[10]A weak exception may be seen in duration. But sentence-initially, significantly longer subjects are found in the item-based analysis only; in the verb position, the contrasted verbs are only longer than verbs of questions in which the object is contrasted, but not in which the subject is contrasted.

6.3.9 Overview for Sentence Mode

An overview of the PostHoc tests comparing the differences for sentence mode for every contrast position is given in Table 6.13.

Table 6.13: Direction and contrast condition of significant differences between statements and questions

	Subjects		Verbs		Objects	
	F1	F2	F1	F2	F1	F2
Duration	–	–	–	–	–	–
Pitch max.	–	–	A<Q: S; A>Q: V	A<Q:S ; A>Q: V	A<Q: S,V,O	A<Q: S,V,O
mean Pitch	A>Q: S	A>Q: S	A<Q: S; A>Q: V	A<Q: S; A>Q: V	A<Q: S,V,O	A<Q: S,V,O
Pitch range	A<S: O	A<Q: S,O; A> Q:V	–	–	A<Q: S,V,O	A<Q: S,V,O
Pitch max. Align.	A<Q: S	A<Q: S	A>Q: V	A>Q: V	A<Q: S,V,O	A<Q: S,V,O
Intensity	A>Q: S	A>Q: S	–	A>Q: V	A<Q: S,V	A<Q: S,V

The part before the colon indicates the direction (A(ssertative sentences = statements) higher [>] or lower [<] than Q(uestions)). The part after the colon indicates contrast positions [S,V, or O] in which the values significantly comply with the statement before the colon

The compilation of significant differences between statements and questions can be explained in two rather simple lines: First, in the position of the subject and the verb, thus sentence initially and sentence-medially, the differences are mainly due to more reliable effects of contrast in statements than in questions. Most cases of effects in these positions show a higher value for the statements, especially for the conditions in which the word is contrasted. And second, sentence - finally, on the objects of the sentences, the fundamental frequency is strongly increased by question encoding (a rise), independently of the position of the contrast.

6.3.10 The Relative Weight of the Parameters

In order to assess the relative weight of the previously analyzed parameters, a logistic regression analysis was run for each constituent. This investigation is interesting as investigations of prominent syllables in corpora rise evidence that for English, duration and intensity are good and sufficient predictors (Kochanski et al., 2005) while for German, duration and fundamental-frequency related parameters, especially the pitch maximum, are good and sufficient predictors of (perceptual) prominence (Wagner, 2002). Following theoretical approaches (Bolinger, 1989; Molnár, 2006; Steube, 2001), one would assume a high correlation of perceived prominence and contrast position, which is in line with the analysis of the perceived main phrase accent (see section 6.2.1). But the investigation of significant differences for six acoustic parameters has shown that there are practically no statistically pertinent effects of contrast in questions. A quest for the relative weight of the acoustic correlates will thus be restrained to statements. Nonetheless, all critical statements were included in the analysis, that means that sentences in which the perceived main accent did not fall on the intended contrast position were not excluded. The complete pattern of prosodic correlates of encoding contextual contrast, and not perceptive prominence, should be more precise in terms of aspects of speech production.

To my knowledge, there is no previous work available, which has studied the relative weight of acoustic correlates systematically in different positions of a sentence. A constituent-wise stepwise backwards logistic regression will filter the six parameters (duration, pitch maximum, mean pitch, pitch range, alignment of the pitch maximum, and intensity) down to the most effective ones. A comparison of their relative weight will give insights not only in a preferred global "German" prosodic usage of these parameters, but reveal eventual position dependent properties, too.

We start with a logistic regression of the six parameters in the sentence initial position, the first syllable in the examined sentences and the subjects of the critical utterances. Table 6.14 shows the final model of the stepwise backwards logistic regression of German statements spoken by German native speakers.

Table 6.14: Coefficients of the final model of a stepwise backwards logistic regression for sentence initial syllables in German statements by native speakers

	Estimate	Std. Error	z value	Pr(> $\|z\|$)	Significance
(Intercept)	-7.081357	1.607855	-4.404	1.06e-05	***
Duration Ratio	0.082365	0.038077	2.163	0.03053	*
Pitch maximum	0.015092	0.004403	3.427	0.00061	***

Residual Deviance Intercept-only model: 183.3; Residual Deviance presented model: 163.4; Deviance: 19.9; P(> χ): .000

We see that the pitch maximum and the relative duration of the subjects are sufficient and significant in their effects of predicting corrective contrast in subject position. As the z-values show, the weight of the pitch maximum is higher than of the duration ratio. Intensity does not carry crucial information.

If we run the same analysis for the sentence medial position, the verbs in the elicited statements, we get completely different results.

Table 6.15: Coefficients of the final model of a stepwise backwards logistic regression for sentence medial syllables in German statements by native speakers

	Estimate	Std. Error	z value	Pr(> $\|z\|$)	Significance
(Intercept)	-11.66015	2.33162	-5.001	5.71e-07	***
Duration Ratio	0.13729	0.05082	2.702	0.006899	**
Pitch maximum	-0.04699	0.02677	-1.756	0.079160	.
mean Pitch	0.07142	0.02707	2.638	0.008341	**
Pitch range	0.02595	0.01708	1.519	0.128729	ns.
Pitch max. Alignment	3.44224	1.00644	3.420	0.000626	***

Residual Deviance Intercept-only model: 183.3; Residual Deviance presented model: 124.9; Deviance: 58.4; P(> χ): .000

First, we have to state that the model needs by far more parameters to get a sufficient model. The acoustic parameter with the highest predictive value is the Alignment ratio of the pitch maximum. This confirms the descriptive pattern that a rising contour (late pitch peak), which is highly correlated with prominence in the perceptive analysis in this position, is the most effective cue also for all of the verbs under contrastive focus. Duration ratio and the

Table 6.17: Ranking of significant parameters for German statements by native speakers

Rank	Subjects	Verbs	Objects
1	Pitch max.	Alignment of Pitch max.	mean Pitch
2	Duration	Duration	Duration
3		mean Pitch	
4		Pitch max.	
5		Pitch range	
6			

other fundamental frequency related parameters (of which only mean pitch is significant) add predictive value to the model. And again, intensity does not play a role.

Looking at the sentence final position, the objects, the logistic regression analysis results in similar parameter weighting than in the sentence initial subject position. Duration and the mean fundamental frequency are sufficient cues for to model corrective contrast. The pitch related parameter again is stronger than duration. And also in the object position, intensity does not play a role in the final regression model.

Table 6.16: Coefficients of the final model of a stepwise backwards logistic regression for sentence final syllables in German statements by native speakers

	Estimate	Std. Error	z value	Pr(> $\|z\|$)	Significance
(Intercept)	-12.236665	2.271318	-5.387	7.15e-08	***
Duration Ratio	0.109983	0.034582	3.180	0.00147	**
mean Pitch	0.035240	0.007447	4.732	2.23e-06	***

Residual Deviance Intercept-only model: 183.3; Residual Deviance presented model: 129.4; Deviance: 43.9; P(> χ): .000

For an overview, the ranking of the acoustic parameters that have significant impact on the regression model in the three positions is shown in Table 6.17

6.4 Summary and Discussion

All in all, the data obtained by the short elicited utterances in the recorded dialogues confirmed the hypothesis derived from the literature.

6.4.1 The Encoding of Sentence Mode

As already predicted from the perceptive data, sentence mode in German is encoded in phrase-final position only. While statements continue a declination and end in a fall or low tone, questions consistently rise at the end of the utterances. There is no statistically significant distinction between statements and questions before the end of the sentences[11]. This is coherent with the literature, especially with the work of Batliner and colleagues (Batliner, 1989a, 1989c; Oppenrieder, 1989a). Both methods, a description of perceived contours and an investigation of properties of the fundamental frequency contribute coherently to these findings. Note that sentence mode is encoded only by the fundamental frequency, there are no significant differences between the two sentence modes in the duration ratio of the constituents and the mean intensity ratio[12].

6.4.2 The Encoding of Contrastive Focus

Correlates of corrective contrast in the three analyzed positions have to be summarized a bit more differentiated:

First, the contextual focus is highly correlated with the location of the perceived main sentence accent. Overall, for 75 % of all sentences the main accent is perceived on the constituent which carries the contrastive focus. While non-matching accents occurred rather rarely in the subject and the object-position, the error rate was higher for verbs, and especially high for verbs in echo-questions. However, the results are similar to those found with the English native speakers. For a discussion of the fact and its consequences, the reader is therefore referred to section 5.2.1.

Second, considering perceived contours on contrasted vs. non-contrasted syllables, the limits of a tonal description for correlates come to light in the quest of correlates of contrast: only in the position of the verbs, there is a high probability of change in the direction of the slope of the contour: if the verb is under contrastive focus, it is preferably realized with a rising contour, while uncontrasted verbs dominantly intoned with a fall. To a certain extent, corrective focus influences the perceived contours also in the sentence final position in statements, in the way of a higher rate of complex rise-fall contours (LHL) for strongly

[11] If there are differences, they can be explained by a stronger encoding of contrast in statements, which has nothing to do with proper acoustic correlates of sentence mode.

[12] The data on absolute duration and intensity has not been analyzed here. See the discussion of the advantages of using normalized data (ratios) in section 2.1.

accented syllables. But the sentence initial constituents contain a rise independently whether they are contrasted or not.

Nonetheless, the perceptive analysis also shows a strength: while the statistical analysis of the correlates of questions did never confirm any valid differences between contrasted and uncontrasted conditions – with the exceptions of a later alignment of the pitch maximum in contrasted verbs – the position of the perceived main accent was strongly influenced by the location of the contrastive focus. Although the steep final rise could be a strong cue for a main accent, in almost 90 % of all subject-contrasted questions, the most prominent syllable was located there. The shift in main accent is also pertinent in contrasted verbs which still gain about 40 % of correct accent assignemnt.

Third, the results of the statistical analyses of differences between contrasted vs. non-contrasted constituents can be summarized by five main points:

There are no reliable differences between contextually contrasted syllables and the same syllables in a non-contrast position in questions, but only in statements.

The pitch maximum is higher and the mean fundamental frequency is increased if a syllable is contextually contrasted in statements, independently of the position in the sentence.

Syllables which are not sentence-initial are longer and louder under contrastive focus than if they are not focused.

In sentence medial-position, the position of the peak is significantly later, too, but not in the other two positions. A significant peak delay for objects, as found in Braun (2004) thus is not replicated[13]. This can be due to the mono-syllabic constituents which were used in the experiment: there is simply not enough time-space for a further delayed pitch maximum. Another argument, raised by Braun (2004) is that subjects used a higher OR a later peak to encode focus: And the height of the fundamental frequency is clearly increased in the stimuli. It is possible that such is sufficient and more approriate to the short places of occurence in the used materials.

And finally, the analysis of the weight of acoustic correlates of contrast in statements shows that for both nominal constituents – the subjects in sentence initial position and the objects in sentence-final position, the height of the fundamental frequency is the most reliable cue for predictions of corrective contrast. In sentence medial position it is the alignment of the peak, reflecting the perceptive fall-rise distinction, which is most effective. Duration

[13] However, numerically, the peak is indeed later in the contrasted condition, it occurs after 34 % of the object, while the non-contrasted conditions peak at 24 % and 25 % of the syllable, respectively (see Table 6.10), but the difference is not approved by the statistical analysis.

adds a significant portion to all regression models of contrast, too. Although there are significant differences in the intensity of contrasted vs. non-contrasted verbs and objects, intensity is never part of significant cues in the final regression models. Hence, the findings in Oppenrieder (1989b); Wagner (2002) which states that intensity is not a reliably correlated parameter to perceived prominence is confirmed.

7 English and German: a Comparison

In chapters 5 and 6, a large number of prosodic correlates of the encoding of sentence mode and corrective contrast for the two languages spoken by native speakers have been presented. The main findings shall now be compared.

As already pointed out in chapter 2 and especially in section 2.4, the Contrastive Analysis Hypothesis is still very influential in the literature of phonological learning. Its key claim is that the learner variety heavily influenced by the native and the target language of the speaker. The learner starts to integrate properties of the foreign language into his native language system and subsequently adapts his L2 representations to the appropriate categories of both language systems separately. Hence, a comparison between the native and the target language of the learner is necessary. Patterns that differ between the languages involve an adaption process, while patterns that are identical would be used for both, the L1 and the L2.

The data obtained from both languages allows for a qualitative and, with restrictions, a quantitative analysis. Qualitatively, the tonal correlates of the perceptive contour analysis can be compared, as well as the differences in the acoustic correlates in the intra-language analyses presented in the two last chapters. The similarities in the data elicitation and the concrete linguistic materials also allow for a cautious statistical assessing of pertinent differences between English and German.

From the large number of qualitative and quantitative correlates of sentence mode and contrastive focus investigated, the following comparison will concentrate on perceptive and measured aspects that have shown effects of sentence mode and focus encoding. Properties that do not show systematic correlates to the linguistic functions in either language will not be mentioned here. Also, if acoustic properties are very much correlated, like the pitch maximum and the mean pitch, a choice was made in the comparison to avoid repetitions.

7.1 Comparison of Tonal Correlates

First, a qualitative comparison of the perceived tonal contours in the two languages shall be presented. It refers to the analysis presented in section 5.2.2 for English native speech and in section 6.2.2 for German[1].

Three main observations shall be emphasized. First, general differences can be pointed out. On the one hand, English native speakers use more different tones in each of the three positions. For example, a complex rise-fall (LHL) is found in English native speech in the subject and verb position, but not in German. It occurs mainly in cases of contrast on the respective syllable. In German this complex combination is found only in the object position of statements: But there the percentage of the LHL tone is even larger than for English (40 % compared to 23 %). On the other hand, simple level low or high tones (L / H) are more frequent in English than in German. This is observed for instance for a high level tone (H) in the subject position, which is used in English in 28 % of all statements and 37 % of all questions, but only in 12 % and 7 % in the same sentence modes in German. Similar can be stated for high level tones in the verb position of questions and low level tones in the object position of statements. To put these findings together, one can state that in German native speech, the preference of uni-directional falls or rises (HL or LH) is more pronounced than in English, which more frequently uses either simple level tones or a complex rise-fall.

Second, considering the tonal correlates of sentence mode encoding: except from the higher frequency of of level tones (H / L) in English, there are no striking differences between English and German in the usage of tones in the subject and the verb position. But, as already emphasized repeatedly, in the object position, English does not distinguish the sentence mode by a different endpoint of the contour, but uses L, HL, or LHL in three quarters of the sentences, irrespective whether they are (echo-)questions or statements. German instead has almost exclusive preferences for sentence mode encoding on the final syllable: questions end in a high (LH, H contours) and statements low (HL, L, LHL contours).

And third, a comparison of the tonal correlates of contrast gives the following results: In the subject position, none of the languages is showing strong differences in the perceived contours between contrasted and non-contrasted syllables. In the verb-position, both speaker groups increased the percentage of rising or high tones in case of contrast, this feature is more pronounced for German compared to English (German: 69 % of the statements and 25 % of the questions; English: 48 % of the statements and 17 % of the questions). But the complex

[1] For more details, the reader is especially referred to Tables 5.4, 5.5, 5.6 for English and Tables 6.2, 6.3, and 6.4 for German.

LHL tone appears only for English verbs under contrastive focus (8 %). Sentence-finally, one can point out the higher percentage of rising-falling (LHL) contours in German (40 %) than in English (23 %) in cases of contrast on these syllables. All in all, the perceived contours as a correlate of the encoding of contrastive focus are rather similar between the languages, but the choice of a contour alternating between contrasted and non-contrasted syllables seems to be more pronounced for German.

7.2 Comparison of Acoustic Correlates

This section concentrates on acoustic correlates of sentence mode and focus which have been statistically tested within the languages in section 5.3 and 6.3.

7.2.1 Comparison of the Encoding of Sentence Mode

The data shows clearly that the encoding of sentence mode is a domain of the fundamental frequency[2].

One of the most striking differences between the utterances of the English and the German native speakers is the pitch register for the echo-questions. While the English questions are produced with a higher fundamental frequency throughout the whole sentence, the German interrogatives do not differ from declaratives until the last syllable. This phenomenon is very clearly shown in Figure 7.1

We see that in sentence initial position, pitch is already about 60 Hz higher for English questions than statements; in sentence medial position, the difference increases to about 100 Hz for uncontrasted verbs and about 70 for contrasted ones. In sentence final position, the mean difference is still more than 40 Hz. In German, there is no difference in fundamental frequency until the last syllable. There, the mean pitch is ca. 100 Hz higher in questions on average.

A two-way repeated measures ANOVA with a between subjects factor of Language (levels: English and German) and the within-subject factor of Sentence Mode (levels: Question and

[2]One may argue that in the object position, the intensity is also significantly higher in questions compared to statements. But first, this is only true if the object is not contrasted (so only statements with contrast on the subject and the verb have a smaller mean intensity ration than statements). And second, the slightly higher intensity can be caused by the heavier pitch movement and the increased fundamental frequency in that position.

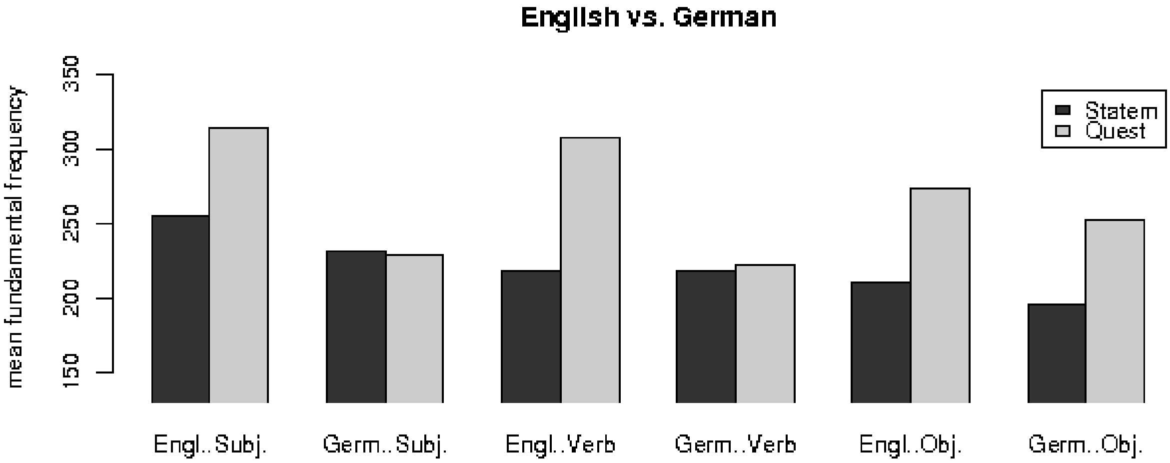


Figure 7.1: Average of the mean fundamental frequency per constituent for English and German statements vs. questions

Statement) confirms these observations[3]: English uses an overall higher pitch than German in the subject and verb position as the significant main effect for Language for the mean fundamental frequency shows[4], but the languages do not differ in the object position. A similar pattern is found for the main effect of Sentence Mode which approves generally higher statements and questions in all positions[5]. The crucial interaction is significant at a 5 % level in the first two positions, on the subjects of the sentences with $F1(1,30) = 53.86$, $p < .001$; and on the verbs with $F1(1,30) = 51.73$, $p < .001$ and approves the higher register for questions in English but not in German. In the object position, the interaction Language*Sentence Mode only approaches significance with $F1(1,30) = 3.697$, $p = .064$. However, one can state the tendency that Sentence Mode affects the German sentences more than the English.

The other difference is the direction of the final syllable of a question: it is dominantly falling or low for English (> 75 %), but almost always rising or high for German (> 95 %). The position of the peak in each syllable can show that fact very clearly (Figure 7.2)

Submitting the data to a mixed repeated measures ANOVA as above, the following results can be shown: In the subject position, we find a significant main effect for Language

[3]The statistical analysis of between languages comparison is limited to an across subjects-aggregation (F1 analysi) because the items were similar (for example only mono-syllabic words were used), but not exactly the same. For example, different verbs had to be used for each language, and also a considerable part of the proper names in the subject and object position had to be changed to avoid an overload of foreign names.

[4]In the subject position: $F1(1,30) = 28.21$, $p < .001$; and $F1(1,30 = 18.47$, $p < .001$ in the verb position

[5]In the subject position: $F1(1,30) = 45.90$, p ¡ .000; in the verb position: $F1(1,30) = 61.93$, $p < .001$; and in the object position: $F1(1,30) = 179.21$, $p < .001$.

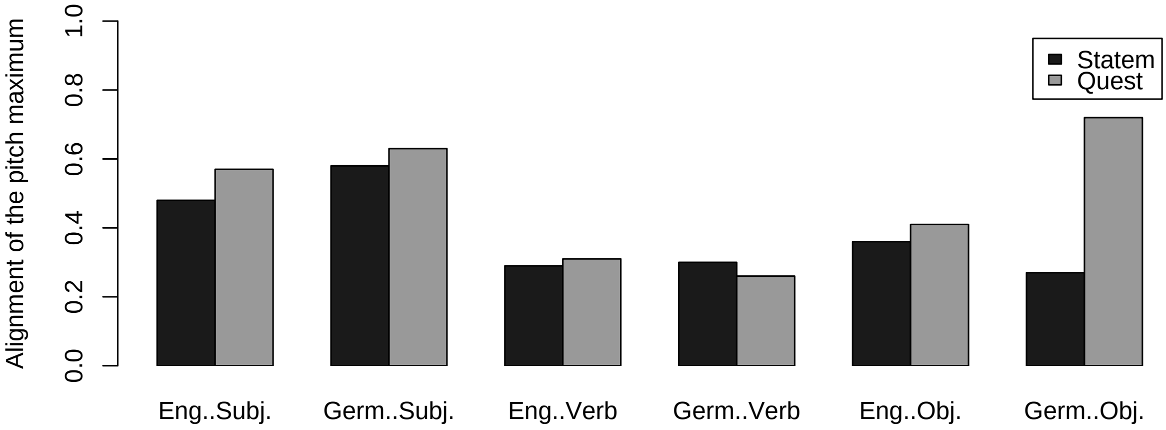


Figure 7.2: Alignment of the pitch maximum for English and German statements vs. questions

(F1(1,30)5.26, $p < .05$, indicating a later peak in German than in English. It can be explained by the consistent rises in German (later peaks), but the more frequent high level onsets of the sentences in English (earlier peak). Sentence Mode is also significant (F1(1,30) = 9.32, $p < .01$) indicating more consistend and pronounced rises for questions overall. The interaction Language*Sentence Mode is not significant. In the verb position, no significant effects are found. But in the object position, the main effect for Language is significant (F1(1,30) = 11.26, $p < .01$) confirming a later peak for German than for English sentences, and the significant main effect for mode approves the later peak in questions (F1(1,30) = 75.23, $p < .00$). The crucial interaction finally shows the stronger effects on the peak position in German than in English (F1(1,29)= 44.48, $p < .001$).

To conclude, the repeatedly stated differences in the encoding of sentence mode between English and German are strongly confirmed: English uses a higher register for the elicited questions throughout the sentences, but German increases the only fundamental frequency at the end of the questions, in tendency stronger than English. The different final contours can be traced by the position of the peak, and clearly, sentence mode encoding has much stronger effects on German objects than on English.

7.2.2 Comparison of the Encoding of Focus

With regard to acoustic correlates of focus, we find several aspects: First of all, none of the languages shows convincing correlates of focus in questions.

Considering the statements only, we see that for both languages duration is systematically increased if a constituent is contrasted. This effect seems to be stronger for English (ca. 20 % longer relative than uncontrasted syllables proportionally to the rest of the sentence (duration ratio)) than for German (ca. 10 %). However, submitting the data to a mixed repeated measures ANOVA with Contrast (contrasted vs. non-contrasted words) as within and Language (English vs. German) as the between subjects factor, the interaction Contrast*Language is not significant in any position of the sentence[6].

A similar picture is found in comparing the usage of pitch maximum to encode corrective contrast.

The pitch maximum is increased for about 50 Hz for contrasted verbs and subjects in both languages (Figure 7.3).

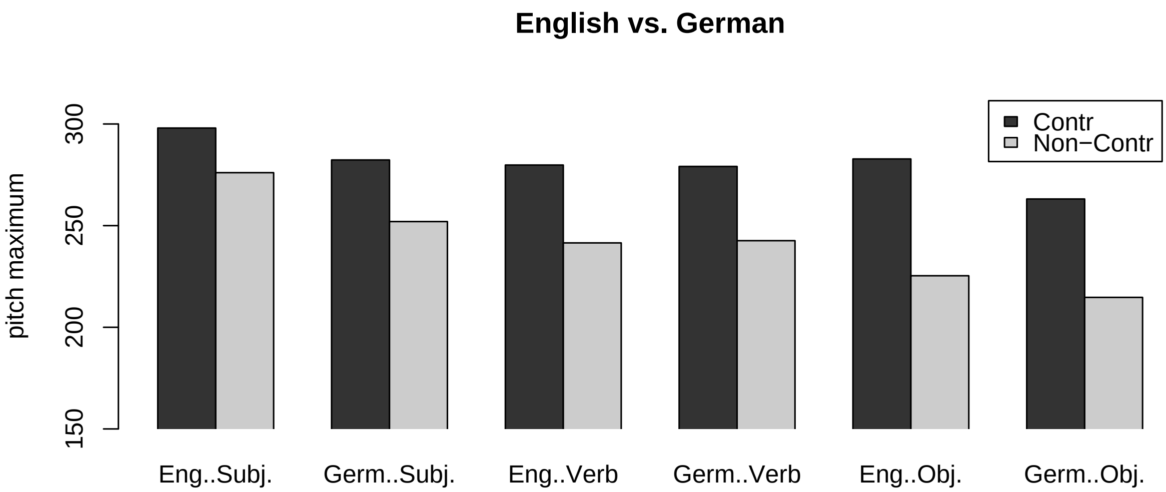


Figure 7.3: Pitch maximum per constituent for English and German statements. The non-contrasted bars represent the mean of both non-contrasted conditions

Note, that, especially in the subject position, this result differs from the data in the analysis of English only (see section 5.3.3): There, no significant main effect for Contrast, and no significant interaction Contrast*Sentence Mode were found. However, restraining the data to the statements, and averaging the non-contrasted conditions (contrast on verbs and contrast on objects in that case) as done in this comparison, contrasted and non-contrasted subjects differ significantly. Therefore, no Interaction between Language and Contrast, and thus no difference in the usage of the pitch maximum to encode corrective contrast between the

[6]The ANOVAs were run in an across-subject aggregation only (F1), as the items (sentences) were slightly different. Note that there is a significant main effect for contrast in all three positions (contrasted words are longer), and there is no difference between the languages in the relative length of the constitutents.

languages was found[7].

As shown with the assessment of the weight of the constituents, intensity plays a stronger role for the distinction between focused and non-focused constituents for English than for German (see sections 5.3.10 and 6.3.10). It is not the case that German does not increase intensity at all to encode contrast, but as the stepwise regression analysis shows, English uses it in a more reliable way. Figure 7.4 displays the mean intensity ratio for contrasted constituents against the averaged values of the non-contrasted words. Clearly, in all positions the difference between contrasted and non-contrasted syllables is larger for English than for German. However, in a joint analysis with a mixed repeated measures ANOVA like described above, the interaction Language*Contrast is significant at a 5 % level only in the object position ($F1(1,30) = 6.29$, $p < .05$). In the subject and verbs position of the sentences, the interaction only reaches around 10 % of error proability (on subjects: $F1(1,30) = 2.820$, $p = 0.103$; on verbs: $F1(1,30) = 3.17$, $p = .085$).

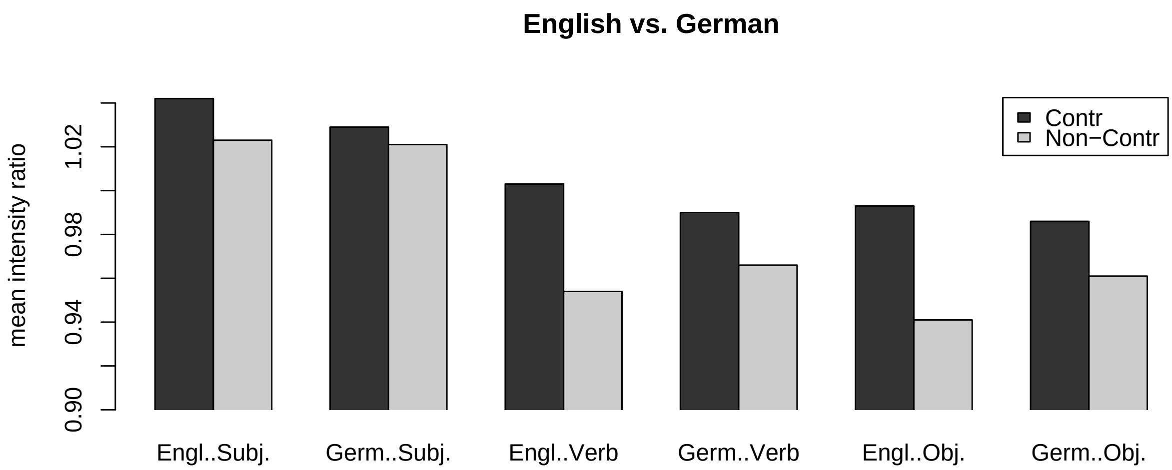


Figure 7.4: Mean intensity ratio per constituent for English and German statements. The non-contrasted bars represent the mean of both non-contrasted conditions

[7] There are no significant differences in the usage of pitch range to encode contrast between the languages, either: both languages show an increased the pitch range for contrasted syllables compared to non-contrasted, but no significant interaction Language*Contrast. The same holds for the alignment of the peak: the average of the peak position is later in contrasted than in non-contrasted conditions, independently of the position, and the languages differ in the general position of the peak on the first and the last syllable of the examined statements, but they do not differ significantly in the usage of the peak alignment for the encoding of corrective contrast.

