universität wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

‘When Irony Sings – Musical Intervals as Prosodic Markers’

verfasst von / submitted by

Simon Anton Lichtenegger

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Philosophie (Mag. phil.)

Wien, 2020 / Vienna 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 190 344 362
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Lehramtsstudium, UF Englisch, UF Russisch
Betreut von / Supervisor	o. Univ.-Prof. Mag. Dr. Nikolaus Ritt

Table of contents

List of figures

1 Introduction

When irony is used in spoken language, interlocutors are expressing conflicting stances simultaneously instead of subsequently as when they are "singing a different tune". Nonetheless, very frequently specific melodies of speech in a literal sense – prosodic tunes – constitute the ways in which speakers express and signal such figures of speech. The use of irony in spontaneous speech therefore represents an ideal space to explore the multifunctional manners in which acoustic and semantic properties of language interact to create meaningful speech acts. The acoustic analysis of speech melodies, as apparent in its terminology, hardly allows researchers to avoid musicological practices of identifying meaningful acoustic units and structures. This interrelation of how words are expressed through the acoustic medium of speech melodies has served as the point of departure for developing theories for the origin and evolution of language and music since ancient Greece. Accordingly, the discussion of the nature of language and music has been approached from many different angles in various fields of linguistic and social studies, yielding the emergence of a vast number of theoretical models but concurrently leading to the accumulation of partly inconsistent and intersecting field-specific terminologies. The ambiguous individualistic constitution, thus somewhat unpredictable nature, of spontaneous conversational speech might be at the root of the multitude of contradictory results by empiric attempts to determine acoustic features of language that can be subsumed under standardized categories. In the case of irony expressed through prosodic cues, such attempts seem to be particularly challenging, since the use of this rhetorical device in spontaneous speech is realized by creating a conflict between the multifunctional structures of pragmatics, semantics and syntax, of the individual moment of expression, involving their own field-specific modalities active in the acoustic realizations of speech. However, the shared physical relations and physiological origins of the acoustic properties of spoken language and music along with the perceptual and structural similarities of creative forms of expression across different forms of human communication, regarding rhetorical devices, as irony, and music, offer a fruitful foundation for investigating how prosody is used to realize specific speech acts. In this master thesis the prosodic properties of ironic uses of language are analyzed regarding their musical properties. Native speakers of English and German uttering sentences in a literal or ironic manner, depending on the varying contexts provided, were recorded, consisting of a group being informed about the suggested ironic properties of the affected sentences and another uniformed group. The acoustic

properties of the syllables of the words that bore the ironic cue, contrasting the contextual information, were analyzed and compared to the equivalent syllables of words in sentences which contained the same lexical and syntactic information, but which were not situated in a context that elicited the use of irony as a rhetorical device for matching the contradicting contextual and literal information. The discussion of this field study is preceded by theoretical considerations divided into three parts, examining the relevant interrelated frameworks and formal correlations of irony and prosody, music and prosody as well as irony and music, respectively. The overarching interest of study can be summarized in the following research question: which discrete musical intervals constitute prosodic markers of irony in spontaneous speech?

2 Prosody as the interface of language and music

Speaking and singing are the most basic and primal modes of expression of the most common forms of human communication, language and music, respectively. The ways in which speech melodies and musical melodies share fundamental functional, formal and structural properties and, particularly, how they can be involved in expressing rhetorical devices as irony will be discussed in the three parts of this chapter.

2.1 Prosody and irony

Both prosody (Di Cristo et al. 1999: 1) and irony (Ben Jannet et al. 2013: 3537) are described as being universal features of languages. While the former constitutes an intrinsic feature of oral language production, irony is used optionally as a rhetorical figure.

2.1.1 Defining prosody

Speech does not only serve the purpose of communicating information but also coordinating the behavior in which such communication takes place (Brown 2017: 13). Prosody is involved in both of these processes being defined as “rhythms and melodies that contribute to language structure and meaning” (Gibbon 2016: 1) transmitting “information about emotion, intention, attentional focus” and “communicative stance” (Brown 2017: 13). The term prosody has been widely, though not undisputedly, used as a synonym of intonation (Di Cristo et al. 1999: 3), a tradition which will be adhered to in this thesis. Prosody is realized in the structures of speech concerning the acoustic properties of duration, loudness and pitch (Nooteboom 2010: 640). The latter feature is defined as the “perceptual correlate” of “the rate of vibration of vocal

chords located in the larynx" (Nooteboom 2010: 642). These physical measures of air pressure disturbances come into effect on all levels of organization of spoken language, from single speech sounds to the ones regarding discourse (Gibbon 2016: 7). Prosodic functions of any level can always be superseded by ones of higher ranking strategies, e.g., lexical stress positions being dispossessed of their individual stress properties in an unaccented syntactic position within a melodic contour of a sentence (Gibbon 2016: 7). Therefore, studying prosodic features on only one level entails the danger of ignoring higher ranking strategies that might have affected the properties of the sample in question. The smallest possible unit in prosodic analysis is the syllable, within or between which pitch movements are realized (Menninghaus 2018). A sequence of such pitch changes moving into a certain direction within a clause constitutes a prosodic contour (Brown 2017: 2). In contrast to the literal devices through which it operates, meaning in intonational units is generally not reached by specific combinations of elements, as in lexical structures, but rather in a holistic manner through the directional shape of such contours. Within such forms the combinatorial literal elements are coordinated in their morphological and syntactic structure by being given accent, boundaries, contrasting values of duration and intensity as well as stance values (Rumsey 2015: 24). Factors that are responsible for the individual shape of the pitch contour of a speaker's utterance are the speaker's individual dialect, mood, stylistic preferences and attitude (Di Cristo et al. 1999: 2). Due to the many further, pragmatic factors that frame individual speech acts, as the hierarchical social distance and the degree of intimacy between interlocutors (Precht 2003: 240), it has been a subject of intense debate, if it is even possible to speak of the existence of standard pitch contours (see 2.3.3.). However, this predicament might have emerged due to the lacking of a systematic cross-domain terminology (Gibbon 2016: 4). Regarding the formal aspects of a pitch contour, its length is typically referred to as corresponding to the length of a clause (Nooteboom 2010: 651), also referred to as an "intonation unit" (Fenk-Oczlon et al. 2009: 203) which shapes the "tonality" (Rumsey 2015: 24) of a sentence. These units are separated by pauses, and their internal rhythmical structure is formed by the number and size of the syllables that they consist of (Fenk-Oczlon et al. 2009: 203), distributing pitch accents, or stress – a process referred to as "tonicity" (Rumsey 2015: 24). The individual "tones" (Rumsey 2015: 24) of an utterance are to be found in the nucleus, the primary stressed syllable and its immediate surroundings. This tone serves as the main reference point of analysis for the qualitative research of this thesis. It is often indicated by duration, loudness or its location in a rhythmic sequence (Rumsey 2015: 30) but can also carry pitch information which refers utterances to the discourse background in which they are

embedded (Gussenhoven 1983: 383). Due to the abundance of factors that can influence the role and form of the nucleus, it has also been argued that even though every language might possess its distinct inventory of local and global pitch movements, on a lexical and phrasal level, respectively, it is virtually impossible to establish standardized categories of empirically practicable pitch movements that occur in spontaneous speech (Gussenhoven 1983: 407). Nonetheless, it is undisputed that the rhythm and the melody of speech carry their own semiotic values, adding to or amplifying the pragmatic, semantic and structural substance of the literal content.

2.1.2 The prosodic framework

Despite the difficulty of creating standardized intonation patterns, many regularities have been detected, since every non-tonal language is considered to possess a lexicon of standard pitch movements which are recognized by native speakers and categorically limited to only its most basic characteristics (Nooteboom 2010: 648). One of the features that affects all languages without exception is caused by the physical limitations of speech production, namely, the correspondence of the duration of an intonation phrase, or pitch contour, to the duration of a clause, since the length of a clause is temporally determined by the timeframe of the breath cycle (Fenk-Oczlon et al. 2009: 220). Besides this physical limitation the most common cross-language regularities appear to be a rising pitch movement located on each accentuated syllable but the last, including all Germanic languages except of English (Di Cristo et al. 1999: 21) and the presence of rising, falling and downstepping pitch contours (Di Cristo et al. 1999: 42). Another regularity regards the final accent: rising contours are used to express demands or protests, while falling contours are used in narrative sentences (Masataka 2009: 16). In English standard pitch contours are found in wh-questions, having a rising contour; informative sentences, displaying a falling contour; and informative sentences that contain a final continuation marker directed at the addressee, producing a minor final rise (Gibbon 2016: 23). Moreover, it has been observed that when the nucleus is not located at the very beginning, most declarative sentences are structured in an arched manner, the so-called "hat-pattern" (Nooteboom 2010: 651), rising before and falling after the nucleus (Brown 2017: 4). However, there are many circumstances that may affect these standard patterns, e.g., when a speaker expresses his uncertainty about the sentence's content in a purely prosodic manner, a declarative sentence usually ends in a final rise (Brown 2017: 8). Furthermore, the regularities identified in pitch contours are always prone to spontaneous alteration, since these boundary marking aspects of prosody also navigate turn-taking in conversations, which can

be rather unpredictable due to interruptions and interventions (Brown 2017: 13). The corresponding prosodic contours can also deviate from their assigned sentences modes, because the semantic and syntactic content may not express a sentence mode directly, e.g., by means of rephrasing a declarative as an interrogative sentence or through signaling affective or emotional states. For other sentence modes as imperatives, vocatives, hortatives and optatives, no clearly distinctive standard pitch movements have been detected (Di Cristo et al. 1999: 24). Concerning certain modalities for expressing affect and stance, however, some features have been found to occur in many languages with relative similarity. Among these consistencies are the uses of rising pitch to differentiate requests from commands, high pitch for surprise, high register for signaling obviousness and delaying for indicating connotative hints, respectively (Di Cristo et al. 1999: 21). On the level of discourse, individual prosodic contours embedded in a plurality of clauses are subjected to the effects of paratones. These intonational units signal the discourse structure of individual speaking turns, expressing degrees of novelty of information in different keys, or pitch levels – from higher key for novel notional direction, mid key for following up information and low key for parenthetical content (Rumsey 2015: 29). Therefore, quite naturally, the higher the level of prosodic rank one investigates, the more items are involved, the harder it is to set standardized measures. On the opposite end of the spectrum of prosodic elements, regarding syllables, regular characteristics are more easily defined. The speech sounds which carry the sonority of syllables, conveying prosodic information, are the vowels. Their shape and number per phrase determine the rhythm of speech (Fenk-Oczlon 2017: 1581). Concerning the shape of vowels, the main indicator for the rhythm of speech are phrasal and sentence stress, or accent, signaled by greater vowel length and loudness than their surroundings (Filippi et al. 2018: 12). The minimal unit for the rhythm of speech is the foot, which consists of one accented and one or more, either preceding or succeeding, unaccented syllables. Due to speech melodies being inherently temporally structured and pitch contours being the global temporal measures of speech, the foot of pitch contours has been claimed to function independently of the boundaries of words (Di Cristo et al. 1999: 22). Generally, accented words are of greater length than unaccented words and words at the beginning and end of phrases are of greater length than at the middle, where the nucleus is mostly placed (Nooteboom 2010: 663). An additional syllable lengthening takes place at the final position of a phrase (Nooteboom 2010: 657). Vowels are also the carriers of the melodic information of a contour, which is inherent in their composition and their ways of production. These speech sounds are distinct from each other in particular frequency peak values, the so-called formants, which are “variable tract

resonances" that "modulate the harmonic series generated by laryngeal oscillations by suppressing some harmonics more than others" (Choi et al. 2007: 9852).

The lowest frequency in the harmonic series of a vowel is the fundamental frequency, from now-on *F0*, which is a direct product of the oscillation in the larynx, "the frequency of vibration of vocal folds" – its "perceptual correlate" (Gibbon 2016: 9) being called pitch. The F0 value of a pitch contour has been termed the most substantial indicator for stress and is thus involved not only in producing the melody but also the rhythm of speech (Di Cristo et al. 1999: 5). The values of the first and second formants, from now on *F1* and *F2*, are determined by how the position of the tongue and the lips condition the shape of the oral cavity as a resonant space, distinguishing one vowel from another (Fenk-Oczlon 2017: 1581). For example, lower vowels have lower F0 than higher ones, while back vowels consist of higher F1 and lower F2 but front vowels consist of lower F1 and higher F2 (Fenk-Oczlon et al. 2009a: 111). Due to these basic physical properties of speech sounds, in addition to the language-specific standard contours of particular pragmatic and syntactic modes, spoken language is realized within specified sets of options of the prosodic framework of a language. Although the individual features of these sets of options can be combined in a limitless number of ways, speakers of a language tend to make such choices within a very confined range owing to cultural conventions (Precht 2003: 240). On the level of tone English is said to possess six F0 arrangements: high, low, rising, falling, rising-falling-rising and falling-rising-falling (Gibbon 2016: 18).

2.1.3 Defining irony

Verbal irony can be defined as "a form of nonliteral language in which speakers communicate implied propositions that are intentionally contradictory to the propositions contained in the words themselves" (Fox Tree 2005: 257). Hence, it occurs when an interlocutor purposely evades the conversational maxim of quality, however, the concept of irony faces similar problems of categorical generalization as prosody (Mauchand et al. 2019: 2). This predicament arises due to the complexity of the pragmatic background of spontaneous speech acts, e.g., interpersonal attitudes, intentions and social hierarchical distance or proximity (Braun et al. 2018: 113). Furthermore, the degree to which speakers share discourse backgrounds can vary greatly based on cultural, educational, generational and societal differences, impeding the listener's capability to decode the speaker's application of rhetorical figures (Gussenhoven 1983: 383). The first issue of categorization emerges when considering

the subtypes of irony, since irony simultaneously serves as an umbrella term and as a label for its own subtype. For example, while irony is described as an interlocutor's examination of a paradoxical situation not belittling the addressee, sarcasm involves the belittling of the addressee (Wilson 2013: 62). Furthermore, when someone is complimented by means of denouncing, it is termed "kind irony", whereas denouncing someone through compliments is referred to as "genuine sarcasm" (Braun et al. 2018: 114).

Sarcasm is further divided into "dripping sarcasm", in which the ironic content is expressed directly thus semantically, and "dry sarcasm", which possesses an ambiguous content, regarding the contextual and semantic information, and is to be recognized by prosodic or gestural modalities (Champagne-Lavau et al. 2018: 166). Thus, additionally to its pragmatically ambiguous function, offering more than one possible interpretation, irony can also be indicated in various manners, linguistically, contextually or extralinguistically (Braun et al. 2018:114). Literally marked forms of irony that can be categorized as particular rhetorical figures include hyperboles, understatements, and sometimes, depending on their wording, rhetorical questions. Contextually marked forms of irony include types of irony and sarcasm that are described above as well as satire (Wilson 2013: 62). Any form of irony can be signified through intonation by ways of applying prosodic contours or extralinguistically by accompanying facial expressions or gestures (Begagic 2013: 46). As a result of its multifaceted nature, the application of irony in spontaneous speech is prone to be misinterpreted, depending on the listener's ability to detect the "contextual incongruity" (Champagne-Lavau et al. 2018: 166) inherent in an ironic utterance. An essential factor that influences an interlocutor's capacity to interpret an ironic utterance in the way it was intended by the speaker is whether the utterance includes a form of criticism or compliment. If criticism is directed towards the receiver of the message, the "asymmetry of affect" (Mauchand et al. 2019: 13) may manipulate the addressee's perception. This phenomenon describes the difficulty of listeners to successfully interpret a speaker's attitude when negative literal content is directed at them. Another complex form of irony, echoic irony, obtains its contextual incongruity from echoing a statement or attitude which refers to a source located outside of what is being discussed or has been stated in the conversation at hand (Feller 2008: 174). Contrary to the definitions above, irony does not necessarily have to be directed at the listener, animating her/him to detect the underlying propositions, or represent a deliberate speech act. These contradictions are apparent in the fact that lying has also been defined as "fake irony" and that, in the case of "unintentional irony", irony can also emerge when the

utterer is unaware of the contextual incongruity of her/his utterance, respectively (Braun et al. 2018: 113).

2.1.4 The framework of irony

One of the requirements for successful dialogue, namely, a minimum comprehension of the conversational counterpart's objectives and psychological state (Hellbernd et al. 2018: 604), is especially important in realizations of irony, since a multitude of meanings and objectives are simultaneously offered to the receiver to produce contextual incongruity (Feller 2008: 177). The higher the degree of shared background knowledge and beliefs between speaker and listener, the easier it is for the listener to detect the relevant cues to recognize applications of irony as such in spontaneous speech (Feller 2008: 179). Listeners tend to prioritize certain cues over others when being the addressee of a polysemous utterance. They first give attention to behavioral cues, involving facial expressions and gestures, then prosodic and, finally, semantic cues (Attardo et al. 2003: 246). In case of the latter, irony is conveyed in the form of a "deadpan delivery", in which the contextual incongruity arises marked by morphological, lexical, syntactical or typographical features (Attardo et al. 2003: 244). In this straight-forward manner the contextual incongruity is most easily deciphered, since the discrepancies are found in the visibly related literal domains of language. When irony is signified by non-literal means, the addressee is first to identify the social purpose of the addresser's speech act, which can be relating either negatively or positively to the interlocutors' social hierarchical distance and degree of intimacy (Mauchand et al. 2019: 2). Such modes can be expressed by prosody in manifold ways.

2.1.5 Prosodic cues for delivering irony

Which types of pitch accents are employed in the pitch contour of an individual utterance is primarily determined by the overarching information structure and pragmatic coordination of a conversation (Gibbon 2016: 18). Regarding the latter, one can distinguish between a "monologic driving force of leadership", in which the speaker tries to impose his views and needs upon the listener, aiming to persuade the her/him to behave or react in a certain way; and a "dialogical driving force" (Brown 2017: 14), where the speaker adapts stylistically and thematically to the interlocutor. After gestural behavior accompanying a speaker's utterance, fittingly termed "gesture prosody" (Gibbon 2016: 18), prosodic cues offer the listener possibilities to interpret the speaker's affective stance and attitudes when they are not superficially detectable (Mauchand et al. 2019: 2). These cues are specific pitch contours that

seem to have innate qualities, since brain damaged patients who are unable to link contextual and non-contextual clues are still able to identify a sarcastic tone of voice (Antoniou et al. 2015: 35). Interestingly, some researchers argue that it is impossible to define or even outline an ironic tone of voice per se due to the multilayer linguistic and pragmatic frameworks that account for the production of individual speech acts (Mauchand et al. 2019: 3), while others speak of an academic consensus of the existence of such (Braun et al. 2018: 113). The proposed features of the specific shapes of contours associated with the expression of irony will be discussed in chapter 2.3.3. However, additionally to sharing a considerable amount of background knowledge and discourse information, familiarity with the individual interlocutor's stylistic linguistic behavior might be necessary to allow for a successful interpretation of ironic utterances. This potential requirement is based on the variances of individual speaker styles which are composed of distinct individual habits of and preferences for applying certain features of intonation (Champagne Lavau et al. 2018: 167). One of the most evident indicators for individual stylistic choices is the expression of affect through prosodic contours. On the one hand, one's educational background regarding register, as the familiarity with different degrees of formality, may influence an interlocutor's ways to express affect prosodically. On the other hand, one's cultural background, defining one's awareness of the degree of acceptability of showing intimacy in conversations, may shape the ways in which interlocutors are willing to express affect in a given situation (Precht 2003: 249). Since irony is frequently used to express forms of criticisms or other interpersonal attitudes, the degree to which specific intonational markers of affect are involved in forming the pitch contours of ironic utterances is not to be ignored (Hellbernd et al. 2018: 605), but attempts to elaborate these prosodic methods of expressing affective stances on the global level of contours have yielded inconclusive results (Grichkovtsova et al. 2011: 1). Nevertheless, it has been shown that the expression of affect in pitch contours can be successfully interpreted by non-native speakers (Grichkovtsova et al. 2011: 1), whereas the same cannot be claimed for sarcasm (Champagne-Lavau et al. 2018: 166). Regarding native speakers, even young children are known to be able to identify intonational indicators of sarcasm (Ben Jannet et al. 2013: 3537). This circumstance may not come as a surprise since it is also children who frequently make use of specific pitch contours, so-called "chanted contours" (Gibbon 2016: 24), to taunt others. These speech melodies can also be seen as a form of irony by applying criticism through chanting a positive statement in a specific way. While specific forms of pitch contours are involved in expressing emotions and attitudes, stance of authority is mainly realized in one's voice quality, constituting another aspect of the

framework that conditions the production of prosodic units (Grichkovtsova et al. 2011: 13). Therefore, specific prosodic contours of pragmatic, sentence or rhetorical modes appear to consist of both, pitch units that have innate qualities based on physical constraints of speech production and the manipulation of these pitch units that amounts to learned qualities based on cultural training.

2.2 Music and prosody

The original Greek meaning of prosody – "a sung song with instrumental music" (Nooteboom 2010: 640) – fittingly illustrates how closely linked melodic pitch contours of music and speech are, which will be elaborated on in this chapter.

2.2.1 Defining music

Depending on the field of study, various perspectives of specifying the inherent features of music have been suggested. From a neurological perspective, music is the "embodying, entraining and transposable intentionalising" of "time in sound and action" and "a social cultural artifact building on the biological inclination for musicality" (Filippi et al 2018: 12); musicality being defined as "a natural, spontaneously developing set of traits based on and constrained by our cognitive abilities and their underlying biology" (Honing 2018: 51). Neurologists further explain music as a "dual coordination system, using both tonality and metric entrainment to engender integration" (Brown 2017: 11). An evolutionary biologist would describe music as a "human cultural universal serving no obvious adaptive purpose" (Masataka 2009: 11), while an acoustic physicist would define it as "an artistic endeavor with substantial intracultural variation and no a priori reason to follow melodic norm" (Iversen et al. 2006: 3041). Ethnomusicology specifies music as "humanly organized sound other than speech" (Brown et al. 2015: 9007), and even more generally it has been termed the "language of emotion" (Besson et al. 1998: 7) or even the "universal language of mankind" (Brown et al. 2015: 8987). Whether one follows the most or the least general definition, the close relation of language and music as primal forms of human communicative behaviors clearly presents itself. As a type of group communication system, music is a result of collective production, where ritualization plays a key role in attaining meaning by using "context specificity and signaling of consequentiality" (Brown 2017: 5), the essential frameworks for coordinating group organization and behavior. The difficulty of explaining musical semantics has been the subject of intense debate, posing the question of similarity of meaning between language and music. Musical expressions simultaneously possess "socio-intentional", "motivational-

structural" as well as "culturally enactive" (Boeckx et al. 2015: 11) functions. The first two of these representing the pragmatic and syntactic qualities, respectively, while the cultural may refer to how these pragmatic and syntactic modalities are learned, ritualized and institutionalized. However, since semantics, as describing processes of shaping units and structures of meaning, pervades all levels of categorization and structuring (Wiggins 1998: 21), the inherent meaning of musical production is not explained by these functions alone. On the levels of categorization and structuring music shows some similarities to the semantics of language, consisting of three systems of meaning: the "extra musical", relating to formal concepts and systems of categorization concerning music such as staves; the "intra musical", referring to the compositional relations of one musical unit to another; and, finally, the "musicogenic", regarding institutionalized traditions of interpretation and ascribing values (Boeckx et al. 2015: 10). Yet these structures are developed and exist within a framework of systematization that is based on the use of verbal language and such might be attributed to any form of artistic expression. However, while the extra musical and musicogenic systems of meaning rely on cultural verbal frameworks to be erected, the intra musical is able to independently attribute meaning to musical units which by themselves merely constitute acoustic signals. By way of referring to the frequency values of sound units to others in sequence or in simultaneity, creating frequency ratios, melodic and harmonic, respectively, relations emerge between single units which are called intervals (Brown 2017: 1). Intervals carry a form of valence, or affective quality, that elicits an emotional response. The type of valence an interval carries relies on its physical properties and is thus elicited in any listener, not requiring any background knowledge. Generally, the more complex a ratio between two frequency values is in terms of its distance to a small-integer ratio, the more tension it creates in its perturbation of air pressure, increasing in dissonance. The closer a ratio between two frequency values is to a small-integer ratio, the less tension between the sound waves of the two frequency values is created, enhancing its consonance. The more dissonant an interval, the more negatively denoted affective states it elicits, and vice versa regarding consonant intervals (Costa et al. 2000: 8). This generating of affect in the process of temporal referencing is exactly what is often referred to as the meaning of music (Wiggins 1998: 21). The intervals used in contemporary musical traditions are determined by musical scale structures, the origin of which is still under debate, ascribing their emergence either to musical theory, constraints of cognitive abilities or cultural preferences (Honing 2018: 52).

2.2.2 Evolutional companions and joint educators

The research on infant perception has provided a number of informative insights into the innateness of musical and speech melodies. At the earliest stages infants do not differentiate between speech and non-speech sounds. Acoustic signals constitute the very first affective cues they are capable to identify, reacting to speech and musical pitch contours indistinguishably. Consequently, the development of the ability to perceive the emotional valence of musical intervals precedes the first processes of language acquisition (Anderson 2005: 49). The first reciprocal communicative behavior infants participate in is "infant-directed speech" (Anderson 2005: 47), in which they, interestingly, are responsible for initiating the shaping of communicative units. It is the caregiver who first unconsciously reacts to the recurring quasi-intonational phrases uttered by the infant in order to comfort her/him as accommodatingly as possible. In turn this copying behavior is then, at the age of six months, mimicked by the infants to produce their first intentionally meaningful speech acts (Masataka 2009: 15). Since the first infant vocalizations constitute by-products of unintentional oral movements when producing coos, the first pitch contours all demonstrate falling shapes, which occur naturally as the vocal folds lose tension when the mouth closes. On the contrary, a rise at the end of a contour requires a willfully produced tension in the vocal folds (Masataka 2009: 15). The identification and fabrication of pitch contours therefore represent the very first encounters of meaningful communicative behavior in a human being's life.

In theories of evolution of language and music three different approaches have been proposed: a separate development of each phenomenon then converging later on; music as an offshoot of language; and, ultimately, language evolving out of vocal music (Fenk-Oczlon et al. 2009: 218). The latter is based on the shared physiological modalities of singing and speaking. Speech and singing are produced with the same oral instruments and their production rates are constrained by the same phrasal boundary – breath taking (Brown 2017: 7). According to the "prosodic scaffold" (Brown 2017: 3) theory, words merely fill out prosodic structures that had been the primary means of language communication into which words gradually evolved into, filling the prosodic units with additional lexical meaning and shaping them in ever more complex ways. This approach is based on different occurrences: the holophrastic nature of prosody, a prosodic contour underlaying every single utterance, even when consisting of a single syllable; the coordinating function of prosodic contours as a framework of every single conversation, signaling speaking turns and the overall informative structure; and the existence of forms of speech that are devoid of meaning but still possess

intonational characteristics, e.g., grammelot (Brown 2017: 15). Furthermore, there are many species of birds, primates as well as other animals who make use of specific melodies as forms of group communication (Brown et al. 2015: 9004).

2.2.3 The universal nature of music and language

Similar to the acoustic signals of speech, instrumental musical sounds constitute complex frequency values, which are made up of their fundamental frequencies and their accompanying overtones (Nooteboom 2010: 644). As much as the size and shape of one's vocal folds and vocal tract are responsible for determining one's individual character of tone of voice, the shape, size and material properties of an instrument determine its characteristic sound. Yet musical instruments are mechanical constructions which are intended to produce discrete frequency values, whereas the human voice is not inherently "tuned" in a specific way. However, the qualities of the human voice as a sound production device might have influenced the development of scales as the structural framework for musical production due to a number of discoveries. In tribal communities of southern Africa a close link between initial phrases of music and speech melodies has been established (Eerola et al. 2009: 238). In more institutionalized musical cultures direct correlations between culture-specific prosodic phenomena and compositional tendencies in the musical tradition of that culture have been detected. For instance, English and French speech differ in their "variability of interval size between successive pitches" (Iversen et al. 2006: 3042), a measure that is reflected correspondingly in compositions of each of the musical traditions. Additionally, the nPVI, the Normalized Pairwise Variability Index computing the "degree of durational contrast between neighboring events in a sequence" (Patel 2005: 61) has been used to compare the rhythm of speech of particular languages with the rhythmical tendencies in the compositional tradition of the respective cultures. In this way the rhythmic properties of speech of British English and French were discovered to correlate with the general rhythmical framework of musical pieces in the respective culture (Iversen et al. 2006: 3034). Japanese percussionists have a tendency to play faster than indicated when note reading, which has been interpreted as reflecting the rhythmical characteristics of Japanese (Fenk-Oczlon et al. 2009: 210). More generally, musical phrases and pitch contours seem to be confined by the same parameters, due to singing being the most primitive and primal form of music production and underlying the same boundary markers, namely, breath taking, as speaking. Due to this physiological boundary, clauses typically consist of five to ten syllables across languages, correlating to a majority of musical phrases comprising six to ten notes (Fenk-Oczlon et al. 2009: 215).

