

Original Paper

Phonetica 2012;69:7–27
DOI: [10.1159/000343171](https://doi.org/10.1159/000343171)

Received: June 3, 2012
Accepted: September 2, 2012

At the Edge of Intonation: The Interplay of Utterance-Final F0 Movements and Voiceless Fricative Sounds

Oliver Niebuhr

Institut für Skandinavistik, Frisistik und Allgemeine Sprachwissenschaft (ISFAS),
Christian-Albrechts-Universität zu Kiel, Kiel, Germany

Abstract

The paper is concerned with the ‘edge of intonation’ in a twofold sense. It focuses on utterance-final F0 movements and crosses the traditional segment-prosody divide by investigating the interplay of F0 and voiceless fricatives in speech production. An experiment was performed for German with four types of voiceless fricatives: /f/, /s/, /ʃ/ and /x/. They were elicited with scripted dialogues in the contexts of terminal falling statement and high rising question intonations. Acoustic analyses show that fricatives concluding the high rising question intonations had higher mean centres of gravity (CoGs), larger CoG ranges and higher noise energy levels than fricatives concluding the terminal falling statement intonations. The different spectral-energy patterns are suitable to induce percepts of a high ‘aperiodic pitch’ at the end of the questions and of a low ‘aperiodic pitch’ at the end of the statements. The results are discussed with regard to the possible existence of ‘segmental intonation’ and its implication for F0 truncation and the segment-prosody dichotomy, in which segments are the alleged troublemakers for the production and perception of intonation.

Copyright © 2012 S. Karger AG, Basel

1. Introduction

1.1. Background

Perceived pitch is the essence of speech melody. Different kinds of pitch movements as well as their timings relative to accented syllables or phrase boundaries convey communicative functions that refer to the associated words, the dialogue partner, or the dialogue itself (note that the term ‘function’ is meant in a broader sense that includes elements of discourse- and information-structure as well as attitudinal and other types of meanings, as there is no need to distinguish between them in the present paper). The phonological structuring and modelling of speech melody focuses on the

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2012 S. Karger AG, Basel
0031–8388/12/0692–0007
\$38.00/0
Accessible online at:
www.karger.com/pho

Oliver Niebuhr
Freyastrasse 10
DE–24939 Flensburg (Germany)
E-Mail niebuhr@isfas.uni-kiel.de

main acoustic counterpart of pitch, the fundamental frequency (F0). This holds true for the Autosegmental-Metrical Model [Pierrehumbert, 1980; Grice and Baumann, 2002; Ladd, 2008] in the same way as for the more hearer- and meaning-oriented IPO Model and Kiel Intonation Model [t Hart et al., 1990; Kohler, 1991, 2006].

From the point of view of such models the segmental string thus appears to be a mere troublemaker for the signalling of melodic functions. Voiceless sound segments interrupt F0, and biomechanical interactions between tongue, jaw, and larynx, as well as changes in the supraglottal air pressure due to vocal-tract constrictions, create so-called microprosodic perturbations in the macroprosodic F0 course intended by the speaker. Bolinger [1972, p. 19] characterized the segmental contribution to speech melody as ‘accidental changes in pitch, the irrelevant quavers’. The conclusion of Kohler [1990] is not as pessimistic as that of Bolinger, as Kohler shows the microprosodic F0 perturbations to be systematic rather than accidental. But, precisely because they are systematic, he maintains the troublemaker perspective and stresses that ‘if these micro F0 differences are ignored [...] the macro intonation pattern changes, the linguistic identity of segments changes, [...] or there is no perceivable change at all’ [Kohler, 1990, p. 73].

So, do we have to find a way to normalize or neutralize the segmental string in order to analyse speech melodies properly? Regarding the habitual efforts of intonational studies to avoid phonetically heterogeneous sound sequences and unvoiced consonants in particular, the answer to this question seems to be ‘yes’. However, not just because everyday conversation typically does contain a considerable amount of segmental heterogeneity and voicelessness, one may wonder whether the question is at all appropriate. By separating intonations and speech sounds along the traditional segment-prosody divide, and by reducing on this basis the role of sound segments in intonation to their possibly systematic, but still troublesome effects on F0, we disregard an important fact: sound segments themselves – or rather their intrinsic and to some extent variable spectral-energy patterns – induce pitch impressions. Simply speaking, the more of the overall acoustic energy of a sound segment is contained in high spectral frequencies, the higher is the pitch impression conveyed by that sound.