7.3 Summary and Conclusions of a Contrastive Analysis

To conclude the investigation of differences between English and German spoken by native speakers one can refer to two main points:

First, considering the encoding of sentence mode, the two languages differ strongly with reference to the final contour in the echo-utterances: English prefers a low ending tone while German uses almost exclusively a high end for the questions. In this case, if one adopts the Contrastive Analysis perspective, if German learners of English would use a final rise to encode the questions, it would be a clear case of transfer from their native language pattern, otherwise they would show a learning effect. Furthermore, the interaction of Language*Sentence Mode for the height of the fundamental frequency in the first two positions of the examined utterances clearly points to the higher overall pitch level in native English than native German questions from the beginning of the sentences on. The learners could, in case of transfer, use a similar mean pitch than in statements, or show an increased level of F_0 if they have learned how the speakers of the target language encode this kind of questions.

Second, the findings of the inter-language difference for contrastive focus encoding shall be summarized. With regard to the perceived contours, except of a more pronounced rate of rising or high tones on contrasted verbs in German than in English, one should certainly note the higher frequency of the complex rise-fall (LHL) on corrected objects in statements in English than in German. Their rate is rather the double (40 % against 23 %).

As for the acoustic correlates, it is clearly a fact that corrective contrast does not take place in a statistically attested manner in the echo questions in either of the languages. Taking into account the statements only, the different final models of the stepwise backwards regression models show that German is relying mostly on duration and fundamental frequency, and English on duration and intensity, especially at the onset and the end of statements. In the verb position, both languages strongly use the position of the peak to encode contrast, but also duration and the other fundamental frequency related parameters: contrary to German, a good model for English includes the intensity, too.

However, testing differences in the usage of acoustic parameters in a joint analysis comparing the correlates of focused and non-focused words in statements, the results indicate that English and German speakers increase the values for contrasted syllables compared to non-contrasted to a similar extent. The only statistically sound difference between the languages is found for the mean intensity ratio on the objects of the sentences: English increases the intensity more than German. Although the differences in the mean intensity ratio indicate a similar effect for the other positions of the utterances, it could not be approved statistically.

8 English Spoken by German Native Speakers

Now that the patterns of the target language (English) and the native language (German) are assessed and analyzed in detail, we can examine the prosodic correlates of sentence mode and corrective contrast for German second language learners in English. The last chapter summarized the main differences between the languages. Conflicts between the languages are especially expected in the location of an increased fundamental frequency to encode questions, and the role of intensity compared to fundamental frequency to mark contextually contrasted syllables.

8.1 Hypotheses

As evaluated in chapter 2 and 3, we have reason to test four different hypotheses.

8.1.1 H_1: Perfect Acquisition

If we assume that the language learners have acquired the prosodic correlates of sentence mode (statements and questions) and the key parameters of encoding corrective contrast, we predict that the L2 patterns do not differ from the native English utterances. We especially expect a higher register for the echo-questions which should be predominantly realized with a final fall. For contrast encoding we expect a reflection of the higher weight of intensity in the encoding, and decrease of the weight of the height of the fundamental frequency in all positions. The role of the alignment of the pitch maximum on verbs is maintained. We do not expect significant effects for contrast in questions. Transfer cannot be excluded for the last two points as there are no differences between the two languages.

There are two arguments in favor for this hypothesis: a) the language learners were proficient learners of English with pronounced experiences in an active oral usage of the target language in a native environment and b) the prosodic correlates are, at least in the case of the higher but falling questions, easily perceivable and frequent in the target language, but common in the native language, too. As seen in the acoustic analysis, also German has a higher intensity on contrasted constituents, but it is less reliably connected to focus. As for this more hidden weight shift of acoustic correlates, one can refer to Mennen (2004) who reported one case which perfectly applied the different positions of the peak for both, the target and the native language.

On the other hand, this hypothesis is implausible: All literature available on second language prosody reports influences of the native language of the learners, just the degree of interferences is different.

8.1.2 H_2: Total Transfer

The second hypothesis assumes total or almost total transfer. This point of view is defended by most older, impressionistic literature, sometimes explicitly (Gårding, 1981) or implicitly (e. g. Lepetit, 1989). It would assume that there is no learning process concerning prosodic patterns. In the experiment, this hypothesis will predict that the prosodic correlates in L2 English are the same as the ones revealed in the speech of German by German native speakers. More precisely, one would assume that questions are not different from statements before the phrase-final constituent, but carry a steep rise on the object. For contrast encoding, only pitch related parameters, together with duration would have a significant predictive value. No effects of contrast are expected in questions. The prediction of a total transfer is nonetheless rather rare in the studies: most of the previous investigations explain their data by partial interferences, which is the next hypothesis.

8.1.3 H_3: Partial Transfer from L1 to L2

The third hypothesis assumes partial transfer from L1 to L2. This point of view is adopted by most of the studies in second language prosody research. Interferences also account for the perception of a foreign accent (e.g. Els & Bot, 1987; Jilka, 2000; Major, 2001) by intonation. One can distinguish two different kinds of L1 - L2 transfer:

The first is transfer of categories: some target language categories are acquired while others are not. In an analogy to segmental speech production learning (Flege, 1987, 2003) this would suggest that the L2 speakers of English have acquired "new" aspects of prosody better than "similar", mainly because there must be a resetting for existing category boundaries, while new distinctive features are added to the existing (L1-based) system easier. Adopted to the design of the Experiments, this hypothesis would predict that German native speakers have learned to connect falling contours to echo- questions: this is a new connection. A "similar" prosodic property would be the different weight of the acoustic dimensions: Intensity and fundamental frequency are correlated, and both, the target language and the native language make use of it. The difference is the higher degree of reliability of intensity in English and of fundamental frequency in German. But subjects may have difficulties in perceiving that, as one could assume a certain trade-off between the different parameters.

There is a second dimension of interferences, a quantitative one: one can assume that the L2 learners try to imitate the patterns of the target language, but are only able to approach them. This phenomenon is found in aspects of quantitatively assessable features of phonetic learning, like VOT for plosives (Flege, 1980; Flege & Efting, 1987) and vowel formants (Flege, 1997; Flege, MacKay, & Meador, 1999). It is also found for prosody, e.g. the alignment of the peak (Mennen, 2004). This suggestion would predict for example a slightly higher fundamental frequency throughout the whole time of (echo-) questions compared to statements; somewhere between the clearly higher register of the target language and the shared register of the native language of the learners. In the case of the different weight of the parameters, partial quantities interferences would predict a higher reliability of the intensity correlates of contrast, without reaching the amount of intensity variation of the English native speakers.

Yet, both dimensions of interferences still assume that there is no pattern in L2 speech that does not origin from another source than the two languages, the native and the target language of the learner.

8.1.4 H_4: L2 Prosody as an Emerging of Universals

The idea that second language prosody might reflect intonational universals is not new. In fact, this account is scattered in the literature of L2 speech production (e. g. Jilka, 2000). The problem with these cited studies is that the hypothesis of a potential interference of universals is never elaborated, and applied only to patterns which "have to remain inexplicable" (Lepetit, 1989) in an interference-based account. This is not satisfying: How can one explain most of the found patterns by interlanguage interferences, but refer to (often not precisely defined) universals, if there are correlates that cannot be accounted for by transfer

? Would it be possible to assume that all second language patterns stem from universal preferences, some are distinguishable from interferences, some are not ?

So, the first problem is to define predictions from presumed "Universals". As discussed in section 2.6, Bolinger's view of prosody as a medium to encode human emotions and attitudes predicts similar functions of prosody in all languages. But he denies the existence of concrete, similar representations of tones for certain functions across languages and speakers. There is the tendency of a higher fundamental frequency for questions, at the end or for the whole utterance, but the place of an accent remains unpredictable.

The only theory available that makes testable predictions about the actual realization of communicative functions like sentence mode and corrective contrast is Gussenhoven's 'Biological Codes' theory (Gussenhoven, 2002, 2004). It has been explored systematically for L2 perception (A. Chen, 2003, 2005; A. Chen et al., 2001; A. Chen, Gussenhoven, & Rietveld, 2002; Gussenhoven & Chen, 2000), and supplied a reasonable basis for L2 speakers' reactions which were not explainable by transfer.

What would it predict for sentence mode and contrast encoding ? The encoding of sentence mode is a domain of the Frequency Code. The person who asks is less powerful than the person who is asked to reply and provide information, thus fundamental frequency should be higher, especially towards the 'edges of an utterance' (Gussenhoven, 2002). So L2 speakers should encode the sentence mode by a final rise , and they may use a higher register, too.

In the comparison above (chapter 7), we saw that both, English and German are in line with the Frequency Code, but in different positions of the sentence. German does not distinguish questions and statements until the final syllable but uses a high rise in almost all the cases. English instead has an overall higher register for the echo-questions. So, a distinction between interferences between the native and the target language and a recourse on the Biological Code will be hard to identify: If the learners use a final rise, this can be also accounted for by transfer, if they use a higher register, it can be explained by perfect acquisition.

The encoding of contrast is a domain of the Effort Code. The Effort Code provides rather precise predictions: The contrasted elements in the sentences will be longer and higher. Extending the original propositions to intensity, contrasted syllables should also be louder, because more physiological effort is needed. All three acoustic dimensions are correlated through the properties of human sound production. From the statistical analysis of English we know that, except from pitch in the initial position, all three dimensions are used. But for the edges of the sentences, the increase of duration and intensity is sufficient for the encoding of corrective contrast. In German, in these two positions of the sentences, pitch height and duration are sufficient predictors of intended corrective contrast. In the verb position, both languages use a reversion of the direction of the pitch contour: a rise for contrastive and a fall for non-contrastive elements. For English, but not for German, an increased intensity is still necessary for a good predictive model in that position. Note that both languages did not distinguish contrasted from non-contrasted syllables in questions. Although it would be easy to rule out transfer (the learners rely dominantly on fundamental frequency like in their native language) or perfect acquisition (they rely dominantly on intensity), again it will be hard to disentangle partial quantitative interferences: If the learners consistently use all three dimensions (duration, fundamental frequency and intensity), it can be due to a transfer of the important role of fundamental frequency and the acquisition of the higher value of intensity in the target language. The Biological Code Theory would be ruled out if the learners did NOT use all dimensions of (physiological) effort. A strong cue for the Effort Code would be an emergence of correlates of focus in the echo-questions: they contain the contrast value of the statements, but none of the native speaker groups has distinguished contrasted from non-contrasted elements in questions.

8.2 Eliciting Correlates of Sentence Moe and Corrective Contrast

To test the hypotheses described above, an experiment with German learners of English speaking English was run.

8.2.1 Subjects

Sixteen female native German speakers took part in the experiment. Eleven of them originated from the region of Saxony, Saxony-Anhalt, or Thuringia. Three came from Northern Germany[1] and two from other places[2]. Their age ranged between 22 and 32 (mean: 26 years). All have spent more than six months (range: 6 to 18 months, mean: 10.25) in sequence in an English speaking environment during the years before the recordings. All reported recent frequent visits to English speaking countries and assessed themselves as fluent speakers of English. Ten of the subjects studied or had studied English at a University, the others reported high personal or professional interest of mastering the English language. Learner

[1]Bielefeld, Münster, Lüneburg
[2]Southwest, Vienna

biography, self assessment and impressions of the experimenter from a short introductory conversation in English preceding the experiment give reason that they all were proficient learners of English.

No speaking or hearing impairment was encountered. They were naive to the purpose of the experiment and paid for their participation.

8.2.2 Materials and Design, Procedure

The same materials and design as in experiment 1 (English native speakers speaking English; see chapter 5, especially sections 5.1.2 ff.), were used. The 96 dialogues were divided into four lists in a Latin square design, so that every speaker spoke each sentence in one condition.The procedure was the same as in experiment 1 and 2: subjects were invited pairwise for the recording sessions which took place in a sound proof booth at the University of Leipzig. The utterances were recorded with a Neumann TLM 103 high quality microphone. The instructions, together with a short conversation preceding the experiment, were given in English. Additionally to the instruction to "act" the dialogues, they were ask to check their parts for unknown vocabulary or unclear contextual meanings. But none of the subject had questions concerning this issue.

All preprocessing was equivalent to experiment 1 and 2: the extracted sentences were tagged for word boundaries, perceptual information was added, and errors by the PRAAT software-package were eliminated.

8.3 Perceptive Analysis

8.3.1 Perceived Accent

First, the correlation between contextually induced contrastive focus and the perceived main accent of the sentences is presented. Table 8.1 shows the percentage of sentences with the main accent on the collectively contrasted constituent.

Overall, in 83 % of all utterances the main accent was perceived in the postion of corrective contrast. While for statements, the number of "incorrect" accent placements amounts to less than 10 %, more than 25 % of all questions had the main accent in another position than the intended. While contrastive accent on subjects and objects rather reliably received the

Table 8.1: Percentage of sentences with perceived main accent in contrastive focus position by sentence mode for English by German native speakers

	CF S	CF V	CF O	mean
Statement	94	85	96	91.7
Question	79	52	90	73.7
mean	86.5	68.5	93	82.7

phrase accent if they were under contrastive focus, verbs in questions received the contrastive accent only in about half of the cases[3].

The overall performance of focus-corresponding accent assignment is similar to the English native speakers with the same materials and even slightly better than the German utterances. The main reason is the better performance for the verb-contrasted questions.

As this study investigates the prosodic encoding of (contextual) corrective contrast in speech production, no utterances were excluded from the following analyses, that means that both, focus-accent corresponding and mismatching sentences were examined.

8.3.2 Distribution of Tones

Equivalent to Experiments 1 and 2, tones were assigned to every constituent following a simplified contour transcription. Each syllable was labeled whether a high (H), a low (L), a rise (LH), a fall (HL) or a combination of these was perceived while listening to it. The analysis was equivalent to the one for the native control group; more details of the principles of the analysis can be found in section 5.2.2.

Tones in Sentence Initial Position

First, the results of the perceptive tone assignment in sentence initial position, the subjects of the sentences, are presented in Table 8.2.

[3] An analysis of the items did not reveal any systematic pattern. Statistically, an exclusion of items in which contextual contrast and perceived accent never matched, was not reasonable as each item was spoken only twice. But systematical effects from lexical biases, e.g. vocabulary problems, can be excluded as there was no sentence that contained errors in all conditions.

Table 8.2: Distribution of perceived tones in subject position for English spoken by German native speakers

		H	HL	L	LH	LHL
statem.	CF S	6.3	4.2		83.3	6.3
	CF V	4.2	25.0	14.6	56.3	
	CF O	18.8	6.3	2.1	72.9	
mean (statem.)		9.7	11.8	5.6	70.8	2.1
quest.	CF S	2.1	2.1		95.8	
	CF V	2.1	18.8	2.1	75.0	2.1
	CF O	6.3	8.3		83.3	2.1
mean (quest.)		3.5	9.7	0.7	84.7	1.4
mean (overall)		6.6	10.8	3.1	77.8	1.7

Contrast conditions are abbreviated as Contrast on the subjects (CF S), verbs (CF V), and objects (CF O) of the sentences. The numbers indicate percentage per condition (row of the table).

We see that, irrespective of the contrast position and sentence mode, the large majority of the sentences start with a rising (LH) contour. In case that the subject is under contrastive focus, the proportion of rising contours is the highest within the two sentence modes. This and the notable proportion of complex rise-falls in statements can be related to effects of focus on the sentence initial position. Another correlate of corrective contrast is not locally relevant, but prepares the pattern in the following constituent, the verb. For the conditions of corrective contrast on verbs (CF V), we find a higher proportion of falling or low contours, already preparing the dominant rising contour on contrasted verbs.

As for sentence mode, one can see a proportion of rising contours which is still higher for questions than for statements (a mean of 84.7 % compared to 70.8 % of all sentences).

Tones in Sentence Medial Position

Table 8.3 presents the perceived tones in the sentence medial position of the elicited target sentences, the verbs.

Table 8.3: Distribution of perceived tones on verbs for English spoken by German native speakers

		H	HL	HLH	L	LH	LHL
statem.	CF S		93.8		6.3		
	CF V		29.2	2.1		62.5	6.3
	CF O		89.6		4.2	6.3	
mean (statem.)			70.8	0.7	3.5	22.9	2.1
quest.	CF S	12.5	85.4		2.1		
	CF V		50.0			50.0	
	CF O	4.2	87.5		2.1	6.3	
mean (quest.)		5.6	74.3		1.4	18.8	
mean (overall)		2.8	72.6	0.3	2.4	20.8	1.0

Contrast conditions are abbreviated as Contrast on the subjects (CF S), verbs (CF V), and objects (CF O) of the sentences. The numbers indicate percentage per condition (row of the table).

The dominant contour in the verb position is a fall: more than 70 % of all utterances were realized with a HL contour. There is only one exception: if the verb is under contrastive focus, it is realized dominantely with a rising contour[4]. The strong preference of a LH tone on contrasted verbs is still increased if we consider utterances with corrective focus and the main accent on the verb: 73 % of all statements and 96 % of all questions carried a rising accent, then. The LH contour thus is an important cue to contrast on the verb, and clearly distinguishes the accented syllables from the non-accented in that position.

As for correlates of sentence mode, one should mention the occurrence of a high level tone (H) in questions, but not in statements.

Tones in Sentence Final Position

In the sentence-final position, the objects of the sentences, one can expect strong correlates of both factors: sentence mode should induce a rising contour for questions, but a falling or low tone for statements. As the object is also assumed to be the default position of the sentence accent, it will be interesting whether contextual contrast triggers stronger cues. The results of the perceived contours annotation are presented in Table 8.4.

[4]In questions, falling and rising contours account for half of the verb-contrasted utterances each.

Table 8.4: Distribution of perceived tones on objects for English spoken by German native speakers

		H	HL	HLH	L	LH	LHL
statem.	CF S	4.2	58.3		14.6	18.8	4.2
	CF V		77.1	2.1	4.2	8.3	8.3
	CF O		52.1	2.1		4.2	41.7
mean (statem.)		1.4	62.5	1.4	6.3	10.4	18.1
quest.	CF S	6.3	8.3	4.2		75.0	6.3
	CF V	4.2	14.6	2.1		72.9	6.3
	CF O	2.1	4.2	2.1		89.6	2.1
mean (quest.)		4.2	9.0	2.8		79.2	4.9
mean		2.8	35.8	2.1	3.1	44.8	11.5

Contrast conditions are abbreviated as Contrast on the subjects (CF S), verbs (CF V), and objects (CF O) of the sentences. The numbers indicate percentage per condition (row of the table).

We see that sentence mode is quite clearly encoded by an inverse direction of the final pitch movement: statements are mostly terminated with a fall (mean 62.5 %), while the large majority of questions receives a final rising contour (79.2 %). A low level tone (L) is not perceived in questions at all, while it is found in at least 14 % of the statements with contrastive accent on the subjects (CF S).

As for contrast, one can clearly state a higher chance that a contrasted object receives a complex rise-fall (LHL) if it is under contrastive focus: more than 40 % of the CF O statements contain this complex and time consuming contour. Contrasted objects in questions only show less other tones than a rise, the following analysis of the acoustic correlates will show whether there is a quantitative difference.

But first, for a better comparison to Experiment 1 and 2, a concatenated table of falling or low contours (L, HL, LHL) vs. rising or high contours (H, LH, HLH) is presented in Table 8.5.

We see that the low/falling vs. the high/rising proportions of perceived tones is inverse for statements and questions. While statements have a low or falling contour in 87 % of all recordings, questions only go down in 14 %. For questions, this proportion is exactly the opposite. So, the inversed proportion of high/rising and low/falling for the two sentence modes

Table 8.5: Phrase final low/falling vs. high/rising tones by sentence mode (summed percentages) for English by German native speakers

	Statements	Questions
low/falling	86.9	13.9s
high/rising	13.2	86.2

is conform with the findings for the German control group (Experiment 2) and not with the equal proportions of low/falling for both sentence modes for the English one (Experiment 1). But it is not the almost exclusive connection between the falling statements and rising contours as found in native German.

8.4 Acoustic Correlates

The statistic analysis contains the same six parameters as for Experiment 1 and 2. Duration ratio, pitch maximum, mean pitch, pitch range, alignment ratio of the pitch maximum, and intensity ratio are examined for each constituent. For more details about the parameters the reader is referred to section 5.3 and especially Table 5.8.

All data were submitted to a constituent-wise within-subject ANOVA with the factors Contrast (three levels: contrast on subjects, verbs, and objects) and Sentence Mode (two levels: statements vs. questions). Pairwise Bonferroni-adjusted comparisons to test significant differences between positions of the corrective contrast within the sentence modes or between the sentence modes within contrast positions.

8.4.1 Overview of the F0 Contours

Before submitting the acoustic correlates of contrastive focus and sentence mode encoding to a detailed, constituent-wise analysis, we should get an idea of the observed fundamental frequency contours over the whole utterances. Figure 8.1 depicts the time normalized mean fundamental frequency for each quarter of each constituent.

At a first glance, the overall pattern of the fundamental frequency resembles German native speech (see Figure 6.1). All sentences start with a frequency between 225 and 250 Hz and end either rather high around 350 Hz in the case of questions (grey lines) or low around

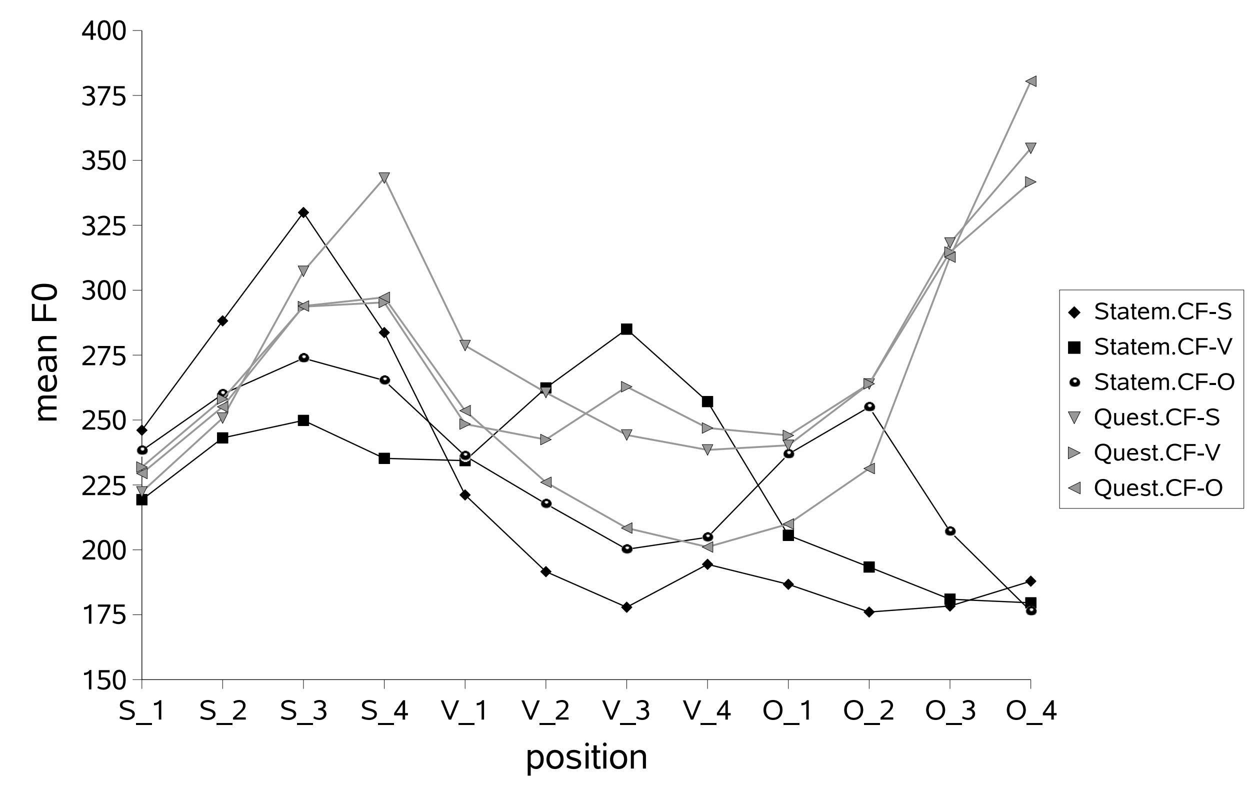


Figure 8.1: Time normalized fundamental frequency contours for English by native speakers of German

175 Hz in the case of statements (black lines). Questions seem to be slightly higher in all positions, unless the constituents in statements carry contrastive accent.

Contrasted syllables in statements exceed the non-contrasted counterparts by far, around 70 Hz or more. For example a contrasted subject (CF-S) peaks around 330 Hz, while the highest non-contrasted statement (contrast on objects: CF-O) peaks at about 260 Hz. In the subject position, the strong excursion approaches the highest question condition (Quest.CF-S), in the verb position they contours for the statements peak even higher than all questions.

The difference between contrasted and non-contrasted constituents is smaller for questions, although the maximum of the contrasted conditions always exceeds the non-contrasted conditions for that syllable.

To sum up: the German learners of English clearly use a strong final rise to distinguish questions from statements, but the questions are somewhat elevated in all of the three positions. We see strong pitch excursions for contrasted syllables in each position of the statements, and some lower but systematic elevation for contrasted syllables in questions. Whether this observations can be confirmed by a statistical analysis will be tested now.

8.4.2 Duration Ratio

The first parameter is the duration ratio. It normalizes over speech rate, thus it does not unveil potentially slower utterances in general, which is assumed to be the case for second language speakers. But the overall speech rate is not of interest in the case of an investigation of acoustic correlates of contrastive focus and sentence mode: the relative lengthening within a speaker community shows whether this parameter is used to encode these functions. Table 8.6 shows the means, the boxplots in Figure 8.2 inform about the distribution.

Table 8.6: Mean duration ratio for English by German native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	31.76	32.24	32.00	30.83	30.31	30.57	37.42	37.46	37.44
CF V	27.10	28.23	27.66	36.18	33.54	34.86	36.73	38.24	37.48
CF O	26.94	29.43	28.18	30.13	30.62	30.38	42.94	39.95	41.44
$\overline{X}_M$	28.60	29.97	29.28	32.38	31.49	31.93	39.03	38.55	38.79

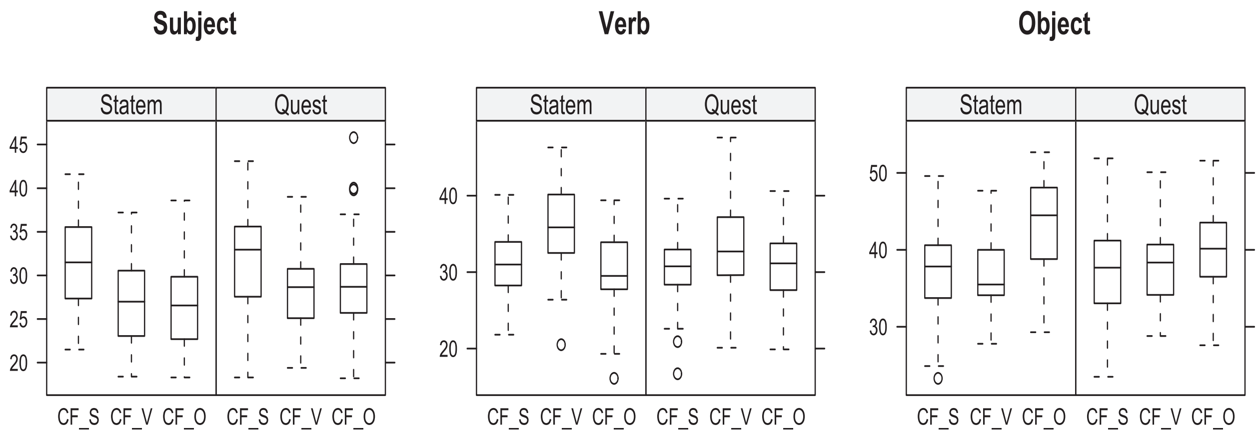


Figure 8.2: Duration ratio for English by German native speakers for each constituent by sentence mode and contrast position

We see that each constituent under contrastive focus is longer than its non-contrasted counterparts. This holds for both sentence modes, statements and questions. The differences generally are larger in declarative mode, but in the sentence initial position (subjects) and the medial position the relative difference to the nearest non-contrasted position is almost 2 percent for the questions; for the objects in interrogatives, contrast does not evoke such strong lengthening due to contrast.