Likewise, the condition that the more syllables constitute one clause, the less phonemes are included in a syllable, also applies to musical phrases in a similar manner, in that the more notes constitute a musical phrase, the smaller the temporal values of these notes (Fenk-Oczlon et al. 2009: 215). Stress, or accent, is also perceived analogically, increased duration and intensity being the main factors in both domains (Filippi et al. 2018: 12). Since vowels, or vowel- like means of sound production, are the main carriers of speech and - regarding singing - musical melodies, these speech sounds possessing their own intrinsic harmonics have been thoroughly analyzed regarding their statistical distribution across languages. Such studies have concluded that the number of vowels used in a language corresponds to the number of intervals used in the musical tradition of that language's culture, e.g., languages with a three vowel system usually only use three intervals in their traditional scale structures (Fenk-Oczlon 2017: 1581). The upper limit of vowels in a language, namely, twelve, is found to be in accord with the upper limit of intervals in a scale (Fenk-Oczlon 2017: 1581), corresponding to the chromatic scale. Scales that include more than twelve intervals allow for standardized variations of these twelve but do not use additional distinguishable, discrete intervallic units (Burns 1999: 222). The lowest number of vowels contained in a language's inventory of speech sounds and the corresponding number of intervals in the musical scale of that language's culture were found to be situated at two. The overall mean of the number of vowels used in a particular language and the number of intervals found in the scales of the musical tradition linked to that language's culture could be established to lie between five and seven. The inherent harmonic values of vowels also tend to determine the direction of melodies in sung musical pieces, especially when vocables, or nonsense syllables, are used, as in many traditional forms of music, e.g., yodeling (Fenk-Oczlon 2017: 1581). Furthermore, some researchers were able to show that the frequency values of vowel phones' first two formants, the ones most important in language perception, include all twelve interval ratios of the chromatic scale. Thus, the harmonic qualities have been named a potential origin for the emergence of musical scale structures (Choi et al. 2007: 9852). The expression of emotions represents another level on which musical and prosodic units share significant features, enacting similar ways of signaling meaning. For example, musical as well as prosodic units utilize similar acoustic techniques to indicate states of happiness: higher tempo, an increase in average pitch and/or sound level, a higher variability of pitch and quick vocal onsets and, lastly, a rising pitch contour (Filippi et al. 2018: 12). Music and language are both "universal and primary forms of communication" (Anderson 2005: 43) that are generationally transmitted but also hold their innate qualities. Since they both stem from utilizing the same

instrument, the vocal devices, they are subject to the same physical limitations and potentials. Hence, when analyzing prosodic contours and musical melodies, the same parameters for distinguishing different types of pitch movements can be applied: directions as up, down, or level, the size in semitones, the rate of change in semitones per second and the duration of individual units (Nooteboom 2010: 649).

2.2.4 Universal boundaries

Music and speech are "two organized sound systems that employ rhythmic and melodic patterns and rule-governed sequences" (Patel 2005: 59), however, they greatly differ in the ways these rules are structured. As discussed in 2.2.1., language and music do not carry and produce meaning in a directly comparable manner, nevertheless, since prosodic contours are an inherent aspect of spoken language, and arguably its evolutional predecessor (see 2.2.2.), the distinction is not straightforward when regarding speech and music. Language's primary communicative purpose lies in the logically encoded transmission of factual information and its logical deduction (Wiggins 1998: 20). Phonetic units become this meaningful factual information by being referenced to an "extralinguistic designated space" (Besson et al. 1998: 5), which is regulated by conventionalized frameworks. Musical sounds, on the other hand, do not require a listener to be familiarized with any conventions to perceive its primary meaning – the evocation of affect. Thus, this affective stimulation results in an immediate emotional response to physical stimuli in the listener and may further cause the listener to reflect upon this experience (Wiggins 1998: 21). Even though such affective meaning might also play a crucial role in the transmission of attitude and stance in spoken language, affect carried by prosodic contours is always linked to the literal content of an utterance. These verbal information units are coordinated and organized in conventionalized manners of composition with relatively fixed rules that presuppose adherence for successful communication, contrary to music, where such rules are of a rather abstract nature (Boeckx et al. 2015: 7). Structural rules in music differ from ones in language in that they are without distinct semantic boundaries. The crossing of musical boundaries does not result in causing a distortion of semantic transfer, but the musical meaning, in the form of affect, is still produced. The infinite variability of musical elements is what makes them meaningful, since the emotional and aesthetic mechanisms of affect rely on creating patterns of tension and release (further discussed in chapter 2.3.2. which can be widely, artistically and innovatively, explored in musical expression (Pearce et al. 2018: 498). Tension and release patterns created by rhythmical and melodic sequences can also be produced by prosodic contours but to a far

lesser extent, since prosodic contours are embedded in a reciprocal relationship of turn-taking between the interlocutors of a conversation, and the standardized inventory of pitch contours of a language (Boeckx et al. 2015: 12). Even though there are turn-taking behaviors in musical performances, so-called "call and response" acts, these forms of musical communicative actions are distinct from the ones in speech in that turns performed by an individual are usually followed or preceded by turns performed by a group – and when they are followed by another individual's performance, the whole performance is embedded in some kind of collective artistic or ritualistic act (Brown 2017: 5). The rhythm of speech is dictated by morphological and syntactic modes in which lexical elements are combined with each other missing one distinguishing feature of musical rhythm, namely, isochrony – "the regularity of timing" (Gibbon 2016: 28). An isochronous beat equips music with its fundamental repetitive metrical structure, producing tension and release patterns in relation to expectations created by the overarching rhythmical framework to elicit emotional responses. The varying metrical structure of speech is achieved through combining phonemes that do not contain isochronous properties that refer to an overarching metrical framework on which its primary meaning depends. As a result of music constituting a group communication system, it has been suggested that isochrony arose from the need for an effortlessly memorable shared framework for social coordination (Burns 1999: 218). Musical as well as speech melodies are "composed of sequential events, variations in pitch unfolding in time" (Anderson 2005: 50). Still the differences in attributing specific qualities to these pitch variations in each realm comprise a substantial discrepancy between music and speech. The pitch units are relative in both, since their intervallic value – their tonal identity in regard to other individual units of the same nature – is determined by the ratio to the preceding or subsequent pitch value (Brown 2017: 11). However, the more institutionalized a culture's musical tradition becomes, the more important is the role that actual discrete frequency values play in a musical culture's structures, since the building of musical instruments requires standardized frequency measures to enable groups of people to play together. The European musical traditions have developed scale structures that are based on fixed physical quantities for measuring frequency ratios and frequency units in order to standardize harmonic relations as well as tone units, respectively (Burns 1999: 243). The basis of every scale is the measure of an octave, representing the starting point of a scale, in form a frequency value in Hertz, times two (Gibbon 2016: 12). The octave is the most consonant of all ratios, followed by the fifth, with a ratio of 3:2, and the fourth, with a ratio of 4:3. These intervals are the most preferred cross-culturally, arguably due to their small integer values (Anderson 2005: 45) – the inclination

towards consonant intervals being attributed to the brain's preference for periodically regular signals (Burns 1999: 241). This preference is already reflected in six months old infants who react to the mistuning of the most consonant intervals (Anderson 2005: 45). However, detecting such deviations play no explicit role in language perception, even though the formants of vowels are made up mostly of musical intervals and musical intervals are frequently found as the components of pitch contours (Adriaens 1991: 64). Therefore, the lacking of tone and interval recognition in speech may also be attributed to the multimodal nature of speech acts, comprising the activation of various perception channels, e.g., the auditory and visual organs, which are limited in their recording and processing capabilities (Burns 1999: 219). Furthermore, the dialogical framework of a conversation creates pragmatic expectations, assigning specific sets of possible reactions, that urge a listener to respond to acoustic input actively and meaningfully within more intricately specified conventions. Therefore, the decoding of acoustic signals employs more complex sets of mental capacities in speech than in music perception – the brain prioritizing the most relevant information for managing interactional behavior, for which the precise identification of individual pitch and interval values is neither necessary nor meaningful. The use of quantifiable recurrent pitch values and their coordination in conventionalized hierarchical frameworks, or scales, constitute the main differences between musical, using discrete, and speech melodies, using indiscrete pitches (Brown 2017: 3). The scale most commonly used is the equal temperament scale made up of twelve equally spaced intervals within an octave. The equal spacing of intervals, instead of only using integer values as in other scale systems, has turned out to be the most practical, since it allows for a standardized construction, and thus tuning, of musical instruments and for the transposing of harmonic and melodic material into different keys. The smallest unit of this logarithmically fixed scale is a semitone, corresponding to one twelfth of an octave (Gibbon 2016: 12). Many non-European cultures use scales including smaller intervals, so-called quartertones, but only as variations of the twelve semitones of a chromatic scales. These quartertones usually do not comprise an exact, or even an approximate, half of the value of a semitone, as their name implies, but are mostly significantly smaller. Such small intervallic units are also not perceptually relevant, since listeners are unable to categorically recognize them and distinguish them from their chromatic variants (Burns 1999: 222). This physiological limitation might also be the foundation for the chromatic structure of standardized scales consisting of twelve semitones.

2.2.5 Intercultural and intracultural variation

Prosodic forms of behavior are socially learned and "shared among all speakers of a dialect" (Rumsey 2015: 20). However, in English the mainstream varieties do not display great degrees of variety of global prosodic features, as contour inventories, but rather of local prosodic features, as phonemic structuring and variability, especially, accent or stress (Grice et al. 2019: 14). In some languages, including English, rhythm of speech and speech rate constitute indicators for dialects (Di Cristo et al. 1999: 38). Significant differences regarding the global shapes of pitch contours occur most notably at the marginal regions of contours, depending on the ways cultural norms and conventions frame turn-taking in dialogical exchanges. Furthermore, the application of markers of attitude and stance possesses various degrees of appropriateness in different contexts, which are determined by the national or regional culture of the language community they are embedded in (Precht 2003: 255).

2.2.5.1 Tonal languages

In contrast to English varieties of the Inner Circle, those of the Outer Circle do exhibit clearly distinctive prosodic contour inventories, especially those varieties where a tonal language constitutes the primary or historical native language of the region. Singaporean English, for example, does not possess the same phrasal structure of prosodic contours as standard varieties, spanning over just one to two content words instead of an entire clause (Grice et al. 2019: 14). In West African varieties and the Hong Kong variety pitch accents with a distinctive tonal character, corresponding to the tonal values used in the native language of the region, are applied to the prosodic contours of English utterances, losing their lexically and semantically relevant meaning that they originally contained. Interestingly, the ways in which these are realized in West African and Hong Kong English resemble each other closely, pointing to inherent physiological properties being responsible for the shape of prosodic pitch contours as much as in non-tonal, or intonational, languages (Grice et al. 2019: 16). From an evolutionary perspective, linguists have pointed out that since the majority of the world's languages are tonal languages, this group of languages very likely also evolved as the first languages from which the intonational languages derived later on (Tillmann 2014: 124). Tonal languages differ from intonational languages in that different levels of pitch or pitch movements differentiate lexical forms. Thus, variations of tone constitute the meaning of words additionally to the varying consonants and vowels. In intonational languages this tonal lexical differentiation may be achieved only through the distribution of stress (Nooteboom 2010: 642). Stress can be realized through the prominent manipulation of a

syllable's duration, intensity or pitch – but the latter, contrary to lexical tone of tonal languages, does not express its own tonal function of lexical significance (Gibbon 2016: 5). The way in which global prosodic contours operate are still very similar in both language groups. For instance, the expression of affect and the signaling of questions and statements are realized in the same fashion (see 2.1.2). The tendency of paratones to have a high initial F0 and a low final F0 is also reflected in tonal languages (Gibbon 2016: 23). The hypotheses that speech evolved from singing and that tonal languages were the first to evolve are also based on findings in the neurological processing of music and speech. It has been demonstrated that people with a high proficiency in tonal languages are also better at discriminating musical and speech melodies than those without. There is also a higher percentage of possessors of absolute pitch, which is the uncommon capacity to identify musical notes without a reference tone, in the populations of tonal language speakers. On the other hand, trained musicians have improved competences of pitch perception in tonal as well as non-tonal languages (Tillman 2014: 127). The phenomena of lexical tone, unique to tonal languages, and prosodic contour, featured by all languages, interact to a certain extent due to syntactic structuring in tonal languages, e.g., the signaling of parts of a sentence simultaneously marking corresponding intonation patterns. Nonetheless, the tone value of a syllable is always expressed regardless of its location in a pitch contour (Di Cristo et al. 1999: 20).

2.2.5.2 Variations between related languages

Regarding the inventories of pitch movements within a language family, every language possesses its peculiarities. These features have such distinct qualities that even young infants are able to distinguish their caretaker's native from other tongues (Iversen et al. 2006: 3043). Within the Germanic language branch there is also considerable variation in a number of prosodic features. The typical pitch range of English within an utterance is about one octave, equaling twelve semitones, whereas in German it is approximately seven and a half semitones (Adriaens 1991: 97). Single pitch movements in English are also generally greater in pitch height, spanning a larger number of semitones, but smaller in length than in German. The peak frequency value in a stressed syllable occurs – about 70 milliseconds – later in German than in English (Adriaens 1991: 106). Other Germanic languages again show different results for the measures presented above. In a typical declarative sentence English is said to possess three lines of declination, referring to parallel declining downward movements on different pitch levels along which a pitch contour orientates itself, while German only possesses two (Adriaens 1991: 97). In most German dialectic variants rising pitch occurs in each stressed

syllable except the last. In American English stressed syllables tend to contain falling patterns, whereas “a step-down in pitch for each stressed syllable between the onset and the nuclear movement” (Di Cristo et al. 1999: 22) is typically produced in British English. Each Germanic language also shows its own tendencies in the quality of duration of final syllables (Di Cristo et al. 1999: 41). In general, Germanic languages are left-headed, having the word stress typically on the first syllable of each word, in contrast to right-headed languages, e.g., Romanic languages, having the word stress on, or towards, the final syllable of each word (Di Cristo 1999 et al.: 24). However, researchers have become cautious about making generalizing classifications concerning the rhythm and melody of speech, terming traditional approaches as *stress* and *syllable timed* “oversimplifications” (Gibbon 2016: 28). As mentioned above, even though there is no doubt about the existence of a specific inventory of pitch contours for every language or dialect, setting up fixed classifications remains to appear futile because of the multifaceted nature of speech acts and the exceptions that have been found to occur regarding even the most basic rules. For instance, in some northern dialects of British English and dialects of the southern regions of the US, declarative sentences typically end in final high pitches, in contrast to the supposedly universal falling pattern (Di Cristo 1999 et al. 39).

2.3 Irony and music

The successful reception effects of music and rhetorical figures as intended by the producer are a matter of aesthetics and thus, arguably to a significant degree, also a matter of subjective perception. It is also a question of subjectivity which musical preferences one develops or if one is able or willing to expose the dissimulation – as the Greek etymological origin of irony implies – with which one is confronted as the perceiver of an ironic utterance. Additionally, the processes of musical perception and the decoding of rhetorical figures which are expressed through intonation share most of their neurological pathways.

2.3.1 Neurological frameworks of music and speech

As discussed in 2.1.3., discourse background and shared background knowledge play a key role in a listener’s success rate in decoding instances of irony. The accumulation of such bodies of knowledge by an interlocutor as well as her/his awareness of discourse factors are based on her/his cultural background and personal experience. In musical perception proficiency in decoding abilities can actually be trained and acquired. For instance, professional musicians are trained to be able to recognize musical intervals, attaining the

ability called relative pitch (Burns 1999: 229). However, it is not possible to attain absolute pitch through training but only through early exposure (Tillmann 2014: 127). Surprisingly, when possessors of absolute pitch identify intervals, they do not apply their absolute pitch competences but instead have to utilize relative pitch strategies if available, implying that there are different brain areas for innate, or at least critical phase-related, and culturally trained perceptive abilities (Burns 1999: 237). Musical meaning is attained through the emergence of an affective reaction in the listener who holds certain structural expectations in her/his mind. The way in which these expectations are triggered or suppressed leads to a resolution producing an emotional response (Pearce et al. 2018: 498). Therefore, similarly to the perception of irony, a listener of music draws from her/his cultural background knowledge to interpret the acoustic signals that are being perceived. However, the expectancies in music as well as prosody are also based on innate physiological qualities (Eerola et al. 2009: 232). Music appears to be "hardwired into the human brain" (Masataka 2019: 11) owing to the fact that young infants react to it intensely. According to the "direct route hypothesis" (Iversen et al. 2006: 3043), prosodic structures are acquired through "tracking patterns in the environment and acquiring implicit knowledge of their statistical properties without any direct feedback" (Iversen et al. 2006: 3043). This process is also referred to as statistical learning, which constitutes the transfer of statistically relevant structures from the short into the long-term memory (Eerola et al. 2009: 262). Statistical learning is crucial in attaining linguistic knowledge, where one internalizes and reproduces repetitive motifs that are pragmatically relevant (Fenk-Oczlon et al. 2009: 217), and in obtaining musical preferences and decoding abilities (Eerola et al. 2009: 243). The inherent mental and corporal restraints of the human body, regarding the auditory and articulatory mechanisms, respectively, further condition the framework of musical and intonational production and reception. For example, the fixity of the phonemic inventory arises due to the mechanical constraints of the speech organs which are represented as "invariant motor commands" (Boeckx et al. 2015: 9) in the brain. Due to these invariants, some aspects of an utterance's pitch contour, e.g., its overall declination, are not deliberately controlled (Nooteboom 2010: 649). In both, speech and music perception, the long-term memory, referencing the material at hand to already stored structures, and working memory are activated. The neural areas of the long-term memory are in separate locations for music and language, but, according to the "shared syntactic integration resource hypothesis" (Patel 2005: 64), the active processing of both occurs in the same neural regions. This hypothesis is supported by the results of studies comparing language and musical aptitude, which seem to promote each other reciprocally (Reiterer et al. 2007). Nonetheless, the ways

in which the working memory is activated differs in a substantial manner, since not even musicians are able to correctly remember the sequence of consecutive intervals, whereas verbal content does not generate such difficulties (Boeckx et al. 2015: 5). In contrast to language perception, the transition of meaning via music, as affect produced by sequences of frequency ratios, does not require definite musical experience, but is processed implicitly also without any cultural training (Tillmann 2014: 126), inasmuch that conventions of harmony are implicitly known even by non-musicians (Patel 2005: 65). Yet these affective properties of acoustic signals utilized in music also play a role in language perception in the form of prosody, anteceding the lexical, syntactic and propositional qualities that are processed in posterior neurological pathways (Boeckx et al. 2015: 12). Learning melodic patterns, including prosodic contours, is an intuitive process functioning through its "biological automaticity", not requiring any psychological negotiation, while acquiring linguistic knowledge is a conscious effort that functions through "structural automaticity" (Anderson 2005: 53). However, the storing processes of these learning processes take different routes. In auditory pattern perception music and speech share the activation of three realms: learning of rules, perception of rhythm and perception grouping (Ten Cate et al. 2014: 9). This grouping of units regards the assorting of the smallest local units, as tones and phonemes, respectively, into larger global structures (Boeckx et al. 2015: 6). From there on, prosodic information is gradually abstracted to eventually converge with linguistic content stored in the long term memory (Hellbernd et al. 2018: 605). This differentiation is also made explicit in that the brain waves measured in harmonics are linked to ones in syntax but not to semantics (Besson et al. 1998: 3). Regarding the processing of prosodic and musical units, it has been found that the more global prosodic units, as pitch contours, are stored in different locations than musical units (Brown 2017: 3), but the decoding and storing of phonemes shares its cerebral position with musical units (Reiterer et al. 2017: 13). The complexity of prosodic phenomena is also portrayed in the findings that showed how socially significant, emotionally labeled prosodic information is processed in different brain areas than neutral prosodic content (Hellbernd et al. 2018: 605).

2.3.2 Building expectations: tension and release patterns

The ways in which physiological, especially neurological, constraints and cultural conventions shape the frameworks of musical and prosodic production and reception form certain expectations and expectancies in a listener. The effects of irony, as a rhetorical figure, and music, as an artistic device eliciting affect, build on the listener's discourse and

background knowledge made up of these physiological and cultural expectations. Contextual incongruity occurs when such expectations mismatch the actions they are referred to, as in the case of music with tensions and release patterns of intervals holding different valence, and in the case of irony with one modal level of a speech act contradicting another. The innate ability to perceive mistunings of the most consonant intervals even in untrained listeners (Burns 1999: 244) and the universal intuitive association of consonant and dissonant intervals with positive and negative valence (Brown 2017: 13), respectively, are some of the physical qualities that create melodic and harmonic expectations. However, cultural influences may also manifest themselves in the physiological domains. For instance, physical arousal is prompted by stylistically unconventional progressions of chords and notes, causing heightened skin conductance and a reduced heart rate in the listener. Such violations of expectations based on conventions can also occur due to a delaying of an expected follow-up in a sequence or a preceding element creating an ambiguous set of expectations (Pearce et al. 2018: 499). Irony operates similarly through confronting the listener with an ambiguous set of possible interpretations of one utterance (Braun et al. 2018: 112). Delaying a part of an utterance is also used as a prosodic mean to mark irony, pointing at a "hinting connotation" (Di Cristo et al. 1999: 27). Finally, the achievement of contextual incongruity inherent to the rhetoric effect of irony can also be based on the violation of stylistic linguistic conventions (see 2.1.4.). In musical perception one distinguishes between "perceptual uncertainty", referring to a listener's set of expectations based on individual preferences and background knowledge, and "stylistic uncertainty" (Pearce et al. 2018: 499), referring to expectations a particular genre brings along. The same parameters could be established for framing instances of irony in spontaneous speech, since, on the one hand, one major aspect of the contextual framework of a conversation within which some kind of incongruity marking irony arises is the individual speech style of an interlocutor (Di Cristo et al. 1999: 2). For instance, listeners are far more likely to deal with the semantic ambiguity of an ironic utterance competently when they are in solidary type of relationship with the speaker than when their relationship is lacking in solidarity (Mauchand et al. 2019: 4). On the other hand, the degree of success of an ironic utterance being interpreted as such also depends on the addressee's background knowledge and awareness of particular discourse practices (Mauchand et al. 2019: 1). The main structural aspect to which melodic expectations regarding prosody refer to, namely, pitch proximity and pitch reversal, are as relevant in music. Based on the former listeners are lead to expect rather small intervals in succession, while the latter guides listeners to expect pitch movements to change direction after the occurrence of large intervals (Pearce et al.

2018: 495). Additionally, the factor of temporal proximity influences a listener's capacities to recognize the sequential structure of melodic events (Eerola et al. 2009: 262). While the many pragmatic functions of a speech act do not allow for much variability in the rhythmic structuring of an utterance (Nooteboom 2010: 661), the isochronous beat structure of music neither permits much temporal mobility (see 2.2.4.). However, while these tension and release patterns are of a predominantly physiological nature in music, in speech they are always strongly linked to the conceptual framework of the literal content (Boeckx et al. 2015: 5). The perceptual abilities in both domains are also constrained by limits of the working and long-term memory. In speech intonation groups do not cause any perceptual problems for the working memory as long as they are confined to a peak of five complex or ten simple syllables, whereas the long-term memory is able to store five to seven different vowels, and variations of them. Equivalently, musical units consisting of six to eleven notes do not overstrain the working memory, while five to seven tones per scale are stored unobstructedly in the long-term memory (Fenk-Oczlon et al. 2009: 220). How important the violations of conventions and the expectations based on conventions are as means of production and reception, respectively, of the aesthetic effect of music is displayed in the phenomenon of out-of-key harmonic progressions correlating to shorter neural response times than in-key harmonic progressions (Pearce et al. 2018: 495). However, the familiarity with tonal material through cultural training plays as important a role in shaping a listener's proficiency in detecting the indented tension patterns by a musical performer (Boeckx et al. 2015: 11), as the awareness of social norms regarding the appropriateness of expressing stance in speech plays in deciphering the intended ironic cues by a speaker (Precht 2003: 240). Additionally, the relationship between a musical performer and her/his audience may be compared to the "comradery" (Braun et al. 2018: 112) emerging between speaker and listener in instances of irony. While the audience of a musical performance is responsible for approving the performance of a musician, e.g., through applause, the addressee of an ironic statement is responsible for enabling the ironic effect to be realized by detecting the communicative intention and approving the violation of conventions. Expectations regarding the rhythm of music and speech are also found to have innate and learned qualities. Two innate procedures, "regularity extraction" and "segmentation into groups", as well as two culturally obtained mechanisms, "hierarchical metric organization" and "hierarchical segmentation organization" (Besson et al. 1998: 2), realize the perception of rhythm. Rhythmic expectations in music are based on its isochronous structure and perceptual constraints. For musical sequences to be most effectively perceived, rhythmic units are to consist of no more than ten pulses.

Correspondingly, intonation units are perceived most effectively when they consist of no more than ten syllables (Fenk-Oczlon et al. 2009: 215). Therefore, the "isochrony hypothesis" (Fenck-Oczlon et al. 2009: 205) has been developed, proposing that any language consists of recurring rhythmic regularities based on its specific morphological and syntactic composition.