In the cases of vowels and diphthongs this phenomenon is known as intrinsic pitch (as opposed to intrinsic F0). Everything else being equal, open vowels like [a] are perceived to be higher pitched than close vowels like [i] and [u] [Chuang and Wang, 1978; Stoll, 1984]. Correspondingly, opening and closing diphthongs with flat F0 can trigger the impression of rising or falling speech melodies [Niebuhr, 2004]. At the opposite end of the range of speech sounds, the pitch impressions caused by obstruents are not based on voice but on noise, more precisely on aperiodic noise. For example, the ability of fricatives to convey pitch impressions is stressed in the following statement of Johnson [2012, p. 157]: ‘When you make a series of fricatives starting from a pharyngeal constriction and moving the constriction forward to the alveolar ridge, you may be able to hear a change in the ‘pitch’ of the fricative’. However, except for some psychoacoustic studies on sibilant pitch [Traunmüller, 1987], aperiodic pitch impressions at the segmental level and their role in speech communication have not been investigated so far.

Instead, research focussed on ‘whisper pitch’, i.e. on global aperiodic phonation at the utterance level. The findings on whisper pitch suggest that aperiodic noise in speech can be brought under conscious control by speakers and is varied in terms of its spectral-energy distribution in such a way that the aperiodic noise patterns compensate to a certain degree for the absence of F0. Most studies investigated the comprehension of lexical tones in languages like Thai and Mandarin. Using very different types

of stimuli that ranged from artificially devoiced words to utterances produced with various types of natural whisper phonation, the studies showed that native listeners identify particularly tones in longer stimuli significantly above chance, and in some cases their performance is indistinguishable from that for normally phonated stimuli [Meyer-Eppler, 1957; Kloster Jensen, 1958; Thomas, 1969; Abramson, 1972; Whalen and Xu, 1992; Higashikawa and Minifie, 1999; Krull, 2001; Nicholson and Teig, 2003; Liu and Samuel, 2004; Konno et al., 2006; Li and Guo, 2012].

A similar conclusion was recently drawn for the intonation language Dutch with regard to the identification of whispered and normally phonated question and statement utterances, which were only distinguishable by the direction of the utterance-final pitch movement [Heeren and van Heuven, 2009]. Compared with the normally phonated questions and statements, listeners were not considerably worse in identifying the sentence mode in the whispered utterances. Acoustic analyses suggested that it was again the spectral-energy distribution in the aperiodic noise that guided the sentence-mode identification in whisper.

Besides the energy distribution across the noise spectrum, another acoustic parameter that is used across languages to signal tones and intonations in unvoiced speech is the overall acoustic energy level of the noise. For example, Whalen and Xu [1992] and Heeren and van Heuven [2009] found that whispered sound sections had a higher acoustic energy level when they were to convey high or rising tones or intonations. These findings are consistent with results of psychoacoustic studies. Houtsma [1995] summarizes a long history of perception experiments starting with Stevens [1935]. They showed within a certain degree of variability (including, among others, speaker-specific effects) that for frequencies ‘above 2000 Hz the pitch tends to rise with increasing intensity’ [Houtsma, 1995, p. 269]. Precisely these frequencies play a major role in the speech signal and particularly in its aperiodic noises.

In some studies like Nicholson and Teig [2003] or Liu and Samuel [2004] the successful transmission of melodic functions through unvoiced speech was additionally due to durational cues. However, unlike spectral-energy distributions and levels, durational patterns do not directly trigger pitch impressions. That duration patterns support the perception of melodic functions is thus an indirect effect based on language experience. For example, more complex F0 patterns are typically linked with longer durations of the corresponding vowels and/or syllables. Having learned this link, listeners are able to infer certain tone or intonation patterns from relative durational patterns in unvoiced speech stimuli.

1.2. Question and Rationale

While aperiodic pitch is obviously exploited by speakers in unvoiced utterances, it is still an open question whether variation in aperiodic noises can also be found (a) at the level of individual segments and (b) in normally phonated utterances, i.e. when the aperiodic noise is surrounded by F0. More specifically, the questions asked are whether the spectral energy characteristics of voiceless fricatives show systematic variation in the contexts of phonologically contrastive melodic elements, and if so, whether this variation can contribute to the perception of speech melody.

These related questions were the impetus for a series of production and perception experiments conducted by Niebuhr [2008, 2009, 2011] and Niebuhr et al. [2011b]. The

experiments aim at bringing the traditionally separated melodies and sound segments closer together, on this basis arguing against the view of (voiceless) segments as mere troublemakers for the elements of the speech melody. The present paper summarizes and extends the production part of this line of research.

The summarized research is phenomenologically oriented. That is, it is concerned with the existence of systematic variation in the aperiodic noise sounds of speech and outlines possible implications for speech perception. It goes beyond the scope of this paper to explain the variation in terms of, for example, intrinsic F0-induced air-stream or resonance conditions, or the speaker's intention to support the function or pitch shape of a melodic element. The search for explanations must be the second step, after possible acoustic and perceptual phenomena have been sufficiently described.