The relative duration of the constituents in statements and echo-questions is similar. The grand means ($\overline{X}_M$) would indicate longer subjects in questions, but slightly shorter verbs and objects in declarative mode.

On the subjects of the sentences, the statistical analysis reveals a significant main effect for Contrast ($F1(2,30) = 22.03$, $p < .001$; $F2(2,46) = 16.10$, $p < .001$. The Post-Hoc Bonferroni-adjusted pairwise comparisons confirm that contrasted subjects are significantly longer than non-contrasted, and that this effects holds for statements AND for questions. The mean duration ratio is higher for questions than for statements, but the main effect for Sentence Mode is significant in the item-based analysis only: $F1(1,15) = 3.06$, p = ns.; $F2(1,23) = 5.22$, $p < .05$. The interaction Contrast* Sentence Mode is not significant: $F1(2,30) = 1.58$, p = ns.; $F2(2,46) = 1.62$, p = ns.

In the verb position, the sentence medial syllable of the sentences, we find a significant main effect for Contrast: $F1(2,30) = 31.73$, $p < .001$; $F2(2,46) = 45.82$, $p < .001$. The Post-Hoc test shows that in statements, contrasted verbs are longer than verbs in both other contrast positions; for questions only the item-based analysis approves a higher value for the CF V condition than both other conditions, while the subject-based analysis confirms a significant longer duration against the CF S condition only. The main effect for Sentence Mode reaches significance in the item-based analysis only: $F1(1,15) = 2.90$, p = ns.; $F2(1,23) = 4.86$, $p < .05$., indicating slightly longer verbs in statements than in questions. The interaction Contrast*Mode also is significant only in the item-based analysis: $F1(2,30) = 1.93$, p = ns.; $F2(2,46) = 4.88$, $p < .05$. If one would like to accept the partially significant interaction, it resulted from the higher difference between contrasted and uncontrasted verbs in statements ($\delta = 6$ %) than in questions ($\delta = 3$ %).

In the object position, the final syllable, we also find a significant main effect for Contrast: $F1(2,30) = 12.73$, $p < .001$; $F2(2,46) = 17.42$, $p < .001$. The pairwise comparisons confirm longer objects in the contrasted condition than both other conditions in statements in the item-based analysis, but only longer than the verb-contrasted condition in the subject-based analysis. Sentence mode is not encoded by duration: $F1(1,15) = 0.28$, p = ns.; $F2(1,23) = 0.53$, p = ns. But the interaction Contrast*Sentence Mode is significant ($F1(2,30) = 4.69$, $p < .05$; $F2(2,46) = 8.02$, $p < .001$, confirming the (stronger) effects of contrast in statements than in questions.

To sum up: we find lengthening effects for Contrast in each of the three positions in statements, and less strong effects in questions. Effects of sentence mode are weak, and probably they can be accounted for by the heavier contrast induced lengthening in statements than in questions from the second syllable onwards.

8.4.3 Pitch maximum

The first of the four parameters related to fundamental frequency is the pitch maximum per constituent. The peak of the fundamental frequency is highly correlated to contrastive focus in the native and the target language of the learners, but there are differences the position in which contrasted constituents are significantly higher, and in the weight of the pitch maximum in a regression model between the languages. German native speech shows consistently a significantly higher pitch maximum for contrasted constituents compared to both non-contrasted constituents in all positions in declarative sentences. For native English, the pitch maximum is significantly higher than both other conditions only in the verb and the object position of statements, but not in the subject position. None of the languages shows a higher pitch maximum due to corrective contrast in the (echo-) questions.

Sentence mode is encoded very different in English compared to German: English native speakers consistently use a higher register to encode echo-questions compared to the declarative sentences; German does not differentiate the height of the fundamental frequency in sentence-initial and sentence-medial position, but ends in a strong rise to encode questions.

Table 8.7 shows the means of the pitch maximum, and Figure 8.3 informs about the distribution of the values.

Table 8.7: Mean pitch maximum for English by German native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	356.2	355.3	355.8	246.9	296.1	271.5	218.1	394.2	307.1
CF V	272.0	324.4	298.2	315.3	301.6	308.5	232.6	381.1	306.8
CF O	292.6	328.4	310.5	256.2	271.0	263.6	282.8	401.7	342.3
$\overline{X}_M$	306.9	336.0	321.5	272.8	289.6	281.2	244.5	392.3	318.7

In all positions and for both sentence modes, the contrasted constituents have a higher mean pitch maximum. But the differences are clearly larger for declarative than for interrogative sentences. With regard to sentence mode encoding, we see that questions have a higher pitch maximum than statements in all positions. But while the values are rather close in the subject (δ = 29 Hz) and verb (δ = 17 Hz), the difference amounts to almost 150 Hz on the objects. So, we can observe an overall slightly higher frequency for questions which end in a high rise. These points have to be confirmed by the statistical analysis.

In subject position, the within subject ANOVAs reveal that there is a significant main effect for Contrast: $F1(2,30) = 20.69$, $p < .001$; $F2(2,46) = 19.11$, $p < .001$. The adjusted com-

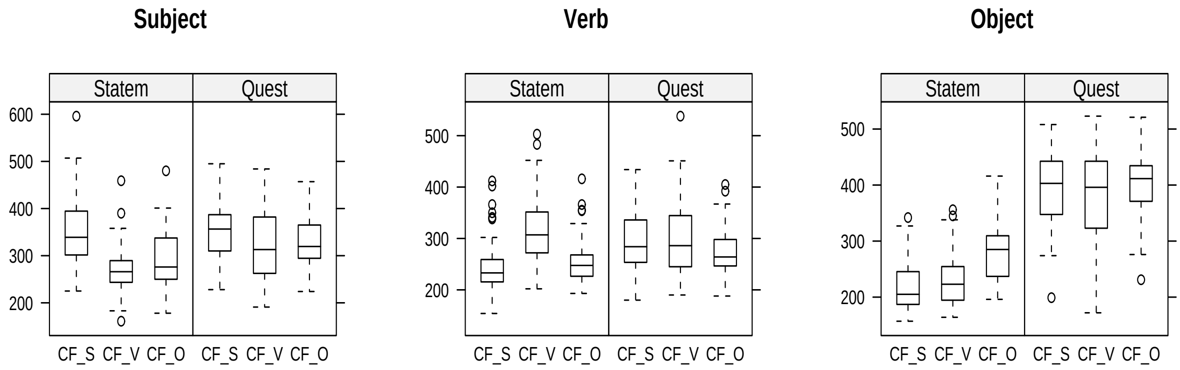


Figure 8.3: Pitch maximum for English by German native speakers for each constituent by sentence mode and contrast position

parisons confirm a significantly higher pitch maximum for contrasted subjects than subjects in both non-contrasted conditions for statements only, while the higher value in questions does not significantly differ between contrasted and non-contrasted conditions. A significant main effect for Sentence Mode F1(1,15) = 8.61, $p < .01$; F2(1,23) = 22.05, $p < .001$ indicates higher values for questions than for statements[5]. The interaction Contrast*Sentence Mode: F1 = 4.51, $p < .05$; F2(2,22) = 5.12, $p < .01$ is significant, too. It can be explained by stronger effects for Contrast in questions on the one hand, and a smaller difference for contrasted subjects in statements to the subjects in questions than the non-contrasted initial syllables in the declaratives.

In sentence medial position, the verbs, the main effect for Contrast is significant: F1(2,30) = 20.58, $p < .001$; F2(2,46) = 14.54, $p < .001$. The main effect for Mode is significant in the item-based analysis only (F1(1,15) = 3.37, $p = 0.086$; F2(1,23) = 7.90, $p < .01$). The interaction Contrast*Mode is clearly significant instead: F1(2,30) = 9.46, $p < .001$; F2(2,46) = 6.46, $p < .01$).Pairwise comparisons show that contrasted verbs are significantly higher than verbs in both non-contrasted conditions in statements. Contrasted verbs in questions are higher only in the item-based comparison with verbs in the object-contrasted condition. The interaction thus results from the stronger effect of contrast in statements.

Sentence finally, the ANOVA shows also a significant main effect for Contrast: F1(2,30) = 12.58, $p < .001$; F2(2,46) = 15.58, $p < .001$. The Post-Hoc comparisons reveal that contrasted objects in statements have a higher pitch maximum than uncontrasted objects in that sentence mode. Sentence Mode shows a highly significant main effect, due to the higher fundamental frequency maximum in questions, irrespectively of the focus condition: F1(1,15)

[5]Pairwise comparisons show that the main effect reflects higher values for the non-contrasted subjects only.

= 193.77, $p < .001$; F2(1,23) = 335.95, $p < .001$. The interaction Contrast*Sentence Mode is significant, too: F1(2,30) = 4.30, $p < .05$; F2(2,46) = 4.89, $p < .05$, confirming that the effects of contrast on the pitch maximum are found for statements only.

To sum up, we can say that contrast induces a significantly higher pitch maximum on each syllable of statements. The numerically higher values for contrasted constituents in questions are not confirmed statistically, except of the 30 Hz differences between a contrasted verb and verbs in object-contrasting sentences.

Sentence Mode shows a clearly up-stepped pitch maximum at the end of the sentences. But also during the first two syllables, the pitch maximum is increased significantly, especially if the constituents of the statements are not contrasted. So, German learners of English do not use the higher register as does the native control group, and the main distinction of fundamental frequency behavior is found at the end of the sentences, like in German. But, contrarily to the German pattern, the learner's maximum of the fundamental frequency is already higher during the first two syllables of the sentences.

8.4.4 Mean Pitch

The mean fundamental frequency is highly related to the pitch maximum presentend above. In fact, the same correlates of corrective contrast and sentence mode encoding can be found. English uses a higher mean pitch for the verbs and the objects of the statements if they are contrasted compared to the non-contrasted. German increases the mean pitch in all positions. None of the languages shows significant effects of contrast in questions. While for German, the mean pitch is one of the highest rated acoustic parameters, it appears only in the verb position of reduced regression models for English.

Sentence mode is encoded by a higher mean fundamental frequency throughout a question in English, but only by a higher end for the interrogatives in German.

The means of the mean pitch are listed in Table 8.8, the distribution can be observed in Figure 8.4.

The values mirror essentially what was already found for the pitch maximum, except the non-occurrence of an increased mean fundamental frequency due to contrast in questions. But the mean pitch is clearly elevated for all contrasted constituents in statements. Questions show an increased mean fundamental frequency from the beginning of the sentences on, but it raises from a mean differences of $\delta = 7$ Hz in the subject position, and a difference of $\delta = 20$ Hz in the verb position to more than 80 Hz at the end of the sentences.

Table 8.8: Mean pitch for English by German native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	291.8	276.3	284.1	191.6	257.5	224.6	181.3	290.0	236.2
CF V	238.5	269.3	253.9	262.2	247.4	254.8	189.8	285.1	237.4
CF O	259.0	267.4	263.2	214.4	224.5	219.4	226.3	275.3	259.8
$\overline{X}_M$	263.1	271.0	267.0	222.8	243.1	232.9	199.1	283.5	244.5

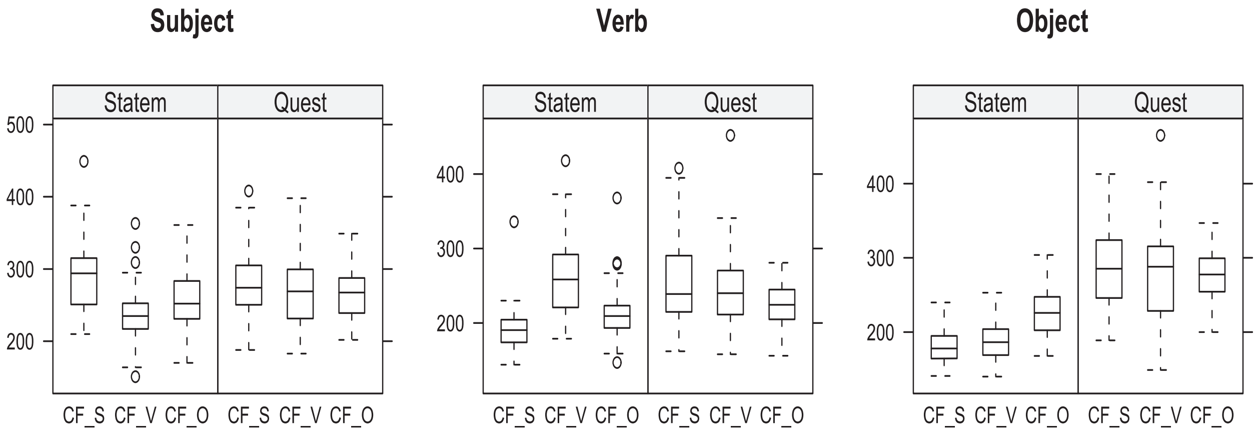


Figure 8.4: Mean pitch for English by German native speakers for each constituent by sentence mode and contrast position

Submitting the data to the ANOVAs for each constituent, one finds a significant main effect for Contrast on the subjects of the sentences: F1(2,30) = 9.69, p < .001; F2(2,46) = 9.92, p < .001. The Post-Hoc analysis reveals that contrasted subjects in statements are significantly higher than the non-contrasted subjects, but no difference is found for interrogatives. The main effect for Sentence Mode is not significant: F1(1,15) = 1.01, p = ns.; F2(1,23) = 3.16, p = ns.The significant interaction Contrast*Sentence Mode: F1 = 7.54, p < .01; F2(2,22) = 8.82, p < .001. confirms the stronger effects for contrast in statements than in questions.

In sentence medial position, the main effect for Contrast is also significant: F1(2,30) = 23.61, p < .001; F2(2,46) = 14.00, p < .001. The Post-Hoc comparisons show that the mean fundamental frequency is higher for contrasted verbs than for the non-contrasted conditions in statements, but not in questions. The main effect for Sentence Mode is also significant: F1(1,15) = 8.40, p < .05; F2(1,23) = 15.28, p < .001, signaling higher mean fundamental frequency for questions. The pairwise comparisons approves that the differences between the non-contrasted conditions are significant. The interaction Contrast*Mode: F1 = 19.45, p < .001; F2(2,22) = 15.00, p < .001 is significant, too, confirming the effect of Contrast in statements but not in questions, or the non-significant differences between the two sentence modes if the verb in statements is contrasted.

In sentence final position, the objects of the sentences, we find a main effect for Contrast, which is significant in the subject-based analysis, but only approaches significance in the item-based analysis: $F1(2,30) = 5.50$, $p < .01$; $F2(2,46) = 2.87$, $p = .067$. The Post-Hoc Bonferroni adjusted comparisons show that in the item-based analysis, the contrasted objects in statements are higher than the objects in both non-contrasted conditions, but in the subject-based analysis, the difference is significant only compared to the objects in subject-contrasted (CF S) statements. The differences in questions are not significant. The main effect for Mode is highly significant in both analyses: $F1(1,15) = 93.26$, $p < 0.001$; $F2(1,23) = 358.78$, $p < .001$, and shows that all objects in questions are higher than their corresponding phrase-final syllables in statements. The interaction Contrast*Mode ($F1(2,30) = 10.60$, $p < .001$; $F2(2,46) = 8.81$, $p < .001$) is significant, confirming the effects of Contrast in statements, but not in questions.

To sum up: the mean fundamental frequency is significantly higher if a contrastive accent should be transmitted in statements, but there are no contrast-induced effects on the syllables in questions. Although the contrasted means slightly exceed the non-contrasted, variance is too high to yield significant results for contrast in questions.

Sentence mode is clearly encoded by a raised final syllable, but already in sentence medial position, the mean fundamental frequency of questions is higher than the (non-contrasted) statements.

8.4.5 Pitch Range

The mean pitch range per syllable is shown in Table 8.9 and the main characteristics of the distribution can be read from the boxplot in Figure 8.5.

Table 8.9: Mean pitch range for English by German native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	126.52	137.9	132.21	79.75	69.56	74.66	57.02	176.27	117.28
CF V	65.92	107.1	86.50	99.46	94.77	97.11	74.45	160.25	117.35
CF O	70.88	109.2	90.03	69.33	76.73	73.03	115.54	205.15	160.34
$\overline{X}_C$	87.77	118.1	102.91	82.85	80.35	81.60	82.34	180.56	131.66

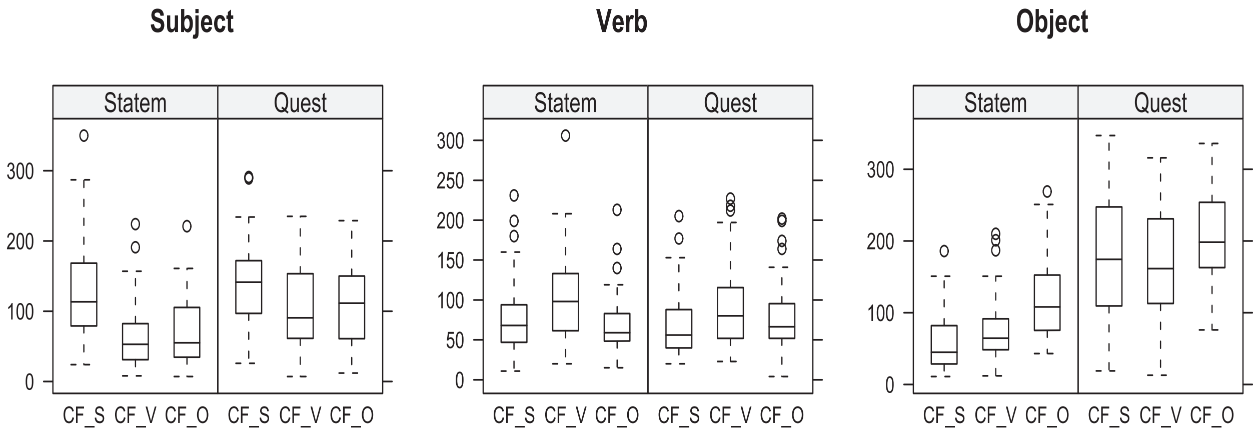


Figure 8.5: Pitch range for English by German native speakers for each constituent by sentence mode and contrast position

The results show that the pitch range is clearly higher if a constituent is contrasted than its non-contrasted counterparts in each of the three positions. But the difference is larger, and the distributions are wider apart from each other in statements than in questions. The interrogatives induce a higher pitch range in the subject and the object position, and no differences on the verbs.

This is confirmed – on the subjects of the sentences – by a significant main effect for Contrast: $F1(2,30) = 17.79$, $p < .001$; $F2(2,46) = 19.17$, $p < .001$. The Post-Hoc tests show that the pitch range for contrasted subjects is higher than for subjects in both non-contrasted conditions in statements. The higher pitch range in questions (about 30 Hz compared to uncontrasted subjects) is not significant in the Bonferroni-adjusted comparisons, except from subjects in verb-contrasted conditions in a item-based analysis. Sentence Mode shows a significant main effect, too: $F1(1,15) = 15.60$, $p < .001$; $F2(1,23) = 23.55$, $p < .001$, confirming a higher pitch maximum–minimum difference in questions. The interaction Contrast*Sentence Mode is not significant: $F1(2,30) = 2.89$, p = ns.; $F2(2,46) = 2.15$, p = ns. This would imply that Contrast has the same effects for statements and questions. Here, the adjusted comparisons are more conservative than the ANOVA in approving significance only between contrasted and non-contrasted subjects in statements.

For the verbs, we find a main effect for Contrast, too: $F1(2,30) = 8.89$, $p < .001$; $F2(2,46) = 6.67$, $p < .01$. The Post-Hoc comparisons reveal that contrasted verbs have a significantly higher pitch range than verbs in both non-contrasted conditions in statements, but verbs in questions differ significantly only from verbs in subject-contrasted (CF S) conditions. There is no significant difference between statements and questions: $F1(1,15) = 0.18$, p = ns.; $F2(1,23) = 0.33$, p = ns. The interaction Contrast* Sentence Mode not significant, either: $F1(2,30) = 1.48$, p = ns.; $F2(2,46) = 1.36$, p = ns. Again, this is a contradiction to the more conservative pairwise Bonferroni-adjusted comparisons, which state a significant differences of contrasted subjects to both non-contrasted conditions in statements but only to one of the non-contrasted conditions in questions.

In the object position, the sentence final syllable, the ANOVA confirms a significant main effect for Contrast: $F1(2,30) = 11.21$, $p < .001$; $F2(2,46) = 23.13$, $p < .001$. Pairwise comparisons bring to light that objects have a higher pitch range in the contrasted condition (CF O) than objects in both other conditions in statements, but in questions they are different from objects in the CF V condition of the only. Sentence Mode is highly significant ($F1(1,15) = 86.90$, $p < .001$; $F2(1,23) = 79.99$, $p < .001$) and confirms a higher pitch range for questions. The interaction Contrast*Mode: $F1(2,30) = 1.59$, p = ns.; $F2(2,46) = 2.98$, $p = 0.061$ is not significant, although the effects of Contrast are stronger in statements than in questions, if one follows the more conservative Bonferroni-adjusted comparisons.

To sum up the findings for pitch range, it is clearly shown that Contrast is encoded by an increased pitch range in all positions of statements. Although the numerical differences show the same tendencies and the interaction between Contrast and Sentence Mode is not significant, the Post-Hoc tests do not confirm a similar strong effect for contrast in questions. This may be attributed to the high variance and the rather small number of items per condition in the experiment.

The encoding of sentence mode is more straightforward: the final rise for questions is clearly mirrored in the pitch range of objects. But also in the subject position, questions tend to rise higher on the initial syllable, and this finding is statistically confirmed if the (initial) subject is not contrasted.

8.4.6 Alignment Ratio of the Pitch Maximum

The alignment ratio of the pitch maximum determines the relative position of the pitch peak within the examined syllables. A later peak is reported to be perceptually equivalent to a higher peak, in German it is occurring especially alternating to the pitch maximum. The final rise for questions will induce a late peak, a falling end of declaratives an earlier peak. The data are presented in Table 8.10 and graphically edited in Figure 8.6.

The ANOVA for the subject position reveals a a significant main effect for Contrast: $F1(2,30) = 13.80$, $p < .001$; $F2(2,46) = 8.27$, $p < .001$. The Post-Hoc comparisons show that contrasted subjects have a later peak than both non-contrasted in the subject-based analysis

Table 8.10: Mean alignment ratio of the pitch maximum for English by German native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	0.6441	0.7326	0.6883	0.1918	0.2088	0.2003	0.3189	0.7098	0.5164
CF V	0.5168	0.5861	0.5515	0.4388	0.3707	0.4048	0.2299	0.7155	0.4727
CF O	0.5306	0.6632	0.5969	0.2777	0.1947	0.2362	0.3350	0.7709	0.5530
$\overline{X}_C$	0.5638	0.6606	0.6122	0.3028	0.2580	0.2804	0.2946	0.7321	0.5140

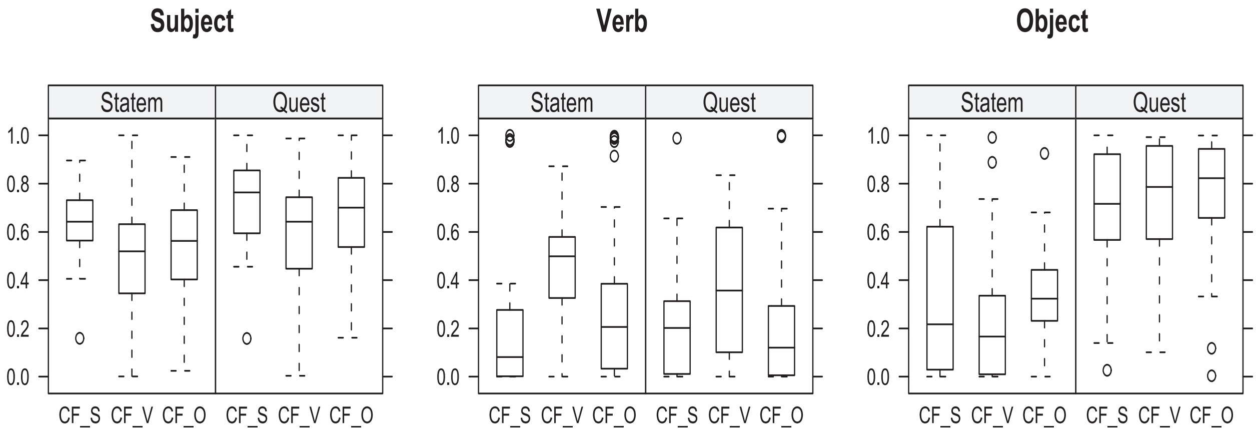


Figure 8.6: Alignment ratio of the pitch maximum for English by German native speakers for each constituent by sentence mode and contrast position

and later than the verb-contrast condition (CF V) only in the item-based analysis for statements. In questions, both analysis show that peak in the CF S condition is significantly later than the peak in the CF V condition. The more consistent difference of a contrasted subject (CF S) to the verb-contrast conditions can be explained by a preparation of the contrast on the verb in the CF V condition. If the verb will be rising – and the chances increase due to contrast encoding – the previous syllable is rather a fall, thus has an earlier peak. The significant main effect for Sentence Mode indicates a later peak for questions: $F1(1,15) = 11.15$, $p < .01$; $F2(1,23) = 18.47$, $p < .001$. The interaction Contrast*Sentence Mode is not significant sentence-initially: $F1 = 0.61$, p = ns.; $F2(2,22) = 0.74$, p = ns.

In sentence-medial position, the verbs of the critical sentences, we find a significant main effect for Contrast again: $F1(2,30) = 18.30$, $p < .001$; $F2(2,46) = 14.86$, $p < .001$. The item-based Post-Hoc tests are supporting a later peak for contrasted verbs than both other conditions in statements and questions. The subject-based Post-Hoc tests assign an error rate of smaller than five percent only in comparison to the verbs in the subject-contrasted condition in statements, and the object-contrasted condition in questions[6]. Sentence mode

[6]CF V - CF S is approaching significance with $p = 0.092$.

as no significant impact on the position of the peak: F1(1,15) = 2.33, p = ns.; F2(1,23) = 1.97, p = ns. Furthermore, there is no interaction Contrast*Sentence Mode: F1(2,30) = 0.54, p = ns.; F2(2,46) = 1.32, p = ns.

The behavior of the pitch peak alignment is very much different in the sentence final position, the objects. There is no significant main effect for Contrast : F1(2,30) = 1.96, p = ns.; F2(2,46) = 2.33, p = ns., thus the location of the pitch maximum in the constituent is independent whether the object is under contrastive focus or not. Of course, questions have a significantly later peak than statements: F1(1,15) = 117.04, p < .001; F2(1,23) = 201.64, p < .001. There is no significant interaction Contrast*Sentence Mode: F1(2,30) = 0.49, p = ns.; F2(2,46) = 0.84, p = ns.

To sum up the effect of contrast and sentence mode encoding of the alignment ratio of the pitch maximum, two points: first, the alignment ratio is clearly used to indicate questions at the end of the sentences, but also already on the initial syllable. Correlates of Contrast can be found most unambiguously only in the sentence medial position, the verb of the examined sentences. The most robust part of the delayed peak differences in the beginning of the utterances can be traced back to a preparation of this effect on the following syllable, the verbs.

Mean Intensity Ratio

We have seen that English native speakers use intensity very reliably and consistently throughout the utterances to encode contrast, while German shows significantly increased loudness only from the second syllable on, and it is not a good predictor of contrast in regression models of contrast encoding in German. In how far the German learners of English used intensity consistently to encode contrast is given in Table 8.11 and Figure 8.7.

Table 8.11: Mean intensity ratio for English by German native speakers

	Subject			Verb			Object		
	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$	Statem	Quest	$\overline{X}_C$
CF S	1.045	1.016	1.030	0.9588	0.9807	0.9697	0.9353	0.9855	0.9604
CF V	1.013	1.018	1.016	1.0023	0.9879	0.9951	0.9357	0.9739	0.9548
CF O	1.022	1.011	1.016	0.9671	0.9798	0.9735	0.9808	0.9872	0.9840
$\overline{X}_C$	1.027	1.015	1.021	0.9761	0.9828	0.9794	0.9506	0.9822	0.9664

The data show a higher intensity for the contrasted constituents in statements. Questions instead do not seem to be influenced by the corrections. The mean values for questions are

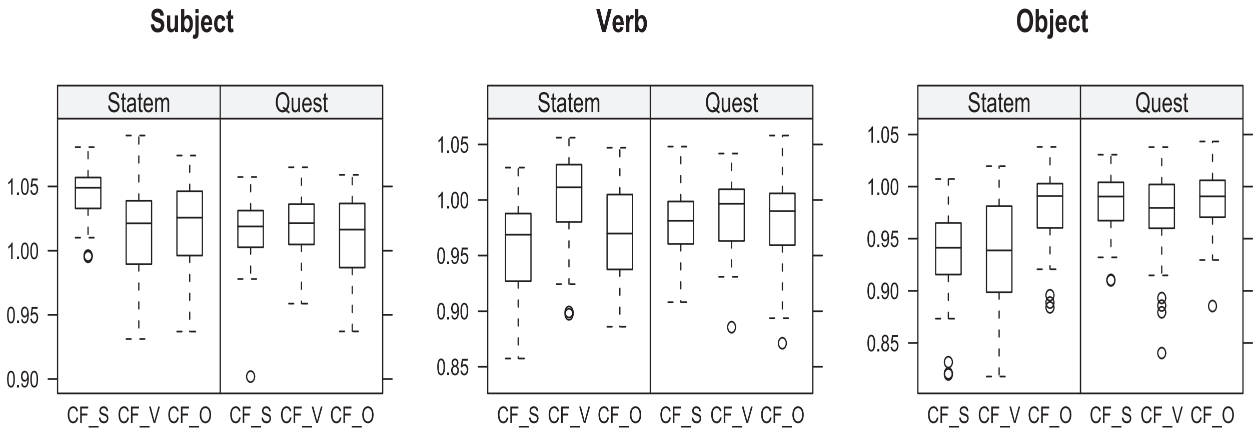


Figure 8.7: Mean intensity ratio for English by German native speakers for each constituent by sentence mode and contrast position

higher for questions than for statements.