2.3.3 How prosody functions as a marker of irony

So far studies have failed to find any empirical indications for the existence of definite musical intervals in prosodic contours of ironic utterances (Choi et al. 2007: 9857). However, there have been no studies that looked at this phenomenon in spontaneous speech (Brown 2017: 17). The results that researchers have been able to gain are abundantly ambiguous, as one would expect regarding the pragmatic complexity of the contexts of individual speech acts. Some studies of ironic utterances in English and German found "inverse pitch obtrusion" (Attardo et al. 2003: 244), referring to the nucleus being pronounced at a lower pitch than its surrounding material, while others observed the opposite, namely, higher but also exaggerated and intensified pitch in the nucleus. Further prosodic indicators that have been detected to mark irony on a more global level include exaggerated patterns of intonation, singsong melody, falsetto and comparably monotonous intonation (Attardo et al. 2003: 245). Some analysts have even established a model describing specific stages of the production of pitch contours marking ironic utterances. From first to last, theses stages consist of using chest voice and creaky voice, head voice and rising pitch and, finally, steady low pitch (Ben Jannet et al. 2013: 3541). Regarding the rhythm of speech, studies have yielded similarly contradicting results. Some have found an increase in speech rate, while others have detected the opposite, a decrease in speech rate (Attardo et al. 2003: 246). The following markers of irony in the rhythm of speech have additionally been discovered: generally wider arrangement of pitch contours, prolonged pauses between words and lengthening of the nucleus. Furthermore, extralinguistic markers of irony have also been identified concerning the speech organs' mechanisms, as increased nasalization and softened tone of voice, or the so-called "gesture prosody" (Gibbon 2016: 8), including the lowering or raising of eyebrows, the wide opening or rolling of eyes, twinkling, grinning, nodding, blank facial expressions and the tongue-in-cheek gesture (Attardo et al. 2003: 246). It has been claimed that prosodic cues for irony are more easily distinguishable when uttered by females than males, linking this occurrence to the comparably superior ability of females to identify emotional expressions in speech (Ben Jannet et al. 2013: 133). The multitude and variety of these results presumably originate from the contextual complexity of irony as a rhetorical figure. In an utterance

expressing irony the speaker communicates “multiple messages by layering propositional and non-propositional information” and provides “prosodic cues in contextually dependent ways that map differentially onto the simultaneously presented info” (Fox Tree 2005: 272). Hence, it depends on the intentions, attitudes and emotions a speaker holds toward the listener as well as the topics discussed and the discourse background of the situation, which means of communicating ironic cues a speaker finds appropriate and is willing to employ for a specific speech act. Dripping sarcasm, for instance, often presupposes either an authoritarian hierarchical relationship between the interlocutors or authoritarian personality traits possessed by the speaker. Generally, sarcastic utterances in English are frequently pronounced with a diminished average F0, reduced variation of F0, decreased rate of speech and more rigid tone of voice (Mauchand et al. 2019: 3). Regarding the interrelation of gestural and prosodic cues of irony, a diversity of results has again hindered researchers from establishing definite categories, since both, gestures and prosodic contours, have been detected to be able to override the other, rendering the subdued domain less meaningful (Attardo et al. 2003: 246). Furthermore, rhythmic and melodic indicators for irony also seem to be able to function independently of each other (Attardo et al. 2003: 270). Since every language has its own inventory of standard pitch contours, one would also assume different prosodic features or means to indicate irony in each language. Within the Germanic languages English and German possess quite similar mechanisms to mark irony, but there are considerable differences between neighboring language branches. For instance, in French, ironic and sarcastic utterances demonstrate a higher overall pitch level than their neutral counterparts. Furthermore, incidences of sarcastic speech also use a wider pitch range, again contrary to Germanic languages. Nonetheless, sarcastic utterances in French share one common mechanism with Germanic languages, namely, the prolonged overall duration of phrases (Ben Jannet et al. 2013: 3541). Interestingly, the modes of decelerating speech tempo, increasing mean F0 but confining F0 variability have also been found to be utilized for the expression of irony in a tonal language, namely, Cantonese (Mauchand et al. 2019: 3). In addition to the mechanisms shared with English, in German irony is frequently expressed with higher intensity, thus with higher volume. Generally, specific intervals used in a prosodic contour that includes markers of irony have not been detected. There is also no consensus on which interval size might actually be perceptually relevant in prosodic contours, however, it is well established that human pitch perception is remarkably accurate (Nooteboom 2010: 646).

3 Field study

An “ironic tone of voice” (Mauchand et al. 2019: 3) has been denied as well as affirmed to exist. On the one hand, researchers emphasize the futility of endeavors to establish such a category, on the other hand, investigators speak of a positive consensus on the question (see 2.1.5.). The field study of this thesis is intended to provide a more profound insight into the musical properties of pitch contours expressing irony. Firstly, a quantitative approach was applied to collect a fair amount of data. Secondly, the data collected was analyzed according to parameters which a majority of findings agrees with. Thirdly, sets of data which corresponded most with these parameters were selected and qualitatively investigated concerning their melodic properties in the form of intervallic values measured in semitones.

3.1 Collecting data

For this field study twenty German native speakers and twenty English native speakers were asked to utter six sentences each, being recorded by the author with a *Zoom H2n* recorder at the rate of 44.1 kHz, which surpasses the recommended minimum of 16 kHz (Gibbon 2016: 10). Before the actual recordings were conducted, the subjects were first asked to read through and sign the standard consent form of the University of Vienna. Afterwards, each subject was asked to fill out a short form indicating their sex, age and information about their linguistic, musical as well as, if available, professional speaking profiles. In the language section subjects were asked about their native tongues, their dialects and their foreign language learning experience in years. The musical profile was established by asking the subjects to provide if they had had any education in singing, how many years of experience they had had in singing and how many hours per week they practiced singing. Subjects were also requested to specify if they had learned any musical instrument and, if so, indicate if they had had any musical education, which instruments they played, how many years of experience they had had in playing them and, finally, how many hours of playing music on instruments they practiced per week. These measures where collected since professional musicians are known to have improved abilities of pitch production and perception in music as well as speech at their disposal (Tillmann 2014: 125). Therefore, musically trained and experienced subjects may also be able to produce specific pitch contours more accurately than non-musicians, regarding the intervallic relation between the pitch values of two neighboring vowels. These competences of trained musicians are reflected in the brain as an increase in grey matter (Patel 2005: 65), as a faster reaction time to prosodic stimuli (Pearce et al 2018:

498) and increased brainstem activity (Brown 2017: 4). Further, speakers were asked to indicate whether they had received any education in professional speaking, if available, how many years of experience they had had in professional speaking and how many hours per week they practiced professional speaking. The subjects were told that professional speaking education included any kind of institutionalized training in public speaking and that professional speaking practice included any type of public speaking in a professional context. These criteria were included to form measures about the speaker's awareness of and training in using intonation intentionally. The recordings were conducted either in the living or working spaces of the subjects or at the facilities of the University of Vienna.

The German native speaker group consisted exclusively of individuals speaking different varieties of Austrian German. The mean age of the participants equaled 41.5 years. The German speaking group consisted of eleven male and nine female participants. The mean age of the male participants amounted to 43.5 years, with the youngest subject being 23 years of age and the oldest 61 years old. The female average was 39.1 years, the youngest being 23 and the oldest 61 years old. The English speaking group included native speakers of different varieties from various British English and American English variants to native speakers from New Zealand, Australia and India. The mean age of the native English group equaled 35.2 years, the youngest being 22 and the oldest 49 years of age. There were ten female and ten male speakers. The mean age of the male speakers amounted to 36.2 years, the youngest subject being 28 and the oldest 49 years old. The average of female speakers was 34.1 years, the youngest being 22 and the oldest 49 years old. These numbers are not to be ignored, since the contour inventories of languages are subject to generational change (Di Cristo et al. 1999: 39). Every language group was further divided into two groups each.

The first group was not informed about the study's goals and was asked to utter six different sentences of which three were contextually marked as ironic and the other three were not. The order of the sentences was randomized with five different types of order overall to get the same amount of results for every sentence. The second group was informed about the study's goal to investigate the differences between pitch contours supporting the literal meaning of a sentence and ones hinting at ironic interpretations. This group was to read three sentences, each of them twice, once in an ironic manner and the other time not marking the utterance with irony. Again different types of order were arranged to gather the same number of results for each sentence. The first half of the informed groups was to start with the ironic version

followed by the non-ironic version of a given sentence, whereas the second half always began with the non-ironic version followed by the ironic version of a sentence. The form off of which the speakers read the sentences was headed by the following instructions (or its German translation, respectively):

> In this survey you are asked to assume 6 different speaker roles which are situated in meticulously specified conversational contexts. You are to utter one single phrase per situation, applying ways of intonation which you perceive to correspond to the given contexts most suitably. Your utterings will be recorded and the data will be saved and evaluated anonymously.

Each of the sentences that the subjects were to perform was preceded by its contextualization through three indicators, namely, the speaker's role, the location and the receiver of the speaker's message. These contextual clues where provided in order to create a relatively confined environment as suggested for studying prosodic cues of irony (Antoniou et al. 2015: 46). The first and third indicated measures tried to suggest the degree of intimacy between the speaker and the receiver, defining the degree of appropriateness of expressing stance and affect. However, the mood of the speaker and the speaker's attitude towards the receiver also determine the ways in which prosodic mechanisms are employed (Di Cristo et al. 1999: 2). The subjective inventory of connotations of an individual is as hard to frame as making predictions about an individual's personal stylistic language profile or a speaker's preferred ways of behavior regarding the expression of affect or emotions. Therefore, in this thesis an inventory of 240 utterances was taken as a quantitative basis from which the most competently performing subjects were selected. These competences constituted measurements that are regarded as reflecting the expression of irony through prosody by a majority of researchers.

Following the three contextual indicators mentioned above a short piece of text, spanning a minimum of two and a maximum of seven sentences, explained the situation in which the sentences, that the subjects were going to read and articulate, were embedded. The situation described a form of behavior displayed, an action conducted or a remark uttered by the receiver of the message. If the sentence was meant in a literal or ironic way was explicitly expressed in the way that the supposed background knowledge of the speaker was provided and how the indicated background knowledge of the speaker correlated to the sentence the subjects were to read and utter after reading the contextualization. Before handing out the

pieces of text, the subjects were informed that the more they tried to put themselves in the shoes of the supposed speaker the more useful the results would be. Furthermore, the participants were asked to use intonation in a manner that they would naturally apply in real life situations. Recordings were only redone, when the subject insisted on doing so, which only occurred in cases of misspeaking. In a few incidents, namely, three times with English native speakers and once with a German native speaker, the wording was altered since the subjects had stated that they would use a different literal content for the same meaning in real life situations. The author did not object to such requests, because these proposals only concerned such minor alterations as the substitution of a single word with another. The whole procedure lasted about fifteen minutes per subject.

The recordings were then analyzed using the Praat plugin *prosogram* (Mertens 2004). This program is designed to produce F0 stylizations that imitate F0 contours as how they would supposedly be perceived as speech melodies by a listener (Iversen et al. 2006: 3036). The difficulty of transforming an F0 contour depicted in a spectrogram into a stylized pitch contour is based on the complexity of establishing the pitch values of "quasi-periodic complex sounds like speech" (Nooteboom 2010: 644) as they are actually perceived by a listener. The single units of a speech melody contain a multitude of frequency values composed of the various, harmonically related, overtones of the F0. The perceived pitch emerges from the overlapping of the frequencies of the various harmonics of a speech signal. Hence, the presence of F0 values is not even a physical requirement for the perception of pitch, e.g., as in telephone conversations, where frequencies below 300 Hz are not transmitted (Nooteboom 2010: 644). The prosogram executes its calculations based on a number of assumptions on how the perception of prosodic contours operates. Among these procedures one finds the division of pitch contours into units corresponding to the size of a syllable and the condensing of the speech signal, which is overloaded with spectral and amplitude fluctuations caused by different acoustic side effects that the operation of the speech organs produces (Iversen et al. 2006: 3036). The latter is determined by two threshold measures. The so-called "glissando threshold" (Iversen et al. 2006: 3036) describes specific sizes of changes in pitch levels which determine if a pitch movement is perceived within the scope of a syllable. The "differential glissando threshold" (Iversen et al. 2006: 3036) measures particular sizes in pitch movement which determine if a change of direction is perceived. Lastly, when the pitch movement is of a subliminal value it is temporally integrated in the stable pitch value of a syllable. The prosogram software applies these two thresholds as well as the

temporal integration of subliminal values in an algorhythm producing discrete pitch values for the vowels indicated in semitones. In rare cases it also generates pitch values for consonants, if so, then mostly for prominent sonorants (Iversen et al. 2006: 3036). By computing the semitone values of acoustic units the prosogram applies a logarithmic fashion of measuring pitch values, which has been defined as the most suitable device for investigating perceived speech (Nolan 2003: 774).

In this analysis the first steps in Praat comprised the following sequential procedures: trimming of the sound files; extracting the trimmed files' visible pitch contours as graphs; manually annotating these graphs; and, finally, drawing the visible pitch contours including their annotations in the basic drawing mode. These graphs were then combined with the graphs generated with the prosogram which were created in the setting "wide, rich, with values pitch targets in semitones" and in the output mode "one strip per file" (Mertens 2004).

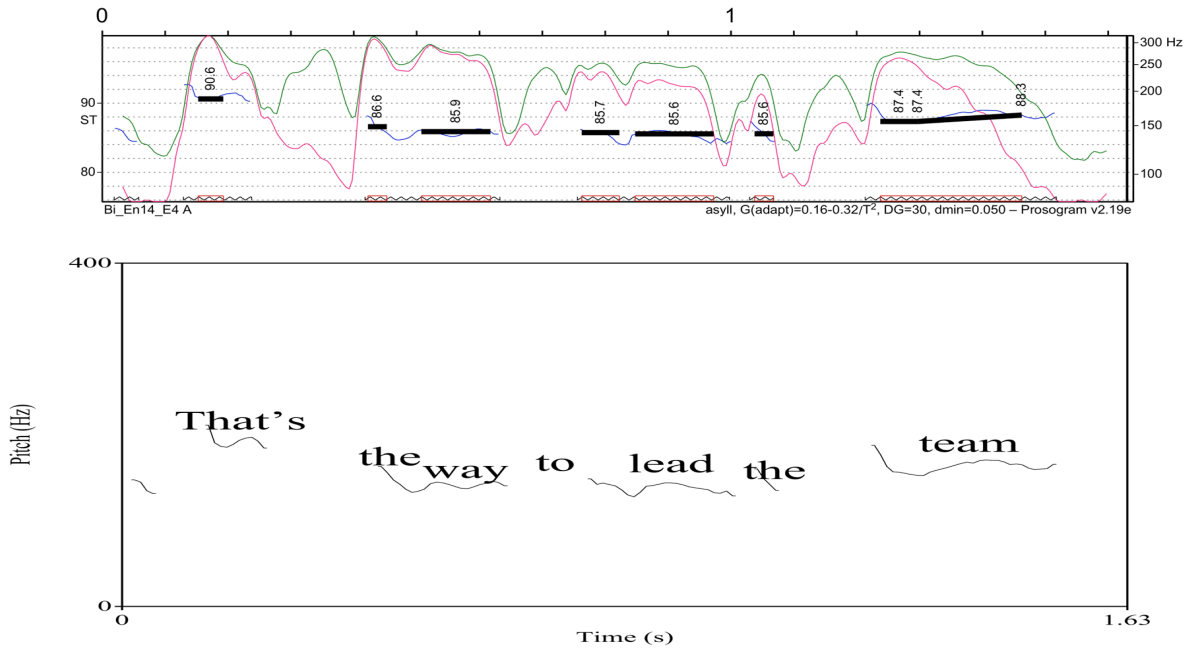


Figure 1. Prosogram graph indicating pitch values in semitones and, below, pitch contour created in Praat.

In figure 1 the prosogram graph is located at the top indicating pitch values in semitones that correspond to the length of a vowel which is illustrated as a thick black bar. The vertical bars on top of the x-axis indicate tenths of a second, while a full second is marked by a written digit. At the bottom the corresponding annotated pitch contour created in standard Praat modalities is shown.

The semitone values of all graphs were listed in various charts to extract specific sets of data separately for English and German native speakers.

Firstly, the initial and final semitone values of each utterance were listed to evaluate the average semitone values with which a speaker begins and ends her/his utterances. The average initial and final semitone values of ironic and non-ironic utterances were calculated separately for each speaker. The variations of pitch levels between the initial and final values were then calculated to determine overall melodic tendencies of pitch movement. The same procedure was repeated taking the highest and lowest semitone values of each utterance to determine the overall pitch range used in individual utterances. Concluding this step, the mean pitch ranges of individual speakers and the differences in mean range between ironic and non-ironic utterances of a speaker were calculated. Secondly, the mean intervallic movement from one syllable to another was calculated for each utterance. The mean intervallic movements in all utterances of a speaker were recorded separately for ironic and non-ironic utterances. Consequentially, the difference between these two means of a speaker was measured. Thirdly, each utterance was examined according to its intervallic values in the nucleus and its immediate surroundings. The number of occurrences of upward and downward intervallic movements in the nuclei was recorded as a mean value comprising all utterances of a speaker and again separately for ironic and non-ironic utterances. Following this process, the mean values for upward and downward movements were extracted for the ironic and non-ironic utterances of a speaker. The differences between the mean values of ironic and non-ironic utterances were then evaluated for each speaker.

Finally, the mean values of each subject were subsumed to calculate the following measures: the divergence in mean size of pitch movements between ironic and non-ironic utterances of a speaker; the number of upward and downward movements in the nuclei per speaker; the discrepancy between the mean sizes of upward and downward movements in the nuclei of ironic and non-ironic utterances of a speaker; the differences in value between a speaker's greatest upward and downward movements in ironic and non-ironic utterances; the divergence between a speaker's pitch range used in non-ironic and ironic utterances; the number of occurrences of final rises in a speaker's utterances; the difference between the mean F0 values of the ironic and non-ironic utterances of a speaker. All of these measures were declared in percentages for each individual speaker. Overall percentages including the values of all speakers were established for each of the four groups of subjects. The results were applied as quantitative measures of selection for the subsequent qualitative analysis, the proceedings and framework of which will be discussed in chapter 4.3.

3.2 Contextualization

The study's literal content comprised twelve different situations consisting of six sentences with two different, either ironic or non-ironic, types of contextualization each. The indicated speaker role, location, and supposed receiver of the respective utterance did not vary between the ironic and non-ironic contexts of an utterance. These frameworks were established in order to be able to directly compare the prosodic contours of a particularly phrased sentence expressed in two different ways, either including or excluding ironic marking, uttered by the same speaker. The wordings of English and German both created by the author were supposed to constitute identically translatable equivalents of each other, pursuing the objective of also being able to draw cross-linguistic comparisons. In order to establish a contextual framework that would allow the subjects to put themselves effortlessly in the position of the role they were to impersonate, it was attempted to create everyday type of situations that anyone might be able to identify with. The reactions of the participants confirmed the success of this attempt. When asked if they had ever used such phrasings, as indicated in the given sentences, or if they would naturally use the indicated phrase or a similar phrase, only ten percent reacted negatively.

The chosen speaker roles belonged to relatively common social categories or roles: father or mother in situation 1; spectator of a basketball game in situation 2; high school teacher in situation 3; car commuter in situation 4; vegetarian in situation 5; and, finally, waiter or waitress in situation 6. The same criterion was applied to the design of the supposed receivers of the messages, which included the following characters: the speaker's son in situation 1; a friend accompanying the speaker to a basketball game in situation 2; a particular student in the speaker's class that she/he teaches at a high school in situation 3; the co-driver of the speaker acting as a car-commuter in situation 4; a friend accompanying the speaker, acting as a vegetarian, to a restaurant in situation 5; and, finally, a co-worker of the speaker impersonating a waiter in situation 6. The indication of specific addresser and addressee roles were designed to confine the subjects' expression of affect to occur within well-defined solidarity parameters. Regarding the contextualization in the texts preceding the actual utterances that were to be performed, the use of adverbs in descriptions concerning the speaker's mood and background knowledge was avoided. This strategy was applied in order not to create any emotional bias in the subjects' but to allow them to produce unprescribed, more authentic expressions of affect. After all, the utterances should ideally imitate incidences

of spontaneous speech. The specification of the receivers of the speaker's utterances and the overall contextualization were intended to create points of reference the subjects could easily relate to in order to successfully imitate spontaneous speech acts. However, the degree of authenticity of these speech acts lacking an actual addressee is subject to debate, since monologic speech acts have been termed impractical for a meaningful investigation of prosody (Feller 2008: 182).

The utterances were composed to include either one single cue hinting at the ironic or one single cue hinting at the literal communicative intention of the speaker in regard to the contextualization of the respective situation. The words of the utterances which contained these cues constituted different parts of speech. Each utterance included a distinct part of speech as the cue word, except for two utterances both containing an adverb. The other parts of speech included an adjective, a demonstrative pronoun, a verb and a possessive pronoun. Concerning the expression of stance and affect, it is important to note that all of the sentences featured an exclamation mark. Moreover, some utterances also included emotional or appraising affect markers (Precht 2003: 242) such as the verb "love" or the adjective "clever", respectively, both also functioning as cue words in the respective utterances. Three sentences contained epistemic stance markers (Precht 2003: 242), including the mental verb "guess" and certainty measures as "surely" and "certainly". The latter two also functioned as the intended cue words of their respective utterances, leaving only one sentence without a distinct stance or affect marker. However, this unmarked sentence stood out by having the location of its nucleus, in contrast to the other sentences, at the initial syllable which comprised the demonstrative pronoun "that".

Affective and epistemic markers were intentionally chosen to produce results which would allow for a comparison of different degrees of affirmation in the analyzed samples. Such a contrastive method has been suggested as a requirement for a meaningful investigation of ironic pitch contours (Braun et al. 2018: 113). The sentence versions that were framed in an ironic context were designed to elicit different types of irony, which will be discussed in chapter 3.3. The non-ironic contexts of these sentences were also drafted to express a variety of speech acts as seen from a pragmatic perspective. The only types of pragmatic classification applied in the following chapter correspond to the so-called "illocutionary acts" (Bach 2003) through which a speaker expresses her/his attitude and stance. The other two levels constitute "locutionary acts" (Bach 2003), describing the acoustic realization of

language, and “perlocutionary acts”, (Bach 2003) characterizing the particular effect or action that a speaker wants to elicit from the listener. The success of the realization of the illocutionary level of speech acts, similar to the effect of irony, relies on the listener’s recognition of the speaker’s expressed attitude. Specifying illocutionary speech acts, a distinction is drawn between constatives, directives, commissives and acknowledgements (Bach 2003). An authoritarian stance was included in the parent/son and high school teacher/student situations. The former typically constitutes an informal, intimate and the latter a formal, reserved type of relationship. A formal but non-authoritarian stance framed the two situations which consisted of interlocutors who are co-workers. One was placed in a rather formal setting at the workplace, potentially in the presence of customers and other co-workers or even a superior, and the other in a less formal setting, with only the two interlocutors present in a car on their way to the workplace. An informal stance outlined the remaining two situations in which two befriended interlocutors constituted addresser and addressee. In one of them they were located at a restaurant in a rather intimate environment, sitting at a table together. In the other situation they were situated at a sports venue in a less intimate environement, sitting among other spectators. In regard to syntax, all but one sentence were declaratives, the exception being an exclamative. Only one sentence proposed a relatively longer speech pause by including a personal address in the form of a given name, namely, “Max” at the very beginning.

The English utterances consisted of an average of 7.6’ syllables per sentence, with a minimum of four syllables and a maximum of twelve syllables. The German utterances contained an average of 9.6’ syllables per sentence, the minimum value representing eight and the maximum thirteen syllables. However, the prosogram failed to calculate the pitch values of vowels in numerous instances, so that the pitch values per recorded utterance differed significantly. Some individual utterances had to be excluded from some of the quantitative measurements since their initial, final or nucleus pitch values were missing. Such failings occur due to low intensity of vowels, devoicing of vowels or obviously divergent F0 values provided by Praat (Iversen et al. 2006: 3039). The combinations of these graphs, a regular Praat and a prosogram chart, were archived together with the recordings and the filled-out forms under the individual speakers and the individual utterances.

3.3 Ironic and non-ironic contexts

In this chapter the introductive contexts followed by the observed utterances, as presented to the subjects, will be discussed in detail. The non-ironic contexts will be referred to with single digits and the letter *a*, while the ironic contexts will be referred to with the corresponding digits and the letter *b*. The German versions are listed in the appendix. The categories describing the various speech acts that are used in this chapter are applied according to Bach (2003).

3.3.1 Situation 1a and 1b

The location of these situations was indicated as "apartment/home of the speaker".

> Your son has struggled with his grades in Maths for quite a while. Coming home from school he informs you that he has just joined a voluntary Maths study group in order to improve. You react to this message with the following statement: *What a clever decision!*
> (1a)

This parent-child situation could either constitutes a speech act of acknowledgement, the parent congratulating the child on an achievement, or a constative speech act, merely affirming the child in having made that decision. The decisive factor which speech act is committed constitute the speaker's attitude and stance towards the importance of this matter resulting in different ways of using intensity and contrasts of prosodic effects.

> Your son has repeatedly received Fs on Maths tests. His teacher offered him to take an oral exam to make up for these results in order to still pass this school year. You have received a call from this teacher telling you that your child did not show up to the arranged exam. As you come home from work, you see your child and immediately address this incident by making the following statement towards him: *What a clever decision!*
> (1b)

The cue word hinting at the contextual incongruity producing the effect of irony is *clever*. The decision to skip an arranged meeting that could have prevented or reduced negative consequences would hardly correspond to intelligent decision-making. The type of irony used here might fall into the category of dry and/or genuine sarcasm (see 2.1.3), since the authoritarian distance between parent and child may allow for the belittling of the addressee without any social risk involved for the addresser. Yet if the speaker does not intend to belittle

or accuse the listener of any wrongdoing but instead teases him jokingly, this utterance will constitute a form or "ironic compliment" (Mauchand et al. 2019: 2).

3.3.2 Situation 2a and 2b

The location of these situations was indicated as "stands at a sports venue".

> You are an avid supporter of the local professional basketball team. Attending a league game with a friend, you witness your home team losing. After the captain of the team has been substituted, his performance not having met expectations, the team manages to recover and eventually win the game. Commenting on the substitute player, who has also scored the most points, and pointing towards him, you make the following remark directed towards a friend sitting next to you: *That's the way to lead the team*!
> (2a)

This situation involving two befriended spectators of a sports event classifying an athlete's performance represents a constative speech act. Contrary to situation 1a, this speech act cannot be classified as an act of acknowledgement in terms of an illocutionary speech act, since it is not the addressee of the utterance but the athlete, who does not participate in this communicative situation, whose performance is being praised.

> You are an avid supporter of the local professional basketball team. Attending a league game with a friend you witness your home team losing. After the captain of the team has failed to score in six consecutive attempts, you make the following remark about this player to your friend: *That's the way to lead the team!*
> (2b)

In this utterance "that" is the cue word. This demonstrative pronoun points to the misperforming of the captain which might not be an ideal example of leadership. The kind of irony produced here may belong to dry and genuine sarcasm, however, the belittling in this instance was not directed at the addressee of the utterance but a third party not involved in the conversation, the athlete that the interlocutors were both observing.

3.3.3 Situation 3a and 3b

The location of these situations was indicated as "in the speaker's classroom".

> There is a very shy student in your class called "Max". He has not engaged in classroom interaction during the entire semester. Due to conversing with his parents, you know that he is interested in geography and knows the name of every European capital. In order to involve him in classroom interaction without

> exposing him to potential ridicule, you ask the entire class if they know the name of the capital of Albania. Since no one has raised his hand, you make the following statement: *Max you can surely answer this question!*
> (3a)

This teacher-student situation could be phrased as a directive, requesting the student to participate in the class discussion, or a commissive speech act, offering the student to take part in the activity. The difference in attitude and stance of the speaker towards the addressee might affect the intensity and styles of contrasting effects of prosodic resources a speaker employs, deciding which of these two speech acts is realized.

> A student called "Max" is not paying attention to your opening monologue of one of your regular classes being held in a school classroom. Instead, he is talking to another student sitting next to him. Addressing this student, you make the following statement:
> *Max, you can surely answer this question!*
> (3b)

The cue word for irony used here is "surely". It seems very unlikely that the student, who is not paying attention to what the teacher is saying, will be able to answer the teacher's question. The authoritarian distance between teacher and student might allow the type of irony being used here to be specified as dry and/or genuine sarcasm, since no significant social risk is involved in the belittling of the addressee.

3.3.4 Situation 4a and 4b

The location of these situations was indicated as "on a freeway".

> It is Monday morning and you are stuck in traffic on your way to work. Luckily, you have started your trip to work earlier than usually, so that you will not be late despite the traffic situation. Your colleague is sitting in the passenger's seat commenting on the bad traffic situation in your city, to which you reply with the following statement: *We are certainly not going to be late again!*
> (4a)

This situation involving two co-workers commuting in a car could cause the speaker to commit three different speech acts: a constative, informing the listener about her/his prediction; a directive, suggesting or urging the listener to refrain from interpreting the outcome of the current state of traffic negatively; or even a commissive, guaranteeing the listener that no negative consequences are to be expected from the current situation.

> It is Monday morning and you are stuck in traffic on your way to work. You realize that you are going to be late for an important meeting. Your colleague is sitting in the passenger's seat commenting on the bad traffic situation in your city, to which you reply with the following statement: *We are certainly not going to be late again!*
> (4b)

The cue word hinting at irony constitutes "not", since it is indicated that the speaker recognizes that they are going to be late. Nevertheless, it is not known if the listener has also recognized this circumstance. Therefore, the speaker is to intentionally employ evident prosodic markers of irony. Due to the fact that the speaker expresses solidarity towards the addressee by using the pronoun "we" referring to herself/himself and the addressee, no belittling and thus no sarcasm are generated in this utterance.

3.3.5 Situation 5a and 5b

The location of these situations was indicated as "at a restaurant".