If the spectral-energy patterns of voiceless fricatives vary systematically depending on the F0 context, then voiceless fricatives at the ends of nuclear pitch accents with high rising or terminal falling utterance-final movements in German are a promising framework to start looking for this variation.

- Pitch accents with high rising and terminal falling movements provide very different F0 contexts. Similar accent contours also occur utterance-internally in prenuclear position. Therefore, in order to stress that the present study focuses on contours at the right edge of an utterance, they are referred to as rising and falling edge contours.
- The communicative functions of rising and falling edge contours are quite well understood in German [Kohler, 2004, 2005, 2006; Niebuhr, 2007; Niebuhr et al., 2010; Niebuhr, 2012], which facilitates their reliable, context-based elicitation.
- It is well known that rising and falling edge contours are truncated under time pressure in German [Gartenberg and Panzlaff-Reuter, 1991; Grabe, 1998; Ohl and Pfitzinger, 2009]. Figure 1 exemplifies truncation as it applies to falling edge contours in utterance-final accented syllables with phonologically short vowels and final obstruents. Under these circumstances, a rising-falling pitch-accent has its terminal F0 fall curtailed and can resemble a high ending pitch-accent rise. Due to the process of F0 truncation, aperiodic pitch may take the same (compensatory) role as in whispered speech. The framework of the present study is thus suitable to foster spectral variation amongst utterance-final fricatives.

Grabe [1998] described the truncation of German edge-contour F0 movements only in the context of utterance-final /f/. Ohl and Pfitzinger [2009] repeated Grabe's [1998] study, but with two different fricatives: /f/ and /ʃ/. It turned out that the degrees of truncation and hence the F0 ranges of the utterance-final edge contours were much more variable than claimed by Grabe [1998]. Crucially, how much of the edge-contour F0 movement was truncated depended on the type of final fricative. There was significantly less F0 truncation before /f/ than before /ʃ/, particularly for terminal falling edge contours.

Provided that the spectral-energy patterns of final fricatives are indeed different in rising and falling edge-contour contexts, it seems possible that the different degrees of truncation in the study by Ohl and Pfitzinger [2009] reflect an interaction between the F0 movement and the final fricatives. If /ʃ/ is a more variable and potent carrier of aperiodic pitch than /f/, then the functional load of F0 in signalling the communicative functions of the edge contours is lower before /ʃ/ than before /f/. In consequence, there

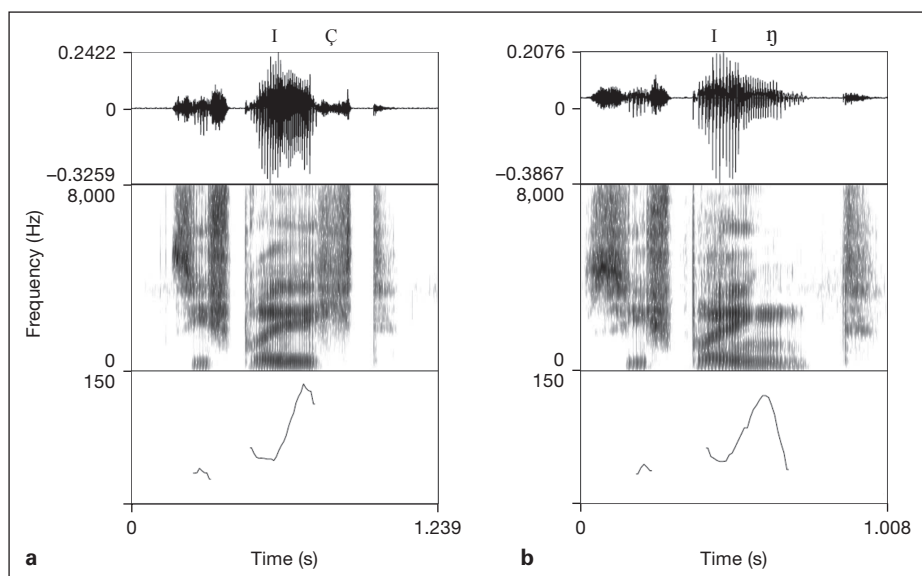


Fig. 1. Utterances *Sie spricht* [ʒi.ʃʁɪçtʰ], ‘she speaks’ **(a)** and *Sie springt* [ʒi.ʃʁɪŋkʰtʰ], ‘she jumps’ **(b)**, produced by the male speaker K.K. with the same edge-contour intonations. They consist of a late pitch accent with an utterance-final terminal fall. The latter is clearly pronounced in the monosyllable *springt*, but strongly truncated in the monosyllable *spricht* [from Gartenberg and Panzlaff-Reuter, 1991]. Audio illustration in audio-fig1.wav provided as online supplement at www.karger.com/doi/10.1159/000343171.

can be more F0 truncation before /j/ than before /f/. The fact that the present production experiment used a set of four phonetically very disparate fricatives allowed taking the possibility of such interactions into account.