In sentence initial position, we find a significant main effect for Contrast: F1(2,30) = 6.26, p < .01; F2(2,46) = 10.03, p < .001. The Post-Hoc Bonferroni-adjusted comparisons reveal that contrasted subjects are louder than non-contrasted subjects in statements. No significant difference is found in questions. Sentence Mode also shows a significant main effect, stating that statements are louder than questions: F1(1,15) = 10.90, p < .01; F2(1,23) = 13.33, p < .001. The significant interaction Contrast*Sentence Mode (F1(2,30) = 6.33, p < .01; F2(2,46) = 11.33, p < .001 confirms the occurrence of effects of Contrast in statements but not in questions.

In the verb position, there is also main effect for Contrast: F1(2,30) = 8.11, p < 0.01; F2(2,46) = 15.78, p < .001. We also find a main effect for Sentence Mode, although it it significant in the item-based analysis only: F1(1,15) = 4.04, p = .063; F2(1,23) = 4.28, p < .05. The interaction Contrast*Sentence Mode is significant (F1(2,30) = 4.64, p < .05; F2(2,46) = 7.29, p < .01, and confirms the stronger effects for Contrast in statements than in questions. As the Post-Hoc tests reveal, it is due to louder contrasted than uncontrasted verbs in statements. No significant differences were found in questions.

In sentence final-position Contrast is significant, too. : F1(2,30) = 12.56, p < .001; F2(2,46) = 12.17, p < .001. Again, the Post-Hoc tests show that only objects in statements are significantly louder if they are contrasted than if they are not. The significant main effect for Mode confirms the higher values for questions than for statements: F1(1,15) = 21.38, p < .001; F2(1,23) = 83.40, p < .001. The significant interaction Contrast*Sentence Mode (F1(2,30) = 6.12, p < .01; F2(2,46) = 10.78, p < .001 approves the pattern found in the pairwise comparisons: intensity differences are significant only in statements, but not in questions.

To sum up the findings for the mean intensity ratio, we see that the German learners of English consistently use intensity to encode Contrast in statements. But Contrast does not induce an increased loudness in questions. Questions are louder than statements in all three positions, especially in the non-contrasted comparisons.

8.4.7 Overview for Contrast

For an overview of the 36 ANOVAs and related Post-Hoc tests, Table 8.12 summarizes the significant differences in the Bonferroni-adjusted comparisons for Contrast.

Table 8.12: Significant differences in the six acoustic parameters between contrasted and non-contrasted constituents by sentence mode for English by native speakers of German.

		Subjects		Verbs		Objects	
		F1	F2	F1	F2	F1	F2
statem.	Duration	++	++	++	++	+	++
	Pitch max.	++	++	++	++	++	++
	mean Pitch	++	++	++	++	++	++
	Pitch range	++	++	+	+	++	++
	Pitch max. Align.	++	+	+	++		+
	Intensity	++	++	++	++	++	++
quest.	Duration	++	++	+	++		
	Pitch max.			+			
	mean Pitch			+			
	Pitch range		+	+	+	+	+
	Pitch max. Align.	+	+	+	++		
	Intensity						

Analyses over Subjects (F1) and over Items (F). '+' means significantly different to one other contrast condition, '++' means significantly different to both other conditions. All significant differences show a higher value for the dependent variable in the contrasted condition than the uncontrasted condition(s).

Second language speakers use very consistently all acoustic parameters to encode contrast in statements. All[7] contrasted constituents are longer than both of non-contrasted conditions. Corrective contrast induces significantly higher values for parameters capturing pitch height (pitch maximum and mean pitch). And contrasted syllables are louder than non-contrasted, irrespectively of their position.

Furthermore, the results for the pitch range on the syllables show a reliable usage of a higher pitch range for contrasted subjects and objects in statements. In the verb position, the pitch range is higher only to verbs in which contrast already was encoded (the subjects), but not compared to verbs for which the contrast encoding will still follow (the objects), because the preparation of the higher fundamental frequency in that position includes a higher rise already for the preceding syllable. The least reliable parameter is the position of the peak (alignment ratio of the pitch maximum) in the recorded utterances. In subject position, the Post-Hoc tests confirm significantly later peaks only compared to subjects in object-contrasted statements in both, the item and the subject-based analysis. A difference between the subjects in subject-contrasted and verb-contrasted statements is significant in the subject-based analysis only. In the verb-position, the case is similar. The Post-Hoc tests approve a significantly later peak for contrasted verbs compared to verbs in the subject contrasted condition for both analysis. A later peak compared to the object contrasted condition is significant only in the item-based Post-Hoc test.

Prosodic correlates of contrast in questions are much less reliably different from non-contrasted syllables. Nonetheless, the number of significant differences is much higher for the second language speakers than for the utterances in the target (Experiment 1) and native language (Experiment 2). For instance the relative duration of the constituents is increased in the subject and in the verb position. The peak of the contrasted syllables in questions is at least later than one out out of the two non-contrasted conditions.

We also have to keep in mind that only significant results from the Bonferroni-adjusted comparisons figure in table 8.12. But the ANOVAs above give raise to insist on their conservative nature: for example for the alignment ratio of the pitch maximum we find a significant main effect for contrast in the subject and the verb position, and no interaction of Contrast*Sentence Mode. This would imply that questions and statements behave similar for contrast encoding. But the Post-Hoc tests confirm significant differences between contrasted and non-contrasted conditions consistently only for statements. A similar pattern is found for pitch range on verbs and objects[8].

[7] With the exception of the duration ratio in objects in the Post-Hoc analysis over subjects.

[8] Contrasted verbs have a larger pitch range than verbs in subject-contrasted questions; Contrasted objects have a higher pitch range compared to objects in verb-contrasted questions.

To conclude, we see that second language speakers consistently use all three dimensions of prosodic correlates (duration, pitch, and intensity) to encode contrast, and distinguish contrasted syllables from non-contrasted in statements. We also state that there are significant differences for contextually contrasted vs. uncontrasted syllables in questions, which we never found in quantity and quality for the utterances in the native and the target-language of the speakers.

8.4.8 Overview for Sentence Mode

Similar to the effects of contrast, the significant results of the Bonferroni-adjusted pairwise comparisons for Sentence Mode will be presented in Table 8.13. The table is slightly more complicated, as it has to indicate the direction of the significant differences. Also, the contrast conditions for which the pairwise comparisons reveal significant differences must be assessed, as they are not equal for all cells of the table.

Table 8.13: Direction and contrast condition of significant differences between statements and questions for English by German native speakers

	Subjects		Verbs		Objects	
	F1	F2	F1	F2	F1	F2
Duration	–	–	S>Q: V	S>Q: V	–	S>Q: O
Pitch max.	S<Q: V,O	S<Q: V,O	S<Q: S	S<Q: S	S<Q: S,V,O	S<Q: S,V,O
Pitch mean	S<Q: V	S<Q: S,V	S<Q: S	S<Q: S	S<Q: S,V,O	S<Q: S,V,O
Pitch range	S<Q: V,O	S<Q: V,O	–	–	S<Q: S,V,O	S<Q: S,V,O
Align. pitch max.	S<Q: S,O	S<Q: S,O	–	–	S<Q: S,V,O	S<Q: S,V,O
Intensity	S>Q: S	S>Q: S	S<Q: S	S<Q: S; S>Q:V	S<Q: S,V	S<Q: S,V

The part before the colon indicates the direction (S(tatements) higher [>] or lower [<] than Q(uestions)). The part after the colon indicates contrast positions [S,V, or O] in which the values significantly comply with the statement before the colon. Significant differences as revealed by Bonferroni-adjusted comparisons in subject-based (F1) and item-based (F2) aggregation.

The largest part of consistent significant differences between statements and questions is found in the object position. Questions have a higher pitch maximum and mean pitch, a larger pitch range and a later pitch peak. This mirrors the final rise for questions, and the declination of statements. Also, the intensity is higher for questions if the object is not under contrastive focus in the statements. The other positions must be regarded more differentiated.

In the subject position, we see that questions have a higher pitch maximum, a higher pitch range and a higher pitch mean, if the subject is not contrasted. If the subject is under contrastive focus, its intensity is even higher for statements than for questions.

Differences in the verb position mostly mirror stronger effects of contrast in statements than in questions. But note that the most prominent cue for contrast on verbs, the later peak, does not differ between statements and questions, even though only half of the questions received a perceivable accent, and thus a rise instead of a fall, to encode contrast, but about 80 % of the verbs in statements.

To sum up, we find that the main location of sentence mode encoding is the final syllable, although fundamental frequency is already indication questions in the initial position. But there, and in the sentence medial position, many of the differences between statements and questions can be accounted for by a more pronounced encoding of contrastive focus in declaratives.

8.4.9 The Relative Weight of the Acoustic Parameters

To compare the weight of the acoustic parameters for the encoding of contrast a stepwise backwards regression analysis has be undertaken. Remember, the perceptively rather hidden difference between the target and the native language of the learners. In subject and object position, native English relies mostly on duration and intensity, while for German, duration and the height of the fundamental frequency are sufficient. In verb position, both languages use the alignment of the pitch maximum as the most important parameter, together with duration and other pitch related parameters. English has to include intensity in that position too, while German does not use it systematically. So, the weight of the acoustic parameters to encode contrast is a conflict between the two languages the German learner of English has to handle.

The analysis of differences has shown that correlates of corrective focus are less consistent for questions than for statements, but in contrast to the native control groups, they are found: for instance a longer relative duration of focused subjects. However, for the ease of a comparison between the speaker groups, the following regression model contains only data of the declarative sentences.

The final model of a stepwise backwards logistic regression is given in Table 8.14 for the sentence initial position, in Table 8.15 for the sentence medial position, and in Table 8.16 for the sentence final position.

Table 8.14: Final model of a stepwise backwards regression analysis for English by German native speakers for the subjects of the critical statements

	Estimate	Std. Error	z value	p	Significance
(Intercept)	-66.540864	14.132561	-4.708	2.50e-06	***
Duration	0.278198	0.059452	4.679	2.88e-06	***
Mean pitch	0.025242	0.006133	4.116	3.86e-05	***
Pitch max. Align.	2.672418	1.523519	1.754	0.0794	.
Intensity	47.500454	12.136782	3.914	9.09e-05	***

Table 8.15: Final model of a stepwise backwards regression analysis for English by German native speakers for the verbs of the critical statements.

	Estimate	Std. Error	z value	p	Significance
(Intercept)	-32.726974	7.129960	-4.590	4.43e-06	***
Duration Ratio	0.216602	0.057813	3.747	0.000179	***
Pitch max.	-0.013844	0.009471	-1.462	0.143808	
mean Pitch	0.039822	0.011874	3.354	0.000797	***
Intensity Ratio	19.905019	6.943467	2.867	0.004147	**

For an overview, the ranking of the acoustic parameters that have significant impact on the regression model is shown in Table 8.17

We can see that in all positions, duration, intensity and the height of the fundamental frequency are the best and sufficient predictors of a regression model. Duration is always the highest ranked, mean pitch and fundamental frequency follow in second and third position. For the sentence final syllables, the mean fundamental frequency AND the maximum of fundamental frequency have significant impact to the regression model, in the other positions mean pitch is the only significant fundamental frequency parameter left.

In the verb position the three parameters of duration, intensity and pitch height even rank out the reliable change in the slope of the pitch movement, namely a rising pitch (and late alignment of the pitch maximum) for accented verbs instead of a default falling contour (and an early alignment of the pitch maximum). The alignment of the pitch maximum is not in the final model of the logistic regression, although in this position, both the target and the native language of the learners use it as the most important parameter in the regression (see Table 5.20 and 6.17).

Table 8.16: final model of a stepwise backwards regression analysis on for English by German native speakers for the objects of the critical statements.

	Estimate	Std. Error	z value	p	Significance
(Intercept)	-32.61975	7.42865	-4.391	1.13e-05	***
Duration Ratio	0.17853	0.04746	3.762	0.000169	***
Pitch max.	-0.05319	0.02443	-2.177	0.029456	*
mean Pitch	0.08269	0.02303	3.590	0.000330	***
Pitch range	0.02477	0.01529	1.620	0.105166	
Intensity Ratio	19.86448	6.62159	3.000	0.002700	**

Table 8.17: Ranking of significant parameters at the final step of a backwise logistic regression model at the three positions for contextual contrast. English statements spoken by German native speakers

Rank	Subjects	Verbs	Objects
1	Duration	Duration	Duration
2	Mean pitch	Intensity	Mean pitch
3	Intensity	Mean pitch	Intensity
4			Pitch max.
5			
6			

9 German English: an interlanguage between German and English ?

Now we can assess the correlates of sentence mode and focus for advanced German learners of English, and compare them to the patterns obtained for the identical materials of native speakers of the target language (English: Experiment1) and similar sentences produced by German native speakers in German (Experiment 2).

9.1 Comparison of the Perceptive Correlates

9.1.1 Perceived Accent and Contrast Position

The percentage of perceived main phrase accent in the positions of corrective contrast is similar, or even slightly higher than for both control groups (see Tables 5.3, 6.1, and 8.1). Thus, one can assume that the learners had no specific problems with the task and the materials, and that they were able to produce a perceivable accent. Especially the main accent on the verbs is more frequently perceived than for both native speaker groups (native English: 77 % in statements, 50 % in questions; native German: 79 % and 38 %; German English: 85 % and 52 %). Certainly, it is not the case that the L2 speakers missplaced the accent more than the native speakers, as observed for example in Jilka (2000).

9.1.2 Perceived Contours

The central findings of the analysis of the perceived contours on each constituent depending on the position of the contrast and the sentence mode can be summarized as follows.

In the subject position, two points shall be noticed. First, the German learners of English tend to use a rise (LH) as frequent as the German native speakers, and do not start with a high level tone (H) as the English native speakers do in about 30 % of the cases. For native German and German English this tone is perceived in about 10 % of all sentences. Second, the L2 speakers use the complex LHL onset as do the native speakers of the target language, but which is not found in native German sentence onsets.

Similar can be stated for the sentence medial position, the verbs of the sentences. The German learners of English do not use a high (H) or low (L) level tone as frequent as the English native speakers, but prefere a movement, thus either rising (HL) tone for non-contrasted or rising (LH) for contrasted verbs as frequent as the German native speakers. But as the English but not the German control group, a certain amount of the complex rise-fall (LHL) is found for contrasted verbs in statements.

In the object position, the similarities between the native German contours and the contours found in the learner's speech are dominant: Non only that sentence mode is encoded by a consistent rise for questions, but also the more frequent usage of a complex LHL tone for contrasted objects in statements follow the patterns found for the German control utterances. The usage of rising contours for statements is more frequent than in the native language of the learners (10 % compared to 2 %), but it does not reach the almost 20 % of the target language controls.

To sum up, the distribution of perceived contours in L2 speech roughly follows the patterns found in the utterances of the native language of the learners. Only for a part of the tonal preferences, a movement towards the patterns of the target language can be assesed.

9.2 Acoustic Correlates

The detailed examination of the acoustic correlates of sentence mode and contrastive focus encoding by native speakers of English and German, and the group of German learners of English has shown a lot of details of their specific preferences. The analysis of the perceived contours is not sufficient for an assessment of gradual differences between the groups, for instance the height of a rise, the overall register of a sentence, or the differences in the weight of the acoustic parameters. But those are necessary to discuss a possible approximation of the learner's speech to the target language, as it is found for aspects of segmental learning, e.g. voice-onset time and vowel-formants (Bohn & Flege, 1997; Flege & Hillenbrand, 1984; Flege & Efting, 1987). Some central aspects shall now be compared between these groups.

9.2.1 The Encoding of Sentence Mode

The data of the three speaker groups has shown that sentence mode is encoded mainly by fundamental frequency[1]. While English natives use a higher pitch throughout the whole sentence in the echo-questions but dominantly end in a fall, German natives use the same height of the fundamental frequency during the first two syllables to rise strongly at the end. In the group specific analysis, the German native learners of English show a significantly increased pitch on the sentence initial syllable, and rise steeply at the end. The mean fundamental frequency and the position of the pitch peak correspond to these two characteristics.

Figure 9.1 summarizes the difference between the mean fundamental frequency of statements and questions for each of the three groups in each position.

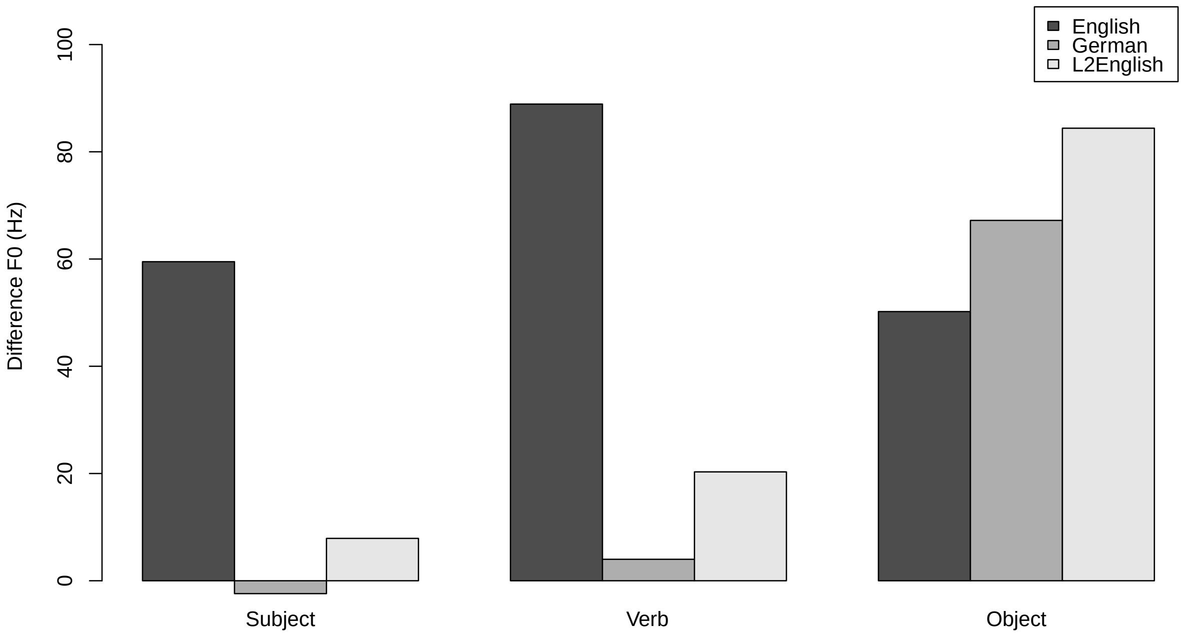


Figure 9.1: Difference in the mean fundamental frequency between statements and questions (mean questions - mean statements)

It is evident that in the first two positions of the sentence, native English shows a clearly increased pitch for questions compared to statements. The values of the learners amount only to about 10 Hz and 20 difference on subjects and verbs, respectively, but the questions are still higher than statements in these positions. There is almost no difference between the two sentence modes for native German utterances. In the object position, all groups show a

[1]All three groups also show an increased intensity for questions in the final position in case the final syllable is not contrasted.

strongly increased mean fundamental frequency, the both native speaker groups up to about 60 Hz, and the L2 speakers even to more than 80 Hz.

Submitting the mean differences between statements and question aggregated over all contrast conditions to a One-Factor ANOVA with Speaker Group as factor and subsequent TukeyHSD tests for pairwise comparisons, these observations are confirmed: native German and L2 English do not differ significantly in the subject and the verb position, while English differs from both. In the object position, L2 English is different from native English, but not from native German. The two native speaker groups do not differ significantly from each other[2]. Note that the findings of the joint analysis presented here differ from the detailed investigation of the speaker groups before. During the first two constituents, native German shows no difference, and native English always shows differences, but for L2 English, a significant main effect for Sentence Mode is found, as well as an interaction Sentence Mode*Contrast. With the Post-Hoc analyses it became clear that L2 English does use a higher pitch for questions than for statements, if and only if the constituents are not contrasted. The grand means used for the joint analysis here comprise the contrasted conditions, and therefore the distance between statements and questions is weakened, and hence closer to the native German pattern in the subject and the verb position of the sentences.

As revealed in the language specific analyses, the other main acoustic parameter encoding the sentence mode is the alignment of the pitch peak, especially at the end of the sentences. Figure 9.2 shows the mean differences between questions and statements in the relative position of the peak within the three constituents for each language.

Clearly, during the first two constituents of the analyzed utterances, none of the speaker groups make use of the relative position of the peak. Only at the end of the utterances, the peak is way later in questions than in statements if the sentences are spoken by German native speakers, both, in their native languages and in English. Native English utterances still do not distinguish the sentence modes by different peak positions[3].

To sum up the results of the statistical comparison of the acoustic correlates of Sentence Mode encoding between the three speaker groups, one has to state that the English utterances produced by native speakers of German do not significantly differ from the native German

[2]Subject position: $F1(2,45) = 24.77$, $p < .001$; native Eng. vs. native Ger.: $p < .001$; native Eng. vs. L2 Eng.: $p < .001$; native Ger. vs. L2 Eng.: p = ns.
Verb position: $F1(2,45) = 32.27$, $p < .001$; native Eng. vs. native Ger. : $p < .001$; native Eng. vs. L2 Eng.: $p < .001$; native Ger. vs. L2 Eng.: p = ns.
Object position: $F1(2,45) = 4.96$, $p < .05$; native Eng. vs. native Ger.: p = ns.; nat. Eng vs. L2 Eng: $p < .05$; native Ger. vs. L2 Eng.: p = ns.

[3]No effects of Sentence Mode in subject and verb position; in object position: $F1(2,45) = 27.62$, $p < .001$; nat. Eng. vs. nat. Ger.: $p < .001$; nat. Eng. vs. L2 Eng.: $p< .001$

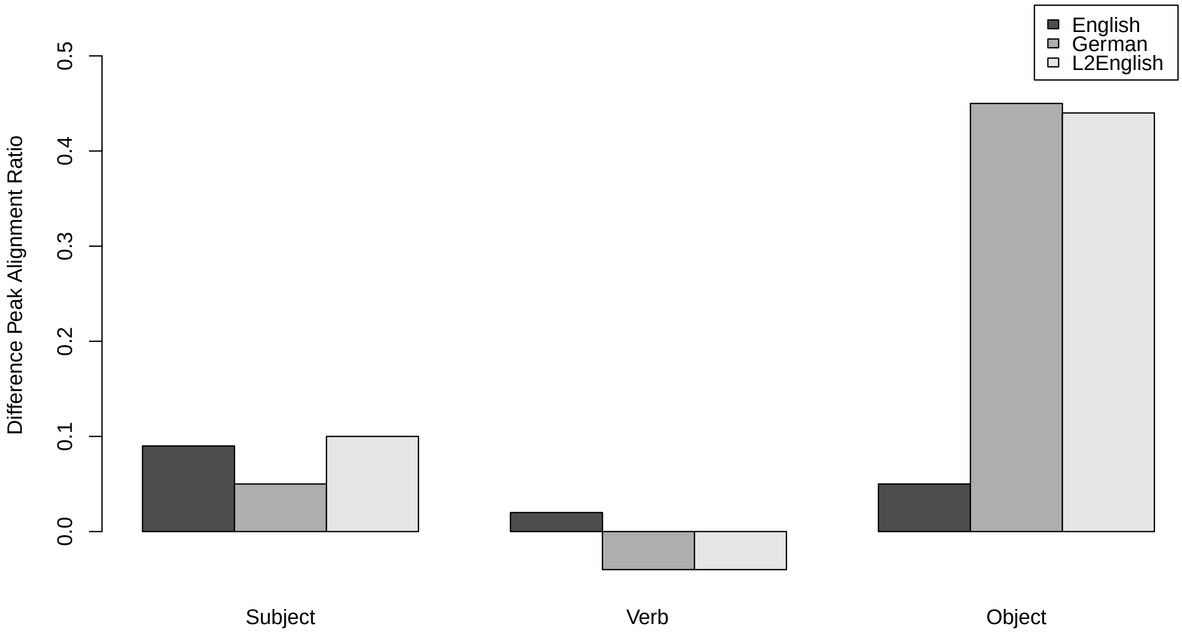


Figure 9.2: Difference in the F0 peak alignment ratio

sentences. However, the seemingly clear result is corroborated in some aspects because it neglects the stronger effects of contrast in L2 English.

9.2.2 The Encoding of Corrective Contrast

The analyses in the previous chapters have shown that corrective contrast is mainly encoded by an increased duration, a higher fundamental frequency and a raised relative intensity. This is similar for all three speaker groups, but the relative weight is different. While for a predictive model of contrast the duration and the intensity ratio are sufficient for English, German relies on duration and fundamental frequency in the sentence initial and sentence final position. L2 English needs all three acoustic parameters for a good reduced regression model. It has also been shown that none of the native speaker groups reliably encoded contrast in the (echo-) questions. A comparison of the patterns of contrast encoding thus will primarily be done with the data obtained for statements only.

First, the differences in duration between contrasted and non-contrasted constituents will be considered. Figure 9.3 shows the differences in the duration ratio.

English by both groups, the native speakers and the language learners, makes a larger difference in length for contrasted compared to non-contrasted constituents than German.

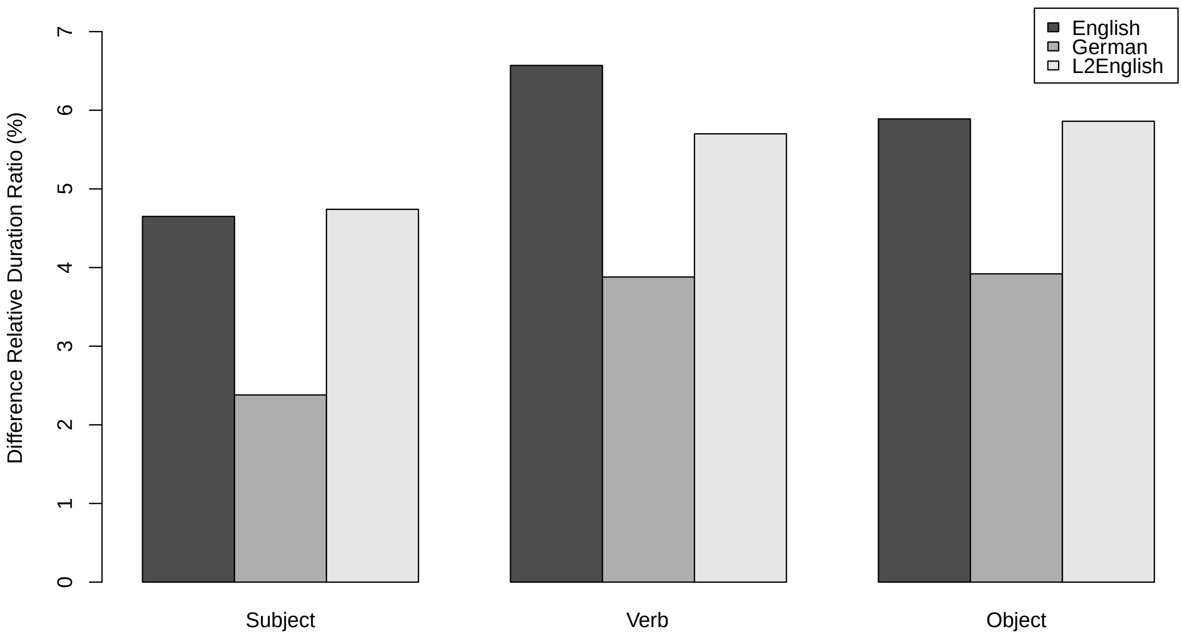


Figure 9.3: Difference in the relative duration ratio between contrasted and and non-contrasted constituents in statements

Native English and L2 English increase the relative duration of a contrasted constituent in the phrase more than two percent further than native German. However, submitting the differences per subject to a One-Factor ANOVA, the differences between the languages are not significant in any of the positions. So, although the mean values indicate a stronger usage of duration to encode focus in English, this point cannot be confirmed statistically. And hence, one has to assume that English, irrespectively of the speaker group, and German, increase the duration for contrasted constituents similarly.