> You have been invited to dinner by a close friend who offers you to share two main dishes in order to enhance the culinary experience. Her/his suggestions for the choice of these dishes are spaghetti and pork chops, the latter being your most favorite dish since childhood. After having been asked what you think about these suggestions, you reply with the following statement: *I love pork chops!*
> (5a)

In this communicative context involving two befriended restaurant visitors the speaker most likely commits a constative speech act, confirming the listener's suggestion to suit the speaker's preferences.

> You have been invited to dinner by a close friend who offers you to share two main dishes in order to enhance the culinary experience. Being one of your closest friends, he/she should be aware that you have been on a vegetarian diet for many years. However, her/his suggestions for the choice of these dishes are spaghetti and pork chops. After having been asked what you think about these suggestions, you reply with the following statement: *I love pork chops!*
> (5b)

The cue word in this ironic context is "love". Being a vegetarian, the speaker most likely does not consider this dish one to which she/he would attribute such a marker of positive affect. Depending on the extent to which the speaker might perceive the erring of the addressee

about the speaker's lifestyle as disrespectful, the utterance may be phrased as a form of sarcasm or teasing.

3.3.6 Situation 6a and 6b

The location of these situations was indicated as "at a restaurant at which the speaker works".

> In your job it is your task to collect all the used glasses from empty tables and return them to the kitchen. You are returning to the kitchen with a tray full of glasses, using only one hand to support the tray, while you are typing on your phone, using the other hand. A co-worker comes out of the kitchen, carrying four bowls of soup. You bump into him since you do not see him, your eyes still fixed on your phone screen. Consequentially, he loses grip of the bowls, which fall onto the floor and shatter. The co-worker looks at the mess despairingly, since it is the second time this has happened to him that day in his first week working there. The owner of the restaurant rushes towards your co-worker, but before the former can say anything you make the following statement: *I guess it is my fault!*
> (6a)

In this situation involving two co-workers, and possibly their superior, a variety of speech acts could be realized. The speaker might be announcing herself/himself as the one to be blamed for this predicament, committing a constative speech act. A directive speech act could also be realized by the speaker permitting the addressee to excuse herself/himself by shifting the blame on the speaker. Through this statement the speaker could also guarantee the addressee a certain form of commitment, volunteering to take responsibility for the accident. The speaker could also commit a speech act of acknowledgement by apologizing to the co-worker or, if the co-workers' superior is already in hearing reach, directly to the superior or to both.

> In your job it is your task to collect all the used glasses from empty tables and return them to the kitchen. You are returning to the kitchen with a tray full of glasses, using both hands to support the tray and moving slowly and carefully since the tray is quite heavy. All of a sudden, a co-worker typing on his/her phone rushes out of the kitchen and bumps into you, so that you lose grip of the entire tray. Consequentially, all glasses fall onto the floor and shatter. Looking up from the mess, this co-worker glares at you angrily but before he/she says anything, you make the following statement: *I guess it is my fault!*
> (6b)

"My" constitutes the cue word in this utterance. The accident has occurred due to the inattention of the addressee, while the speaker has conducted her/his work task especially mindfully. Hence, the speaker is very unlikely to consider her/his own behavior as the cause of this accident. The utterance comprises a speech act of defense against an expected

accusation and might therefore be considered as expressing sarcasm, belittling the addressee in her/his supposed intention of accusing the speaker of being responsible for this accident.

3.4 The selection of successful performers

Even in relatively constricted types of contextualization, as illustrated in 3.3.1. to 3.3.6., in which the contextual incongruity is clearly indicated in the way the description of the context and the wording of the utterance contradict each other, multiple different speech acts and types of irony can be realized. Which acts are realized may also depend on the individual personality traits and the mental state of a speaker, influencing how emotionally involved she/he might become about the topic or the situation of a given conversation (Bach 2003). Furthermore, the degree to which one considers applying irony appropriate in a particular context or towards a particular addressee is also determined by the individual stylistic language profile of a speaker (Nooteboom 2010: 670).

Due to this multitude of aspects framing the production of ironic utterances, this thesis was designed to apply various measures of quantitative selection to focus on the utterances of speakers who were most successful at producing ironic utterances. Ideally, this selection process should have been conducted by applying measures of selection gained from a survey in which another group of subjects had assessed the recordings according to the conveyed ironic effect. Unfortunately, such an endeavor would have been beyond the scope of this thesis. Hence, criteria based on measuring the speakers' success in using prosodic means to express irony and interpreting the given texts were used to select the most suitable samples for the concluding qualitative analysis. The success in expressing irony via intonation was measured only according to criteria that described melodic properties of pitch contours. Furthermore, only prosodic markers of irony that were agreed on by a multitude of findings were used as the selective criteria in this analysis. Other indicators of irony which concern only the rhythmic properties of prosody, as increased syllable duration and decelerated speech rate (Champagne-Lavau et al. 2018: 166), were not included as selective factors. This exclusion was conducted due to this analysis' focus on the melodic properties displayed as musical intervals, not the rhythmical properties of prosody. Comparing ironic with non-ironic utterances, the following features of prosodic signals for irony are the most agreed upon: a decrease in mean F0 (Champagne-Lavau et al. 2018: 166); a tendency towards melodic monotony realized by smaller intervallic movements (Braun et al. 2018: 115); and, finally, a reduction of pitch range used over a whole contour (Braun et al. 2018: 116). These three

features were established as the selection criteria for the qualitative analysis. When a speaker's overall performance evaluated in mean values did not correspond to these measures, that speaker's utterances were not taken into account for the qualitative analysis. Two further measures were set up in order to assess whether a subject correctly correlated the contextualizing pieces of text with the utterances to detect the ironic cues. One of these criteria was to determine if the subjects expressed the correct sentence type. When it was clearly evident that a subject phrased an utterance as a question, by applying a significant final rise in the respective pitch contour, the utterance was marked as failing to produce the correct sentence type. The second criterion was established by evaluating the overall occurrence of the directions of intervallic movements in the nuclei for each group. The speaker's overall performance in this regard was then referenced to the overall mean value. However, in many utterances the prosogram failed to produce a semitone value for one or more vowels often at crucial locations as the initial, final or nucleic syllable. In such instances the mean values of all the critical features were still evaluated with all the given measurements. The most important measures for the qualitative analysis constituted the semitone values occurring in the nucleus as well as the pre- and post-nucleus vowels. A group of three subsequent pitch values represents the minimum sequence of a melodic unit in music, since a sequence of two pitch values merely constitutes a musical interval. Furthermore, a sequence of three units also corresponds to the minimum structure of an arpeggio, which constitutes the consecutive instead of simultaneous production of the individual notes of a basic harmonic unit – a chord.

4 Results and discussion

The quantitative analysis of the 240 recordings is presented in 4.1., comparing the overall findings according to the first three criteria that are described in 3.4. The other two criteria which rated the overall task fulfillment of speakers are discussed in 4.2. The first groups denote the subjects that were not informed about the study's purpose of investigating irony, whereas the second groups refer to the subjects that were informed about the study's aim. It was assumed that by being informed about the study's purpose the second group would produce irony through prosodic means more consciously and thus serve as a form of control group.

4.1 General tendencies

All data discussed in this chapter are taken from the graphs produced by the Praat plugin prosogram. The semitone values were evaluated with a Texas Instruments TI-30X IIS calculator and arranged in charts corresponding to the measurements discussed in the following subchapters. The criteria in this chapter will from now on also be referred to as the "general criteria", since they have already been established as prosodic markers of irony in prior research.

4.1.1 Mean fundamental frequency

The first feature that was evaluated constituted the mean F0 of the pitch contours, for which a decrease is known to be found in ironic utterances of various Germanic languages, including German and English (see 3.4.). This measure was confirmed by 70% in the first and second English groups. The main difference between the results of the first and second English group was the distribution of peak frequency values. In the first group 80% of the mean peak pitch values of non-ironic utterances were higher than the ones of ironic utterances, whereas in the second group this occurrence amounted to 70%. The difference in semitones between the mean F0 of the higher non-ironic utterances' contours and the mean F0 of the lower ironic utterances' contours added up to about 2.7 semitones in the first group and 2.46 semitones in the second group. The greatest differences in mean F0 between higher non-ironic utterances and lower ironic utterances amounted to 5.81 and 6.2 semitones, respectively, in individual speakers. For the speakers whose mean F0 of ironic utterances was higher than that of non-ironic utterances, the mean difference in semitones accounted for 1.86 semitones in the first but only 0.62 semitones in the second group. The latter value being close to the perceptual threshold of human interval detection (see 2.2.4.) may indicate that other errors regarding task fulfillment, as the phrasing of an utterance as a question, might have caused the results of higher mean F0s in ironic utterances in the second group. Additionally, the greatest individual differences of the speakers having higher mean F0 in ironic utterances amounted to 2.91 semitones in the first but only 1 semitone in the second group. In the first group almost all speakers that produced a higher mean F0 in non-ironic utterances also had higher mean pitch peaks and higher mean minimum pitches in their non-ironic utterances. Only one of the three speakers producing a higher mean F0 in ironic utterances delivered a higher mean pitch peak in non-ironic utterances. The second group included two speakers that produced a higher mean pitch peak and a higher minimum pitch in ironic utterances, respectively, although they

generated a higher mean F0 in non-ironic utterances. In this group there were also two speakers that provided a higher mean pitch peak and a higher minimum pitch in non-ironic utterances, respectively, although they produced a higher mean F0 in ironic utterances. The mean pitch peak of speakers that showed higher pitch peaks in ironic utterances amounted to a difference of only 1.45 semitones to the mean pitch peaks of their lower non-ironic utterances, whereas the mean pitch peak of speakers having higher pitch peaks in non-ironic utterances was 4.2 semitones higher than the mean peak of their lower ironic utterances. Furthermore, in 70% of each group the mean minimum pitch values of non-ironic utterances were higher than the ones of ironic utterances, by a mean of 1.89 in the first and 2.49 semitones in the second group. The ironic utterances which displayed a higher minimum pitch level possessed a mean difference of 2.07 in the first and 1.24 semitones in the second group. These minor differences in cases where the subjects' means did not correspond to the expected general criterion may again imply a disturbance of data by erroneous task fulfillment.

However, the findings in the German groups' utterances displayed some more significant contrasts than the English groups. The mean F0 of non-ironic contours was higher than that of ironic contours in 80% of the first group's subjects but in only 40% of the second group. The mean F0 of these higher non-ironic contours was 1.78 semitones larger in the first and 0.93 semitones larger in the second group than the ones of the lower ironic utterances. The greatest measures of this criterion found in individual speakers amounted to 4.12 semitones in the first and 1.66 semitones in the second group. The mean F0s of the higher ironic utterances displayed a greater difference to lower non-ironic utterances, namely, 2.33 semitones in the first group and 1.4 semitones in the second group. The maximum difference of an individual speaker with a higher mean in ironic than non-ironic utterances constituted 2.71 in the first and 2.3 in the second group. The first group included two speakers who delivered a higher mean pitch peak and two other speakers who supplied a higher minimum pitch in their ironic utterances, even though the greater mean F0 was featured in their non-ironic utterances. In the second group only two of the four subjects who exhibited a higher mean F0 in their non-ironic utterances also revealed a higher mean pitch peak in their non-ironic utterances. Furthermore, one of the speakers showing a higher mean pitch peak and a higher mean F0 in non-ironic utterances did not produce a higher minimum pitch in non-ironic utterances than in ironic utterances, in contrast to the other three showing higher mean F0 in non-ironic utterances. There was also one speaker among the subjects producing a higher mean F0 in ironic

utterances that still delivered a higher mean pitch peak in the non-ironic utterances, and another who delivered a higher mean minimum pitch in non-ironic utterances. The speakers that exhibited higher pitch peaks in ironic utterances did that by a difference of 2.8 semitones in the first group and 1.75 semitones in the second group. The difference between the mean pitch peaks of ironic and non-ironic utterances of the speakers delivering higher peaks in non-ironic utterances amounted to 3.18 semitones in the first and 1.52 semitones in the second group. The mean minimum pitch differential between ironic and non-ironic utterances of speakers with higher mean minimum pitch in ironic utterances was 1.65 semitones in the second but only 0.47 semitones in the first group. This measure constituted 2.06 in the first but only 0.66 in the second group in speakers with a higher mean minimum pitch in non-ironic utterances. These occurrences of values close to 0.5 semitones may repeatedly imply the emergence of some differentials as results of erroneous task fulfillment. Interestingly, while the German control group produced significantly smaller mean pitch peak differentials than the first group, the English control group produced significantly larger mean pitch peak and mean minimum pitch differentials than the first group. Regarding speakers with higher measures in non-ironic utterances, the overall differentials between non-ironic and ironic utterances mean F0, mean pitch peak and mean minimum pitch were significantly higher in English than in German. However, concerning speakers with higher values in ironic-utterances, the German subjects produced significantly higher differentials in mean F0 and mean pitch peak but not mean minimum pitch. Furthermore, since the German control group included a majority of speakers with higher mean F0, mean pitch peak and mean minimum pitch in ironic utterances – in contrast to the English control group – one might come to the conclusion that the ways in which prosodic contours hinting at irony are realized in these languages are substantially different. These strategies seem to be of a quite variable nature in both languages, depending on the contextual embedding in which a particular speech act is situated.

4.1.2 Monotony in irony

Smaller mean interval movement within an utterance has been established as a key feature of prosodic contours that express irony (see 3.4.). There was a critical variation of this measure in the English groups. While 60% of the first English group produced smaller mean interval movements in ironic utterances, 90% of the second group corresponded to this measure. Furthermore, regarding the former, a negative mean differential of 0.8 semitones to the mean intervallic movement in non-ironic utterances was found, whereas the latter produced a

difference of about one semitone. The mean intervallic movements of the subjects producing greater intervallic movement in ironic utterances constituted a difference of 0.58 in the first and 0.63 semitones in the second group. These small measures suggest that the greater mean intervallic movements of ironic utterances found in some speakers might originate from erroneous task fulfillment, e.g., the phrasing of the indicated utterances as questions instead of declaratives. Furthermore, the maximum differential between a speaker's mean intervallic movements was 2.05 in the first and 2.25 semitones in the second group regarding speakers with a higher intervallic movement in non-ironic utterances. However, it only amounted to 1.03 and 0.63 semitones, respectively, regarding speakers with a higher intervallic movement in ironic utterances. There were no such significant discrepancies between the first and second German group. 70% of both German groups produced smaller mean intervallic movement in ironic utterances. The mean differential to the mean intervallic movement in non-ironic utterances of these subjects amounted to negatives of 0.5 in the first and 0.83 semitones in the second group. The subjects who displayed a higher mean intervallic movement in ironic utterances produced a mean difference of 0.37 in the first and 0.51 semitones in the second group. The maximum differentials also did not show any critical disparity between the two German groups and were also significantly lower than in the English groups. The maximum differential of a speaker producing a higher mean intervallic movement in non-ironic utterances amounted to 1.21 semitones, while the maximum differential of a speaker producing a higher mean intervallic movement in ironic utterances equaled 0.98 semitones.

4.1.3 Reduced pitch range

The reduction in pitch range has been observed as a typical feature of prosodic contours expressing irony (see 3.4.), a measurement which was confirmed in this study's analysis. In the first English group about 77% employed a higher mean pitch range in their non-ironic utterances. The mean difference in range between the non-ironic utterances with greater range and the lower ironic utterances with smaller range amounted to about 2.5 semitones, with 3.8 semitones as the maximum individual value. The mean differential between the ironic utterances with a higher pitch range and the corresponding non-ironic sentences added up to only 0.84 semitones, with a maximum individual value of 1.15 semitones. The smallest pitch movement in non-ironic utterances was larger than the smallest interval value in ironic utterances in 55.5%, with a mean differential of 1.76 semitones and a maximum of 3.05. This measure included the two speakers who had a greater range in ironic utterances but excluded four speakers with a greater range in non-ironic utterances. Likewise, the mean smallest

interval which was higher in ironic utterances amounted to a difference of 1.36 semitones with a maximum of 3.06 semitones. All but two speakers with a greater range in non-ironic utterances also had significantly higher maximum pitch movements in their non-ironic utterances, with a minimum mean difference of 1.52 and a maximum of 4.2 semitones to the maximum intervals in ironic utterances. This measure also included the two speakers with a greater range in ironic utterances. For those speakers that produced a higher mean maximum pitch movement in ironic utterances the difference amounted to a mean of 0.73 semitones and a maximum of 0.96 semitones. In the second group even 90% utilized a greater range in non-ironic utterances with a mean differential of 3.34 semitones and an individual maximum differential of 11.7 semitones when compared to ironic utterances. The only speaker providing a greater range in ironic utterances produced a mean differential of 1.2 semitones. All subjects of this group generated a higher mean maximum pitch value in non-ironic utterances, ranging from 0.5 to 9.92 semitones greater than the mean maximum pitches in ironic utterances with a mean differential of 4.52 semitones. Only two speakers with a greater range in non-ironic utterances produced mean minimum pitches that were greater in ironic than in non-ironic utterances, with a mean of about 1 semitone. The others, including the speaker with a greater range in ironic utterances, produced higher mean minimum pitches in non-ironic utterances, with a mean differential of 1.83 semitones.

The results in the German groups were quite similar. However, while in the first group 80% used a greater range in the non-ironic utterances, only 70% did so in the second group. The mean differential to the ranges in ironic utterances amounted to 2.1 and 4.66 semitones, the mean differential of higher ironic to lower non-ironic utterances' ranges to 1.26 and 1.32 semitones, respectively. Regarding the maximum intervals of individual speakers, 80% of each group had higher values in their non-ironic utterances, including a mean differential of 2.92 semitones in the first and 2.97 semitones in the second group and individual maximums of 6.56 and 6.25, respectively. The few speakers producing a higher maximum pitch movement in ironic utterances displayed a mean differential, to the maximum intervals in ironic utterances, of 1.8 in the first and 1.65 semitones in the second group. However, while 60% of the first group produced higher mean minimum intervals in non-ironic utterances, with a mean differential of 1.5, only 40% did so in the second group, with a mean differential of 2.74 semitones. When speakers produced a higher mean minimum pitch movement in ironic utterances, it amounted to mean differentials of 1.21 in the first and 0.69 semitones in the second group. These similar measurements of the pitch range as well as the maximum and

minimum intervals between non-ironic and ironic utterances in the German groups might be attributed to a smaller inventory of possible contour variations when compared to English, or the lower variety in speakers of different dialects than in the English groups.
The initial pitch values were also compared in order to investigate if the reduced pitch range of ironic utterances may already be signaled by speakers – hence also detected by the listener if he/she is familiar with the typical tone of voice and intonational behaviors of the speaker – at the very beginning of an utterance. The English groups produced relatively diverging results, with 42.9% of the first group's and 75% of the second group's mean beginning pitches being higher in non-ironic utterances. The the mean differential between higher initial pitches of non-ironic utterances and lower initial pitches of ironic utterances constituted 2.41 semitones in the first and 6.13 semitones in the second group. However, the mean differential between higher ironic and lower non-ironic initial pitches amounted to only 0.78 in the first and 0.65 semitones in the second group. In regard to the global contour, the distance between the initial and final pitch values was greater in non-ironic utterances than in ironic utterances in 57.1% in the first and 62.5% in the second group, with mean differentials of 4.16 and 2.97 semitones, respectively. When the distance between the initial and final pitch values was greater in ironic utterances, the mean differentials added up to 1.56 and 2.61 semitones in the first and second group, respectively.

The German groups also delivered varying outcomes, with 55.5% of the first and 75% of the second group using higher mean initial pitches in non-ironic utterances. The mean differentials to the lower initial pitches of ironic utterances equaled 1.43 and 2.95 semitones, respectively. The mean differentials between higher initial pitches in ironic utterances and lower initial pitches in non-ironic utterances equaled 3.9 in the first and 1.9 semitones in the second group. The distance between the initial and final pitch values was greater in non-ironic utterances in 75% of the second but only 33.3% of the first group, with mean differentials of 4.04 and 3.42 semitones, respectively. When the distance between initial and final pitch was greater in ironic utterances, the mean differentials amounted to 4.74 in the first and 1.56 semitones in the second group. The tendency exhibited in the second groups to produce significantly higher mean initial pitch values in non-ironic utterances than ironic utterances may allow listeners to recognize a pitch contour that marks irony already at the start of an utterance, when they are familiar with the tone of voice of the speaker.

4.2 Erroneous task fulfillment

The criteria regarding the composition of pitch contours that signal irony which are agreed upon by the majority of empirical evidence in this field of study have been confirmed in chapter 4.1. with the exception of reduced mean F0 in German. These measurements, including this exception as a new criterion for the German subjects, were applied first in the selection process for the subsequent qualitative analysis of single utterances' pitch contours. The speakers whose utterances did not display the indicated features were excluded as subjects of further qualitative observations. In the second stage of this selective procedure the pitch contours were analyzed according to criteria that determined if the speakers had correctly understood the instructions and identified the cue words and punctuation marks as indicated.

4.2.1 Final rise disturbances

The final pitch values and their immediate surroundings were observed to demonstrate if the speakers phrased their sentences according to the almost universal falling pattern of English and German declarative sentences without markers of continuation, protest, demand or uncertainty (see 2.1.2.). Occurrences of production of these markers in an utterance were considered an erroneous task fulfillment by the subject. This judgement was based on the explicit contextual constraints that the pieces of text provided before every utterance. These texts' function was to suggest a specific speaker-listener relationship by indicating speaker and listener roles, and to describe prior actions of the listeners to which the speakers were to react to. Furthermore, a specific degree of formality of setting was determined by indicating the location. Finally, a certain attitude and stance that were to be embodied by the speaker were implied by providing explicit information about the speaker's and/or listener's discourse and background knowledge. Out of the 120 utterances that were uttered by the English speaking subjects 13, equaling 10.8%, could not be analyzed regarding this criterion. In these utterances the software did not produce any final pitch values, resulting in 55 utterances left in the first and 52 in the second group. Remarkably, about 47.3% of the first and 40.4% of the second group's utterances ended in a final rise. A final rise was considered as such even when the pitch movement equaled a value below the perceptual threshold of 0.5 semitones (see 2.2.4.), since an upward pitch movement at the end of an intonation unit still represents a purposeful physical action (see 2.2.2.). Three situations were never produced with a final rise, namely, 3b in the first and 1b and 5a in the second group. Two situations were produced

exclusively with final pitch rises, namely, 3a and 3b in the first group. The speaker with the most final rises in the first group produced four out of six utterances with a final rise. The lowest value of two final rises in six utterances was found in two speakers. In the second group the lowest number of final rises produced was one, found in three speakers, and the highest was four, found in one speaker.

The prosogram failed to produce final pitch values only for seven utterances of the German groups, resulting in 58 usable utterances in the first and 55 in the second group. About 32.8% of the utterances in the first and 29.1% of the utterances in the second group were produced with a final rise. In the first group there were two speakers who did not produce a single final rise but one speaker who produced a final rise in every single one of the utterances. In the second group the maximum number of final rises per speaker was three, produced by two speakers, but there were three speakers who did not produce a single final rise. However, all of these three speakers also produced utterances for which the prosogram was unsuccessful in extracting a pitch value for the last syllable, ranging from a minimum of one to a maximum of three utterances per speaker. There was only one utterance in the first group, namely, the one included in situation 1a, but three utterances in the second group, the ones of situations 1a, 2b and 3a, which did not include a final rise. The lower levels in final rise production in the second groups in both languages indicate the effect that the level of background knowledge and discourse awareness may have had on the speakers' confidence in producing intentional speech acts. The emergence of the relatively high numbers of final rises might have been caused by a variety of factors. The experiment was designed to collect data of spoken language that were supposed to resemble spontaneous speech acts. Nevertheless, however well framed and contextualized the speech acts might have been, their actual realizations could not avoid lacking one of the most relevant features of irony's purpose and function in spoken conversation. This key feature constitutes the elicitation of a response in the listener, guiding her/him towards a particular interpretation of the speaker's utterance (see 2.3.2.). The listener present at the experimental situations of this study was always the author of the study, thus also of the utterances and the accompanying pieces of texts, himself. Hence, the success of an authentic production of a prosodic contour depended on the author's explanations and the imagination of the subjects. Cases in which the comprehension of the contextualizing pieces of text constituted a problem arose in all groups. Such incidents were dealt with by the author responding in different ways to the inquiries of subjects. In the first groups, which were not to be informed about the specific empiric goals of the study, the replies were

confined to elaborations on the instructions and information about the speaker and listener roles. Replies to inquiries about the stance and attitude were reduced to informing the subjects that they were to execute and personify these traits according to their own interpretation. The second groups were explicitly informed about the ironic and non-ironic framework of the respective utterances. However, the possibility that subjects could not authentically identify with a given speaker role might have posed a problem as much as their individual evaluations of the context as being appropriate for the use of irony. There were also several instances in which the subjects asked the author whether he was content with their performance immediately after producing an utterance. These requests for affirmation may reveal that the subjects might have found themselves in a certain state of feeling in which they felt as though they were being pressured to deliver a performance, desiring to meet the expectations of the author. Such mood states might have led the subjects to use uncertainty markers as final rises in the final syllables, signaling their state of uncertainty regarding whether they had successfully performed the task at hand. Furthermore, since only three subjects, belonging to the German speaking groups, had either professional speaking or singing experience, most may have not been used to being recorded speaking. Additionally, the enacting of specific speaker roles while being expected to produce specific or even predetermined, instruction-based prosodic features might have intensified the potential inconvenience for unexperienced public speakers. Such circumstances may have further manipulated the prosodic contour of an utterance. For instance, it has been shown that mental tension might also manifest itself as forms of bodily tension including tension in the vocal folds affecting an individual's tone of voice (Hellbernd et al. 2018: 610).

4.2.2 Stress and nucleus

Regarding stress placement and the shape of the nucleus in prosodically marked ironic utterances, a great variety of opposing features has been discovered (see 2.3.3.). Due to this predicament, the nucleus was analyzed according to its direction of movement. The dominant feature that was determined by overall quantitative measures comprised either an upward or downward pitch movement passing into the nucleus. In regard to overall mean measures, the mean occurrence of upward or downward interval steps into the nuclei produced by a speaker was used to further select the most competent speakers for the qualitative stage of this analysis. The intervallic value of the nucleus was evaluated by contrasting the pitch value of the syllable preceding the nucleic syllable with the pitch value of the nucleic syllable. One utterance, situation 2, was not included in this criterion, since the utterance was designed to

have its nucleus located in the initial syllable. In the first group of English speakers the occurrences of upward and downward movements varied considerably between utterances as well as speakers. One half of the utterances' nuclei was accented by an upward movement and the other half by a downward movement. There were two speakers using an upward movement in all of their utterances' nuclei but also one speaker who used exclusively downward movements. In two cases ironic utterances were articulated with an upward movement more often than their non-ironic counterparts, namely, in situation 1b and 1a with 60% and 20%, respectively, and situation 5b with 80% and 5a with 50%. In other two instances the number of upward movements in the nucleus was even between the corresponding non-ironic and ironic utterances. These situations were 3a and 3b with each 40% and 6a and 6b with each 20%. Situation 4 was the only one including an exclusive occurrence of upward movements in the non-ironic versions, opposing 80% of upward movements in the ironic versions. The intervallic movements were greater in the non-ironic nuclei in 60% of the utterances.