Let us now consider the height of the fundamental frequency. The assessment of the weight of the parameters withing the speaker groups has shown that German relies stronger on pitch height to encode corrective contrast than English spoken by native speakers. For instance, the final model of the stepwise backwards logistic regression shows that the mean pitch or the pitch maximum together with duration are sufficient for an appropriate modeling of contrast in German in the subject and object position, while the height of the fundamental frequency is not necessary for a good model in English (see sections 5.3.10 and 6.3.10). However, restraining the analysis to statements which – contrary to questions – show clear focus encoding, there is no significant interaction between Contrast and Speaker Group in a comparison of the pitch maxima for contrasted and non-contrasted constituents between the two languages (see section 7).

Assuming strong influences from the native or the target language of the L2 speakers as it is put forward by the Contrastive Analysis Hypothesis, the prediction is that the usage of the pitch maximum to encode focus in English as a second language would not differ from either of the two native speaker groups. However, as Figure 9.4 shows, the learners increase the fundamental frequency stronger than both, the native speakers in the first and second position of the analyzed statements.

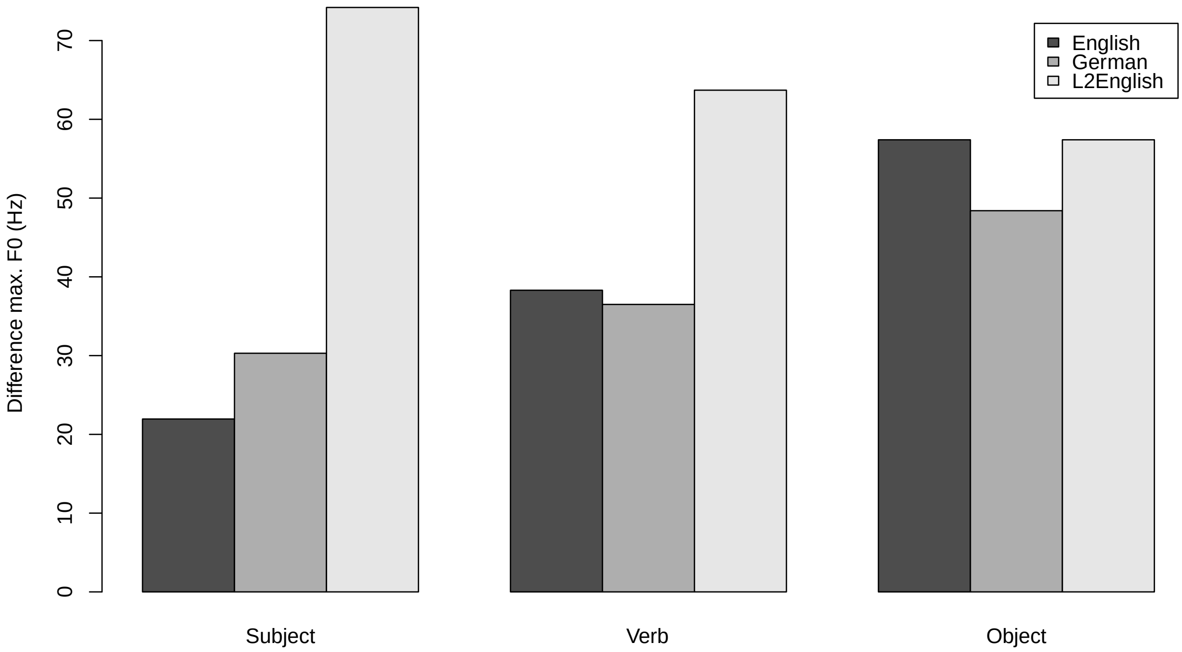


Figure 9.4: Difference in the pitch maximum between contrasted and and non-contrasted constituents in statements

In subject position, L2 English uses a pitch maximum that is more than 70 Hz higher for contrasted than for non-contrasted constituents, while native English shows a mean difference of 22 Hz, and native German 30 Hz. In the verb position, the differences amount to 64 Hz for the learners, and 40 and 37 for native English and German, respectively. For the first constituent, a One-Factor ANOVA with a factor of Speaker Group is significant ($F1(2,45) = 6.16$; $p < .01$) and a subsequent TukeyHSD confirms significant differences between L2 English and both native languages. In the verb position, a similar ANOVA does not approve significant effects of Speaker Group despite the clear differences for the means[4]. In the object position, all groups show a very similarly increased height of the pitch peak.

To summarize this point, the height of the pitch maximum on the first syllable of the statements is not derivable from neither the native nor the target language of the learners.

[4]In fact, the differences reach only 10 % of error probability ($F(2,45) = 2.49$; $p = .094$)

In the other positions, the statistical analysis does not approve differences between the languages.

Finally, the data of the third acoustic dimension, the intensity, shall be compared between the three speaker groups. Within the speaker groups, it has been shown that both, native English and L2 English need the mean intensity ratio for a good predictive model of contrast in a final backwards logistic regression model, where it is completely absent for German native speech. A comparison between the correlates of intensity for contrasted vs. non-contrasted constituents in statements has shown that the two languages spoken by native speakers differ only in the object position significantly, and reach an error rate of less than ten percent in the other two position, indicating that the intensity ratio of contrasted constituents is larger in English than in German. In a perspective of the Contrastive Analysis Hypothesis one would predict that English spoken by German learners would use intensity somewhere between the two native speaker groups.

Indeed, as Figure 9.5 shows, this is what we find in the verb and object position of the statements. English spoken by native speakers increases the relative intensity more than English spoken by German native learners and they raise the intensity more than German by native speakers. In the subject positions, all languages show a relatively small additional relative intensity for contrasted syllables, but L2 English a little stronger than native English and that a little stronger than native German.

Submitting this data to a One-Factor ANOVA to test significant differences between the languages, one has to state that only in the object position, there is a significant difference at the usual 5 % level ($F1(2,45) = 3.41$; $p < .05$). A TukeyHSD Test shows that native English differs from native German, but L2 English does not differ from either language. In the other positions, the use of the mean intensity ratio does not differ significantly between the speaker groups.

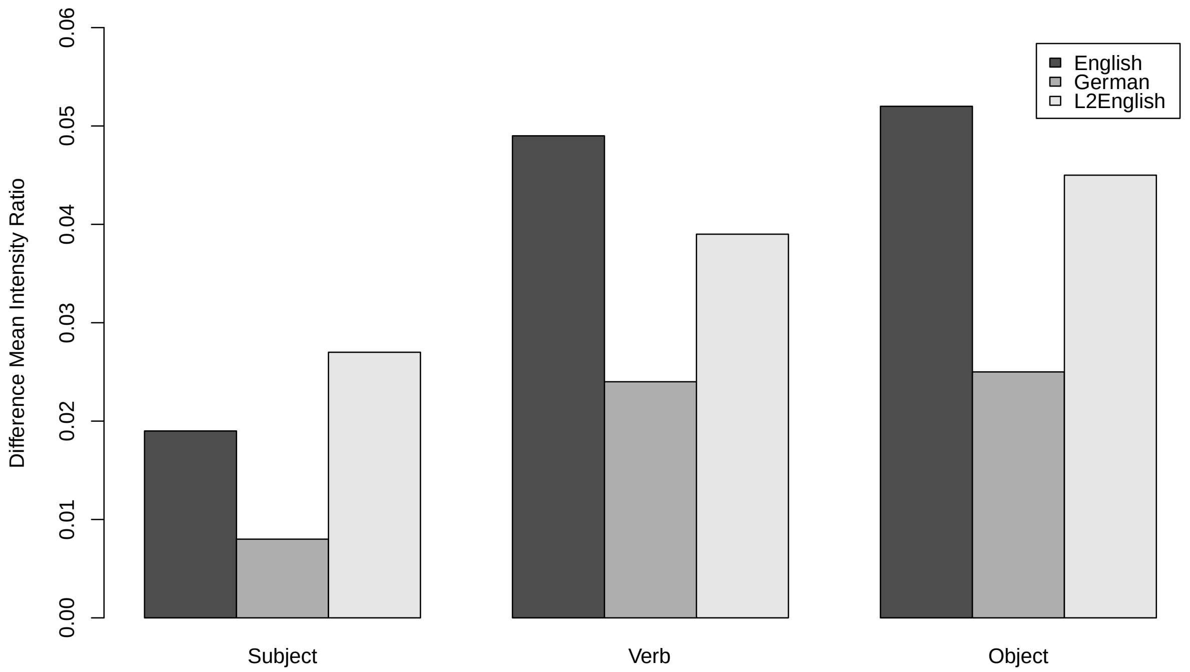


Figure 9.5: Difference in the mean intensity ratio between contrasted and and non-contrasted constituents in statements

10 Explanations of the Encoding of Sentence Mode and Contrast in L2

In the beginning of chapter 8 four hypotheses of possible sources of a prosodic pattern of second language prosody were set up.

The first hypothesis, perfect acquisition, can be rejected clearly. For example, the German learners of English did not use a falling contour to encode echo-questions, and a comparison of the grand means of the mean pitch in statements and questions clearly showed that native English and L2 English are significantly different. Additionally, fundamental frequency remained a significant predictor in a reduced regression model to predict corrective contrast from acoustic data.

The second hypothesis, total transfer, can be rejected, too. English by German native speakers and native German differ in several aspects. For instance, L2 English increases the fundamental frequency for echo-questions during the first two constituents significantly if one compares the non-contrasted conditions[1]. Total transfer is not supported by some aspects of contrast encoding, too: For example, in the assessment of the weight of the acoustic parameters, intensity has to remain in a final reduced regression model, but it is not found for the German native speakers.

So, the two other hypotheses, partial transfer, and a recourse on the Biological Codes remain, which shall be analyzed in some more detail now.

10.1 Partial transfer

The last chapter, the comparison of the three speaker groups, is written entirely in the perspective of an assessment of interferences of the native or of the target language.

[1]However, taking the grand means of all contrast conditions, no significant differences between questions and statements were found during the first two constituents in the sentences of the L2 speakers.

Many aspects of the L2 prosodic encoding of sentence mode and contrast can be derived from transfer from German. General sentence mode encoding in L2 English does not differ from the prosodic pattern in native German: The main difference between statements and echo-questions in L2 English and German is the strong rise on the last syllable of the questions, while native English speakers chose a higher register from the beginning of the sentences on.

There are further indications of transfer from the analysis of the perceived contours: for example the preference for unidirectional pitch movements, as mentioned in section 9.1.2, and a gradient approximation to the pattern found in the target language as the percentage of perceived final rises in statements – the distribution of contours of the German learners of English is just about in the middle between the rate of the native German and the native English speakers.

Patterns of the encoding of contrast could be interpreted as interferences from the native language, too. For example, fundamental frequency remains a significant predictor in the assessment of the weight of the acoustic parameters – while the increased role for intensity is learned from the target language. A nice example is also found in the comparison on the effects of contrast encoding on intensity of the objects of the statements: The degree of the increasing of intensity in native English is significantly higher than in native German, while English spoken by native German learners does not differ from either language.

In sum, these facts can be interpreted in a way that the prosodic encoding of sentence mode and corrective contrast in L2 starts from the native language inventory and partially approaches the patterns of the target language.

Generally, this would be in line with the Flege's speech learning model (see 2.4). However, if one maintains the categorization that connecting a falling tone to questions is a 'new' aspect and the increased role of intensity for contrast encoding is a 'similar' entity in the inventory[2], the results clearly oppose Flege's assumptions. The learners continue to use the final rise to

[2]One could argue about that categorization: It is possible to treat falling tones in questions as 'similar', as falling contours occur also in German (especially statements), and intensity as a 'new' predictive parameter for contrast encoding. However, although falling tones are not a new category like a non-occurring phoneme, the CONNECTION between falling contours and questions is new, in a sense that they do not occur in German (echo-)questions, but are frequent in the English echo-questions and in many other kinds of questions, too (Bartels, 1997). And, the mechanism of categorical perception that hinders the distinction of similar phonemes in the native and the target language cannot take place in this case: falling and rising contours cannot be confused. Categorical perception mechanisms could interfere in the case of the increased role for intensity in contrast encoding, instead: If one assumes a category of "accentedness" it is obviously correlated with an increased duration, fundamental frequency and intensity of the respective syllable, in both, English and German (at least in the verb, and object position of the statements). As there is only one category, namely "accentedness" the different weight of the acoustic parameters that encode it is likely to be ignored.

distinguish questions from statements, and add intensity to the reliably increased acoustic correlates of contrastive focus.

10.2 Biological Codes

We have seen that partial transfer can explain many of the patterns of the prosodic encoding of sentence mode and corrective contrast in the speech of German learners of English. However, there are some aspects that can not be covered with the assumption of an approximation to the target language with interferences from the native language:

For instance, the number of acoustic parameters that show significant differences between contrasted and non-contrasted conditions in questions (and in statements) is higher than for any of the native speaker groups (compare Table 5.15 for English, Table 6.12 for German, and Table 8.12).

Furthermore, the difference in the pitch maximum between contrasted and non-contrasted subjects[3] in statements is much higher than for both native speaker groups. None of these findings can be explained by an increased number of accentes in general (Grosser, 1997), as these would weaken the differences between contrasted and non-contrasted conditions; neither a higher number of missplaced accents (Jilka, 2000) can account for this phenomenon[4].

In addition to these points addressing only singular acoustic correlates in certain positions, we have seen that is is rather complex to assess the origin of the patterns in the chapters above: some encoding preferences have to be learned (e.g. the increased intensity in the beginning of the utterances, higher questions than statements from the verb on), some must be transfer from the native language (final rise in questions, reliable usage of fundamental frequency for contrast), but there is no straightforward rule that could account for the different sources in the three positions of the sentences that would describe pattern in the acoustic parameters.

More likely, the L2 pattern can be summarized by a general principle: *use everything that is presumably intelligibly for my partner!*. The L2 speaker reccurs on every prosodic property that he thinks that is correlated with the functions he wants to transmit. This includes a strong rise at the end of interrogatives, as well as a (slightly) increased fundamental frequency

[3]On verbs, the significance level is only around 10 %, but the difference in maximum of the fundamental frequency is strongly elevated compared to non-contratsted verbs, too.

[4]The number of “missplaced” accents (contextual focus and accent are not in the same place) is certainly not higher for the learners than for the native speaker groups. In fact, the groups do not differ to a notable ammount (compare Tables 5.3, 6.1, and 8.1)

from the beginning of the (echo-)questions on. Contrast is encoded by a strong lengthening, a higher fundamental frequency, and an increased intensity of the syllable. Furthermore, in positions that allow a correlation with a change of the contours (e.g. the rise instead of the fall on verbs in both sentence modes, or the complex rise-fall (LHL) in the object position of statements), the language learners tend to use it more consistently than the other groups. A simple principle of a cooperativity for the encoding of communicative functions seems to be more plausible than a forcedly differentiated distinction of transfer, acquisition, and approximation for each parameter in each position for each function.

Now, the question arises where the linking of certain patterns and communicative functions comes from. If one wants to avoid complex interactions of transfer, acquisition, and position that have to be generated at each instance of encoding, one can refer to the principles stated in the Biological Codes. As already stated, the second language prosodic patterns are completely in line with the predictions of Gussenhoven's theory: sentence mode encoding is a function of the Frequency Code, and questions (including echo-questions) lead to an increased pitch, especially towards the end. The encoding of contrast is a function of the Effort Code: Syllables that are intended to be perceived more prominently than others need more effort in the production which leads to an increased duration and a higher fundamental frequency. Effort is also needed to produce a rise instead of a fall in the sentence medial position, and a complex rise-fall instead of a simple fall at the end of statements. Extending Gussenhoven's predictions of the Effort Code to the parameter of intensity, it is obvious that a higher intensity correlates with a higher effort of the pulmonary tract.

The hypothesis that second language prosody processing is not necessarily explained best by an assumption of an interlanguage between the native and the target language of a learner is not new. It is also confirmed by research on perception (A. Chen et al., 2001; A. Chen, 2005; Makarova, 2001), and it has been claimed for production already 20 years ago: Mairs (1989, p. 271) concludes: "in short, neither a direct transfer of stress rules from the learner's native language nor a stress system derived from the native language by means of some reasonable and motivated changes in the native language stress rules accounts for the interlanguage data. Thus the source of Spanish speakers' stress rules must be sought in some other component of their linguistic competence." The Biological Codes provide such an (innate) linguistic competence.

This study only provides data about one group of learners, namely proficient German learners of English. However, it is tempting to predict that the Biological Codes are able to explain the patterns found with any learner of any native and target language – including those whose language specific preferences are further apart from each other and from the predictions of the Biological Codes. Makarova (2001) and A. Chen et al. (2001) have shown that rising nonsense syllables are judged as questions, independently whether the native language of the raters shows that correlation. Further research has to show whether the Biological Codes provide valid predictions for the prosodic patterns of any learner which is independent of the concrete native and target language. In case this hypothesis could be verified by further research, the Biological Codes would provide a description of something like the *Basic Variety* (to pick up Klein and Perdue's 1997 term) of second language prosody.

To resume the attempts of the explanation of the encoding of sentence mode and corrective contrast in the speech of German learners of English, it has been shown that most of the patterns can be explained by partial transfer: the German learners start with their native language patterns, and partially were able to acquire the conventions of the target language. But the characteristics of the L2 prosody can equally be captured by instantiations of the Biological Codes. All the prosodic patterns are fully in line with the predictions of the Frequency and the Effort Code. The question of which is the real source of L2 prosody must be forwarded to further research in which interferences and the Biological Codes would provide more diverging outcomes. Of course, a decision about the source of the L2 prosodic pattern has influences for models of prosody production, as the following chapter will show.

11 L2 Prosody and Models of Speech Production

In chapter 4 the relevant characteristics of three models of speech (prosody) production were introduced. The goal of this chapter is to integrate the findings of the previous experiments into the models. The discussion will start with Fujisaki's model of prosody generation. Later, the results should be integrated into Xu's model of prosody production and finally, the in-depth-model 'Speaking' - model by Levelt (and de Bot's extention to L2 speech production) will be treated.

11.1 L2 Prosody and Fujisaki's Model of Prosody Generation

As already stated in section 4.3, Fujisaki's model of prosody generation (Fujisaki et al., 1980a; Fujisaki & Hirose, 1982, 1984) and its extensions (e.g. Mixdorff, 2002) is not a psycholinguistic model, but is mainly used for computational speech synthesis. However, as it is very precise in the description of prosodic patterns, and sucesfully applied to a number of languages, it should be mentioned here. Fujisaki describes prosody by two components, the phrase component which captures the overall declination and the pitch register of a speaker, and the accent component. The original model only treats the fundamental frequency, and Mixdorff (2002) implemented timing and duration algorithms. An integration of intensity has not yet been undertaken, but one can imagine that it would be very comparable to the fundamental frequency generation model: There is just another dimension of the phrase and and the accent component, which describes the overall intensity curve of an utterance (phrase component) and the intensity peaks on relevant syllables (accent component).

First, the findings of the two native speaker groups shall be discussed in the light of the model. The encoding of the echo-questions of the English native group could be transcribed by either a higher phrase component or accent components for all syllables/constituents of questions compared to statements. The average German native speakers distinguish questions from statements by a peak on the accent component on the last syllable of questions. Corrective contrast is computed by an additional accent componet on the respective syllable of statements for both native speaker groups, with a slightly lower value for the English native group on the subjects of the sentences. Although corrective contrast does not show *significantly* increased pitch in questions, the values are higher for focused syllables than non-focused, and a Fujisaki Model would be able to mirror that by having accent components also for focused words in questions. However, one cannot assume *one* value for the accent component for corrective contrast that is applicable to statements and questions, because the difference between focused and non-focused syllables is much weaker in questions. This point cannot be treated internally of a Fujisaki-Model, but one has to distinguish questions and statements before adding specific values for the accent components of focused and non-focused syllables. As this implies non-parallel processing of these linguistic functions, it is a problem for the Fujisaki-Model.

Second, Ignoring this point for a moment, for the L2 speakers the Fujisaki Model would be especially tempting in a perspective of the Contrastive Analysis: it provides exact numbers of the accent component for the native speaker groups, and, if we stick to a strict Contrastive Analysis, it should be possible to compute the values for the learners from the values of the native speaker group. The most radical (but testable) approach would be an attempt to define the values for x and y in a formula like: *L2 prosody = x % of L1 prosody + y % of target language prosody.*

Mixdorff (1996) did not go so far, but he compared timing and height of accent and phrase components of German and Japanese native speakers to those of native Japanese learners of German. He concludes that "Japanese speakers partly carry over their intonation habits to German" (Mixdorff, 1996, p. 1472). Two of the arguments for such conclusion are the transfer of falling intermediate phrase boundaries, and the percentage of the usage of falling or rising intonation for yes/no and wh-questions. While the German native speakers used a rising contour at the end of 98 % of (non further specified) yes-no questions, but only in 4 % of the *wh*-questions, and the native speakers of Japanese used a fall in about half of the questions of both types, the Japanese learners of German used rising tones at the end of 68 % of the yes/no questions and 73 % of the wh-questions. Mixdoff did not evaluate other explanations than transfer, but for example the case of the question contour in wh-questions cannot be explained by transfer from L1 nor an approximation to the target language. More likely, it can be accounted for by an influence of the Frequency Code. Furthermore, Mixdorff has to state "great variations in performance" among the learners, depending on their level of acquisition, but also on the lenght of the sentences, words, and certain properties of words.

Taking into account the high variation among speakers, even in a very limited set of communicative functions and words as for the materials above is a weakness of the rather static Fujisaki Model approaches. Additionally, it would not be sufficient to establish a specific accent component for corrective contrast, but a Fujisaki-based approach has to take into account at least the proven interactions with the sentence mode of the utterances, as well as the position. All in all, a Fujisaki - based approach of prosody generation certainly does not model human speech production neither in L1 nor in L2, but may produce reasonable results in speech synthesis.

Usually the quality of a speech synthesis model is tested a perceptive analysis of model-based generated speech. It would be interesting if it was possible to generate prosody with a specific "foreign accent" by using the characteristic phrase and accent components of the native and the target language of the artificial learner.

11.2 L2 Prosody and the PENTA Model

The PENTA model (Xu, 2005) was described in section 4.4. It is a recent model of human intonation production. It assumes a parallel encoding of communicative functions with specific encoding schemes which are modified by so-called melodic primitives for which the strength, a variable which defines how quickly (or if ever) a certain pitch target is approached, is the most relevant for this study. In a final step the information is integrated in a process of target approximation.

The model is able to capture the findings of the native speaker groups to a large part. The different patterns of the encoding of echo-questions (a higher register in English, and a final rise in German) can result from different schemes. Similar can be stated for the cases of an alternation of contours for contrasted syllables (e.g. a rise instead of a fall for sentence medial contrast, or the usage of a complex rise-fall for focused syllables at the end of statements). Also, an increased duration and pitch range are previewed by the model. The PENTA model is perfect to explain the weaker effects of contrast in questions: In case a syllable is focused in a question, the strength of the encoding scheme for the function focus is lower than if it occurs in a statement[1].

[1]Liu and Xu (2005) argue that focus in questions results in similar alternations than focus in statements, and put forwart the idea of a strict parallel processing of the respective encoding schemes. However, they did not investigate a communicative function of (whatever kind of) focus, but the task for the subjects was to "emphasize any word that was surrounded by square brackets" (Liu & Xu, 2005, p. 76)

One problem is that the large variance in both contours and acoustic correlates is difficult to explain with the PENTA model : there is certainly no unique correlation between a communicative function and one encoding scheme. The correlations must contain further systematic rules like the position dependency, and some random part of choice for each syllable and each scheme and each strength in the process of target approximation. But all in all, the PENTA - model is an interesting approach which should be refined by further research – and of course integrate intensity in the melodic primitives.

As for the pattens in L2 speech, PENTA is also capable to explain a large amount of the findings. One possibility would be to include a mechanism of language choice. It could be located as a tag of the encoding schemes or as a (separate) decision point between the communicative functions and the encoding schemes. For the first, one would assume one kind of lexicon with entries that contain information about the language they are used for. Learning could take place very similar to the learning of a phonemic inventory. From strong influences of the native language, probably as an effect of a non-perception of acoustic differences due to perceptual assimilation mechanisms, a learner would stepwise approach the correct properties of the encoding schemes and the links between communicative functions and encoding schemes of the target language. Such schedule for example would explain the L2 performance for the usage of intensity for contrast encoding. More appropriate for the general observation that a part of the patterns found in L2 can be explained by transfer, and another part by acquisition, but yet no regularities of the choice were found, would be an assumption of a separate decision mechanism: randomly, or guided by rules that still have to be defined, the learner opts for the inventory of one or the other language. In these cases, one would assume that a learner has something like a concurrence of prosodic lexica and is lead to one or the other encoding scheme for the same communicative function. Transfer could be a result of an incorrect (or impossible) choice.

Xu explicitly mentions the possibility of "universal" encoding schemes, too. The findings of the investigation of the encoding of sentence mode and contrastive focus in L2 are also explainable as a recurrence on properties of the Biological Codes (see chapter 10. Hence, one other possibility of explaining the results for the L2 speakers would be to assume a concurrence of encoding schemes of the L1, L2, and universals, and analogous processes of choice: The universal preferences are used if the schemes of L1 do not seem to be appropriate in the perspective of the speaker, but the L2 patterns are not accessible in fluent speech.

There is another possibility how the findings of the L2 experiment could be explained within the PENTA model: The specific patterns of German learners of English could also be accounted for by differentiating the Melodic Primitives only, but no language choice. The

L2 speakers use their native encoding schemes, but strongly increase the strength of the encoding schemes. Thus, duration, pitch (and intensity) are increased compared to the L1, but fundamental contours like rising questions, and a complex rise-fall on objects under contrastive focus are not touched.

A decision between these three possibilities needs further research with more different L1-L2 language pairs.

11.3 L2 Prosody and the Speaking Model

This last section will pick up some aspects of Levelt's thourough 'Speaking' Model of human speech production and de Bot's extensions to L2 production presented in section 4.1 and 4.2. Both authors argue that the planning of the sentence mode takes place in the language-independent process of macroplanning in the Conceptualizer. De Bot assumes that the language to be produced is already selected there, too. Markers of information structural properties (especially focus) are set in the subsequent micro-planning. The still abstract information is collected in the preverbal message. Following de Bot, the point of decision between the potential languages to be produced is when the preverbal message enters language specific Formulators. There, after grammatical encoding, phonological encoding takes place, and the information about sentence mode and focus of the preverbal message is passed to the surface structure. Information of the surface structure, together with the metrical and segmental spellout and information about attitudes and emotions summarized as 'intonational meaning' are integrated incrementally for each syllable in the 'Prosody Generator'. Neither Levelt nor de Bot have detailed assumptions about the processes therein. But de Bot proposes a 'largely language specific' Prosody Generator: well known difficulties in the correct production of second language prosody should result from interferences in a unique, non-language specific Articulator.

How can the results of the experiments above be implemented, and do they add information to reshape the assumptions of the model?

I want to emphasize four points: They deal with the specific translation of information of the surface structure or intonational meaning: Variances of the metrical or segmental spellout can be ruled out as these were controlled by the read words of the utterances. The contexts for the utterances were designed to rule out variances in the surface structure. This leads to the first point:

Even though there was no reason for a choice of a different contour or extension of a acoustic parameter, variance within and between speakers of every speaker group was high. As Levelt predicts, there is no “intonational lexicon” with a fixed contour-meaning relationship. Contrarily, the findings suggest that principally the whole set of “canonical contours” is available for each position, focus condition, and sentence mode. However, it has also been shown that there are preferences in the choice of a perceivable contour, and differences in the usage of acoustic parameters that depend on the focus-condition, the sentence mode, and the position within an utterance. These preferences can be categorical like the usage of a rise instead of a fall for contrasted verbs or the final rise for questions by German native speakers, or gradual as a higher peak for contrasted subjects and objects in statements.

Second, the weaker effects of contrast in questions give rise to assume a certain hierarchy of information on the surface structure. The non-local, phrase wide information that the utterance to be produced is a question seems to be higher ranked than the information about corrective contrast on a constituent. It has to be tested whether this ranking is generalizable, or whether it is restricted to echo-questions: for the materials in the experiment, the focus structure is already clearly expressed in the preceding statements.

A third point is that Levelt’s assumption of strict incrementality is difficult to hold for prosody: He claims that the speaker knows what has been uttered, but the choice of an actual tone is not influenced by what will come next. On the one hand, this explain for instance deaccentuation effects in English, and is needed to account for spontaneous rephrasing, and probably to explain parts of the variance in the choice of tones. On the other hand, there are some instances in which preparation effects are clear: for example if the verb is spoken with a rise instead of the usual fall, the chance to have a falling or low preceding syllable (the subjects of my utterances) is increasing. This effect is overwritten by the strong rise in the beginning of echo-questions in English, but visible for the German native speakers, and especially clear for the German learners of English. It is a non-local preparation effect that occurs before the actual focus encoding on the verb takes place. It indicates that the position of a focus diacritic f on the surface structure must be accessible before the prosodic pattern of the actual syllable is generated.

And finally, the results give rise to a discussion of the relation of the “intonational meaning” and the surface structure. The sentences that have been analyzed and the contexts kept lexical sources of information (metrical and phonological spellout) and the information on surface structure constant within a group. However, considerable variance has been found in the prosodic patterns within and between speakers. Is the “intonational meaning” that carries information about attitudes and emotions a potential candidate of explanation ?

The emotional load of the utterances in contexts could not be controlled, and none of the subjects was a professional actor who may have learned standardized ways of encoding a certain amount and/or kind of emotion. Of course, some speakers or dialogues induced more emotional involvement than others, but the data was normally distributed so there was no reason to exclude certain speakers and/or dialogues from the analysis. Therefrom follows that the grade of variation has to be considered as “normal”.

The argumentation can be pushed further by assuming that the intonational meaning does not only account for the variation, but also for the preferred encoding patterns of the communicative functions. This would be in line with Bolinger’s perspective that prosody is mainly encoding emotions and attitudes instead of linguistic, and especially grammatical functions. Remember that the surface structure issues from Grammatical Encoding in the formulator. The echo-questions are charged with incredulity and surprise, and ask for confirmation of a fact that was already encountered. And corrective contrast induces emphasis, or expresses an attitude of special interest on the respective syllable whose amount is not predictable from the linguistic content and context alone. Hence, it is possible to assume a prosody generator without information of the surface structure, or at least that information of the ’intonational meaning’ component is more valuable than information of the surfuace structure for the Prosody Generator.

Such a position could become even more evident if we consider the encountered patterns of second language prosody. As stated in section 10.2, a realization of encodings plans directly derivable from the Biological Codes is a very good candidate for a complete explanation of L2 prosody. The Biological Codes, and more specifically the Frequency Code and the Effort Code, work with an assumption of grammaticalized encoding patterns from unconscious, rather emotional sources. If further research can prove that second language prosody production is better explained with an application of the Biological Codes, de Bot’s problem of assuming language specific Formulators – which the Prosody Generator is part of – but not being able to account for the discrepancy between the morphosyntactic, lexical, and phonological acquisition on the one side, and the prosodic performance on the other side, could be solved. The actual solution, pushing the source of interferences and low performance down to processes of the Articulator is not satisfactory because the actual ways of encoding do not pose an problem of articulation. For example, German speakers are obviously able to produce utterances with an higher overall level of the fundamental frequency with a fall at the end to imitate English echo-question patterns, and intensity is used to encode contrast in German as well. The L2 speakers use a different plan, and thus the problem is to be located within the Formulator.

To come to an end of the discussion, one can summarize that many aspects of a detailed model of second language prosody production have to be investigated further. Two main points have to remain open because the current models of prosody production cannot explain a large part of the data. The first is to bring more evidence on the source of L2 prosody: One has to test L1-L2 language pairings in which transfer and the recourse on the Biological Codes would provide more diverging predictions. This would bring more light into the relationship of concrete, language specific information on something like Levelt's surface structure, and a more universal aspect subsumed in the "intonational meaning". The other aspect has to do with the time course of intonational planning, more specifically the parallelism or hierarchical organization of the encoding of local (like contrast) and non-local (like sentence-mode) functions. It has been show that a simple addition of encoding schemes is not appropriate for the data, but this study can only state that for the echo-questions, sentence mode encoding is stronger than the focus structure. More differentiated materials should test whether the reason is that non-local functions tend to overwrite local functions, or, what is more probable, that the relative importance in the concrete situation favors an emphasis on the question mode and not a repetition of the focus structure of the preceding statement.

References

Alter, K., Mleinek, I., Rohe, T., Steube, A., & Umbach, C. (2001). Kontrastprosodie in Sprachproduktion und -perzeption. *Linguistische Arbeitsberichte*, *77*, 59-79.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgements of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*(4), 529-555.

Archibald, J. (1998). *Second language phonology.* Amsterdam: John Benjamins.

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, *26*, 3-25.

Backman, N. (1977). Learner intonation. A pilot study. In C. Henning (Ed.), *Proceedings of the second language research forum.* Los Angeles.

Bartels, C. (1997). *Towards a compositional interpretation of English statement and question intonation.* Unpublished doctoral dissertation, University of Massachusetts Amherst, Amherst / MA.

Bartels, C. (1999). *The intonation of English statements and questions: A compositional interpretation.* Garland Science.

Batliner, A. (1989a). Eine Frage ist eine Frage ist keine Frage. Perzeptionsexperiments zum Fragemodus im Deutschen. In H. Altmann, A. Batliner, & W. Oppenrieder (Eds.), *Zur Intonation von Fokus und Modus im Deutschen* (p. 87-109). Tübingen: Niemeyer.

Batliner, A. (1989b). Fokus, Modus und die grosse Zahl. Zur intonatorischen Indizierung des Fokus im Deutschen. In H. Altmann, A. Batliner, & W. Oppenrieder (Eds.), *Zur Intonation von Modus und Fokus im Deutschen* (p. 21-70). Tübingen: Niemeyer.

Batliner, A. (1989c). Wieviel Halbtöne braucht die Frage ? Merkmale, Dimensionen, Kategorien. In H. Altmann, A. Batliner, & W. Oppenrieder (Eds.), *Zur Intonation von Fokus und Modus im Deutschen* (p. 111-162). Tübingen: Niemeyer.

Baumann, S., Becker, J., Grice, M., & Mücke, D. (2007). Tonal and articulatory marking of focus in German. In *Proceedings of icphs xvi* (p. 1029-1032). Saarbrücken.

Baumann, S., Grice, M., & Benzmüller, R. (2001). GToBI - a phonological system for the transcription of German intonation. In S. Puppel & G. Demenko (Eds.), *Prosody 2000. Speech recognition and synthesis* (p. 21-28). Poznan: Adam Mickiewicz University, Faculty of Modern Languages and Literature.

Baumann, S., Grice, M., & Steindamm, S. (2006). Prosodic marking of focus domains - categorical or gradient. In *Proceedings of SpeechProsody* (p. 301-304). Dresden.

Beckman, M. E. (1986). *Stress and non-stress accent.* Dordrecht: Foris Publications.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In H. C. Nusbaum (Ed.), *The development of speech* *perception: The transition from speech sounds to spoken words* (p. 167-224). Cambridge / MA: MIT Press.

Best, C. T. . (1995). A direct realist perspective on cross-language speech perception. In W. Strange & J. J. Jenkins (Eds.), *Cross-language speech perception* (p. 171-204). Timonium / MD: York Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*(2), 775-794.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception on approximants. *Journal of Phonetics*, *20*, 305-330.

Biemans, M. (2000). *Gender variation in voice quality.* Utrecht: LOT.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5.9/*, 341-345.

Boersma, P., & Weenink, D. (2007). *Praat: doing phonetics by computer (V.4.6.34)* (Computer program). Retrieved October 19 2007, from http://www.praat.org.

Bohn, O.-S., & Flege, E. (1997). Perception and production of a new vowel category by adult second language learners. In A. James & J. Leather (Eds.), *Second language speech: Structure and process* (p. 53-73). Berlin / New York: de Gruyter.

Bolinger, D. (1957). *Interrogative structures of American English* (Vol. 28). Tuscaloosa: University of Alabama Press.

Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language*, *37*(1), 83-96.

Bolinger, D. (1972). Accent is predictable (if you're a mind reader). *Language*, *48*(3), 633-644.

Bolinger, D. (1978). Intonation across languages. In J. Greenberg (Ed.), *Universals of human language : Phonology* (Vol. 2, p. 471-524). Stanford: Stanford University Press.

Bolinger, D. (1982). Intontation and its parts. *Language*, *58*(3), 505-533.

Bolinger, D. (1987). Echoes reechoed. *American Speech*, *62*(3), 261-279.

Bolinger, D. (1989). *Intonation and its uses.* London: E. Arnold.

Braun, B. (2004). *Production and perception of contrastive and non-contrastive themes in German.* Doctoral dissertation, Universität des Saarlandes, http://www.phon.ox.ac.uk/b̃ettina/thesis.html.

Broselow, E., Chen, S.-I., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition*, *20*, 261-280.

Chen, A. (2003). Language dependence in continuation intonation. In *Proceedings of the 15th international congress of phonetic sciences.*

Chen, A. (2005). *Universal and language-specific perception of paralinguistic intonational meaning.* Utrecht: LOT.

Chen, A., Gussenhoven, C., & Rietveld, T. (2002). Language specific uses of the Effort Code. In *Speech prosody 2002: Proceedings of the first international conference on speech prosody* (p. 47-57). Aix en Provence.

Chen, A., & Mennen, I. (2008). Encoding interrogativity intonationally in a second language. In *Proceedings of the 4th international conference on speech prosody.* Campinas, Brasil.

Chen, A., Rietveld, T., & Gussenhoven, C. (2001). Language-specific effects of pitch range on the perception of universal intonational meaning. In *Proceedings of the 9th Eurospeech* (p. 1403-1406).

Chen, Y., Robb, M. P., Gilbert, R., Harvey, & Lerman, J. W. (2001). A study of sentence stress production in Mandarin speakers of American English. *Journal of the Acoustical Society of America*, *109*(4), 1681-1690.

Ching, M. K. (1982). The question intonation in assertions. *American Speech*, *57*(2), 95-107.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America*, *77*(6), 2142-2156.

Corder, S. P. (1994). A role for the mother tongue. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language learning* (p. 18-31). Amsterdam: John Benjamins.

Cruttenden, A. (1997). *Intonation.* Cambridge: Cambridge University Press.

DeKeyser, R., & Larson-Hall, J. (2005). What does the Critical Period really mean. In *Handbook of bilingualism: Psycholinguistic aproaches* (p. 88-108). New York: Oxford University Press.

Delattre, P. (1965). *Comparing the phonetic features of English, German, Spanish and French.* Heidelberg: Groos.

Dietrich, R. (1990). Zur Form und Bedeutung der Kontrastintonation im Deutschen. *Linguistische Berichte*, *129*, 415-430.

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, *80*(2), 402-415.

Els, T. van, & Bot, K. de. (1987). The role of intonation in foreign accent. *The Modern Language Journal*, *71*(2), 147-155.

Elsner, A. (2000). *Erkennung und Beschreibung des prosodischen Fokus.* Doctoral dissertation, Friedrich-Wilhelms-Universität Bonn, http://hss.ulb.uni-bonn.de/diss_online/phil_fak/2000/elsner_anja/.

Fant, G., & Kruckenberg, A. (1989). Preliminaries to the study of Swedish prose reading and reading style. *Quaterly Progress and Status Report*, *30*(2), 1-80.

Féry, C. (1993). *German intonational patterns.* Tübingen: Niemeyer.

Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Language Learning*, *30*, 117-134.

Flege, J. E. (1987). The production of 'new' vs. 'similar phones in a foreign language: Evidence for the effects of equivalence classification. *Journal of Phonetics*, *15*, 47-65.

Flege, J. E. (1997). English vowel production by Dutch talkers: more evidence for the 'similar' vs. 'new' distinction. In A. James & J. Leather (Eds.), *Second language speech: structure and process* (p. 12-52). Berlin: de Gruyter.

Flege, J. E. (2003). Second language segmental production and perception. In N. Schiller & A. Meyer (Eds.), *Assessing constraints on second-language segmental production and perception* (p. 319-355). Berlin: Mouton de Gruyter.

Flege, J. E., & Efting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*, 67-83.

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America*, *76*(3), 708-721.

Flege, J. E., & Hillenbrand, J. (1987). Limits on phonetic accuracy in foreign language speech production. In G. Ioup & S. H. Weinberger (Eds.), *Interlangugae phonology: The acquisition of a second language sound system* (p. 176-203). Cambridge / Mass.: Newbury House Publishers.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native italian speakers' perception and production of english vowels. *Journal of the Acoustical Society of America*, *106*(5), 2973-2987.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent. *Journal of the acoustical society of America*, *97*(5), 3125-3134.

Fries, C. (1964). On the intonation of yes-no questions in English. In D. Abercrombie, D. Fry, P. McCarthy, N. Scott, & J. Trim (Eds.), *In honor of Daniel Jones* (p. 242-254). London: Longman.

Fujisaki, H. (1980). From discrete linguistic units to continous speech characteristics - a functional formulation of articulatory and phonatory dynamics. *Research on human information processing*, *1*, 1-11.

Fujisaki, H., & Hirose, K. (1982). Modeling the dynamic characteristics of voice fundamental frequency witch application to analysis and synthesis of intonation. In *Papers for the working group on intonation* (p. 57-70). Tokyo.

Fujisaki, H., & Hirose, K. (1984). Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan*, *5*(4), 233-241.

Fujisaki, H., Hirose, K., & Sugito, M. (1980a). Comparison of word accent features in English and Japanese. *Research on human information processing.*, *1*, 27-36.

Grabe, E. (1998). *Comparative intonational phonology: English and German.* Wageningen: Ponsen en Looien.

Grabe, E. (2004). Intonational variation in urban dialects of English spoken in the British Isles. In P. Gilles & J. Peters (Eds.), *Regional variation in intonation* (p. 9-31). Tübingen: Niemeyer.

Gårding, E. (1981). Contrastive prosody: a model and its application. *Studia Linguistica, 35*(1-2), 146-165.

Grice, M., & Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte, 191*, 267-298.

Grice, M., Baumann, S., & Benzmüller, R. (2005). German intonation in autosegmental-metrical phonology. In S.-A. Jun (Ed.), *Prosodic typology* (p. 55-83). Oxford: Oxford University Press.

Grice, M., & Benzmüller, R. (1995). Transcription of German intonation using ToBI-tones – the Saarbrücken system. *Phonus, 1*, 33-51.

Grosser, W. (1982). *Intonatorische Interferenzerscheinungen Deutsch-Englisch: Eine experimentallinguistische Untersuchung.* Wien: VWGÖ.

Grosser, W. (1997). On the acquisition of tonal and accentual features of English by Austrian learners. In A. James & J. Leather (Eds.), *Second language speech: structure and process* (p. 211-228). Berlin: De Gruyter.

Gussenhoven, C. (2002). Intonation and interpretation: phonetics and phonology. In *Proccedings of Speech Prosody* (p. 47-57). Aix en Provence.

Gussenhoven, C. (2004). *The phonology of tone and intonation.* Cambridge: Cambridge University Press.

Gussenhoven, C., & Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. In *Proceedings of ICSLP* (p. 91-94). Bejing.

Gut, U. (2000). The production of emphasis by German learners of English. In *Proceedings of New Sounds* (p. 155-157). Amsterdam.

Gut, U. (2003). Non-native speech rhythm in German. In *Proceedings of the 15th ICPhS* (p. 2437-2440). Barcelona.

Gut, U., & Beyerl, P. S. (2004). Measuring the reliability of manual annotations of speech corpora. In *Proceedings of speech prosody* (p. 565-568). Nara.

Haan, J. (2002). *Speaking of questions: An exploration of Dutch question intonation.* Utrecht: LOT.

Hirschfeld, U., Reinke, K., & Stock, E. (Eds.). (2007). *Phonetik Intensiv: Aussprachetraining.* München: Langenscheidt.

Höhle, T. (1982). Explikationen für "normale Betonung" und "normale Wortstellung". In W. Abraham (Ed.), *Satzglieder im Deutschen: Vorschläge zur syntaktischen, seman-*

tischen und pragmatischen Fundierung (p. 75-154). Tübingen: Gunter Narr.

Höhle, T. (1992). Über Verum-Fokus im Deutschen. In J. Jacobs (Ed.), *Informationsstruktur und Grammatik* (p. 112-141). Opladen: Westdeutscher Verlag.

Iwata, S. (2003). Echo questions are interrogatives ? Another version of a metarepresentational analysis ? *Linguistics and Philosophy*, *26*, 185-254.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar.* Cambridge, MA: MIT Press.

Jacobs, J. (1988). Probleme der freien Wortstellung im Deutschen. *Sprache und Pragmatik: Arbeitsberichte*, *5*, 8-37.

Jenner, B. (1976). Interlanguage and foreign accent. *Interlanguage Studies Bulletin*, *1*(2-3), 166-196.

Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent.* Doctoral dissertation, University of Stuttgart.

Jilka, M., & Möhler, G. (1999). Intonational foreign accent: speech technology and foreign language teaching. In *Proceedings of the ECSA workshop on speech technology and language learning* (p. 115-118). Marholmen.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language processing. *Cognitive Psychology*, *21*, 60-99.

Jong, K. de. (1994). Initial tones and prominence in Seoul Korean. In *Working papers in linguistics* (Vol. 43, p. 1-14).

Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *IRAL (International Review of Applied Linguistics in Language Teaching*, *28*(2), 99-118.

Kaunzner, U. A. (1997). *Aussprachekurs Deutsch: ein komplettes Übungsprogramm zur Verbesserung der Aussprache für Unterricht und Selbststudium.* Heidelberg: Julius Groos.

Kelm, O. (1987). An acoustic study on the differences of contrastive emphasis between native and non-native spanish speakers. *Hispania*, *70*(3), 627-633.

Kiss, K. É. (1998). Identificational focus vs. information focus. *Language*, *74*, 245-273.

Klein, W., & Perdue, C. (1997). The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research*, *13*(4), 301-347.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: fundamental frequency lends little. *Journal of the Acoustical Society of America*, *11*(2), 1038-1054.

Kohler, K. (1987). Categorical pitch perception. In *Proc. xith international congress of phonetic sciences* (p. 331-333).

Kohler, K. (2004). Categorical speech perception revisited. In *Poster presented at From*

Sound to Sense: 50+ years of discoveries in speech communication. MIT.

Ladd, D. R. (1996). *Intonational phonology.* Cambridge: Cambridge University Press.

Lado, R. (1957). *Linguistics across cultures.* Ann Arbor: University of Michigan Press.

Lado, R. (1964). *Language teaching: A scientific approach.* New York: McGraw-Hill.

Lemke, S., Graubner, D., & Lüssing, P. (Eds.). (2006). *Sprechwissenschaft /Sprecherziehung: Ein Lehr- und Übungsbuch.* Berlin: Peter Lang.

Lepetit, D. (1989). Cross-linguistic influence in intonation: French/Japanese and French/English. *Language Learning*, *39*(3), 397-414.

Lepetit, D., & Martin, P. (1990). Étude différentielle intonative français/anglais. *IRAL (International Review of Applied Linguistics in Language Teaching*, *38*(2), 135-152.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Lewis, D. (1969). *Convention.* Cambrige / Mass.: Harvard University Press.

Lieberman, P. (1967). *Intonation, perception and language.* Cambridge, MA: MIT Press.

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, *62*, 70-87.

Lyons, J. (1977). *Sematics.* Cambridge: Cambridge University Press.

Mairs, J. L. (1989). Stress assignment in interlanguage phonology: an analysis of the stress system of spanish speakers learning english. In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (p. 260-283). Cambridge: Cambridge University Press.

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology.* London: Lawrence Erlbaum Associates.

Makarova, V. (2001). Perceptual correlates of sentence-type intonation in Russian and Japanese. *Journal of Phonetics*, *29*, 137-154.

McAlister, R., Flege, J. E., & Piske, T. (1999). The acquisition of Swedish long vs. short vowel contrasts by native speakers of English, Spanish and Estonian. In *Proceedings of icphs 99* (p. 1357-1360). San Francisco.

Mennen, I. (2004). Bi-directional interferences in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, *32*, 543-563.

Mixdorff, H. (1996). Foreign accent in intonation patterns – a contrastive study applying a quantitative model of the F0 contour. In *Proceedings of the ICSLP* (p. 1469-1472.).

Mixdorff, H. (1998). *Intonation patterns of German - model-based quantitative analysis and synthesis of f_0 contours.* Doctoral dissertation, Technische Universität Dresden, http://www.tfh-berlin.de/˜mixdorff/thesis/index.html.

Mixdorff, H. (2002). *An integrated approach to model German prosody.* Dresden: TUD-press.

Möbius, B., Pätzold, M., & Hess, W. (1993). Analysis and synthesis of German f_0 contours by means of Fujisaki's model. *Speech Communication*, *13*, 53-61.

Molnár, V. (2006). On different kinds of contrast. In V. Molnár & S. Winkler (Eds.), *The architecture of focus* (p. 197-233). Berlin: Mouton de Gruyter.

Neufeld, G. G. (1987). On the acquisition of prosodic and articulatory features in adult language learning. In G. Ioup & S. H. Weinberger (Eds.), *Interlangugae phonology: The acquisition of a second language sound system* (p. 321-332). Newbury House Publishers.

Ohala, J. J. (1983). Cross-language use of ptich: An ethological view. *Phonetica*, *40*, 1-18.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilisation of f0 of voice. *Phonetica*, *41*, 1-16.

Oppenrieder, W. (1989a). Deklination und Satzmodus. In H. Altmann, A. Batliner, & W. Oppenrieder (Eds.), *Zur Intonation von Fokus und Modus im Deutschen* (p. 245-266). Tübingen: Niemeyer.

Oppenrieder, W. (1989b). Fokus, Fokusprojektion und ihre intonatorische Kennzeichnung. In H. Altmann, A. Batliner, & W. Oppenrieder (Eds.), *Zur Intonation von Modus und Fokus im Deutschen* (p. 267-280). Tübingen: Niemeyer.

Parker, F., & Pickeral, J. (1985). Echo questions in English. *American Speech*, *60*(4), 337-347.

Pechmann, T. (1984). *Überspezifizierung und Betonung in referentieller Kommunikation.* Doctoral dissertation, University of Mannheim, Kleve.

Peters, J., Gilles, P., Auer, P., & Selting, M. (2002). Identification of regional varieties by intonational cues: An experimental study on Hamburg and Berlin German. *Language and Speech*, *45*(2), 115-139.

Pierrehumbert, . (1980). *The phonology and phonetics of English intonation.* Doctoral dissertation, MIT, Cambridge, MA. (Unpublished PhD - Thesis)

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (p. 271-311). Cambridge, MA: MIT Press.

Pike, K. (1945). *The intonation of American English.* Ann Arbor: University of Michigan Press.

Pürschel, H. (1975). *Pause und Kadenz: Interferenzerscheinungen bei der englischen Intonation deutscher Sprecher.* Tübingen: Niemeyer.

Quirk, R., Greenbaum, S., Leech, G. N., Svartvik, J., & Crystal, D. (1985). *A comprehensive grammar of the English language.* London: Longman.

Reyelt, M., Grice, M., Benzmüller, R., Mayer, J., & Batliner, A. (1996). *Prosodische Etikettierung des Deutschen mit ToBI* (Report No. 154). Universität Braunschweig, Universität Saarbrücken, Universität Stuttgart, Universität München.

Rietveld, T., Haan, J., Heijmans, L., & Gussenhoven, C. (2002). Explaining attitudinal ratings of Dutch raising contours: Morphological structure vs. the Frequency Code. *Phonetica*, *59*, 180-194.

Rossi, M. (1971). Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole. *Phonetica*, *23*, 1-33.

Santen, J. van, & Shih, C. (2000). Suprasegmental and segmental timing models in Mandarin Chinese and American English. *Journal of the Acoustical Society of America*, *107*, 1012-1026.

Searle, J. (1969). *Speech acts: An essay in the philosophy of language.* Cambridge: Cambridge University Press.

Selkirk, E. O. (1984). *Phonology and syntax : the relation between sound and structure.* Cambridge / MA: MIT Press.

Shah, A. (2004). Production and perceptual correlates of spanish-accented English. In *Proceedings of Sound to Sense* (p. 79-84). Boston.

Shen, X. S. (1990). Ability of learning the prosody of an intonational language by speakers of a tonal language: Chinese speakers learning French prosody. *IRAL (International Review of Applied Linguistics in Language Teaching*, *28*(2), 119-134.

Shen, X. S. (1993). The use of prosody in disambiguation in Mandarin. *Phonetica*, *50*(4), 261-271.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., et al. (1992). Tobi: a standard for labeling English prosody. In *Proceedings of the ICSLP* (p. 867870). Banff.

Silverman, K. E. A., & Pierrehumbert, J. B. (1990). The timing of prenuclear high accents in English. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I* (p. 72-106). Cambridge, UK: Cambridge University Press.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition.* Oxford: Blackwell.

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, *31*(4), 649-489.

Steube, A. (2001). Correction by contrastive focus. *Theoretical Linguistics*, *27*(2/3), 215-249.

Syrdal, A. K., & Jilka, M. (2004). Acceptability of variations in question intonation in natural and synthetisized American English. *Journal of the Acoustical Society of America*, *115*(5), 2543.

Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics.* Berkeley: University of California Press.

Trim, J. (1988). Some contrastive intonated features of british english and german. In J. Klegraf & D. Nehls (Eds.), *Essays on the english language and applied linguistics on* *the occasion of gerhard nickel's 60th birthday* (p. 235-249.). Heidelberg: Julius Groos.

Ueyama, M. (2000). *Prosodic transfer: An acoustic study of prosodic transfer of L2 English vs. L2 Japanese.* Doctoral dissertation, University of California, Los Angeles, Los Angeles.

Ueyama, M., & Jun, S.-A. (1998). Focus realization in Japanese English and Korean English intonation. *Japanese and Korean Linguistics*, *7*, 629-645.

Uldall, E. (1964). Dimension of meaning in intonation. In D. Abercrombie, D. B. Frey, P. McCarthy, N. Scott, & J. Trim (Eds.), *In honor of Daniel Jones: Papers contributed on the occasion of his eightieth birthday* (p. 271-279). London: Longman.

Wagner, P. S. (2002). *Vorhersage und Wahrnehmung deutscher Betonungsmuster*. Doctoral dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn.

Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly*, *4*(2), 123-130.

Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton.

Wennerstrom, A. (1994). Intonational meaning in English discourse. *Applied Linguistics*, *15*(4), 399-420.

Willems, N. (1982). *English intonation from a Dutch point of view.* Berlin: Mouton de Gruyter.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f_0 contours. *Journal of Phonetics*, *27*, 55-105.

Xu, Y. (2005). Speech melody as articulatory implemented communicative functions. *Speech Communication*, *46*, 220-251.

Yeou, M. (2004a). Effects of focus, position and syllable structure on F0 alignment patterns in Arabic. In *Proceedings of JEP-TALN.* Fez.

Yeou, M. (2004b). The realization of accentual focus in Moroccan learners of English. In *Specch prosody 2004* (p. 85-88). Nara.

A English Materials

Sam killed Frank

CF0

A I missed the first part of that thriller on TV last night. Why were the police in that apartment?

B Well... Sam killed Frank.

A What did you say ? Sam killed Frank ?

B Yes, and the police arrested Sam right there.

CF1

A I missed the start of that thriller on TV last night. Was it Rocco who killed Frank?

B Oh, not that wasn't Rocco. Sam killed Frank !

A What ? Sam killed Frank ? I thought they liked each other.

B That's right. But Frank knew about Sam's past, so Frank feared he would tell the police about him...

CF2

A What's all the commotion in the film just now? I missed a part. I saw that Frank fell into the water and Sam came to his rescue...

B But Sam didn't help Frank. Sam killed Frank !

A What ? Sam killed Frank ? But it looked like he was helping him.

B Yes, but Frank knew too much about Sam's past, so Frank killed him to keep it covered up.

CF3

A I missed the first part of that thriller. Sam seems to be in quite a lot of trouble. Did he kill that police officer ?

B Oh no, not the police officer. Sam killed Frank !

A What ? Sam killed Frank ? Why would he kill his best friend ?

B Frank knew too much about Sam's past, so Frank killed him to keep it covered up.

Flo saw Kurt

CF0

A Do you have any news from home ?

B Oh, yes. Flo saw Kurt.

A Really ? Flo saw Kurt? I thought that they both moved away a long time ago ?

B Yes, but they met up accidentally at Heathrow Airport.

CF1

A How do you know that Kurt's wife is having a baby in September ? Did you meet him last week?

B No, I didn't see him myself. Flo saw Kurt.

A What ? Flo saw Kurt ? How could that be? Isn't she overseas?

B She met up with Kurt at Heathrow Airport last week.

CF2

A How does Flo know that Kurt's wife is having a baby in September ? Did she call him last week ?

B No, she couldn't call anybody. Her mobile was stolen. Flo saw Kurt !

A What ? Flo saw Kurt ? But she has been living in the US since last year...

B Yes, but she was in England to meet some customers. So she dropped by to see his old friend...

CF3

A How does Flo know that Kurt's wife is having a baby in September? Did she meet up with her ?

B No, Flo didn't see her. Flo saw Kurt !

A What ? Flo saw Kurt ? How come ? I thought Kurt is working abroad this year ?

B Yes, but they met up at the Airport restaurant at Heathrow.

Clark fouled Steve

CF0

B What's that commotion on the football field ? What happened ?

A Clark fouled Steve.

B What did you say ? Clark fouled Steve ?

A Yes, and it was a bad foul. He kicked him right in the ankle.

CF1

B I hear that Steve got injured in a horrible foul in yesterday's match. I bet it was the new defender, Franco, right?