In the second group of English speaking subjects this ratio amounted to 90%. In this group approximately 73.5% of the nuclei were produced by an upward movement from the preceding pitch value. The prosogram failed to produce a pitch value for the nucleic syllable four times. Regarding the performance of the individual speakers, two produced all of their utterances with nuclei consisting of upward intervallic movements, while on the other end of the spectrum one speaker produced an upward movement in only one of the respective nuclei. Two situations displayed an even amount of upward movements in both, the non-ironic and ironic situations. These were situation 4 with each 100% and situation 6 with each 60%. The remaining three situations all showed a higher occurrence of upward movements in the ironic versions, with ratios of 60 to 40% in situation 1, 60 to 25% in situation 3 and 100 to 80% in situation 5. This contrast between the results of the first and second group may be attributed to a number of methodological issues. The speaker roles in situation 1, a parent talking to his son, and 3, a teacher talking to a student, might not have constituted simple tasks for the subjects. Producing an authentic speech act in these situations requires the impersonating a specific attitude towards a listener who is subject to certain expectations of obedience to the speaker. Subjects who had never been attributed with these types of powers of authority might have found impersonating such speaker roles difficult. The situations which displayed results of a 100% contained speaker and listener roles about which it is safe to assume that every

subject had experienced them before, as the commuting co-workers in situation 4 and the befriended restaurant guests in situation 5.
The German groups also produced significant inconsistencies in the use of intervallic movements in the nuclei. The overall measures were quite inconclusive with about 54% of the utterances consisting of upward intervallic movements, similarly to the English groups' approximate 62%. The latter value closely resembles the figure of the frequency of upward movements in the first German group with about 63%. In this group the prosogram did not succeed in producing pitch values for the nucleus in only one instance. The lowest number of upward movements of an individual speaker was one represented by only one subject. The maximum of this measure was five produced by a speaker whose utterances included only one omission by the prosogram but who otherwise exclusively produced upward moving nuclei. Two situations showed the same amount of nucleic upward movements in their non-ironic and ironic versions, namely, situation 3 with two thirds and situation 4 with half of their utterances. A predominance of upward movements in one of the versions, either non-ironic or ironic, was found in two situations each. In situations 1 and 2 the non-ironic utterances' nuclei included more upward intervallic movements than the ironic utterances, with ratios of 75 to 66.6% and 100 to 66.6%, respectively. The ironic utterances included more upward nucleic movements than their non-ironic counterparts in situations 5 and 6, with ratios of 80 to 40% and 100 to 16.6%, respectively. In 70% of the situations the non-ironic nucleic pitch movements were greater than in the ironic versions, the same measurement was also found in the second group. In the latter group of native German speakers the prosogram also failed to produce a pitch value for the nucleus in only one utterance. Additionally, there were four utterances in which the pitch value of the nucleus was level with its preceding pitch value, not rendering an intervallic movement. There were two speakers who did not use an upward movement in the nucleus in any of their utterances.

In contrast to all other groups, the maximum number of nucleic upward movements produced by an individual speaker was four, which could be attributed to three subjects. Four utterances included a higher number of upward intervallic movements in the non-ironic versions than in their ironic equivalents. These instances consisted of situation 1 with a ratio of 50 to 0%, situation 3 with a ratio of 75 to 16.6%, situation 4 with a ratio of 75 to 60% and situation 5 with a ratio 83.3 to 25%. The only sample including a higher measurement of upward intervallic movements in its ironic version constituted situation 6 with a ratio of 33.3 to 0%. These significant differences between the occurrences of upward intervallic movements in the

nuclei of non-ironic and ironic versions found in the first and second groups insinuate that the marking of utterances as holders of rhetoric content through intonational manners is subject to a rather complex set of influences. The form of a prosodic cue indicating the ironic meaning of an utterance seems to depend on the utterance's embedding in its pragmatic and social-hierarchical context. These contexts may influence the way in which the intensity of affect applied in a speaker's utterance. An individual prosodic contour combines various elements from distinct sources, as a culture specific standard inventory of pitch contours and the standard modes of expressing different emotions emerging from universal physiological states. The standard pitch contours of a language constitute the frames within these expression of emotions can be realized. Hence, the prosodic marking of an utterance as ironic may be carried out, on the one hand, by violating expectations regarding the ways in which affect – as in the form of attitude, authority or emotion – is expressed in an utterance. On the other hand, such instances of irony may violate expectations concerning which type of pitch contour is used to shape the way a sentence mode – as a declarative, question or imperative – is structured to express a specific speech act. Since the expression of affect is always situated within a prosodic contour, such expectations are simultaneously held and violated, respectively, regarding both levels – the level of expressing affect and the level of shaping the prosodic contour of a clause. Where the particular ironic cue of an individual utterance is placed and how it is shaped, violating formal expectations, thus depends on the speaker's awareness of the situational appropriateness of expressing affect in a particular context. The degree of this appropriateness in a particular context is shaped by the authoritative distance between speaker and listener and the dialogical network of interactive responses between the interlocutors. The subjectivity of these factors alternating from speaker to speaker and from situation to situation may constitute the main reason for the difficulty of predicting an ironic tone of voice.

4.3 Qualitative analysis

The quantitative findings presented in 4.1. were found to confirm the relatively well established, but in no case undisputed, criteria for defining the properties of prosodic contours marking irony in most cases. Hence, these general criteria were further applied in selecting the most "prototypical" utterances for a qualitative analysis. Due to the even distribution of the twelve utterances in groups of six to each speaker and the varying results between the individual utterances of speakers meeting the criteria of 4.1. and 4.2., some utterances had to be chosen for a qualitative analysis that either did not meet all these criteria or that were

performed by a speaker who originally did not pass one of the two stages of selection. One situation, namely, situation 2, was excluded from the qualitative analysis, because the nucleus of the respective utterances of this situation was located at the very first syllable where the majority of speakers did not produce any significant intrasyllabic pitch movements, contrary to the author's expectations.

4.3.1 English speaking groups

In the English groups the three criteria discussed in chapter 4.1. were fulfilled by four speakers in the first group and six in the second group. These numbers indicate that specific prosodic markers for irony may indeed be a part of language's standard inventory of pitch contours. This assumption is based on the fact that the subjects of the second group who had been informed about the research purpose regarding irony fulfilled these criteria in a higher number, probably applying prosodic methods to mark irony with explicit intention. This advantage in background knowledge over the first group might have been balanced out by a higher pressure to perform a more specific task than the first group, which can be shown in the still relatively high number of final rises in the second group's utterances.

4.3.1.1 The uninformed subjects

Out of these four subjects in the first group one produced a majority of final rises in the utterances as well as not a single upward moving nucleus, opposing the overall tendencies regarding task fulfillment. Another one of these speakers provided more downward than upward movements in the nuclei. The remaining two speakers displayed measurements of a 100% and 60% upward nucleic movements and 50% and 33.3% final rises, respectively. Removing the utterances which were formed with a nucleus moving downward or a final rise, no matching utterances were left to be directly compared. This result demonstrates the magnitude of the difficulty of defining applicable practical criteria for framing pitch contour inventories. However, for the purpose of showing how many varieties of prosodic features one utterance may contain between the performances of two different speakers, situation 6a performed by the two speakers mentioned lastly will be illustrated below. Due to the belonging to the first group of English speakers, the subjects will be called speaker 1a, a female subject aged 49, and speaker 1b, a female subject aged 31, from now on.

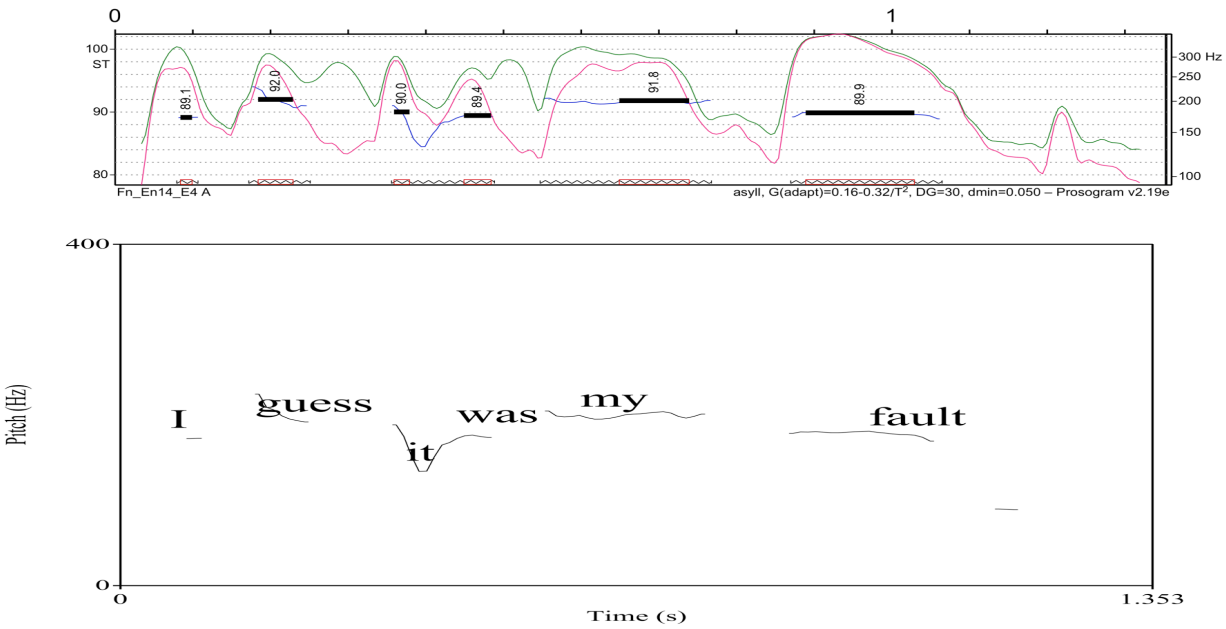


Figure 2. Situation 6a, speaker 1a, non-ironic.

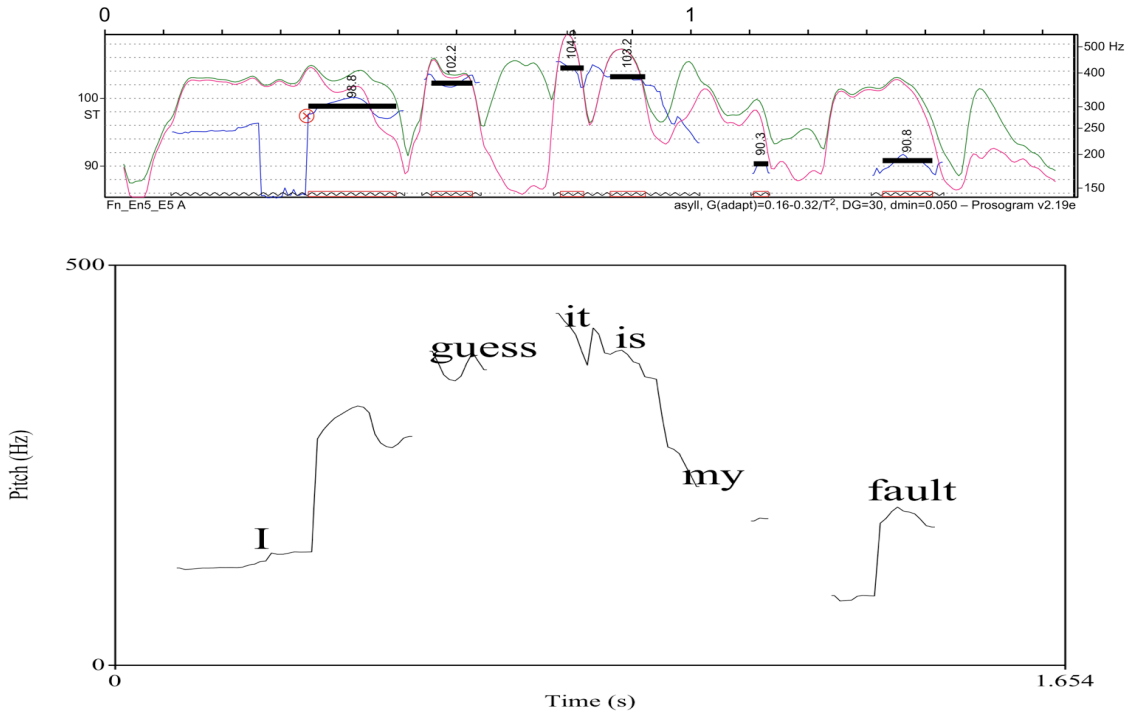


Figure 3. Situation 6a, speaker 1b, non-ironic.

The first striking difference in these two samples is the duration of the entire utterances, with speaker 1a only barely but speaker 1b clearly exceeding the one second mark. Disregarding speaker 1a's erroneous wording, it is still apparent how factors regarding the individual personal background and attributes may have influenced the subjects' interpretation of the situation and production of the prosodic contour. The most significant distinct feature can be

found in the location where speaker 1b produced her most prominent drop of pitch, equaling 12.9 semitones. In the equivalent position speaker 1a produced only a slight rise of 2.4 semitones. This difference may be attributed to different speech acts being committed by the subjects. Speaker 1b may have performed the utterance as a speech act of acknowledgement, already apologizing to the approaching superior, whereas speaker 1a may have merely affirmed the co-worker that he/she will take responsibility for the accident, committing a constative speech act. However, besides such expressions of stance, the extent to which an individual subject would rate the state of affairs, or the potential consequences thereof, as a social threat – or as affecting her/his own emotional state – may also affect the subjects' prosodic behaviors. Such estimations most likely depend on the personal background and individual character traits of a particular speaker.

4.3.1.2 The informed subjects

Due to the high number of subjects of the second group of English speakers affirming the general criteria, some situations could be selected for a direct comparison of speakers' ironic and non-ironic versions of an utterance, namely, situations 5a and 5b as well as 1a and 1b. For the descriptions of the utterances belonging to situation 5 the speakers will be referred to as speaker 2a, a female subject aged 35, and 2b, a male subject aged 34, since they were part of the second group of English speakers.

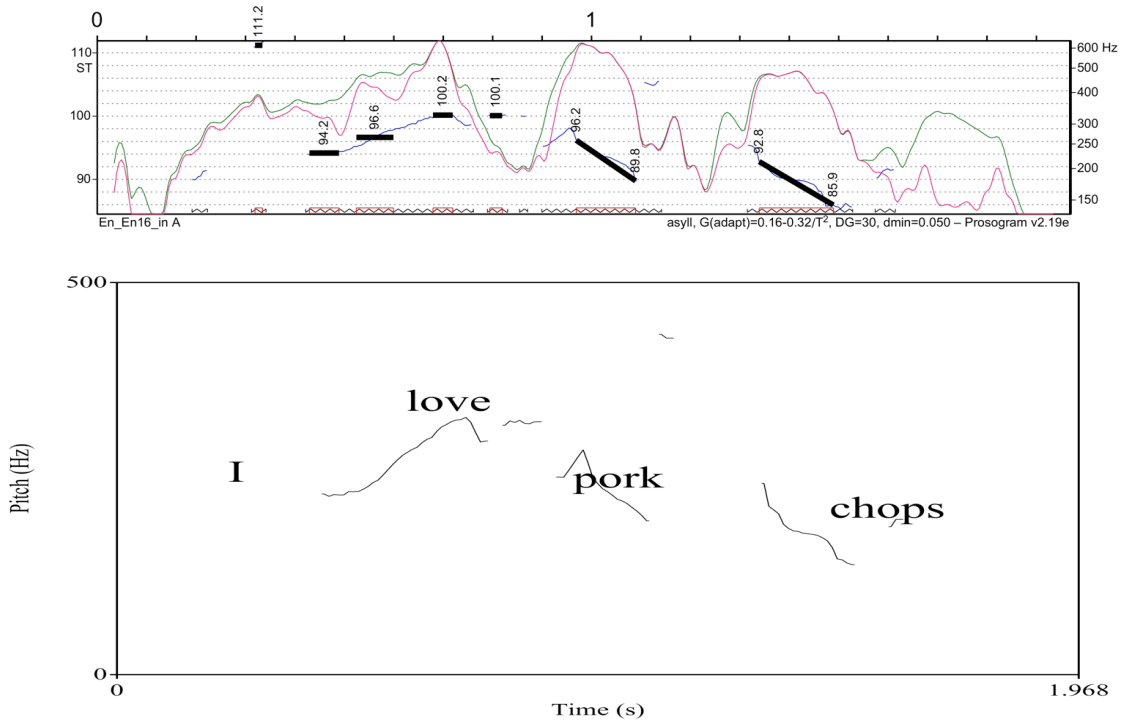


Figure 4. Situation 5a, speaker 2a, non-ironic.

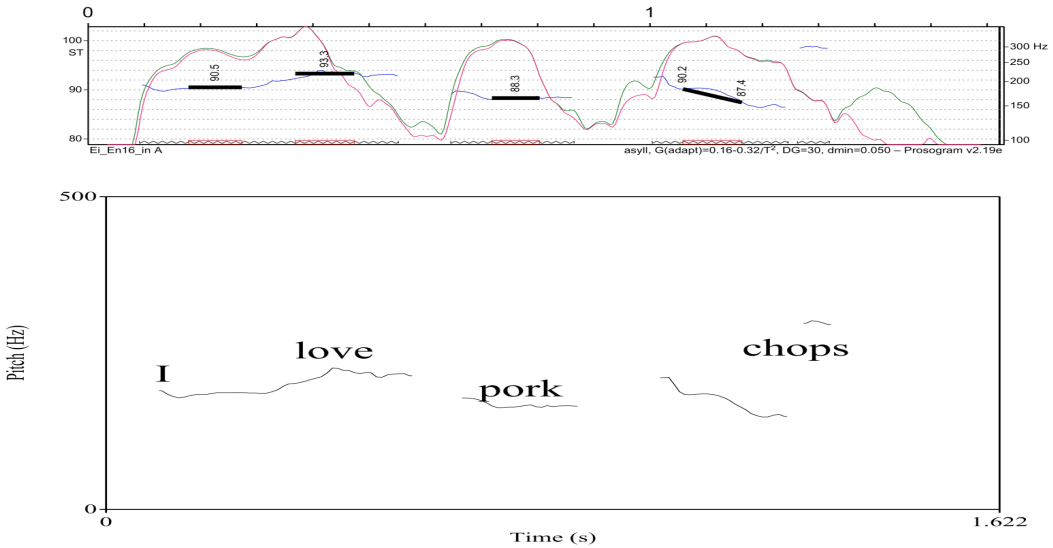


Figure 5. Situation 5b, speaker 2a, ironic

Figure 4 and 5 depict the non-ironic and ironic, respectively, utterances of situation 5 produced by speaker 2a. The three general measurements for irony, namely, reduced mean F0, decreased pitch range and monotonous pitch movement can be clearly seen in contrasting figure 5 to figure 4. Disregarding the obviously misplaced, erroneous first value of 111.2 semitones, the highest pitch value of the non-ironic utterance is 6.9 semitones greater than its ironic counterpart. Additionally, the lowest pitch value of the non-ironic utterance is 1.5 semitones smaller than its ironic counterpart, and the largest intervallic movements constitute 6.9 and 5 semitones, respectively. The mean intervallic movement constitutes 3.76 in the non-ironic and 3.13 semitones in the ironic version.

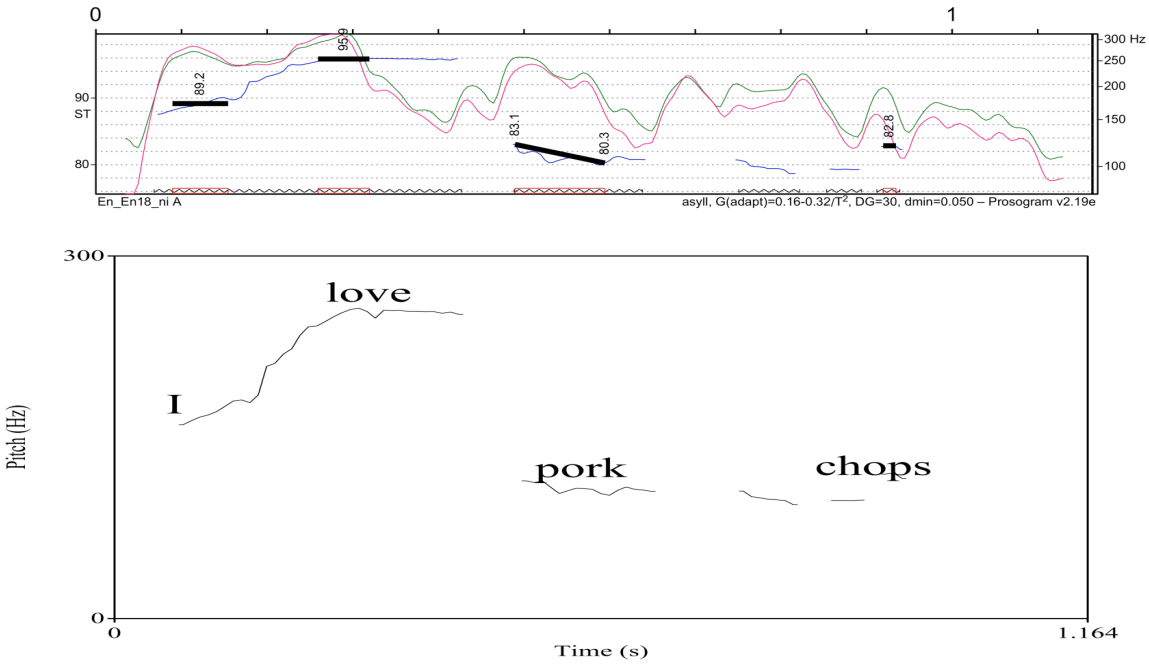


Figure 6. Situation 5a, speaker 2b, non-ironic.

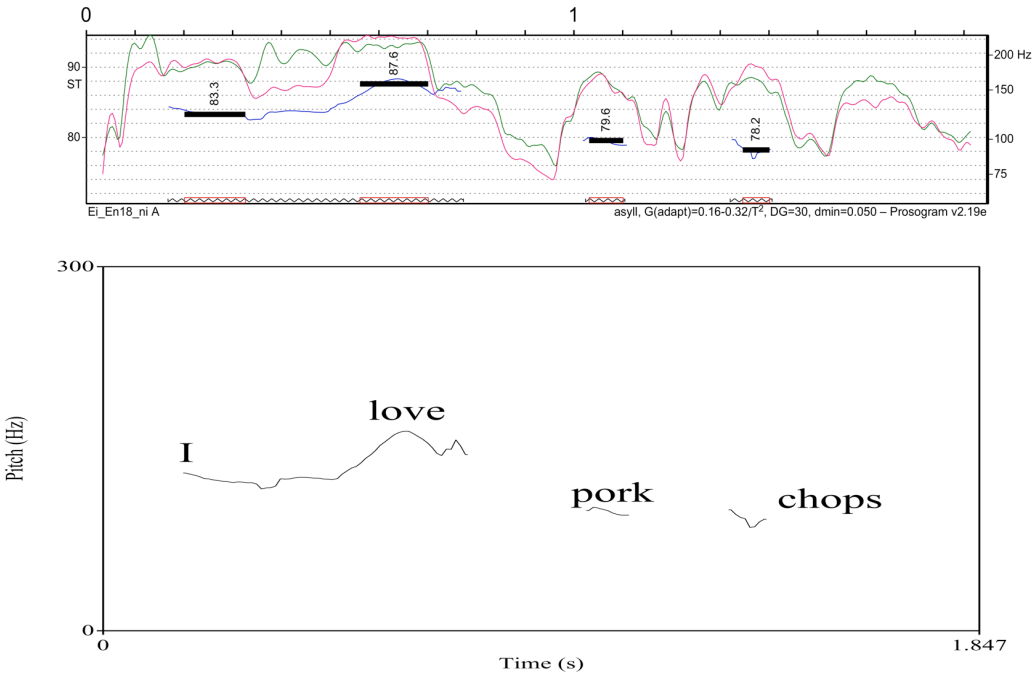


Figure 7. Situation 5b, speaker 2b, ironic.

In this comparison the last pitch value depicted in figure 6 was not considered a final rise. It rather appears as an erroneous evaluation by the software, since it appears clearly after the vowel value of the corresponding syllable when considering its annotation as well as the equivalent location in figure 4 and 5. In addition to the three general strategies of marking irony also applied by speaker 2a, speaker 2b marked the ironic utterance rhythmically by lengthening single vowels and also decelerating the overall speech tempo. The latter strategy

can be seen in the articulation of the last syllable in the proximity of the 1.5 seconds mark in the ironic version, which is in contrast to the last vowel's location in the equivalent non-ironic version remaining close the 1 second mark. Considering the multitude of pitch values produced in figure 4, one can assume that these measurements were generated as a result of the prosogram evaluating the high intensity used to physically amplify the prominence of the emotional affect marker "love" as multiple vowels.

Comparing this graph with its annotation and the other graphs belonging to this situation type, it can be assessed that the pitch values of 94.2 and 96.6 semitones constitute pitch values of the liquid /l/ in "love". In high intensity environments the software often calculated pitch values for this type of consonant. Furthermore, the pitch value of 100.1 semitones can be further framed as a pitch value that was calculated from an acoustic signal produced as a side effect of articulation. It can be classified as such since it significantly diverges from the general intensity contour and the corresponding segmental structure of the annotation. The missing pitch value for "I" may be established by referencing the location of the annotation for "I" to the short unvalued contour segment of the black F0 curve – onto which the prosogram usually projects the individual pitch values in the form of bars – located at 0.2 seconds. Since this contour segment spans between the marks of 90 and 92 semitones, one might propose a mean value of 91 semitones for the initial vowel value of this utterance.

When this proposed initial pitch value is taken as the reference point for the very first interval of this utterance, and the three additional erroneous pitch values not corresponding to distinct vowels are excluded, one can establish an interval of 9.2 semitones passing into the nucleus and 10.4 semitones from the nucleus to the end of its succeeding vowel. This melodic structure of the non-ironic version of situation 5 uttered by speaker 2a is in opposition the its ironic counterpart with intervals of 2.8 semitones passing into and exactly 5 semitones passing out of the nucleus. Interestingly, not only do the intervals of the ironic version correspond to roughly a third and a half of the intervals of the non-ironic version, they all constitute specific musical intervals to which particular types of more or less culture-independent valence and consonance can be attributed. 9 semitones and 10 semitones, representing the rounded pitch values of the non-ironic utterance of speaker 2a, correspond to a major sixth and a minor seventh, respectively. The former constitutes an interval that is perceived as a consonant in major but not in minor chords, and to which positively denoted affective values as kindness, satisfaction and gratification are generally attributed. Similarly,

the latter pitch movement represents an interval that is commonly associated with expressive values of warmth, love and romantics (Costa et al. 2000: 9). The rough approximates of the ironic version, a minor third consisting of 3 and a perfect fourth consisting of 5 semitones, are generally perceived as intervals expressing heaviness, sadness or discouragement, and indifference, coldness or firmness, respectively (Costa et al. 2000: 9). Speaker 2b produced intervals of 6.7 semitones passing into the nucleus and of 15.6 exiting the nucleus of the non-ironic utterance, and of 4.3 semitones entering and 8 passing out of the nucleus of the ironic version. While the exiting interval of the ironic utterance again constituted about a half of its non-ironic equivalent, the intervallic step from the nucleus in the ironic utterance forward represented roughly two thirds of the equivalent interval of the non-ironic utterance. The rounded pitch values of the non-ironic utterance, 7 semitones and 16 semitones, represent a perfect fifth and a major third (having subtracted the included octave, representing 12 semitones), respectively. The former interval is perceived as expressing positively denoted affective values as balance, love and certainty. Likewise, the latter is known to elicit sentiments as clearness, joy and hope. The rounded intervallic values of the ironic utterance represented again a major third with 4 semitones and a minor sixth with 8 semitones, the latter being associated with emotions as melancholy, worry and pity (Costa et al. 2000: 9). These results imply that the valence of musical intervals might play an important role in shaping pitch contours either enforcing or invalidating the literal content of an utterance. Whether it does one or the other seems to depend on the degree of correspondence between the denotative values of the literal content and the emotional valence of the prosodic content. The non-ironic utterances expressing genuine positive affect were produced with melodic contours consisting of intervals that contain positive valence by both speakers. On the contrary, both speakers created contextual incongruity in the ironic utterances by expressing literal content including a positively denoted affective marker through a prosodic contour that was made up of intervals of negative valence. The one exception, a major third used by speaker 2b as the initial interval, might be attributed to shortcomings in production accuracy. This assumption is based on the fact that speaker 2a, who remained without exception in this regard, indicated having twenty years of experience in singing, practicing singing ten hours per week as well as having undergone singing education. Contrarily, speaker 2b specified having neither training nor experience in signing and not practicing singing or any other musical activity. These differences in personal background may have played a decisive role in these performances as musically experience has been shown to improve pitch production and

perception in speech (see 3.1.). For the analysis of situation 1, the subjects will be referred to as speaker 2c, a 29-year-old male subject, and speaker 2d, a 44-year-old female speaker.

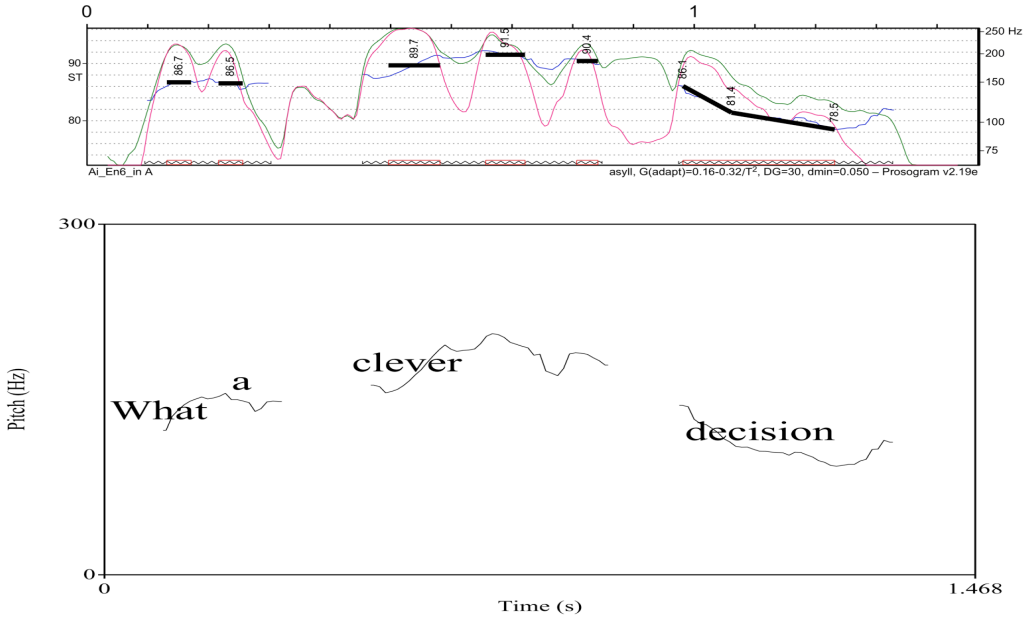


Figure 8. Situation 1a, speaker 2c, non-ironic.

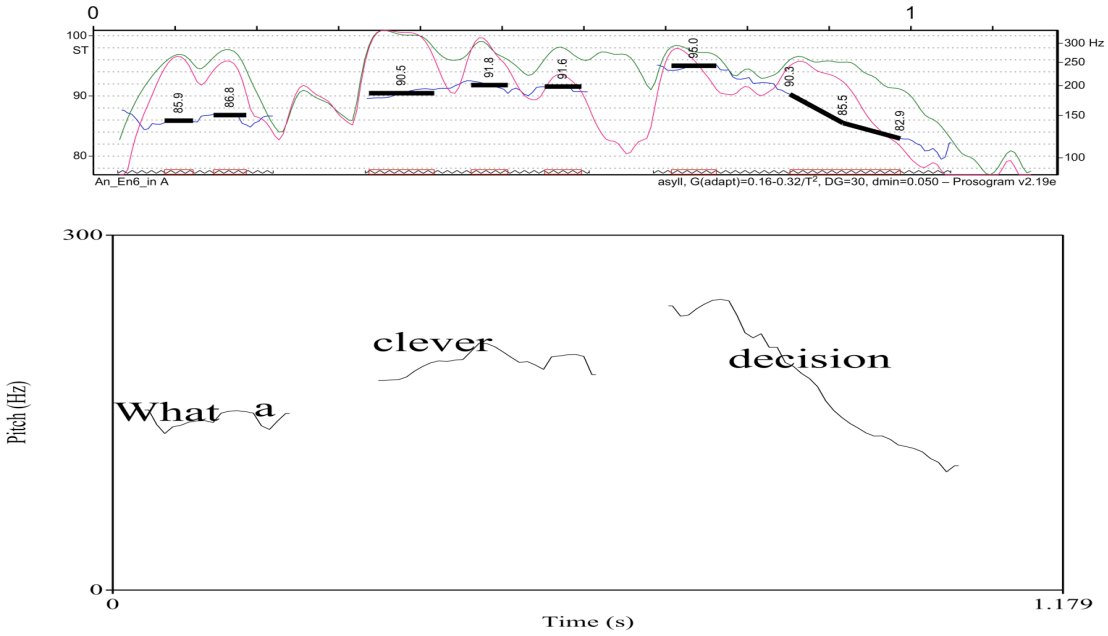


Figure 9. Situation 1b, speaker 2c, ironic.

In this comparison of the non-ironic, found in figure 8, and the ironic, shown in figure 9, utterances of situation 1 by speaker 2c, one can see the violation of one of the general criteria of ironic utterances, namely, a reduced mean F0, and a shift of the nucleus to the last word in the utterance. Furthermore, contrary to the samples analyzed before, the ironic utterance is of significantly greater length. However, due to these utterances being exclamative sentence

types, the location of the nucleus does not follow established standard pitch contour formations, which have only been consistently categorized for declaratives and interrogatives (see 2.1.1.). There is relatively little movement in the syllables before the last word in both versions, with a surprisingly similar mean intervallic movement of 1.58 semitones in the non-ironic and 1.53 semitones in the ironic version. However, the non-ironic version then includes a downward movement of 4.3 semitones, whereas the ironic utterance's pitch contour passes into the last word with an upward movement of 3.4 semitones. The former interval represents a major third being universally perceived as arousing joyous sentiments, while the latter, a minor third, is conceived as transmitting heaviness and lament (Costa M. et al. 2000: 9). These occurrences may support the idea that the valence carried by musical intervals that are contained within prominent parts of a prosodic contour constitute an explicitly meaningful layer of communicating affect, which may either substantiate or negate the meaning of the literal content. In speaker 2c's non-ironic utterance the positively valent interval following a relatively monotonous sequence of intervals may have been chosen to assure the potential listener that the literal content is to be interpreted literally, the positive valence of the interval reinforcing the denotation of the literal content. Moreover, in this way this interval of positive valence might have marked the utterance as a whole as a benevolent speech act of acknowledgment, congratulating the listener on an achievement (Bach 2003). This assumption is supported by the fact that this relatively monotonous sequence preceding the prominent interval was also found in the ironic version, where it was followed by a negatively valent interval. This different choice of interval possessing distinct affective properties may have been applied to signal that the preceding literal content, being invalidated by the negative sentiment of this interval, is not to be interpreted literally.

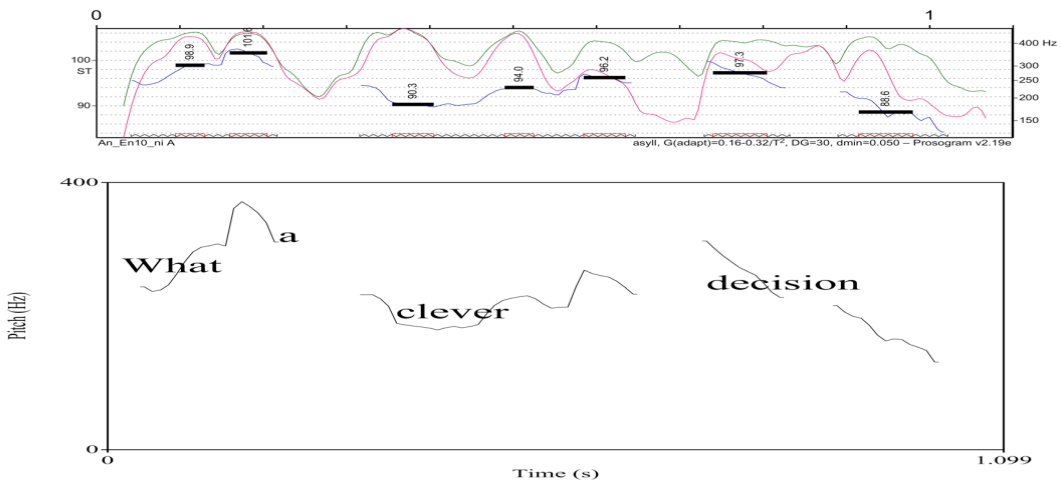


Figure 10. Situation 1a, speaker 2d, non-ironic.

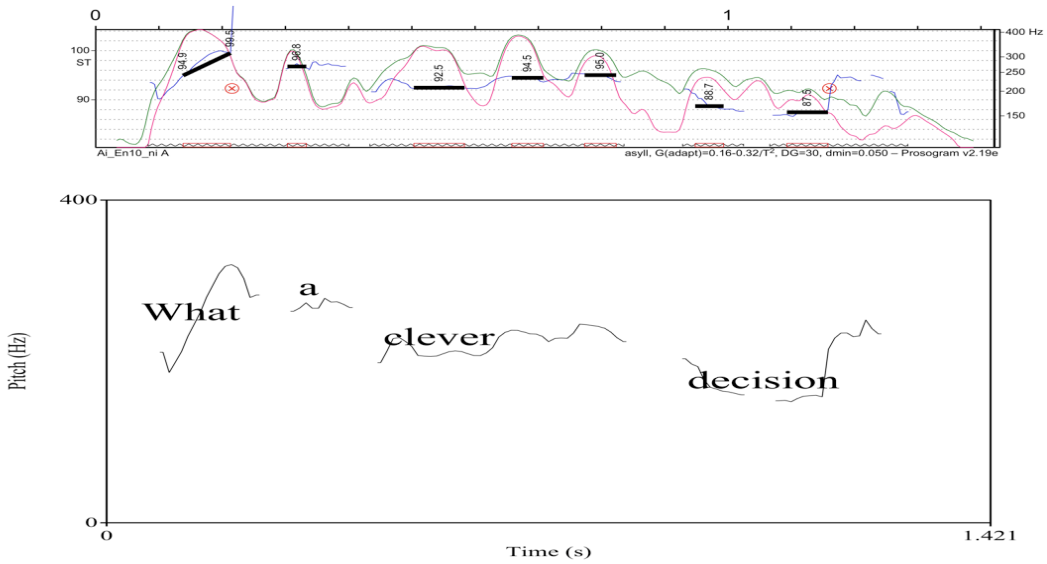


Figure 11. Situation 1b, speaker 2d, ironic.

In contrast to speaker 2c, speaker 2d utilized all the three general strategies of marking irony through prosody, in addition to the protracting of the ironic utterance. Furthermore, whereas speaker 2c used a significant pitch rise passing into the last word in the non-ironic utterance and a pitch drop in the ironic utterance, speaker 2d used these strategies inversely, depicted in figures 10 and 11, respectively. This discrepancy might have arisen due to differences in dialectical background, speaker 2c having been raised bilingually and speaking New Zealand English and speaker 2d being a bearer of the Geordie dialect. In the non-ironic utterance depicted in figure 10 there is no significant rise from the last pitch value of the second to last word to the final word, but a major second, equaling 1 semitone, which produces an emotionally neutral effect if not stressed (Costa et al. 2000: 8). Due to this occurrence one may consider more global melodic indicators as the initial pitch intervals of the second to last and the final word. The interval resulting from the ratio between these pitch values, exactly 7 semitones, correspond to a perfect fifth which is associated with positively denoted sentiments as balance, love and certainty. The rounded value of the following intervallic movement of 8.7 semitones, equaling 9 semitones, corresponds to an interval of a major sixth which also elicits positive emotional responses as gratification, effusiveness and radiance (Costa et al. 2000: 9).

In the ironic version there is a substantial intervallic movement between the second to last and the final word, which renders an analysis starting at this interval most suitable. The rounded

value of this interval of 6.3 semitones, accounting for 6 semitones, corresponds to an augmented fifth which is a dissonant interval eliciting negatively denoted sentiments as hostility and destructiveness (Costa et al. 2000: 8). This pitch shift is followed by an intervallic movement of 1.2 semitones. The rounded value of this interval step, constituting 1 semitone, is more prominent, thus also more meaningful, in this sequential context than in the sequence discussed in the non-ironic utterance, since it is expressed in the final syllable of the utterance and is thus subject to lengthening processes (Di Cristo et al. 1999: 36). A single semitone corresponds to the musical interval of a minor second which is the smallest intervallic step of the chromatic scale and known to elicit negatively denoted sentiments as roughness, fear and anger (Costa et al. 2000: 9). Hence, even though speaker 2d utilizes inverse strategies to position the ironic cue in the pitch contour, considering the direction of intervallic movements in regard to speaker 2c, the factor of intervallic consonance seems to function identically.

Dissonant intervals with physical properties that elicit negative emotional responses are used to signal a negation of the literal meaning of the verbal content of an utterance, while consonant intervals with physical properties that evoke positive sentiments are utilized to assure the listener of the positive, direct meaning of an utterance's literal content. These indications were not only found in the local context of the nucleus and its neighboring syllables but also in a more global environment, from the beginning of one word to another. In utterances that contain sequences with only small intervallic steps that have no individual harmonic significance unless under stress – intervals smaller than a minor third, the minor and major second consisting of 1 and 2 semitones, respectively – musical intervals may serve as affective boundaries of larger stretches of a prosodic contour, e.g., from one initial pitch value of a word to another of the following word, as shown in the non-ironic utterance of speaker 2d.

4.3.2 German speaking groups

The general criteria for marking irony through intonation were met by six speakers in the first group of German speakers. However, in the second group one of these three criteria was excluded as a selective factor, namely, mean F0, since 60% of the speakers displayed a higher mean F0 in their ironic utterances. This inverted ratio may indicate that native speakers of German, or at least speakers of Austrian German, might utilize different strategies to produce intonational cues for irony than native speakers of English.

4.3.2.1 The uninformed subjects

Out of the six subjects of the first group one produced no final rises, two produced one final rise, another produced two final rises and the remaining one produced only final rises. The number of upward intervallic movements in the nucleus reached from one to five occurrences per speaker, with a mean of 3.7 upward interval steps per speaker. Since the utterances of the first group of native speakers of English that were discussed above constituted non-ironic statements, the utterances that were performed by subjects of the first group of German speakers will be the ironic statements of situation 5b shown in figures 12 and 13. The speakers that were chosen for the following analysis were selected by the lowest number of occurrences of final rises. However, the speaker with no final rises was the only subject who requested further information about the contexts or the wording of the utterances multiple times. This uncertainty was reflected in the prosogram's failing in producing pitch values in many instances of this subject's utterances. Hence, the two next-ranking subjects, both having produced only one final rise, were chosen. The subjects will be referred to as speaker 11a, a 29-year-old male subject, and speaker 11b, a female subject aged 37.

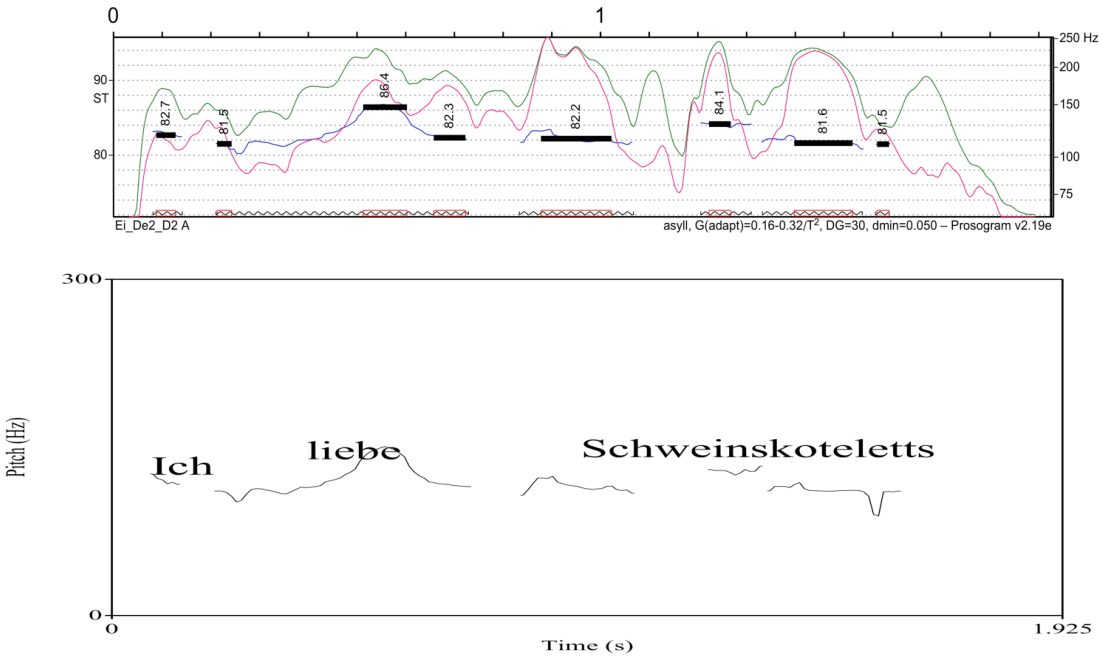


Figure 12. Situation 5b, speaker 11a, ironic.

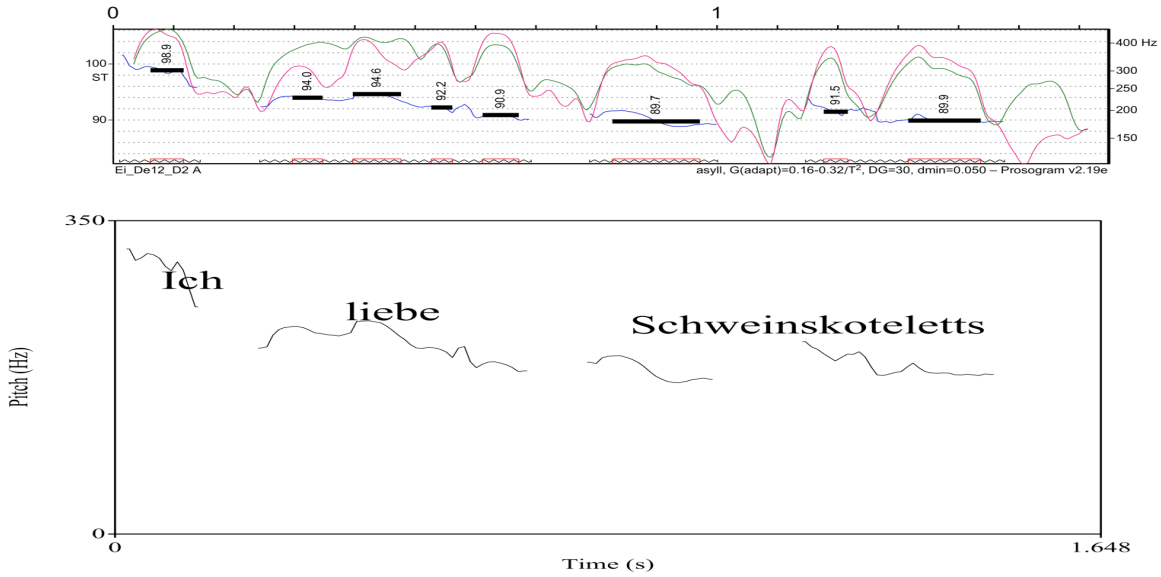


Figure 13. Situation 5b, speaker 11b, ironic.

In the utterance of speaker 11b, depicted in figure 13, there are four pitch values for the word *liebe* 'love' containing only two syllables, so that the initial value was most likely calculated for the liquid /l/. Speaker 11a produces intervals of 4.9 semitones passing into and 4.1 semitones exiting the nucleus. The rounded value of the former, equaling 5 semitones, corresponds to a perfect fourth. This interval comes into effect as rather negatively denoted sentiments as hardness, coldness and indifference (Costa et al. 2000: 9). These stances fittingly describe opposites of the denotations of the affective marker, *liebe* 'love', that is used in this utterance and might thus be an indicator for the purposeful application of this musical interval. However, it is important to note that due to its consonant properties the interval of a perfect fourth can also produce rather positively denoted sentiments as achievement, contrary to the other intervals associated with negative emotional responses. Nonetheless, it elicits neutral sentiments as firmness even more frequently. This ambiguousness emphasizes the importance of the preceding intervals in attributing meaning to its successor in a sequence by creating harmonic and melodic tension and expectations of release, especially with the interval of a perfect fourth. In this case the perfect fourth seems to be of negative valence, since its preceding interval constitutes 1.2 semitones, the rounded value of which corresponding to a single semitone. The intervallic movement exiting the nucleus in speaker 11a's ironic utterance, depicted in figure 12, corresponds to a major third, an interval associated with positively denoted sentiments (see the analysis above). This interval may have been chosen to alleviate the negative affective value of the preceding intervals in order to signal the lack of hostile intention, marking the utterance as an instance of irony rather than sarcasm.

In contrast to the beginning of speaker 11a's utterance, speaker 11b's pitch contour begins with the highest pitch value of the entire utterance. From there, the subject produced an interval of again 4.9 semitones – corresponding to the musical interval of a perfect fourth – passing into the word containing the ironic cue, exactly as speaker 11a, but instead of an upward a downward movement was executed. The following intervals in the second word consist of a sequence of rounded intervallic movements of a major second and two minor seconds, which are the least harmonically relevant and perceived neutrally if not accented (Costa et al. 2000:8). Interestingly, both speakers utilized the same intervallic movement as the most prominent interval to shape their prosodic utterances that include a contextual incongruity, allowing for an interpretation of the utterance as an instance of irony, even though their intervallic steps moved into opposing directions due to different locations of accentuation.

4.3.2.2 The informed subjects

In the second group of German speakers four subjects conformed to two of the three general criteria and additionally met the newly established, inversed criterion of lower mean F0 in non-ironic utterances. The situations chosen for the analysis of this chapter constitute situation 3 and 4, since they both have not yet been subjected to a qualitative analysis. Due to the varying distribution of the twelve utterances among the subjects having to perform six utterances each, situation 3a is found only once among these four selected speakers. Thus, an utterance of a speaker who was not among the four mentioned subjects, conforming to the three initial general criteria for marking irony through prosody, was chosen to complete the samples for this qualitative observation. The speakers whose utterances of situation 3a will be investigated will be referred to as speaker 22a, a 23-year-old female subject, and speaker 22b, a female subject, aged 34.

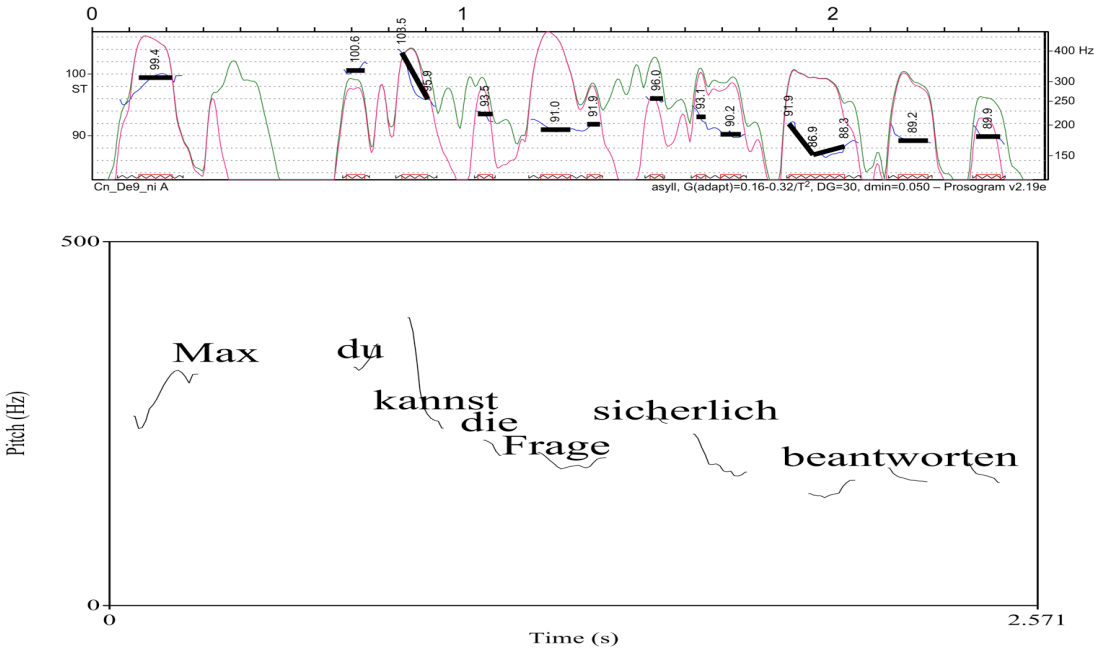


Figure 14. Situation 3a, speaker 22a, non-ironic.

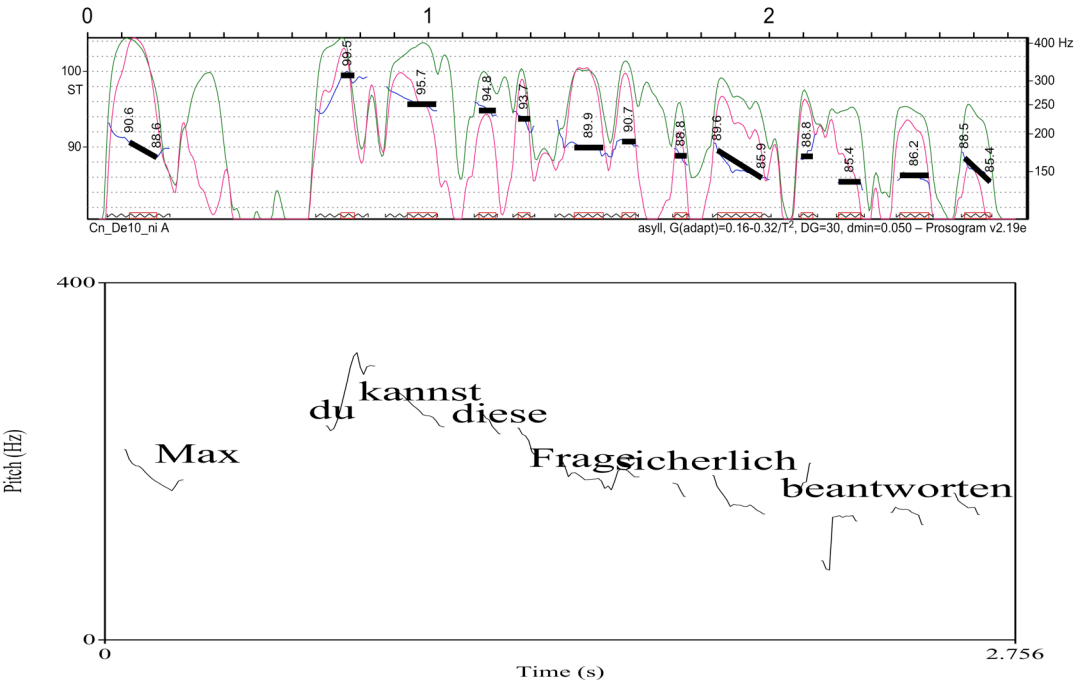


Figure 15. Situation 3a, speaker 22b, non-ironic.

The word – *sicherlich* 'surely' – that will be analyzed in these statements is the one that contains the ironic cue in the ironic equivalents of the non-ironic utterances depicted in figure 14 and 15. Interestingly, speaker 22a, exhibited in figure 14, used an upward interval step of 4.1 semitones moving into this word. The rounded value of this pitch movement corresponds to the musical interval of a major third eliciting positively denoted sentiments (see 4.3.2.1.). Speaker 22b, shown in figure 15, did not use the same interval passing into this word but

shortly after in the subsequent intervallic movement. This shift of prominence is not to be ignored since it is the most prominent pitch value of the entire pitch contour, being attributed a perceivable pitch glide from the second to the third syllable of this word. This downward intervallic movement of 3.7 also sets itself apart from the preceding, harmonically irrelevant interval units of a major second and a minor second, the rounded equivalents of 1.9 and 0.8 semitones, respectively. This occurrence might support the tendency detected in the English groups that musical intervals of positive valence are used to reinforce the literal meaning of an utterance. These results seem to reinforce the implication formed in the analysis of speaker 11a and 11b that the direction of the intervallic movement can go either way, upward or downward, still retaining its specific valence. Furthermore, regarding the different positions of this interval in speaker 22a's and speaker 22b's utterances, another conclusion can be drawn. The position of a prominent musical interval may occur varyingly depending on rhythmical and stress placement choices and does not inhibit the interval's specific affective value of either reinforcing or denying the literal meaning of an utterance.