A No, Franco was nowhere near him. Clark fouled Steve !

B What ? Clark fouled Steve ? But Clark is usually such a fair player.

A Yes, I don't know why he kicked him so hard in the ankle.

CF2

B I just missed that last minute of the game. Why are Clark and Steve lying on the field ? Did Clark block Steve ?

A No, Clark didn't block Steve. Clark fouled Steve !

B What ? Clark fouled Steve ? He is usually such a fair player !

A Yes, but he kicked him hard right in the ankle...

CF3

B Who is the poor guy that Clark fouled so badly on the football field. That's Franco, isn't it ? I can tell by the black hair.

A No, that's not Franco. Clark fouled Steve !

B What ? Clark fouled Steve ? But Steve has blond hair, not black.

A Maybe he dyed it. But that's Steve, I'm sure...

Ken saved Fran

CF0

A Do you have news about the sailboat accident ?

B Yes, I do. Ken saved Fran.

A Really ? Ken saved Fran. Oh, lucky team.

B Yes, they are. They may still even be able to finish the race.

CF1

A Do you have news about the sailboat accident ? I heard that Kerry saved Fran who was drowning.

B Yes, Fran was saved, but not by Kerry. Ken saved Fran !

A What ? Ken saved Fran ? I didn't know that Ken was with the boat nearby ?

B Well, he joined them at the very last moment. ..

CF2

A Why is Fran so angry with Ken ? Did he leave her alone in the sinking sailboat ?

B But Ken didn't leave her. Ken saved Fran !

A What ? Ken saved Fran ? But why is she so angry anyway.

B Because Ken caused the sailboat accident in the first place and they had to forfeit the race...

CF3

A Do you have news about the sailboat accident ? I heard that Ken saved Kerry from the sinking ship.

B Ken got somebody out of the water, but not Kerry. Ken saved Fran !

A What ? Ken saved Fran ? I didn't realize that Fran was part of the sailing crew..

B Oh, yes he was. He joined them just before departure...

Seth found Carl

CF0

B Did you hear about the avalanche accident in the mountains ? Weren't some of our colleagues involved in that? I heard that some of them were rescued successfully.

A Yes. Seth found Carl.

B What did you say ? Seth found Carl ?

A Yes. Carl was lucky that he had his red jacket on, so he was easy to find...

CF1

B I heard that Christian found Carl after that horrible avalanche. Is that true?

A Yes, but it was not Christian who found him. Seth found Carl !

B What ? Seth found Carl ? I didn't know that he was part of the rescue team ?

A Yes, in fact he joined them two years ago.

CF2

B I heard that Seth rescued Carl after the avalanche. Is that true ?

A Well not exactly. He didn't rescue him. Seth found Carl.

B What ? Seth found Carl ?

A Right. Once the snow had settled, Seth found Karl because he could see his red jacket.

CF3

B I heard that Seth found Christian after the avalanche accident. Is that true?

A No, Seth didn't find Christian. Seth found Carl !

B What ? Seth found Carl ? I didn't even know that Carl was among the hiking team. Is he OK ?

A Oh yes, he is very lucky. He just had a broken leg.

Fred kissed Sue

CF0

B Oh, and what happend later that night?

A Well, Fred kissed Sue.

B Really ? Fred kissed Sue ?

A Oh yes, and they're really in love now.

CF1

B Who was the guy who kissed Sue last night. Wasn't that Harry?

A Oh no, it wasn't Harry who kissed Sue. Fred kissed Sue !

B What ? Fred kissed Sue ? I thought he didn't like her.

A Well, apparently he changed his mind about her.

CF2

B I heard that Fred met Sue in the disco last night. Did you see them talking to each other ?

A Yes, but they were not only talking. Fred kissed Sue !

B What ? Fred kissed Sue ? I thought that he was Mary's boyfriend...
A That's what I thought too. But he kissed her right on the dance floor...

CF3

B Who was the girl that Fred kissed last night ? Was that Marion?
A Oh no, not Marion. Fred kissed Sue !
B What ? Fred kissed Sue ? I thought she was Christian's girlfriend.
A Well, yes, but Fred doesn't seem to care...

Luke met Ben

CF0

A Do you have any news about our class mates from the school reunion ?
B Oh, yes. Luke met Ben.
A What ? Luke met Ben ? And what did he say ?
B Well, he told me that half of the guys are already married. Alfred, Tino, Harry...

CF1

A Do you know who has news about Ben. Has Alfred maybe met him at the bowling club ?
B No, Alfred hasn't seen Ben for years. Luke met Ben !
A What ? Luke met Ben ? But Luke never goes bowling, does he ?
B Oh, no: but he met Ben at the train station, on Monday morning.

CF2

A How does Luke know that Ben has married Tina ? Did he call him on his birthday ?
B Oh, no. He didn't call him. Luke met Ben !
A What ? Luke met Ben ? How could they meet ? Ben is in the US, and Luke is working in London, isn't he ?
B Well, they ran into each other at Heathrow. Ben had to see some clients in London...

CF3

A How does Luke know that Tina got married last week ? Did he meet her at the tennis club yesterday ?
B No, he hasn't seen Tina for weeks. Luke met Ben !
A What ? Luke met Ben ? I thought Ben is living in London now.

B Yes, but they met at a school reunion last Saturday.

Bert missed Lou

CF0

A Good to see you once again. I heard that some of your family was overseas last year. How did it go ?

B Well, all right, all right. Bert missed Lou.

A What ? Bert missed Lou ? But didn't she return home every 2 weeks ?

B Yes, but that was not enough for Bert...

CF1

A How did your sons get along now that their mother is working in the US ? Didn't little Brian miss Lou very much ?

B Brian was OK. Bert missed Lou !

A What ? Bert missed Lou ? But he has never been such a mama's boy... and he is 14 now...

B Maybe: But still he wanted to talk to her every day...

CF2

A Bert must have been a very happy man since Lou left him. They were fighting all the time...

B Well. But he wasn't happy at all. Bert missed Lou !

A What ? Bert missed Lou ? Why that ?

B I don't know. He said that he loved her despite all the problems they had...

CF3

A Didn't Bert miss his mother since she left home to live with her new boyfriend.

B Well, he was alright without his mother. But the cat ! Bert missed Lou !

A What ? Bert missed Lou ? Bert missed the cat more than his mother ?

B Well, he certainly asked more about the cat than about his mom...

Mark booked Lars

CF0

B What about the musicians for the big wedding ?

A Well... Mark booked Lars.

B Really ? Mark booked Lars ? That should be nice ...

A Yes, we hope so. His music is fun for everyone.

CF1

B Who booked Lars and his Jazz-Band for your wedding ? Did Ernest have that crazy idea ?

A No, Ernest really didn't want him. Mark booked Lars !

B What ? Mark booked Lars ? But Mark usually prefers Rock-Music to Jazz..

A I don't know. But he quite liked Lars's music the other day...

CF2

B That's great ! Lars is playing at Mark's wedding. Did he volunteer to play for free ?

A No, no. Mark booked Lars !

B What ? Mark booked Lars ? How can he afford it ?

A Well, Mark said : good music is always worth the extra money...

CF3

B Who will play at Mark's barbecue. Did he book Ernest, the jazz-man ?

A No, he didn't book Ernest. Mark booked Lars !

B What ? Mark booked Lars ? I didn't know that Mark likes Rock'n'Roll ...

A Yes, usually he prefers smooth piano-jazz. But for the party, Lars is more appropriate...

Liv brought Mel

CF0

B How did our friends get to the castle yesterday ?

A Well... Liv brought Mel.

B What ? Liv brought Mel ? I didn't know that they were supposed to go there...

A Actually yes. They were because I invited them last week...

CF1

B How did Mel get to the castle ? Did Caroline bring her ?
A No, it wasn't Caroline who brought her. Liv brought Mel !
B What ? Liv brought Mel ? But she doesn't even have a driver's license !
A Actually, yes. She gave her a ride on her motorcycle...

CF2

B I heard that Mel doesn't have a car now, but she wanted to be at the party until quite late last night. Did Liv drive Mel home ?
A No, Liv didn't drive Mel home. Liv brought Mel !
B What ? Liv brought Mel ? But she wanted to drive her home too, didn't she ?
A Yes, but she was too drunk to drive.

CF3

B Who did Liv bring to the party ? Did she come with Caroline ?
A No, Liv didn't bring Caroline. Liv brought Mel !
B What ? Liv brought Mel ? But they don't live in the same city, do they ?
A No, they don't, but Liv was staying in Cambridge for a job interview, anyway.

Bess loves Matt

CF0

B I didn't get that movie. Why is everybody so unhappy ?
A Bess loves Matt.
B What ? Bess loves Matt ? I must have missed that bit.
A Well, Bess pretends that she doesn't love him, but everything she does is for his love...

CF1

B I didn't get that movie. Angela loves Matt and ...
A No, Angela doesn't. Bess loves Matt !
B What ? Bess loves Matt ? But she betrayed him in the end..
A Yes, because he loves Angela...

CF2

B I didn't get that movie. Why does Bess hate Matt so much ?

A But Bess doesn't hate Matt ! Bess loves Matt !

B What ? Bess loves Matt ? But she will betray him in the end...

A Yes, because he only has eyes for Angela, and Bess is jealous...

CF3

B I didn't get that movie. Bess loves Charles and...

A No, Bess doesn't love Charles. Bess loves Matt !

B What ? Bess loves Matt ? But she betrays him in the end..

A Yes, because HE only has eyes for Angela and not Bess...

Mike lost Bill

CF0

A How was your trip to Italy ?

B It got off to a bad start. Mike lost Bill.

A What ? Mike lost Bill ? How come ?

B They were both in a crowded museum and then Bill went to look for something. And we were only able to find him 2 hours later...

CF1

A I heard that you had a really bad incident during your holidays in Italy. Bill got lost in a crowd. Was it you who lost him ?

B No, not me ! Mike lost Bill !

A What ? Mike lost Bill ? But he usually keeps a careful eye on his son...

B Yes, but they were queuing at a museum, and Bill suddenly wandered away..

CF2

A I've heard about the trouble during your holiday. Bill got lost. Was it Mike who found him again ?

B No. Mike lost Bill !

A What ? Mike lost Bill ? But he usually keeps a careful eye on his son...

B Yes, but in the crowd at a museum, Bill suddenly wandered away ...

CF3

A I've heard that Mike lost one of your children during your holiday in Italy. Was it Melanie who got lost ?

B No, Mike didn't lose Melanie. Mike lost Bill!

A What ? Mike lost Bill ? But Bill is such a shy little boy. He usually stays right beside his father

B Yes, but they were in a crowd at the museum, and Bill suddenly followed an ice cream vendor...

Pam knows Dave

CF0

A It's quite friendly at this welcome party, isn't it?

B Oh, yes. Pam knows Dave.

A Really ? Pam knows Dave ? How come ?

B Well, I think they were class mates in elementary school..

CF1

A Who of you knows Dave, the owner of the flat ? Do you know him ?

B No, I don't know him. Pam knows Dave !

A What ? Pam knows Dave ? But Pam only recently moved here ?

B Yes, but they met at the tennis club...

CF2

A Look ! Pam and Dave are having an intense conversation at this party. Didn't they just meet tonight ?

B No, no. Pam knows Dave !

A What ? Pam knows Dave ? But didn't she just recently move here ?

B Yes, but they were class mates at elementary school...

CF3

A How come that Pam is invited to your party ? Does she know you when you worked at the linguistics department ?

B Oh, no. Pam doesn't know me. Pam knows Dave !

A What ? Pam knows Dave ? But he only recently moved here.

B Oh, yes, but Dave was one of her classmates in elementary school...

Pete dressed Nick

CF0

B What happend at the final show at fashion school ?
A Pete dressed Nick.
B Oh, really ? Pete dressed Nick ? I didn't know that he is doing men's fashion, too..
A Oh, he only had one leisure suit to show...

CF1

B Who made Nick's outfits at the fashion show ? Did Myriam dress him?
A No, Myriam didn't dress Nick. Pete dressed Nick !
B What ? Pete dressed Nick ? I didn't know that Pete was doing man's fashion, too.
A Oh, he only had that one leisure suit to show.

CF2

B Who has done Nick's fabulous haircut at the fashion show ? Was that Pete ?
A Oh, no Pete didn't do Nick's haircut. Pete dressed Nick !
B What ? Pete dressed Nick ? I didn't know that he is doing clothes...
A Well, he only had that one leisure suit to show.

CF3

B Who was Pete's model at the fashion show ? Did he dress Walter ?
A No, Pete didn't dress Walter. Pete dressed Nick !
B What ? Pete dressed Nick ? I didn't know that Nick is modelling, too.
A Well, he only had that one leisure suit by Pete to show...

Neil draws Pat

CF0

B What are they always doing in the art studio ?
A Neil draws Pat.
B Really ? Neil draws Pat ? I didn't know that Neil is a sketch artist..
A Well, he wanted to give it a try...

CF1

B Who is the guy who will draw Pat's portrait ? Is it Jerry ?

A No, he doesn't like Jerry's style. Neil draws Pat !

B What ? Neil draws Pat ? I didn't know he has the skills ...

A Well, he wanted to give it a try ...

CF2

B It's been two weeks since Neil hired Pat. Will his sculpture be ready ?

A No, because Neil doesn't sculpt Pat. Neil draws Pat !

B What ? Neil draws Pat ? I thought he wanted a sculpture

A Well, but Pim decided that a sculpture was too expensive...

CF3

B Whose portrait has Neil been drawing all these weeks ? Is that Jerry ?

A No, that's not Jerry. Neil draws Pat !

B What ? Neil draws Pat ? I don't see any ressemblance, looks like Jerry, doesn't it ?

A Well, but he wanted to have something more abstract. And Neil is an absolute beginner, after all...

Nat payed Dirk

CF0

A What was done to take care of the band at the wedding ?

B Nat payed Dirk.

A Really ? Nat payed Dirk ? I thought the band had been paid when they were booked ?

B Oh, no, they wanted to handle it after the performance...

CF1

A Who payed Dirk, the musician at your wedding party ? The bride father is in charge, isn't he ?

B Oh, he didn't want to pay for that kind of music. Nat payed Dirk !

A What ? Nat payed Dirk ? But he has no money at all.

B Yes, but he insisted on paying him all the band's expenses...

CF2

A How did Nat get Dirk to play for her wedding ? Did he do it as a favour ?

B No, no. Nat payed Dirk !

A What ? Nat payed Dirk ? But he is broke..

B Yes, but for the wedding, he wanted something really special...

CF3

A The band is leaving now. Has Nat already payed the manager ?

B No, Nat didn't pay the manager. Nat payed Dirk !

A What ? Nat payed Dirk ? Who in God's name is Dirk ?

B Dirk is the band's piano player and in charge of the finances...

Dolph needs Paul

CF0

B Why did that guy from the Programming Department call ?

A Dolph needs Paul.

B What did you say: Dolph needs Paul ? What for ?

A Only Dolph knows the specifications of the client's system environment for the software.

CF1

B Who of the guys from the Programming Services department has called Paul away so urgently ? Does Lionel need him ?

A No, Lionel doesn't need Paul. Dolph needs Paul !

B What ? Dolph needs Paul ? But Dolph is an expert himself...

A Yes, but he has some problems with the system environment he cannot solve on his own...

CF2

B Dolph has to dismiss several people from the programming services department to cut down costs. Do you think Paul will will have to go too ?

A Oh,no, Dolph won't fire Paul. Dolph needs Paul !

B What ? Dolph needs Paul ? Please, just about anybody knows how to script a database surface...

A Yes, but Paul knows many of the clients. He's good for sales, too.

CF3

B Who do you think Dolph needs to implement that new function into the software. How about Lionel ?

A I don't think Dolph needs Lionel. Dolph needs Paul !

B What ? Dolph needs Paul ? But Paul isn't better than Dolph in programming..

A Yes, but that's not the problem.. The most important thing is Paul knows the customer, and this is where his advice is needed....

Doug pulled Ned

CF0

A How were your winter holidays in the mountains. I heard you played all day on sledges ?

B Yes, it was big fun. Doug pulled Ned.

A Really ? Doug pulled Ned ? He must have been tired in the evening. Ned is 7 years older than him...

B He was, but they had fun...

CF1

A Wow, that hill is quite high. Who of you pulled little Ned the whole way up. Was that you ?

B No, that wasn't me. Doug pulled Ned !

A What ? Doug pulled Ned ? But Doug is 2 years younger than Ned.

B Yes, and he is quite tired right now...

CF2

A Oh, Doug and Ned have done a very long walk in the snow ! Did they really walk the whole way ?

B Oh no. Ned didn't walk at all. Doug pulled Ned !

A What ? Doug pulled Ned ? I didn't know that they had a sledge..

B Oh, they borrowed one from the neighborhood...

CF3

A Oh, Doug is coming back from the walk in the snow with his children. Did he pull little Clarence on the sledge ?

B Oh no, he didn't pull Clarence. Doug pulled Ned !

A What ? Doug pulled Ned ? But he should be old enough to walk by himself, shouldn't he ?

B Yes, but he is such a lazy boy. He always wants to be pulled when there is a sledge...

Tim warned Ron

CF0

A Does everyone know not to leave valuables in the car in this neighborhood ?

B Yes. Tim warned Ron.

A Really ? Tim warned Ron ? But I saw his wallet lying on the seat.

B Oh, I should tell him myself then before it's too late.

CF1

A Wasn't it Ron whose car was stolen last night ? Didn't you warn him not to leave it in that neighborhood ?

B It wasn't me who warned him. Tim warned Ron !

A What ? Tim warned Ron ? Wasn't Tim away yesterday ?

B Yes, but he came home early and saw Ron's car...

CF2

A Who conviced Ron to sign that terrible insurance policy. Was it Tim ?

B No, Tim didn't convince Ron. Tim warned Ron !

A What ? Tim warned Ron ? But Tim works for that company himself...

B Yes, but he knew the loophole in that contract - and Tim is a good friend of his..

CF3

A Why did Elizabeth and her husband sign that terrible insurance policy. Didn't Tim warn her about it ?

B He didn't warn Elizabeth. Tim warned Ron !

A What ? Tim warned Ron ? But why didn't Ron say anything to her ?

B Maybe he did. But Elizabeth signed the policy anyway ...

Walt riled Tess

CF0

B What's that argument in the office about ?

A Walt riled Tess.

B Really ? Walt riled Tess ? What did he do ?

A Oh, I think he wants two more forms to be filled out for the marking procedure for every student...

CF1

B Who made Tess curse so loudly. Did that annoying secretary rile her again ?

A No, it's not the secretary who riled her. Walt riled Tess!

B What ? Walt riled Tess ? What could he do to make her so upset ?

A Well, he claimed that Tess had to fill out two more forms for every student to get the marking procedure to be correct..

CF2

B Tess is still upset. Couldn't Walt calm her down a little ?

A No, he couldn't calm her at all. Walt riled Tess!

B Really ? Walt riled Tess ? What has he done ?

A He wanted her to fill out 2 more forms for every student...

CF3

B What's the commotion from the assistants' office ? Did Walt rile Eric again ?

A No, this time he didn't rile Eric. Walt riled Tess !

B What ? Walt riled Tess ? But Tess usually does a good job and is always ready in time.

A Yes, but now Walt wants Tess to do some extra work before she goes home...

Wim trapped Ralph

CF0

A Is the chess match over ?

B Yes, almost. Wim trapped Ralph.

A Really ? Wim trapped Ralph ? I didn't think he could do that ...

B Well, but Ralph forgot about the queen...

CF1

A Who could trap Ralph in the big chess tournament last Saturday. I bet that Albert got him at last.

B No, Albert couldn't do it. Wim trapped Ralph !

A What ? Wim trapped Ralph ? But Wim isn't usually such a good player...

B Yes, I know. But he had a great day...

CF2

A How is the chess match between Wim and Ralph going ? Wim should be about to lose. Ralph is the best of our group !

B Oh, no, Wim is not loosing against Ralph now. Wim trapped Ralph !

A What ? Wim trapped Ralph ? Was Ralph not paying attention ?

B Yes, he only had eyes for his queen and didn't see the king's trap coming...

CF3

A I bet that Wim trapped Albert in this big chess tournament, didn't he ?

B Oh, no. Wim didn't get Albert.. But look: Wim trapped Ralph !

A What ? Wim trapped Ralph ? But Ralph usually plays much better than Albert ...

B Yes, but today he is great in playing

Ted rammed Wayne

CF0

B How did that car accident between our friends happen yesterday ?

A Ted rammed Wayne.

B What did you say ? Ted rammed Wayne ? His big car against that little car ?

A Yes, but no one was hurt... They were lucky - only Wayne's car is totalled...

CF1

B Do you have any more news about the car accident on Oxford Road yesterday. I heard that Harry rammed Wayne in his Audi.

A Yes, but it wasn't Harry who rammed Wayne. Ted rammed Wayne !

B What ? Ted rammed Wayne ? But it was Harry's car, wasn't it ?

A Yes, but Ted was driving. Harry let him borrow the Audi for the evening...

CF2

B They say that Ted scraped Wayne's car because the road was so slippery after the storm.

A I don't think so ! It was more than a little scratch. Ted rammed Wayne !

B What ? Ted rammed Wayne ? But I saw Wayne driving to work in his car this morning ?

A Really ? That's impossible. His car was totalled...

CF3

B Do you have any more news about the car accident on Oxford Road yesterday ? I've heard that Ted rammed Harry in his Audi .

A Yes, but Ted didn't ram Harry. Ted rammed Wayne !

B What ? Ted rammed Wayne ? But it was Harry's car, wasn't it ?

A Yes, but Harry let Wulf borrow his car for the evening...

Rick wants Tom

CF0

A Have you heard anything about the director's assistant position at the theater ?

B Yes. Rick wants Tom.

A What did you say: Rick wants Tom ?

B Yes, someone said that Rick made his decision.

CF1

A Have the director and the producer finally agreed on the main character in the new film ? I heard that Enrico wants Tom.

B No. It's not Enrico who wants Tom. Rick wants Tom !

A What ? Rick wants Tom ? I thought that Rick didn't want to work with Tom again. They had a lot of troubles with their last project, didn't they ?

B I don't know. But Tom is perfect for that part...

CF2

A Have you heard anything about the cast for that new play they want to put on in September ? I heard that Rick doesn't want Tom to play the main part again.

B No, no. Rick wants Tom !

A What ? Rick wants Tom ? Why did he change his mind ?

B I think Rick changed his mid after he saw Tom in the last production...

CF3

A Have you heard anything about the cast fort the new play they want to put on in September ? I've heard that Rick wants Francis to play the main part, isn't it ?

B No not, Rick doesn't want Francis. Rick wants Tom !

A What ? Rick wants Tom ? But they had so much trouble with the last project.

B Maybe, but Tom is the best man for that part, certainly...

Ray tricked Will

CF0

B What's all that cursing during the poker game about ?

A Ray tricked Will.

B Really ? Ray tricked Will ? I can't imagine that happening.

A Yes, nobody thought that he had a straight in his hand...

CF1

B I heard that Will was tricked last night playing poker. Did Carlos trick him ?

A No, no. Carlos is not the man to trick Will. Ray tricked Will !

B What ? Ray tricked Will ? I didn't think he had a good poker face.

A Oh, he does. He won more than 20 pounds...

CF2

B Ray surely didn't have a chance to win the poker game against Will. He is an absolute beginner and so honest.

A Oh, no. Ray tricked Will !

B What ? Ray tricked Will ? I didn't think that he would have the skills for that...

A Nobody would have thought it. But he is quite good at poker...

CF3

B I've heard that Ray tricked someone in the poker game. Did he trick Carlos ?

A Oh, no. Ray couldn't trick Carlos. Ray tricked Will !

B What ? Ray tricked Will ? But Will is one of the most experienced poker players around, isn't he ?

A Yes, but Ray has a great poker face

B German Materials

Ken sucht Franz

Ken is looking for Franz

CF0

A Was ist denn da drauen im Garten los ?

B Ken sucht Franz.

A Wie bitte, das hab ich jetzt nicht verstanden ? Ken sucht Franz ?

B ja, j

A Verstecken ist ihr neuestes Lieblingsspiel.

CF1

A Franz hat sich schon versteckt. Wer muss ihn denn jetzt suchen ? Ist Barbara schon dabei ?

B Nee, Barbara hat keinen Bock mehr. Ken sucht Franz !

A Was ? Ken sucht Franz ? Seit wann spielt er denn beim Verstecken mit ?

B Er kam gerade von zu Hause.

CF2

A Was machen denn die beiden Jungs da hinten auf dem Spielplatz ? Sie rennen nun schon die ganze Zeit hintereinander her. Spielen sie Fangen ?

B Nee, nee, Fangen spielen sie eigentlich nicht. Ken sucht Franz !

A Was ? Ken sucht Franz ? Da msste sich Franz aber auch mal verstecken...

B ja, ja, das hat er vorhin auch. Aber nun scheinen sie eher zu streiten.

CF3

A Wer ist denn dran nach Barbara zu suchen? Ist Ken schon dabei ?

B Nee, der sucht nicht Barbara. Ken sucht Franz !

A Was ? Ken sucht Franz ? Spielt Franz berhaupt mit beim Verstecken ?

B Ja, er kam gerade dazu.

Falk sah Kurt

Falk saw Kurt

CF0

A Gibts was neues von gestern ?

B Oh ja. Falk sah Kurt.

A Ach was: Falk sah Kurt ? Wie soll denn das gehen ? Ich dachte Falk wre im Urlaub und Kurt studiert doch dieses Jahr in Marseille.

B Sie sind sich in Frankfurt am Flughafen begegnet...

CF1

A Hast du schon von Kurt gehrt, dem neuen Superstar ? Ich glaube, Albert hat ihn gestern im Konzert gesehen.

B Nee, Albert war das nicht. Falk sah Kurt !

A Was ? Falk sah Kurt ? Aber Falk mag doch so eine Musik gar nicht.

B Er hat aber Freikarten bekommen...

CF2

A Was haben Falk und Kurt denn abgemacht ? Sie wollten doch gestern telefonieren ?

B Nee, telefoniert haben sie nicht. Falk sah Kurt !

A Was ? Falk sah Kurt ? Aber Kurt wohnt doch in Gttingen.

B Kann sein. Aber Falk sagte, sie seinen sich zuflig am Bahnhof in Eisenach begegnet.

CF3

A Hast du schon von Marcello gehrt, dem neuen Superstarstar. Ich glaube, Falk hat ihn gestern im Konzert gesehen.

B Nee, doch nicht den Marcello. Falk sah Kurt !

A Was ? Falk sah Kurt ? Seit wann steht er den auf Schlagersnger..

B Er hat Freikarten bekommen...

Klaus fährt Sven

Klaus gives Sven a ride

CF0

B Weit du, wie unsere Freunde zu unserer Hochzeit kommen?
A Klaus fhrt Sven.
B Wie bitte ? Klaus fhrt Sven ?
A Ja genau, sie kommen doch beide aus Mannheim.

CF1

B Wer bringt denn Sven auf unsere Hochzeit? Fhrt ihn Gerald?
A Nee, Gerald kann nicht ! Klaus fhrt Sven !
B Was ? Klaus fhrt Sven ? Seit wann hat er denn einen Fhrerschein?
A Er hat ihn doch erst letzte Woche gemacht und will nun dauernd Auto fahren.

CF2

B Werden Klaus und Sven dann noch zu dieser Berghtte wandern?
A Nee, wandern werden sie nicht. Klaus fhrt Sven !
B Was ? Klaus fhrt Sven ? Warum knnen sie denn nicht gehen?
A Sven hat sich gestern anscheindend den Fu verstaucht.

CF3

B Mensch, wir mssen noch daran denken, wer die Kleinen auf unsere Hochzeit mitbringt. Kann Klaus den Gerald fahren?
A Nee, nicht den Gerald. Klaus fhrt Sven !
B Was ? Klaus fhrt Sven ? Und Svens Eltern fahren dann ohne ihn, oder wie ?
A Ja, sie kommen erst spt am Abend, weil Svens Mutter noch einen Termin hat.

Sam kennt Frank

Sam knows Frank

CF0

B Hast du schon das Neueste gehrt? Sam kennt Frank.

A Was? Sam kennt Frank ? Wo haben die sich denn kennen gelernt ?
B Sie waren wohl zusammen auf der Schule.
A Hm, kann sein, sie sind ja beide aus Hannover...

CF1

A Was macht denn Frank auf der WG-Party hier ? Wer ldt denn diesen Idioten ein ? Kennt Arno ihn berhaupt ?
B Nee, Arno kennt ihn nicht. Sam kennt Frank !
A Was? Sam kennt Frank ? Aber Sam ist doch gerade erst aus Hannover hierher gezogen !
B Naja, aber er und Frank waren wohl zusammen in der Schule...

CF2

A Was ist denn mit Sam und Frank los? Jetzt sind wir schon fast eine Woche zusammen auf Fahrradtour und die beiden haben noch kein Wort miteinander gewechselt. Sind sie denn so schchtern?
B Nee, nee, das liegt nicht daran. Sam kennt Frank !
A Was? Sam kennt Frank ? Und trotzdem reden sie nicht miteinander?
B Ja, ja. Sie haben sich wohl mal gleichzeitig in Anna verliebt. Und seit sie mit Frank zusammen ist, redet Sam nicht mehr mit ihm...

CF3

A Was macht denn Arno auf Sams Party hier? Kennt ihn Sam berhaupt ?
B Nee, Arno kennt Sam gar nicht. Sam kennt Frank !
A Was ? Sam kennt Frank ? Aber Frank studiert doch in Marburg !
B Aber er und Frank gingen wohl zusammen auf die Schule. Und Frank ist der Bruder von Arno ...

Sepp fand Karl

Sepp found Karl

CF0

B Hast du Neuigkeiten vom Lawinenunglck?
A Ja. Sepp fand Karl.
B Hab ich das richtig verstanden ? Sepp fand Karl ?

A Ja, glcklicherweise. Er fand ihn, noch bevor es dunkel wurde.

CF1

B Ich hab gehrt, dass man Karl nach dem Lawinenunglck gefunden hat. Hat ihn Alfred mit all seiner Erfahrung gefunden ?

A Nee, Alfred lag trotz seiner Erfahrung daneben. Sepp fand Karl !

B Was ? Sepp fand Karl ? Aber der ist doch erst seit diesem Winter bei der Bergwacht!

A Aber er hatte die richtige Nase, so wie es scheint.

CF2

B Bitter: schon wieder ein Lawinenunglck. Aber ich hab gehrt, dass der Sepp den Karl gerettet hat.

A Naja, das mit dem retten stimmt leider nicht ganz. Sepp fand Karl ! Das ist richtig...

B Wie.. ? Sepp fand Karl ?

A Als Sepp ihn gefunden hat, war Karl leider schon tot. Da kam jede Rettung zu spt...

CF3

B Ich hab gehrt, dass Sepp mal wieder eine gute Nase gehabt hat. Er hat anscheinend Alfred in der Lawine gefunden.

A Nee, nee. Den Alfred hat er nicht gefunden. Sepp fand Karl !

B Was ? Sepp fand Karl ? War Karl berhaupt mit in der Wandergruppe?

A Ja, er hat sich am Morgen spontan dazu entschlossen. Alfred blieb zu Hause im Hotel...

Fred küsst Sue

Fred kisses Sue

CF0

B Hey, was passiert denn da drben?

A Wieso ? Fred ksst Sue.

B Ach was ! Fred ksst Sue ? Wer htte das gedacht !

A Ich hab' sie letzte Woche schon gesehen...

CF1

B Hey, schau mal, da drben im Cafe. Ist das nicht der Manni, der da eben Sue ksst ?

A Nee, nee, das ist doch nicht Manni! Fred ksst Sue !
B Was ? Fred ksst Sue ? Seit wann sind die denn zusammen ?
A Ich hab sie letzte Woche schon gesehen...

CF2

B Hey, schau mal, da drben im Cafe... Fred und Sue unterhalten sich ja ziemlich angeregt...
A Aber jetzt unterhalten sie sich nicht mehr.. Fred ksst Sue !
B Was ? Fred ksst Sue ? Da hat er sie aber verdammt schnell erobert...
A Stimmt, aber bei seinem Charme...

CF3

B Hey, schau mal, da drben im Cafe. Wen ksst denn da der Fred ? Ist das nicht Anna?
A Nee, das ist doch nicht Anna. Fred ksst Sue !
B Was ? Fred ksst Sue ? Seit wann denn die ? War er nicht mit Anna zusammen ?
A Ich hab' ihn und Sue letzte Woche schon gesehen...

Lutz mag Babs

Lutz likes Babs

CF0

A Was sagt denn die Gerchtekche aus unserer alten Klasse ?
B Ich hab da was gehrt... Lutz mag Babs.
A Das hab ich jetzt nicht verstanden. Lutz mag Babs ? Oder was hast du gesagt ?
B Ja, genau. Er ldt sie stndig ein mit ihm auszugehen...

CF1

A Babs ist schon eine ziemliche Zicke ! Auer ihrem Bruder kann sie doch keiner leiden.
B Das stimmt nicht. Lutz mag Babs !
A Was ? Lutz mag Babs ? Aber die haben sich doch frher auch immer gestritten.
B Ja, aber seit sie zusammen Handball spielen, kommen sie gut miteinander aus.

CF2

A Warum ist denn Lutz immer so fies zu Babs ? Sie haben sich wohl total zerstritten...
B Oh nein. Lutz mag Babs !

A Was? Lutz mag Babs ? Warum hnselt er sie dann dauernd ?

B Naja, er ist halt ein pubertierender Junge... das mit dem nett sein, das muss er noch lernen...

CF3

A Lutz ist schon ein ziemlicher Eigenbrdler. Er kann irgendwie keinen Menschen leiden...

B Oh, mit einer Ausnahme... Lutz mag Babs !

A Was ? Lutz mag Babs ? Aber die streiten sich doch dauernd.

B Naja, aber nach dem Handball will Lutz immer noch mit Babs ein Eis essen gehen...

Bernd malt Lou

Bernd portrays Lou

CF0

A Wieso ist denn das Atellier die ganze Zeit verschlossen?

B Bernd malt Lou.

A Was ? Bernd malt Lou ?

B Ja, er hat sich fr diesen Auftrag seit drei Tagen hier eingemietet.

CF1

A Sag mal... Lou lsst sich doch von Rainer portraitieren, oder ?

B Nee, Rainer malt sie nicht. Bernd malt Lou !

A Was ? Bernd malt Lou ? Der hat doch noch kein einziges Portrait gemalt !

B Das stimmt schon, aber er kennt Lou ziemlich gut und deshalb wollte sie sich von ihm portraitireren lassen.

CF2

A Bernd und Lou sind jetzt schon den ganzen Nachmittag in eurem Foto-Studio. Macht er so aufwndige Fotos von ihr ?

B Nee, Bernd fotografiert sie nicht. Bernd malt Lou !

A Was ? Bernd malt Lou ? Kann er denn berhaupt malen?

B Er glaubt es jedenfalls: Er hat letztes Jahr einen Kurs an der Volkshochschule besucht.

CF3

A Bernd ist jetzt schon den ganzen Nachmittag in seinem Atellier. Malt er an dem Portrait von Anna weiter?

B Nee, Annas Bild ist schon fertig. Bernd malt Lou !

A Was ? Bernd malt Lou ? Wie kam Lou denn auf diese Idee: sich portraitieren zu lassen ?

A Sie hat den Entwurf von Anna gesehen und gesagt: das will ich auch !

Mark braucht Lars

Mark needs Lars

CF0

B Was sollte denn der Anruf von der Baustelle ?

A Mark braucht Lars.

B Wie bitte ? Mark braucht Lars ?

A Ja, Mark hat keine Ahnung wie man eine Asphaltfarbe mischt.

CF1

B Mensch, ich brauche dringend jemand der mir diese Funktion programmieren kann. Wo ist denn Lars ? Braucht Arco ihn gerade ?

A Nee, Arco braucht ihn heute nicht. Mark braucht Lars !

B Was ? Mark braucht Lars ? Aber Mark kann doch auch programmieren.

A Aber diesmal mssen sie wohl zu zweit daran arbeiten.

CF2

B Mark muss anscheinend schon wieder Leute entlassen. Ist Lars davon denn auch gefhrdet ?

A Nee, den kann er nicht entlassen. Mark braucht Lars !

B Was ? Mark braucht Lars ? Aber was kann Mark denn, dass du dir da so sicher bist.

A Lars ist doch mit Sicherheit der beste Programmierer in der Firma.

CF3

B Mark muss doch morgen die neue Software prsentieren. Soll ich ihm Arco noch zur Untersttzung schicken ?

A Nee, mit Arco kann er heute nichts anfangen. Mark braucht Lars !

B Was? Mark braucht Lars ? Ich wusste gar nicht, dass Lars auch programmieren kann.

A Das nicht, aber er kennt den Kunden und wei was der will...

Leif bringt Mel

Leif brings Mel

CF0

B Was gibt es neues von unseren Gsten ?
A Leif bringt Mel.
B Wie bitte ? Leif bringt Mel ? Sprich doch nicht immer so leise !
A Ja, ist ja gut. Sie kommen zusammen aus Mnchen.

CF1

B Wie kommt denn Mel auf unser Fest ? Bringt Hartmut sie mit ?
A Nee, Hartmut kann sie nicht mitnehmen. Leif bringt Mel !
B Was ? Leif bringt Mel ? Aber sie kommen doch gar nicht aus derselben Stadt ?
A Trotzdem, Leif ist am Freitag vorher wohl dienstlich in Mnchen.

CF2

B Wird Leif Mel von der Party abholen ?
A Nein, abholen wird er sie nicht. Leif bringt Mel !
B Was ? Leif bringt Mel ? Aber er arbeitet doch normalerweise um diese Zeit noch !
A Kann sein, aber er hat wohl Urlaub nchste Woche...

CF3

B Wen fhrt denn Leif zum Schulfest heute abend ? Bringt er Fatima ?
A Nee, Fatima bringt er nicht. Leif bringt Mel !
B Was ? Leif bringt Mel ? Und wer fhrt dann Fatima ?
A Keine Ahnung, aber das bleibt wohl an uns hngen...

Maik lobt Bill

Maik commends/compliments Bill

CF0

A Was passiert denn gerade im Bro vom Chef?
B Maik lobt Bill.

A Nochmal ? Maik lobt Bill? Oder was hast du gesagt ?

B Ja, genau. Bill hat wohl einen groen Auftrag an Land gezogen.

CF1

A Hast du es schon gehrt: Bill hat einen groen Auftrag fr uns an Land gezogen. Das wird die Chefs aber freuen. Da wird ihn Ernst aber dick belobigen ...

B Nee, nicht Ernst. Maik lobt Bill !

A Was ? Maik lobt Bill ? Der hat doch noch nie jemanden gelobt.

B Aber Maik war letztens auf eine Fortbildungskurs "Mitarbeiterfhrung". Vielleicht nicht ganz umsonst ...

CF2

A Warum muss denn Bill schon wieder zu Maik. Schimpft der Chef schon wieder ?

B Nee, hr mal, der schimpft ganz und gar nicht. Maik lobt Bill !

A Was ? Maik lobt Bill ? Das hat er ja noch nie gemacht ...

B Aber Bill hat wohl einen groen Auftrag an Land gezogen...

CF3

A Warum mussten denn Bill und Hans schon wieder zu Maik, unserm Chef? Wird Hans endlich mal dafr gelobt, dass er so viel arbeitet?

B Nee, Hans ist diesmal nicht dran. Maik lobt Bill !

A Was? Maik lobt Bill ? Der ist doch der faulste Mitarbeiter auf Erden.

B Aber letztens hat er wohl einen wichtigen Auftrag gerettet...

Bess liebt Matt

Bess loves Matt

CF0

A Hast du schon das Neueste gehrt ? Bess liebt Matt.

B Wie bitte ? Bess liebt Matt ? Hast du wirklich Bess liebt Matt gesagt ? Das ist ja hochinteressant !

CF1

B Wie war das gleich wieder in dem verrckten Theaterstck ? Vera liebt Matt und...

A Nee, Vera macht sich nichts aus Matt. Bess liebt Matt !

B Was? Bess liebt Matt ? Aber sie verrt ihn doch dann !

A Sie verrt ihn, weil ER Vera liebt.

CF2

B Wie war das gleich wieder in dem verrckten Theaterstck ? Bess verrt Matt, weil sie ihn abgrundtief hasst?

A Nee, sie hasst ihn ganz und gar nicht. Bess liebt Matt !

B Was? Bess liebt Matt ? Warum verrt sie ihn dann ?

A Weil er wiederum nur Vera liebt.

CF3

B Wie war das gleich wieder in dem verrckten Theaterstck ? Bess liebt Thomas und ...

A Nee, sie macht sich nichts aus Thomas. Bess liebt Matt !

B Was? Bess liebt Matt ? Aber sie verrt ihn dann doch !

A Ja, aber nur weil er Vera liebt und sie nicht heiraten will.

Ned pflegt Dyke

Ned tends Dyke

CF0

A Oh, wir sind aber wenig heute zum Fuball spielen. Wo sind denn die anderen ?

B Dyke pflegt Nils.

A Was ? Dyke pflegt Nils ? Ist Nils denn krank ?

B Oh, ja. Die Grippe hat ihn voll erwischt.

CF1

A Oh, das tut mir leid, dass Nils so krank ist. Kann seine Mutter zu Hause bleiben um ihn zu pflegen ?

B Nein, seine Mutter kann nicht. Dyke pflegt Nils !

A Was ? Dyke pflegt Nils ? Hat der denn Zeit fr sowas.

B Ja, er hat diese Woche ohnehin Urlaub.

CF2

A Ned spricht oft von Dyke, dass er ihm so geholfen hat, seit seinem Unfall. War er es, der Nils operiert hat ?

B Nein, operiert hat ihn ein anderer. Dyke pflegt Nils !

A Was ? Dyke pflegt Nils ? Gibt es dafr nicht die Krankenschwestern ?

B Ja natrlich auch. Aber Dyke kmmert sich fast die ganze Zeit um ihn, seit er zu Hause ist ...

CF3

A Wen soll Dyke denn pflegen, wenn er aus dem Urlaub kommt ? Kann er sich um Quentin kmmern ?

B Nein, Quentin geht nicht. Dyke pflegt Nils !

A Was ? Dyke pflegt Nils ? Nils liegt doch auf der Intensivstation. Ich dachte, Dyke darf nicht dort arbeiten?

B Doch. Vor seinem Urlaub hat er endlich die Fortbildung bestanden. Und Nils kann jede Hilfe gebrauchen...

Pam neckt Dave

Pam banters Dave

CF0

A Was ist denn da hinten los?

B Pam neckt Dave.

A Wie bitte ? Pam neckt Dave ? Oder was hast du gesagt ?

B Ja, ja, deshalb kichern die beiden so...

CF1

A Wer kitzelt denn Dave schon die ganze Zeit mit einem Grashalm. Ist das nicht Sonja, die ihn neckt?

B Nein, das ist doch nicht Sonja. Pam neckt Dave !

A Was ? Pam neckt Dave ? Ich dachte die knnen sich gar nicht leiden.

B Wer wei, wer wei...

CF2

A Oh Mann, Pam und Dave rgern sich schon wieder...

B Nein, ich glaube nicht, dass sie sich wirklich gegenseitig rgern. Pam neckt Dave !

A Was ? Pam neckt Dave ? Dafr sieht es aber ganz schn brutal aus...

B Aber es ist bestimmt nicht wirklich bse gemeint...

CF3

A Sag mal, wen kitzelt Pam denn die ganze Zeit mit einem Grashalm ? Ist das Erich, den sie da neckt ?

B Nein, doch nicht den Erich. Pam neckt Dave !

A Was ? Pam neckt Dave ? Ich dachte die knnten sich gar nicht leiden!

B Naja, sie scheinen sich ja wieder etwas anzunhern...

Pete dankt Nick

Pete thanks Nick

CF0

B Hast du mitbekommen, was die auf der Bhne gerade machen?

A Ja. Pete dankt Nick.

B Wie bitte ? Pete dankt Nick ? Es ist so laut hier.

A Ja, alles habe ich auch nicht verstanden, aber ich glaube es geht darum, dass Pete die Abifeier so schn organisiert hat.

CF1

B Wer wird denn Nick im Namen aller Abiturienten fr die Organisation der Abifeier danken. Soll das Stefan machen?

A Nein, Stefan muss sich da fr den Auftritt umziehen. Pete dankt Nick !

B Was ? Pete dankt Nick ? Aber Pete wollte doch auf keinen Fall auf der Bhne was sagen !

A Naja, aber einen Satz wird er schon loswerden knnen...

CF2

B Warum bekommt denn Nick ein teures Buch von Pete, dem Schulsprecher. Wird ihm fr sein exzellentes Abitur gratuliert ?

A Nein, das ist keine Gratulation. Pete dankt Nick !

B Was ? Pete dankt Nick ? Wofr denn ?

A Nick hat doch die Abiturfeier organisiert...

CF3

B Wem dankt denn Pete im Namen aller Abiturienten fr die Organisation der Abifeier ? Stefan hat dafr doch am meisten gemacht, oder ?

A Das stimmt. Aber das ist nicht Stefan auf der Bhne. Pete dankt Nick !

B Was ? Pete dankt Nick ? Aber der hat doch fast gar nichts gemacht !

A Tja, ich wei auch nicht, warum gerade er das teure Buch erhlt...

Neil droht Pim

Neil threatens Pim

CF0

B Hey, was ist denn da vorne los ?

A Neil droht Pim.

B Was ? Neil droht Pim ? Warum denn das ?

A Ach, das wei ich auch nicht, irgendwas muss wohl passiert sein.

CF1

B Hey, was machen denn die zwei da hinten. Ist das nicht Erik, der sich da so drohend vor Pim aufbaut ?

A Nein, das ist nicht Erik. Neil droht Pim !

B Was ? Neil droht Pim ? Aber die waren doch die besten Freunde !

A Keine Ahnung, aber irgendwas muss passiert sein.

CF2

B Mann, sind die beiden aggressiv da vorne. Prgeln sich Neil und Pim ?

A Nein, prgeln tun sie sich nicht. Neil droht Pim!

B Was ? Neil droht Pim ? Das sind aber wohl schon fortgeschrittene Drohungen, oder ?

A Keine Ahnung, ob wir da doch mal dazwischen gehen ?

CF3

B Vor wem hat sich den Neil so drohend aufgebaut. Ist das nicht Erik ?

A Nee, das ist nicht Erik. Neil droht Pim !

B Was ? Neil droht Pim ? Wieso denn das?

A Keine Ahnung, aber irgendwas muss wohl passiert sein.

Nat prüft Dirk

Nat examines Dirk

CF0

A Ist das Bro besetzt ?
B Ja. Nat prft Dirk.
A Was hast du gesagt ? Nat prft Dirk ?
B Ja sie haben wohl den Termin verschoben.

CF1

A Bei wem lsst sich denn Dirk in Literatur prfen ? Geht er zu Helena ?
B Nein, Helena prft dieses Semester nicht. Nat prft Dirk !
A Was ? Nat prft Dirk ? Aber Dirk hatte doch gar keinen Kurs bei ihr.
B Keine Ahnung, aber es geht dieses Semester nicht anders.

CF2

A Was macht denn Dirk schon so lange im Bro von Nat? Sprechen sie ber seine Magisterarbeit ?
B Nein, die hat er doch schon lngst abgegeben. Nat prft Dirk !
A Was ? Nat prft Dirk ? Hat er denn schon alle Scheine zusammen ?
B Es scheint so: das Prfungsamt hat ihn zugelassen.

CF3

A Ach ja, Nat muss ja heute den ganzen Tag Prfungen abnehmen ! Ist im Moment nicht gerade Helena bei ihr ?
B Nein, Helena war schon heute mittag dran. Nat prft Dirk !
A Was ? Nat prft Dirk ? Aber der war doch gar nicht in ihrem Seminar.
B Naja, aber das Prfungsamt hat ihn zugeteilt ...

Dolph nervt Paul

Dolph annoys Paul

CF0

B Was ist denn hier fr eine komische Stimmung ?

A Ach: Dolph nervt Paul.

B Was hast du gesagt ? Dolph nervt Paul ?

A Ja, es hat vorhin schon einen kleinen Eklat gegeben...

CF1

B Was kuckt denn Paul so sauer. Nervt ihn Winnie mal wieder?

A Winnie ist diesmal unschuldig. Dolph nervt Paul !

B Was ? Dolph nervt Paul ? Die beiden verstehen sich doch sonst so gut !

A Ja, aber Dolph erklrt nun schon zum hundertsten Mal, wie toll sein neues Handy ist ...

CF2

B Warum kuckt Paul denn Dolph so belustigt an. Erzhlt Dolph mal wieder Witze ?

A Nein ich glaube nicht, das Paul Dolph lustig findet. Dolph nervt Paul !

B Was ? Dolph nervt Paul ? Aber Dolph kann doch eigentlich gut Geschichten erzhlen.

A Das schon: Aber nicht die gleiche Geschichte zum hundertsten Mal ...

CF3

B Warum habt ihr denn Dolph einfach so stehen gelassen ? Nervt er Winnie mal wieder ?

A Nein, Winnie hat sich diesmal nicht beschwert. Dolph nervt Paul !

B Was ? Dolph nervt Paul ? Aber Paul ist aber doch der ruhigste von allen...

A Aber heute ist ihm der Kragen geplatzt: immer dieselben Witze zu unpassendsten Momenten.

Walt ruft Tess

Walt summons Tess

CF0

A Hrst du das auch ?

B Was ?

A Walt ruft Tess.

B Was ? Walt ruft Tess ? Nee, wirklich nicht. Du musst dich verhrt haben...

CF1

B Hr mal, wer ruft denn da nach Tess ? Ist das Vater ?

A Nee, Vater ist das nicht. Walt ruft Tess !

B Was ? Walt ruft Tess. Was will denn der schon wieder von ihr.

A Keine Ahnung. Aber er hrt und hrt nicht auf...

CF2

B Ist Walt nach drauen gegangen um nach Tess zu suchen ?

A Nee, er hat keine Lust auf Suchen. Walt ruft Tess !

B Was ? Walt ruft Tess ? Es wre aber das erste Mal, dass sie darauf hrt...

A Wir werden es ja sehen..

CF3

B Wieso schreit denn Walt denn die ganze Zeit ? Ruft er Lisa ?

A Nee, doch nicht Lisa. Walt ruft Tess !

B Was ? Walt ruft Tess ? Was will er denn von ihr ?

A Keine Ahnung, aber es scheint dringend zu sein...

Wim traut Ralph

Wim weds Ralph

CF0

B Bist du nchsten Sonntag nicht dabei?

A Bei was?

B Wim traut Ralph.

A Was ? Wim traut Ralph ? Ralph heiratet ?

B Ja, er und Susanna haben sich spontan dazu entschlossen.

CF1

A Wer wird denn die Hochzeitsmesse fr Ralph halten ? Traut ihn Konstantin ?

B Nee, Konstantin ist in Rom nchste Woche. Wim traut Ralph !

A Was ? Wim traut Ralph ! Ist er denn berhaupt schon zum Pfarrer geweiht ?

B Ja, am nchsten Sonntag ist er so weit...

CF2

A Sag mal, wird Wim morgen Ralph eigentlich die Beichte abnehmen ?

B Nein, nein. Das hat er schon gestern gemacht. Wim traut Ralph!

A Was ? Wim traut Ralph ? Mit wem denn ?

B Mit Susanne. Die beiden haben sich spontan dazu entschlossen.

CF3

A Sag mal, Wim muss doch morgen eine Hochzeitsmesse halten. Traut er dabei seinen Jugendfreund Konstantin ?

B Nee, Konstantin war letzte Woche. Wim traut Ralph !

A Was ? Wim traut Ralph ? Mit wem denn ?

B Mit einer Susanne. Die beiden haben sich spontan dazu entschlossen.

Raul trägt Will

Raul carries Will

CF0

B Der kleine Will hat sich doch gestern abend den Knchel verletzt. Fhrt ihn Raul nun hier auf die Almhtte?

A Nein, rauf fahren kann er ihn da nicht. Raul trgt Will !

B Was ? Raul trgt Will ? Schafft er das denn berhaupt ?

A Er wird ihn in die Kraxe packen ...

CF1

B Oh, wen muss denn Raul hier den Berg hoch tragen ? Knnte Mathias sein, oder ?

A Das ist doch nicht Mathias ! Raul trgt Will!

B Was ? Raul trgt Will ? Ich dachte Will wre schon mit seiner Mutter gekommen ?

A Nee, ich glaube der konnte einfach nicht mehr...

CF2

B Was kommt hier denn dahergelaufen ? Das sieht aber komisch aus...

A Oh ja. Raul trgt Will.

B Was ? Raul trgt Will ?

A Ja, Raul trgt Will. Wir mssen fragen, ob was passiert ist !

CF3

B Oh, wer muss denn hier Will den Berg hoch tragen ? Ist das nicht Mathias ?

A Nein, das ist doch nicht Mathias ! Raul trgt Will !

B Was ? Raul trgt Will ? Aber Raul ist doch mindestens zwei Kpfe kleiner als Will !

A Ja, aber Kraft hat er ...

Tim winkt Ron

Tim waves at Ron

CF0

B Guck mal Klara ! Tim winkt Ron.

A Was ? Tim winkt Ron ? Ach ja, jetzt seh ich es auch...

CF1

A Oh, kuck mal. Wer winkt denn da Ron gerade ? Ist das Klara ?

B Nee, das ist doch nicht Klara. Tim winkt Ron !

A Was ? Tim winkt Ron ? Woher kennen die sich denn ?

B Sie kennen sich jedenfalls schon lange, hat mir Ron vorhin gesagt...

CF2

A Was macht denn Tim da oben. Ruft er Ron ?

B Ach nein. Zum Rufen ist es hier zu laut ! Tim winkt Ron !

A Was ? Tim winkt Ron ? Da msste Ron aber auch mal zum Fenster gucken.

B Na klar: jetzt hat er ihn gesehen...

CF3

A Oh, kuck mal. Wem winkt den Tim gerade auf der Strae ? Ist das Klara ?

B Nee, das ist doch nicht Klara ! Tim winkt Ron !

A Was ? Tim winkt Ron ? Haben Ron und Klara die selbe Jacke ?

B Scheint so: der Sommerschlussverkauf bei H & M ist ja gerade vorbei...

Ted rügt Wolf

Ted reprehends Wolf

CF0

B Was passiert denn da im Bro vom Chef?
A Ach: Ted rgt Wolf.
B Was ? Ted rgt Wolf ? Warum denn das?
A Wolf hat gestern einen groen Bock geschossen...

CF1

B Wer von den Chefs beschimpft denn den armen Wolf so lautstark. Ist das Enzo?
A Nee, Enzo ist auf Dienstreise. Ted rgt Wolf!
B Was ? Ted rgt Wolf? Aber Ted hielt doch immer groe Stcke auf Wolf..
A Aber Wolf hat ziemlich Mist gebaut. Der ganze Auftrag fr China ist futsch !

CF2

B Warum ist denn Wolf schon wieder bei Ted ? Wird er schon wieder befrdert ?
A Nein, ganz und gar nicht. Ted rgt Wolf !
B Was ? Ted rgt Wolf ? Wieso denn das ?
A Ted ist ziemlich sauer, weil Wolf den Auftrag fr China verbockt hat.

CF3

B Wen brllt denn Ted hier schon seit 10 Minuten an ? Ist das wieder Enzo, der irgendwas falsch gemacht hat ?
A Nee, Enzo ist diesmal unschuldig. Ted rgt Wolf !
B Was ? Ted rgt Wolf ? Was hat der denn ausgefolfressen?
A Naja, er hat den Auftrag fr China verbockt...

Rick will Tom

Rick wants Tom

CF0

A Was gibt es denn Neues bei den Bewerbungen um einen Regieassistenten ?

B Rick will Tom.

A Wie bitte ? Rick will Tom ? Hab' ich das richtig verstanden ?

B Ja, ja: Rick glaubt, dass er am besten mit Tom arbeiten kann...

CF1

A Haben sich der Intendant und der Regisseur endlich auf einen Regieassistenten geeinigt ? Ich habe gehrt, dass Tom die Stelle bekommt, weil Dieter ihn unbedingt will.

B Nein, Dieter ist gegen ihn. Rick will Tom !

A Was ? Rick will Tom ? Aber bei den Proben haben die sich ganz schn gezofft...

B Naja, aber Rick meinte, dass Tom gut fr seine Kreativitt ist. Was auch immer das heit...

CF2

A Ich habe gehrt, dass der Vertrag von Tom nicht verlngert wird, weil Rick nicht weiter mit ihm zusammenarbeiten will.

B Oh, nein, ganz im Gegenteil . Rick will Tom !

A Was ? Rick will Tom ? Beim Treffen mit dem Intendanten hrte sich das aber anders an.

B Was wei ich, aber zu mir sagte er, dass er Toms Vertrag nicht verlngern kann, weil die Mittel gekrzt wurden. Aber er setzt alles dran ihn zu behalten...

CF3

A Wen will denn Rick als neuen Regieassistenten ? Ich glaube, er hat sich sehr gut mit Dieter verstanden.

B Nein, Dieter will er nicht. Rick will Tom !

A Was ? Rick will Tom ? Mit dem hat er sich aber bei den Proben nur gezofft !

B Aber Rick meinte, dass das gut fr seine Kreativitt sei ...

C Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt bernommenen Gedanken sind als solche kenntlich gemacht. Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts habe ich Untersttzung von folgenden Personen erhalten:

Ich habe nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Die Arbeit wurde zuvor weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Behörde vorgelegt. Sie war weder Bestandteil eines ruhenden Verfahrens noch wurde sie in einem gleichartigen Promotionsverfahren als endgltig nicht bestanden erklärt. Die Arbeit ist vorher auch noch nicht verffentlicht worden.