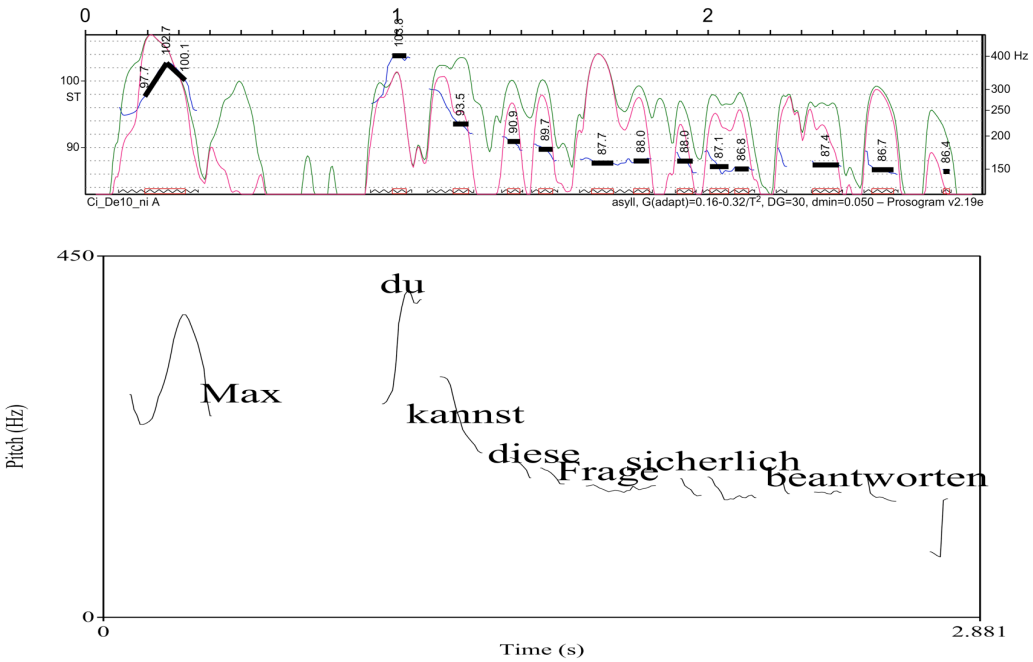


Figure 16. Situation 3b, speaker 22b, ironic.

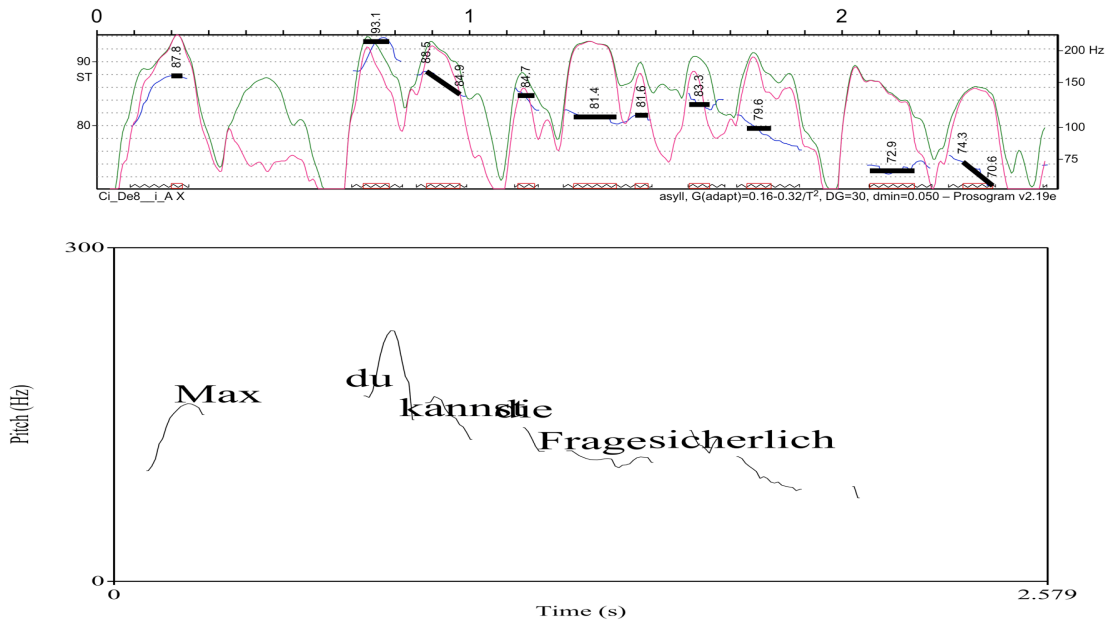


Figure 17. Situation 3b, speaker 22c, ironic.

In figure 16 the ironic utterance of situation 3 performed by speaker 22b is depicted. Figure 17 shows the ironic utterance of speaker 22c, a 54-year-old male subject. Speaker 22b did not use harmonically relevant intervals in the word that was intended to constitute the cue word for irony by the author, producing intervals from 0 to 0.9 semitones. However, one may still consider the most prominent interval of this utterance as a pitch unit possibly containing relevant prosodic information. This interval constitutes the greatest intervallic movement of this utterance, a drop of 10.3 semitones from the word du ‘you’ to its neighboring syllable. The rounded value of this interval, corresponding to a minor seventh, is in stark contrast to the interval the same speaker used at the equivalent location in the non-ironic utterance, corresponding to a major third also moving downward. The former interval is associated with negative emotional responses contrary to the latter interval. Hence, in this instance the ironic cue might thus be contained in the personal pronoun du ‘you’, similarly to the instance of speaker 11b’s utterance, depicted in illustration 13, also containing a significant downward movement in a syllable representing a personal pronoun. These results seem to confirm that musical intervals present in prominent parts of a prosodic contour in German fulfill the same functions in regard to the verbal content of utterances as in English.

In the analysis of situation 4 the speakers will be referred to as speaker 22d, a male subject aged 57, and 22e, a 55-year-old male subject.

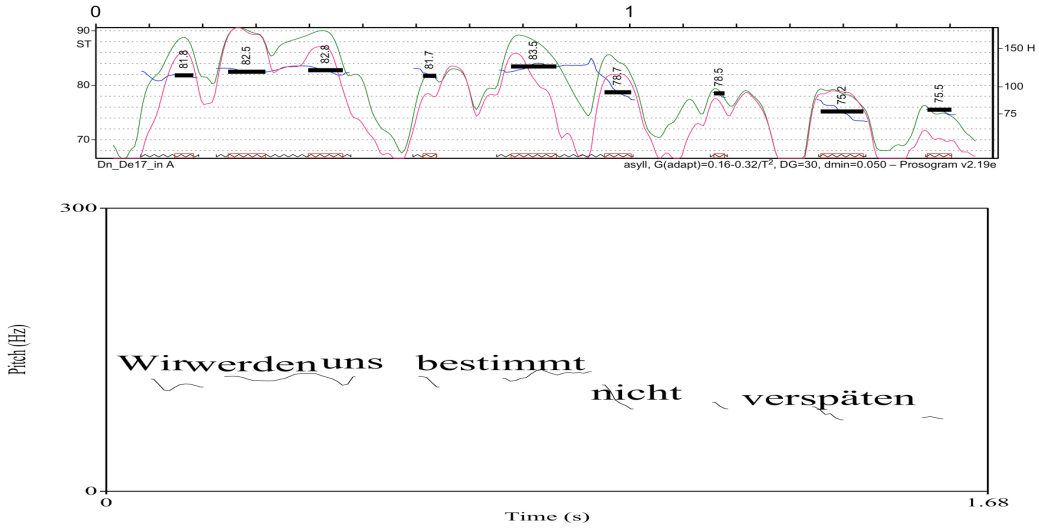


Figure 18. Situation 4a, speaker 22d, non-ironic.

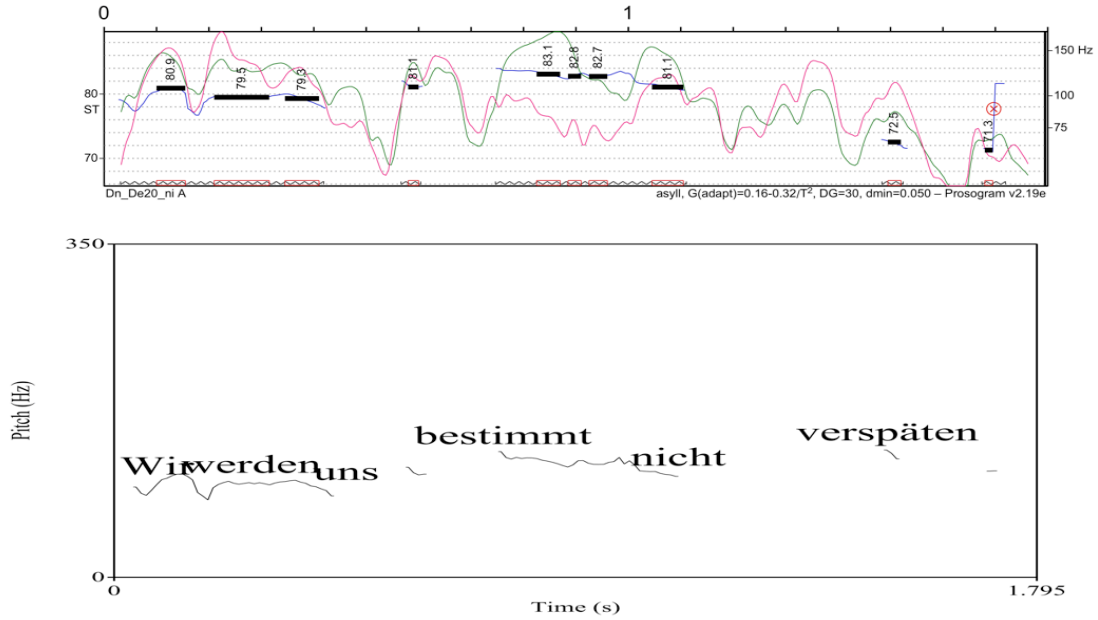


Figure 19. Situation 4a, speaker 22e, non-ironic.

The non-ironic utterance of speaker 22d, depicted in figure 18, contains a downward moving interval of 4.8 semitones into the cue word of its ironic equivalent, namely, *nicht* 'not'. The rounded value of this interval corresponds to the musical interval of a perfect fourth, which is the most ambiguous of the musical intervals of the chromatic scale regarding its affective substance, as discussed in the analysis of speaker 11a and 11b. The subsequent interval of 0.2 semitones corresponds to the level musical interval called a unison (Costa et al. 2000: 8), which may play an important role in this melodic structure since it is not part of a longer sequence of shorter intervals but is followed by a larger interval of 3.3 semitones. Thus, it

might have been used to specify the ambiguous affective valence of the preceding perfect fourth as bringing into effect either its positive or negative affective force. Due to a unisonous interval inducing emotional states or stances as insistence, peace and smoothness, it may have reinforced the perfect fourth to be perceived as an indicator for a literal interpretation (Costa et al. 2000: 8). In the non-ironic statement of speaker 22e, depicted in figure 19, a sequence of rather small intervallic movements prior to the word under investigation includes pitch values representing the articulation of the nasal /m/ of the word *bestimmt* 'certainly'. The following pitch drop of 1.6 semitones into the cue word appears to be of less significance compared to the prominent downward interval step out of it. The latter interval of 8.6 semitones which corresponds to a major sixth, eliciting sentiments as satisfaction and gratification, might have been used to signal the positivity of the literal content by releasing the tension created by the prior sequence of relatively small and short intervals. Hence, speaker 22d and 22e may have applied similar techniques of signaling the verbal content as intended to be interpreted according to its literal denotation through a prominent downward movement. The way in which these methods were realized differed in their respective proportions since their environments were arranged differently regarding the size of the surrounding intervals and the shape of the global pitch contour.

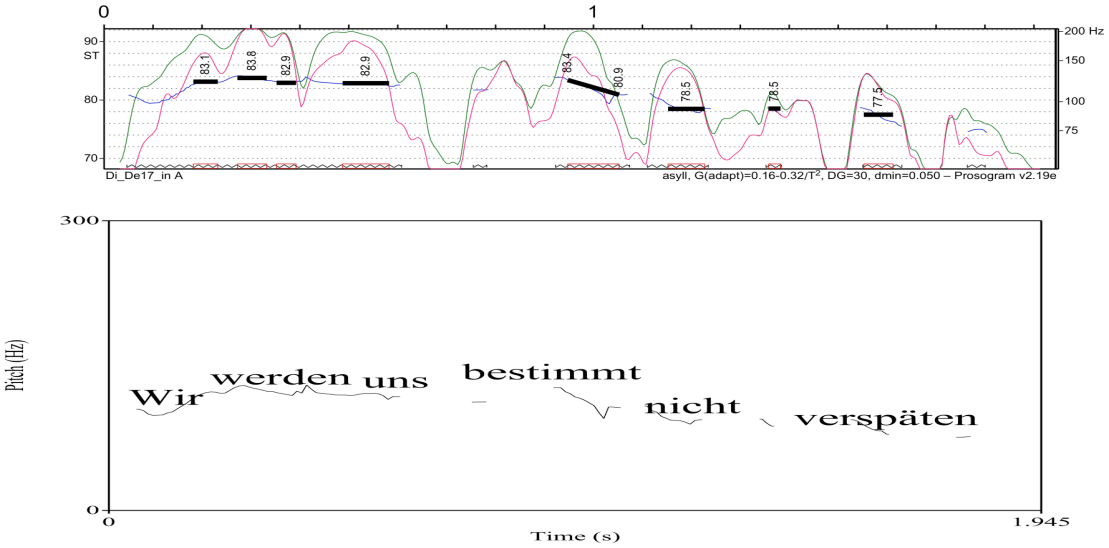


Figure 20. Situation 4b, speaker 22d, ironic.

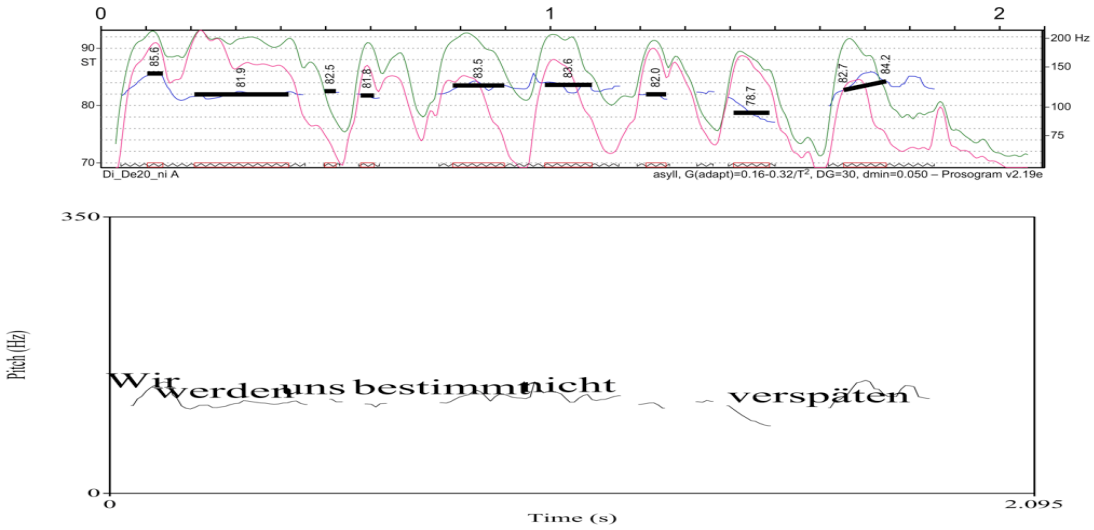


Figure 21. Situation 4b, speaker 22e, ironic.

In figures 20 and 21 the ironic utterances of speaker 22d and 22e, respectively, are portrayed. The ironic utterance of speaker 22d met all the criteria of marking irony established for the second group of German speakers, with the mean F0 higher in the ironic version. Speaker 22d's pitch contours of both versions are quite similar to each other, significant differences are found in the overall pitch height of sequences of relatively closely neighboring pitch values but not in the intervallic movements themselves except of one instance. There is an intrasyllabic pitch movement of 2.5 semitones in the final syllable of *bestimmt* 'certainly', which could have been interpreted as the cue word. This interval's rounded value of 3 semitones represents the musical interval of a minor third, which is linked to evoking negatively denoted sentiments as discouragement and sadness (Costa et al. 2000: 8). Due to the interval's prominence in the pitch contour, even the rounding down to an interval of 2 semitones, constituting the musical interval of a major second, would have marked the verbal content with negative affect, since a major second is known to evoke negative sentiments as displeasure when stressed (Costa et al. 2000: 8). Additionally, the succeeding interval also constituted a major second, enforcing its negative valence. Therefore, this intervallic movement might have been applied to mark the respective literal meaning as incongruent with the prosodic medium through which it is realized. The ironic utterance of speaker 22e met all the relevant criteria of this group except of the intervallic movements actually producing a slightly greater mean value than in its non-ironic equivalent. Furthermore, this ironic utterance stands in contrast to its non-ironic equivalent in the way that the prominent drop towards the final word is reduced to 3.3 semitones, which may have thus functioned similarly to the minor third interval in 22d's ironic utterance. The two preceding intervals also

do not contain an interval step higher than 1.6 semitones. However, the intervallic movement from the initial to the final pitch value of the final word constitutes an upward motion of 5.5 semitones. Rounding up, this pitch movement corresponds to the musical interval of an augmented fourth, evoking sentiments as disdain or aversion (Costa et al. 2000: 8), whereas rounding down, it represents a perfect fourth, which holds a more ambiguous affective content. Although in this case it might lean towards its negative aspects coming into effect in regard to the preceding intervals' negative valence. Therefore, in these utterances local as well as global prosodic features of prominent segments of a prosodic contour of an ironic statement were shown to correspond to musical intervals to which negative valence is attributed.

5 Conclusion

Analyzing prosodic markers of irony entails a number of methodological obstacles that begin at the first step of an analysis – the establishing of a theoretical framework. The lack of coherent terminological frameworks in both the study of prosody and the study of irony present the researcher with a multitude of perspectives that partly overlap with and partly diverge from each other. Yet this plurality may be used to the analyst's advantage, if the different approaches are examined for their commonalities regarding the empirical and theoretical results they produce in diverse fields of study. For instance, considering the acoustic properties of prosody from a musicological perspective uncovers the functionality of speech as a direct medium of conveying affect and emotions that can be illustrated in physical measures. Likewise, observing irony from this very angle discloses some details about the complex multifaceted relations of language and the types of mediums through which it expresses and produces meaning. The consideration of as many perspectives as possible is also necessary for establishing a multitude of points of orientation in order to increase the degree of predictability in investigating an ambiguously multilayered phenomenon as spontaneous speech. For instance, the inclusion of definite pragmatic categories allows one to create effective methods of data collection that aim at observing spontaneous speech acts produced from fixed sets of text more reliably. However, there is a fine line between too little and too much control in designing simulations of spontaneous speech. When the production context of a study provided to the subjects is so constrained that the subjects are not able to identify enough with a given speaker role to impersonate the respective utterance as if it originated from their own mental processes, the results do not carry much significant meaning

in terms of the study's goal of analyzing physical dimensions that reflect intuitive aspects of language use. The degree of liberty that subjects should be granted in interpreting the instructions of such a study is hard to define, since a particular degree of similarity between a subject's interpretation of a given text and the certain type of interpretation correlating to the author's intention is required for controlling individual instances of production to turn out usable for the comparison with others. The extent to which an individual subject perceives a given instruction as authentic will probably always remain unpredictable to a certain degree, due to the impossibility of creating basic contextual frames that anyone regardless of age, gender, personal experience and knowledge may identify with.

The difficulty of designing an experimental environment in which spontaneous speech can be replicated relatively authentically has presented itself in the results of the quantitative evaluation of the samples collected in this study. For instance, one of the measures that were predetermined to be applied in the selection process for the qualitative investigation had to be modified in the course of this analysis in one of the four groups, due to its figures diverging significantly from the prearranged parameter. This condition might have arisen due to a variety of factors that relate to the literal as well as the contextual framework of the study. Situation 2 with its utterances' nuclei located in the first syllable proved to be unfit for a further qualitative analysis. This particular design of that situation probably caused the software's relatively frequent failing in producing a pitch value in the nuclei of these utterances resulting in whole utterances not containing any significant pitch values. This risk might have been circumvented by including an arbitrary interjection prior to the first word in order to provide a reference point from which an intervallic step into the nucleus could be evaluated. Another methodological improvement could constitute the investigation of utterances that are exclusively performed by subjects who possess the respective speaker roles in real life. Such an endeavor would require a careful selection of subjects according to profession or social status and the design of the study's texts to be tailored correspondingly. The extent to which the subjects were informed by the instructions and the author about the study's aims may have influenced their individual commitment to perform a given speaker role. This manipulative influence might have been avoided by replacing the author as the overt instructor of the experiment in order to reduce the potential of subjects feeling as though they were under the pressure to perform, changing their tone of voice fundamentally. Despite the shortcomings of the study's design, all samples selected for the qualitative analysis proved to share significant features even cross-linguistically, about which one may

draw compelling conclusions about the musical qualities of pitch contours. Whenever the nucleus was shaped by a prominent intervallic movement into or out of the respective syllable, the rounded value of that movement corresponded to a musical interval which carried a specific form of valence. The pitch movements used in the nuclei of non-ironic statements comprised musical intervals that contained a valence that is associated with positively denoted sentiments. On the contrary, prominent interval steps found in ironic-utterances' nuclei comprised musical intervals that possess negative valence, eliciting negatively denoted emotional responses in a listener. The direction of the movements varied and seemed to depend on its overall location in the respective utterance. When there was no prominent interval in the nucleus itself, an interval functioning in the same manner occurred on a more global scale, stretching from the initial to the final syllable of the respective word or arising at word boundaries. In cases where a musical interval of neutral or indefinite, ambiguous affective valence occurred in the nuclei it was succeeded by an interval which carried a definite, either positive in non-ironic or negative in ironic statements, valence, dissolving its initial indefiniteness and being attributed the valence of its subsequent interval. There was only one case in which an interval of positive valence occurred in a prominent syllable of an ironic utterance.

These tendencies imply that musical intervals may indeed constitute an independent level of meaning in intonation, on which not only the transmission of affect but also of indicators of the degree to which the explicit literal content includes additional implicit meanings by referring to contextual clues take place. However, these musical intervals occurring in speech are not as precisely formed as intervals in musical performance since their individual units do not relate to fixed pitch values in order to attain meaning. Moreover, the relatively unpredictable sequential order of turn-taking in conversations may not allow for a detailed analytic processing of the acoustic material by the listener. The detection of musical intervals would require the listener to employ cognitive abilities way beyond the prioritized decoding of pragmatic and syntactical modalities expressed through a prosodic contour. These domains have to be indispensably attended to in order to successfully participate in a conversation, in contrast to musical information. The large number of pitch values occurring in short amounts of time within a rhythmic framework of irregular pulses may add to the brain's inability to attend to the analytical decoding of musical properties of speech. However, the universal physical properties of frequency ratios of acoustic signals, the attributes of either consonance or dissonance of an interval, seem to function as reference points for marking stances and

affects in prosodic contours, especially in locations of prominence. Speakers appear to utilize the production of specific musical intervals intuitively, which does not come as a surprise since the decoding of the valence of musical intervals constitutes an innate ability. Additionally, musical intervals form the fundamental units of speech production as the harmonics of vowels, the formants, which are unconsciously and arbitrarily produced as resonances by the physical shape of the vocal tract.

The results of this thesis may serve as a starting point for further investigations into the musical properties of speech and how they interact with the literal meaning of utterances. Additionally, the documented methodological errors arising in the course of the conduction of the study may serve as guidelines for improved methods for future investigations into this topic.

6 References

Adriaens, L.M.H. 1991. “Ein Modell deutscher Intonation: eine experimentell-phonetische Untersuchung nach den perzeptic relevanten Grundfrequenzänderungen in vorgelesenem Text.” Eindhoven: Technische Universiteit Eindhoven.

Anderson, J.W. 2005. “The natural ways of musicality, language, and self-regulation in infants”. *Psychomusicology* 19(1), 41-58.

Antoniou K., Clin E., Deliens G., Kissine M. and E. Ostashchenko. 2015. “Context, facial expression and prosody in irony processing. *Journal of Memory and Language* 99, 35-48.

Attardo S., Eisterhold J., Hay. J. and I. Poggi. 2003. “Multimodal Markers of Irony and Sarcasm”. *International Journal of Humor Research* 16(2), 243-260.

Bach, K. 2003. “Speech Acts and Pragmatics”. In: Devitt, M. and R. Hanley (eds.). *Blackwell Guide to the Philosophy of Language*. Hoboken, NJ: Wiley-Blackwell.

Begagic, M. 2013. “Semantic preference and semantic prosody of the collocation make sense”. *Jezikoslovlje* 14.2-3, 403-416.

Besson M. and Friederici A.D. 1998. “Language and Music: A Comparative View”. *Music Perception* 16(1), 1-9.

Boeckx, C. and R. Asano. 2015. “Syntax in language and music: what is the right level of comparison?”. *Frontiers in Psychology* 6, 1-16.

Braun A. and A. Schmiedel. 2018. “The phonetics of ambiguity: A study on verbal irony”. In: Thaler, V. and E. Winter-Froemel (eds.). *Cultures and Traditions of Wordplay and Wordplay Research*. De Gruyter. 111-136.

Brown, S. 2017. “A Joint Prosodic Origin of Language and Music”. *Frontiers in Psychology* 8, 1-20.

Brown, S., Currie, T.E., Sakai E. and P.E. Savage. 2015. “Statistical universals reveal the structures and functions of human music”. *Proceedings of the National Academy of Sciences of the USA.* 112(29), 8987-92.

Burns E. M. 1999. “Intervals, scales and tuning”. In: Deutsch D. (ed.). *The Psychology of Music 2ed*. San Diego, CA: Academic Press. 271–300.

Champagne-Lavau M., Kein M. and E. Riviere. 2018. “Using context and prosody in irony understanding: Variability amongst individuals”. *Journal of Pragmatics* 138, 165-172.

Ben Jannet, M., Champagne-Lavau M., d’Imperio M., Loevenbruck H. and M. Spini. 2013. “Prosodic cues of sarcastic speech in French: slower, higher, wider.” *14th Annual Conference of the International Speech Communication Association* Aug 2013, Lyon France, 3537-3541.

Choi, J., Purves D. and D. Ross. 2007. “Musical intervals in speech”. *PNAS* 104(23), 9852-9857.

Costa, M., Ricci Bitti P. and L. Bonfiglioli. 2000. “Psychological Connotations of Harmonic Musical Intervals”. *Psychology of Music* 28, 4-22.

Di Cristo, A.F. and D. Hirst. 1999. *Intonation Systems: A Survey of Twenty Languages*. Cambridge University Press.

Eerola, T., Louhivuori, J. and E. Lebaka. 2009. “Expectancy in Sami Yoiks revisited: The role of data-driven and schema-driven knowledge in the formation of melodic expectations.” *Music Science* 13, 231–72.

Feller, S. 2008. “Irony as a Rhetorical Device in Dialogic Interaction”. In: Weigand E. (eds.). *Dialogue and Rhetoric*. Amsterdam: John Benjamins.

Fenk-Oczlon, G. 2017. “What Vowels Can Tell Us about the Evolution of Music”. *Frontier Psychology* 8, 1581.

Fenk-Oczlon G. and A. Fenk. 2009. “Musical pitch in nonsense syllables: correlations with the vowel system and evolutionary perspectives”. In: Eeerola P.-S., Eerola T., Himberg T., Louhivuori J. and S. Saarikallio (eds.). *Proceedings of 7th Triennial Conference of the Europaean Society for the Cognitive Sciences of Music*. Jyväskylä: European Society for the Cognitive Sciences of Music. 110–113.
Fenk-Oczlon G. and A. Fenk. 2009. “Some parallels between language and music from a cognitive and evolutionary perspective”. *Music and Science* 13, 201–226.
Filippi P., Ravignani, A. and B. Thompson. 2018. “The Evolution of Musicality: What Can Be Learned from Language Evolution Research?”. *Front Neurosci* 12.
Fox Tree. J. E. 2005. “Is there an ionic tone of voice?”. *Language and Speech* 48(3), 257-277.
Gibbon, D. 2016. “Rhythm and Melody: Aspects of Language and Music”. Jinan University, Guangzhou.
Grice, M. Sneed German J. and P. Warren. 2019. “Intonation systems across varieties of English”.
Grichkovtsova I., Lacheret A. and M. Morel. 2011. “The role of voice quality and prosodic contour in affective speech perception”. *Speech Communication*.
Gussenhoven, C. 1983. “Focus, mode and the nucles”. *Journal of Linguistics* 19(2). 377-417.
Hellbernd, N. and D. Sammler. 2018. “Neural basis of social communicative intentions in speech”. *Social Cognitive and Affective Neurosciences*, 604-615.
Honing, H. 2018. “On the biological basis of musicality”. *The Neurosciences and Music* 6, 51-56.
Iversen, J.R., Patel A.D. and J.C. Rosenberg. 2006. “Comparing the rhythm and melody of speech and music: The case of British English and French”. *The Journal of the Acoustical Society of America* 119(5), 3034-3047.
Masataka, N. 2009. “The origins of language and the evolution of music: A comparative perspective”. *Physics of Life reviews* 6(1), 11-22.
Mauchand, M., Vergis N. and M.D. Pell. 2019. “Irony, Prosody, and Social Impressions of Affective Stance”. *Discourse Processes* 56(1), 1-17.
Mertens, P. 2004.“The Prosogram: Semi-Automatic Transcription of Prosody based on a Tonal Perception Model”. In Bel B. and I. Marlien (eds.). *Proceedings of Speech Prosody 2004*, Nara (Japan), 23-26 March.
Menninghaus, W. 2018. “Poetic speech melody: A crucial link between music and language”. *PloS One* 13(11).
Nolan, F. 2003. “Intonationonal equivalence: an experimental evaluation of pitch scales”. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*, 771-774.
Nooteboom, S. 2010. “The prosody of speech”. In: Hardcastle, W.J., Gibbon. F.E. and J. Laver (eds.). *The Handbook of Phonetic Sciences 2ed*. Hoboken, NJ: John Wiley & Sons. 640-673.
Patel, A.D. 2005. “The Relationship of Music to the Melody of Speech and Syntactic Processing Disorders in Aphasia”. *Annals NY Academy of Sciences* 1060, 59-70.
Pearce, M. and M. Rohrmeier. 2018. “Musical Syntax II: Empirical Perspectives”. In: Bader, R (ed.). *Springer Handbook of Systematic Musicology*. Germany: Springer-Verlag. 487-505.
Precht, K. 2003. “Stance moods in spoken English: Evidentiality and affect in British and American conversation”. *Text* 23(2), 239-257.
Rumsey, L. 2015. “Da-DA-da-da-da: intonation and poetic form”. *Thinking Verse* 5, 15-49.
Reiterer, S.M., Schneider, P., Seither-Preisler, A. and S. Turker. 2017. “When Music Speaks: Auditory Cortex Morphology as a Neuroanatomical Marker of Language Aptitude and Musicality”. *Frontiers in Psychology* 8, 1-17.
Ten Cate, C. and M. Spierings. 2018. “Rules, rhythm, and grouping: auditory pattern

perception by birds". *Animal Behaviour* 30, 1-9.
Tillmann, B. 2014. "Pitch processing in music and speech". *Acoustics Australia* 42(2), 124-130.
Wiggins, G. 1998. "Music, Syntax, and the Meaning of 'Meaning'". *Proceedings of the First Symposium on Music and Computers*, 18-23.
Wilson, D. 2013. "Irony comprehension: A developmental perspective". *Journal of Pragmatics* 59, 40-56.

7 Abstract

7.1 English abstract

Acoustic signals act as transmitters of affect in music as well as speech. In this diploma thesis the ways in which units similar to musical sounds may structure and coordinate the production of speech melodies were investigated. The empirical research conducted comprised two stages. The first consisted of collecting data and quantitatively evaluating this data by referring it to prior findings. The results of this evaluation served as a selective filter for choosing the most competently performing subjects for further qualitative examination. The collected data consisted of the recordings of ironic statements and their non-ironic equivalents produced by twenty native speakers of German and twenty native speakers of English. The speakers were split into two groups per language. The subjects were provided with pieces of text which included six utterances that the subjects were to perform as well as the corresponding situations contextualizing the utterances. The first groups were not informed about half of the utterances contradicting the given contextual information, whereas the second groups were explicitly informed about it. The recordings were analyzed using "prosogram", a plugin for the speech analysis software Praat, in order to be able to compare the prosodic contours in musicological units, namely, semitones. The intervallic values forming the pitch contours calculated by the prosogram were compared to musical intervals regarding their valence and affective functions. The tendencies that were found comprised the use of musical intervals of positive valence in prominent syllables to reinforce the literal meaning of an utterance's verbal content and the use of musical intervals of negative valence in prominent syllables to signal some form of nonliteral meaning of an utterance's verbal content. The findings of this thesis highlight the entanglements of music and speech as primary forms of human communication that can be physically measured and may therefore provide valuable subject matter for further investigations into the musical properties of intonation.

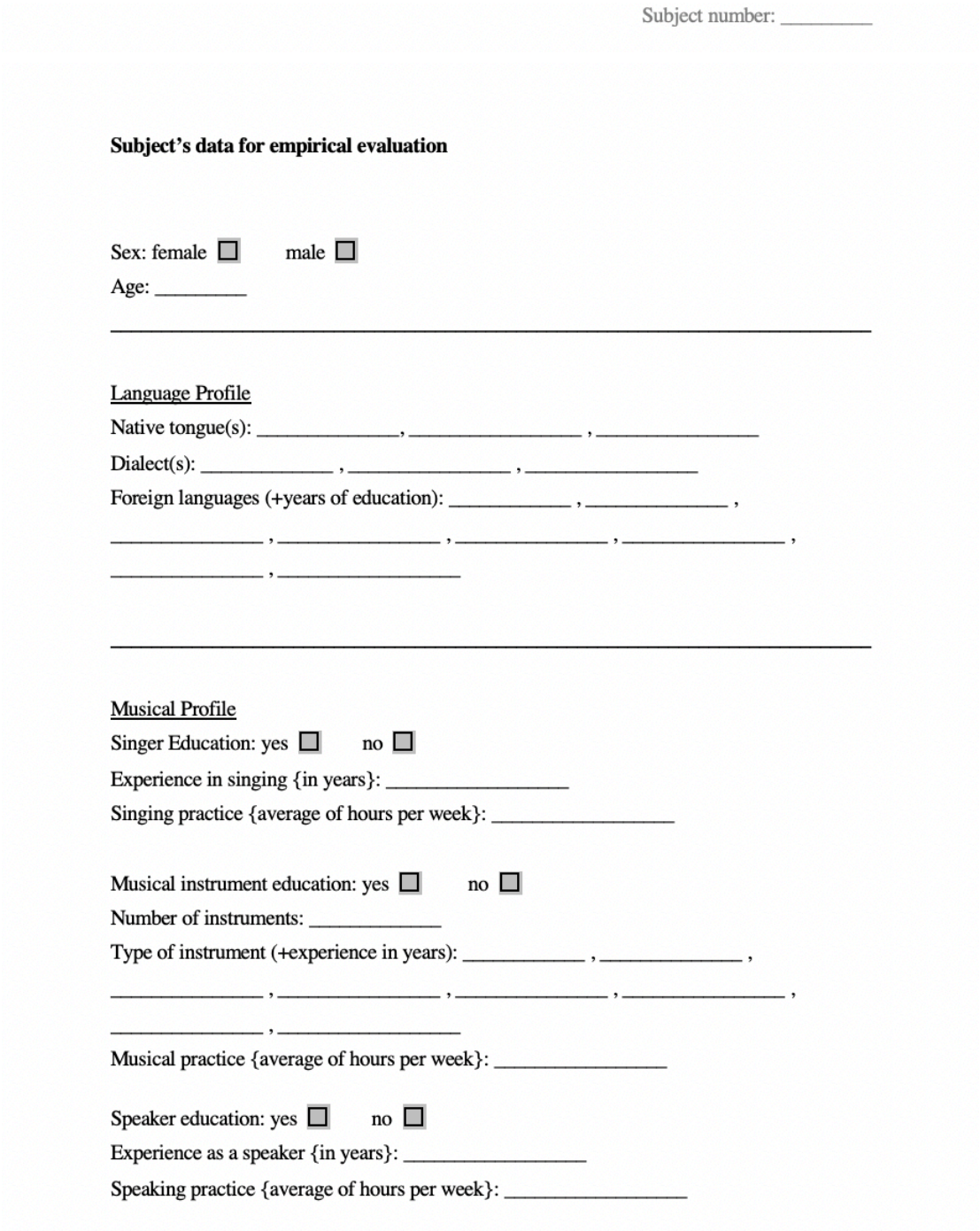
7.2 German abstract

Akustische Signale fungieren als Transmitter von Emotion in Musik wie auch gesprochener Sprache. Die Art und Weise mit welcher eine, einem musikalischen Klangbild ähnelnde, Intonationseinheit die Strukturierung und Koordination in der Sprachproduktion beeinflusst, war Gegenstand dieser Diplomarbeit. Das durchgeführte empirische Verfahren umfasste zwei unterschiedliche Phasen. Die erste bestand aus dem Sammeln von Daten und der quantitativen Auswertung dieser Daten in Bezug auf Messwerte früherer Recherchen. Die Resultate dieser Auswertung dienten als selektiver Filter, um für die qualitative Analyse qualifizierte Probanden zu ermitteln. Die gesammelten Daten enthielten Aufnahmen von ironischen Aussagen und Aussagen mit demselben wörtlichen Inhalt, jedoch ohne jegliche ironische Wertigkeit. Die Aufnahmen wurden ausgeführt von jeweils zwanzig deutsch-und englischsprachigen Personen, welche die kontextualisierte Sprache als ihre Muttersprache beherrschten. Die Teilnehmer wurden innerhalb ihrer Sprachzugehörigkeit erneut in zwei verschiedene Gruppen eingeteilt. Den Probanden wurden kurze Texte vorgelegt, welche sechs verschiedene Situation beinhalteten, die die Hintergründe der dazugehörenden Aussagen darstellten.

Die jeweils ersten Gruppen wurden im Gegensatz zu den zweiten Gruppen, nicht darüber informiert, dass die Hälfte der Aussagen der gegebenen kontextuellen Informationen widersprach. Die Aufnahmen wurden mit einem Plug-in, dem „Prosogram“, für das phonetische Analyseprogramm Praat ausgewertet, um die Intonationsprofile mit musikologischen Einheiten, den Halbtonwerten, beschreiben zu können. Die errechneten Intervallwerte der Intonationsprofile wurden in Bezug auf ihre affektiven Funktionen und Wertigkeiten untersucht. Die Ergebnisse dieser Untersuchung zeigten eine Tendenz zur Verwendung von musikalischen Intervallen von positiver konnotierter Wertigkeit in markanten Silben zur Bekräftigung der wortwörtlichen Bedeutung des verbalen Inhalts einer Aussage. Des weiteren konnte auch die Verwendung von musikalischen Intervallen von negativer affektiver Wertigkeit in markanten Silben zur Kennzeichnung von implizierten Bedeutungen einer Aussage festgestellt werden. Die Ergebnisse dieser Diplomarbeit heben physikalisch messbare Verstrickungen von Musik und Sprache als ursprüngliche menschliche Kommunikationsform hervor und ebnen damit den Weg für weitere Analysen auf diesem Gebiet.

8 Appendix

8.1 Data sheet for English native speakers



Subject number: _____

Subject's data for empirical evaluation

Sex: female ☐ male ☐

Age: _____

Language Profile

Native tongue(s): _____, _____ , _____

Dialect(s): _____ , _____ , _____

Foreign languages (+years of education): _____ , _____ ,

_____ , _____ , _____ , _____ ,

_____ , _____

Musical Profile

Singer Education: yes ☐ no ☐

Experience in singing {in years}: _____

Singing practice {average of hours per week}: _____

Musical instrument education: yes ☐ no ☐

Number of instruments: _____

Type of instrument (+experience in years): _____ , _____ ,

_____ , _____ , _____ , _____ ,

_____ , _____

Musical practice {average of hours per week}: _____

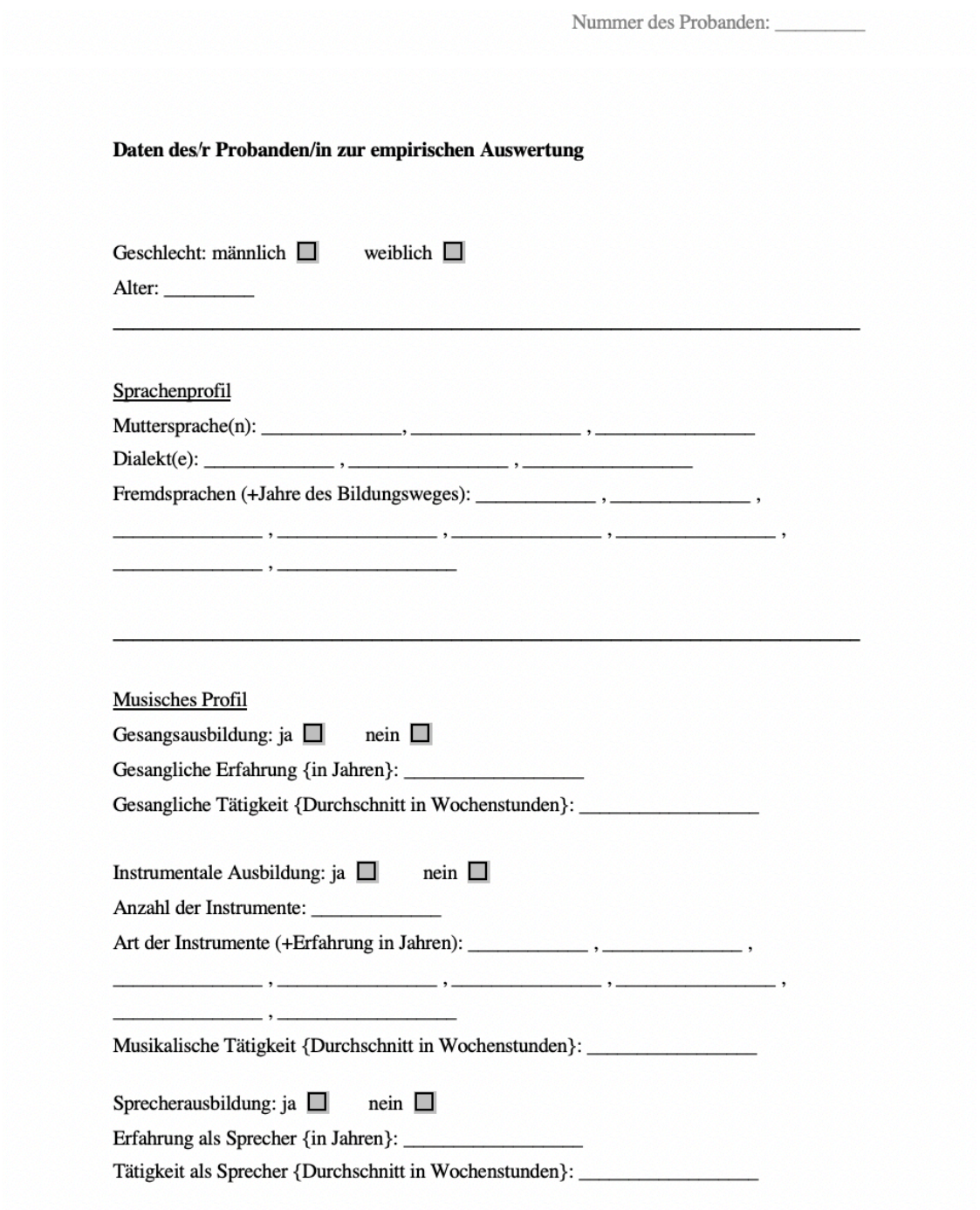
Speaker education: yes ☐ no ☐

Experience as a speaker {in years}: _____

Speaking practice {average of hours per week}: _____

Figure 22. Data sheet for English native speakers

8.2 Data sheet for German native speakers



Nummer des Probanden: _____

Daten des/r Probanden/in zur empirischen Auswertung

Geschlecht: männlich ☐ weiblich ☐

Alter: _____

Sprachenprofil

Muttersprache(n): _____, _____ , _____

Dialekt(e): _____ , _____ , _____

Fremdsprachen (+Jahre des Bildungsweges): _____ , _____ ,

_____ , _____ , _____ , _____ ,

_____ , _____

Musisches Profil

Gesangsausbildung: ja ☐ nein ☐

Gesangliche Erfahrung {in Jahren}: _____

Gesangliche Tätigkeit {Durchschnitt in Wochenstunden}: _____

Instrumentale Ausbildung: ja ☐ nein ☐

Anzahl der Instrumente: _____

Art der Instrumente (+Erfahrung in Jahren): _____ , _____ ,

_____ , _____ , _____ , _____ ,

_____ , _____

Musikalische Tätigkeit {Durchschnitt in Wochenstunden}: _____

Sprecherausbildung: ja ☐ nein ☐

Erfahrung als Sprecher {in Jahren}: _____

Tätigkeit als Sprecher {Durchschnitt in Wochenstunden}: _____

Figure 23. Data sheet for German native speakers

8.3 Situation 1a and 1b

Sprecherrolle: Vater/Mutter
Ort: im Haus/ in der Wohnung der/s Sprechers/in
Empfänger der Sprecherbotschaft: der Sohn des/r Sprechers/in

Kontext: Ihr Sohn hat seit einiger Zeit Schwierigkeiten, gute Ergebnisse bei Mathematik Schularbeiten zu erzielen. Von der Schule nach Hause kommend, erzählt er Ihnen, dass er gerade einer freiwilligen Mathematik Lerngruppe beigetreten ist, um sich zu verbessern. Sie reagieren auf diese Botschaft mit der folgenden Aussage:

Botschaft: **"Was für eine schlaue Entscheidung!"**

Sprecherrolle: Vater/Mutter
Ort: im Haus/ in der Wohnung der/s Sprechers/in
Empfänger der Sprecherbotschaft: der Sohn des/r Sprechers/in

Kontext: Ihr Sohn hat zum wiederholten Mal ein Nicht Genügend auf seine Mathematik Schularbeit erhalten. Sein Lehrer bot ihm an, eine mündliche Kompensationsprüfung zu absolvieren, um das Schuljahr noch positiv abschliessen zu können. Soeben haben sie von diesem Lehrer einen Anruf erhalten, in welchem er Sie darüber informierte, dass Ihr Sohn nicht zu dem verabredeten Termin für diese Prüfung erschienen war. Als sie von der Arbeit nach Hause zurückkommen und Ihren Sohn erblicken, kommentieren sie diesen Vorfall umgehend mit der folgenden Bemerkung in seine Richtung:

Botschaft: **"Was für eine schlaue Entscheidung!"**

8.4 Situation 2a and 2b

Sprecherrolle: Zuseher/in bei einem Basketballspiel
Ort: Zusehertribüne einer Sporthalle

Empfänger der Sprecherbotschaft: ein/e Freund/in der/die den/die Sprecher/in begleitet

Kontext: Sie sind ein begeisterter Unterstützer Ihrer lokal ansässigen Profibasketballmanschaft. Während eines Ligaspiels dem Sie beiwohnen, sehen Sie Ihre Mannschaft in Punkten zurückliegen. Nachdem der Kapitän der Mannschaft ausgewechselt worden ist, seine Leistung nicht den Erwartungen entsprechend, schafft es die Mannschaft, sich zurückzukämpfen und das Spiel letztendlich für sich zu entscheiden. Sie machen den folgenden Kommentar über den Spieler der für den Kapitän eingewechselt wurde und die meißten Punkte im Spiel erzielte in Richtung Ihrer/s Freundes/in:

Botschaft: **"So muss man eine Mannschaft führen!"**

Sprecherrolle: Zuseher/in bei einem Basketballspiel
Ort: Zusehertribüne einer Sporthalle
Empfänger der Sprecherbotschaft: ein/e Freund/in der/die den/die Sprecher/in begleitet

Kontext: Sie sind ein begeisterter Unterstützer Ihrer lokal ansässigen Profibasketballmanschaft. Während eines Ligaspiels dem Sie beiwohnen, sehen Sie Ihre Mannschaft in Punkten zurückliegen. Nachdem der Kapitän der Mannschaft zum sechsten Mal in Folge ein Wurfversuch misslingt, äußern Sie den folgenden Kommentar über diesen Spieler in Richtung Ihrer/s Freundes/in.

Botschaft: **"So muss man eine Mannschaft führen!"**

8.5 Situation 3a and 3b

Sprecherrolle: Oberstufenlehrer/in
Ort: in Ihrem Klassenzimmer
Empfänger der Sprecherbotschaft: ein bestimmter Schüler

Kontext: In Ihrer Klasse gibt es einen schüchternen Schüler namens Max. Er hat das ganze Semester über kaum aktiv am Unterricht teilgenommen. Aufgrund eines Elterngesprächs wissen Sie, dass er sehr sich sehr für Geographie interessiert und den Namen jeder

europäischen Hauptstadt kennt. Um ihn mehr in den Unterricht zu involvieren, ohne die Gefahr ihn in eine unangenehme Situation zu bringen, fragen Sie die ganze Klasse nach dem Namen der Hauptstadt Albaniens. Da niemand aufgezeigt hat, machen sie folgende Aussage:

Botschaft: **"Max, du kannst diese Frage sicherlich beantworten!"**

Sprecherrolle: Oberstufenlehrer/in
Ort: in Ihrem Klassenzimmer
Empfänger der Sprecherbotschaft: ein bestimmter Schüler

Kontext: Ein Schüler mit dem Namen "Max" schenkt Ihrem eröffnenden Monolog zu einer Ihrer Regelstunden in Ihrem Klassenzimmer keine Aufmerksamkeit. Stattdessen unterhält er sich mit einem Schüler der neben ihm sitzt. Diesen Schüler ansprechend machen sie die folgende Bemerkung:

Botschaft: **"Max, du kannst diese Frage sicherlich beantworten!"**

8.6 Situation 4a and 4b

Sprecherrolle: Pendler/in
Ort: auf einer Autobahn
Empfänger der Sprecherbotschaft: Beifahrer/in

Kontext: Es ist Montagmorgen und Sie stecken im Stoßverkehr am Weg zur Arbeit fest. Glücklicherweise sind Sie an diesem Tag früher losgefahren als gewohnt und werden trotz der schlechten Verkehrssituation nicht zu spät kommen. Ihr/e Kollege/in sitzt im Beifahrersitz und macht eine Bemerkung über die schlechte Verkehrssituation in der Stadt, worauf Sie mit der folgenden Aussage antworten:

Botschaft: **"Wir werden uns bestimmt nicht verspäten!"**

Sprecherrolle: Pendler/in

Ort: auf einer Autobahn
Empfänger der Sprecherbotschaft: Beifahrer/in

Kontext: Es ist Montagmorgen und Sie stecken im Stoßverkehr am Weg zur Arbeit fest. Sie erkennen, dass Sie sich zu einem wichtigen Treffen verspäten werden. Ihr/e Kollege/in sitzt im Beifahrersitz die schlechte Verkehrssituation in der Stadt kommentierend, worauf sie mit der folgenden Aussage antworten:

Botschaft: **"Wir werden uns bestimmt nicht verspäten!"**

8.7 Situation 5a and 5b

Sprecherrolle: Restaurantgast
Ort: in einem Restaurant
Empfänger der Sprecherbotschaft: ein/e Freund/in der/die den/die Sprecher/in begleitet

Kontext: Sie sind zu einem Abendessen von einem/r Ihnen nahestehendem/n Freund/in eingeladen worden. Diese/r bietet Ihnen an, sich zwei Hauptspeisen zu teilen, um die kulinarische Erfahrung auszuweiten. Ihre/Seine Vorschläge für diese zwei Hauptspeisen beinhalten unter anderem Schweinekotelette, welches Ihr Lieblingsgericht von Kindheit an ist. Nachdem Ihr/e Freund/in Sie nach Ihrer Empfindung über diese Speisewahl fragt, antworten Sie mit der folgenden Aussage:

Botschaft: **"Ich liebe Schweinekotelette!"**

Sprecherrolle: Vegetarier/in
Ort: in einem Restaurant
Empfänger der Sprecherbotschaft: ein/e Freund/in der/die den/die Sprecher/in begleitet

Kontext: Sie sind zu einem Abendessen von einem/r Ihnen nahestehendem/n Freund/in eingeladen worden. Diese/r bietet Ihnen an, sich zwei Hauptspeisen zu teilen, um die

kulinarische Erfahrung auszuweiten. Da der/die Einladende eine/r Ihrer engsten Freunde/innen darstellt, sollte ihm/ihr bewusst sein, dass Sie seit vielen Jahren als Vegetarier/in leben. Nichtsdestotrotz, beinhalten ihre/seine Vorschläge für diese zwei Hauptspeisen unter anderem Schweinekotelette. Nachdem Ihr/e Freund/in Sie nach Ihrer Empfindung über diese Speisewahl fragt, antworten Sie mit der folgenden Aussage:

Botschaft: **"Ich liebe Schweinekotelette!"**

8.8 Situation 6a and 6b

Sprecherrolle: Kellner/in
Ort: in einem Restaurant in welchem Sie arbeiten
Empfänger der Sprecherbotschaft: Mitarbeiter/in

Kontext:
In Ihrem Beruf als Kellner/in ist es Ihre Aufgabe, all die benutzten Gläser von den leeren Tischen aufzusammeln und diese zur Küche zurückzubringen. Sie bewegen sich gerade zurück zur Küche mit einem vollen Tablett Gläser, wobei Sie nur eine Hand benutzen um diese zu halten, da Sie mit Ihrer anderen Hand Ihr Handy bedienen, um eine private Nachricht zu verfassen. Ein Mitarbeiter kommt aus der Küche, vier Schüsseln Suppe tragend. Sie bemerken ihn nicht, den Blick auf das Handy gerichtet, und stoßen mit ihm zusammen. Folglich fallen dem Mitarbeiter alle Schüsseln aus der Hand, welche dann am Boden zerspringen. Der Mitarbeiter betrachtet den Scherbenhaufen verzweifelt, da es bloß seine erste Woche als Angestellter ist, und ihm ein ähnliches Missgeschick schon passiert ist. Der Besitzer des Restaurants eilt auf ihn zu, doch bevor dieser etwas sagen kann, machen sie die folgende Aussage:

Botschaft: **"Ich schätze, das ist meine Schuld!"**

Sprecherrolle: Kellner/in
Ort: in einem Restaurant in welchem Sie arbeiten
Empfänger der Sprecherbotschaft: Mitarbeiter/in

Kontext: In Ihrem Beruf als Kellner/in ist es Ihre Aufgabe, all die benutzten Gläser von den leeren Tischen aufzusammeln und diese zur Küche zurückzubringen. Sie bewegen sich gerade zurück zur Küche mit einem vollen Tablett Gläser, wobei Sie beide Hände benutzen und langsam schreiten, um diese schwere Last sicher zurückzubringen. Plötzlich eilt ein Mitarbeiter auf seinem Handy tippend aus der Küche und stößt mit Ihnen zusammen, so dass Sie das Tablett aus den Händen verlieren. Alle Gläser fallen zu Boden und zerspringen. Von dem Scherbenhaufen aufblickend, starrt Sie der Mitarbeiter zornig an, bevor dieser jedoch etwas sagen kann, machen Sie die folgende Bemerkung:

Message: **“Ich schätze, das ist meine Schuld!”**