1.3. Outline of Variables

The present study focuses on the direct triggers of pitch impressions in fricatives and is hence concerned with their acoustic-energy levels and distributions. Durational measurements were neglected at this stage (which does not mean of course that durational patterns are irrelevant).

As for the spectral-energy distributions, it has been shown that differences in the place of articulation of fricatives can be systematically represented in terms of the spectral centre of gravity (CoG, also known as centroid or first spectral moment), with lower CoG values for back and higher CoG values for front fricatives [Jassem, 1979; Forrest et al., 1988; Jongman et al., 2000; Gordon et al., 2002; Li et al., 2007; Niebuhr et al., 2011a; inter alia]. Moreover, the CoG measure has proven to be sensitive enough

- to represent smaller cross-linguistic differences within the same broad place and manner of segmental articulation (cf. Hamann and Sennema [2005] for /f/ in German and Dutch),

- to mirror the speaker-specific patterns of ‘alveolar’-to-‘postalveolar’ sibilant assimilations in French and English [Niebuhr et al., 2011a], as well as the dialect-specific degrees of retroflex fricative articulations in Mandarin [Chang, 2011], and
- to capture phonologically distinctive differences in the rounding of fricatives, as in Gordon et al. [2002].

It was reasonable to assume in view of the previous findings above that the CoG measure is also capable of revealing variations in spectral-energy distribution that result from different edge-contour contexts. With additional regard to the studies by Meyer-Eppler [1957], Kloster Jensen [1958], Traunmüller [1987], Higashikawa and Minifie [1999; *inter alia*] it is furthermore reasonable to predict that changes in CoG due to edge contours are positively correlated with changes in the perceived pitch of the fricatives. That is, a higher aperiodic pitch impression is reflected in a higher CoG measurement.

2. Experiment

2.1. Method

2.1.1. Target Sounds

The voiceless fricatives that served as target sounds covered all phonological places of articulation that can occur in syllable-final position in German: /f/, /s/, /ʃ/ and /x/. The labiodental fricative /f/ can hardly vary in its articulation (although the tongue is free to move for /f/, these secondary articulations have only a minor impact on the aperiodic noise – and hence on the aperiodic pitch, cf. below). The aperiodic noise generated by /f/ has a relatively low acoustic energy. In contrast, /s/ and /ʃ/ are high-energy fricatives that are less constrained in their places of articulation and produced with different tongue shapes, see Narayanan et al. [1995] for English and Mooshammer et al. [2007] for German. In relation to the typically narrow-grooved shape of the apical part of the tongue for /s/, /ʃ/ is based on the wider-grooved shape that concerns the laminal part of the tongue. This gives /ʃ/ additional freedom to vary its place of articulation while maintaining the characteristic sibilant quality. Moreover, the production of /ʃ/ in Standard Northern German typically involves lip rounding [Russ, 2010] and thus another dimension along which allophonic variation takes place. The aperiodic noise of /x/ has a medium acoustic-energy level and varies allophonically between [ç] and [χ] [Kohler, 1995]. Similar to /ʃ/, adding some degree of lip rounding also characterizes the realizations of /x/, particularly those of [χ], which occur after back vowels and are the target fricatives of /x/ in the present study.

The aperiodic pitch impression conveyed by fricative sounds is primarily determined by the length of the front cavity of the mouth between the noise source and the lips [Heinz and Stevens, 1961; Johnson, 2012]. It follows that variation in place of articulation and lip rounding will have the largest effects on the aperiodic pitch of a fricative. Additionally, higher energy levels of aperiodic noise sounds like fricatives in speech are fairly consistently correlated with higher aperiodic pitch impressions. Variation in the energy level of fricatives also involves lip rounding, since the concomitant protrusion of the lips directly affects the radiation surface of the mouth.

In summary, the intrinsic characteristics of /f/, /s/, /ʃ/ and /x/ cover a wide range of aperiodic noises and pitches. With regard to the possible interaction between the degree of F0 truncation and the capability of the following fricative to convey and vary aperiodic pitch (see section 1.2), it is further reasonable to assume that German /f/, /s/, /ʃ/ and /x/ are not equal. The variation in the quality of the aperiodic noise and hence the aperiodic pitch range should be greatest for /ʃ/ and smallest for /f/; /s/ and /x/ should take an intermediate position, with slight advantages for /x/ for which variation in lip rounding is part of its allophonic profile.

2.1.2. Target Words

Eight monosyllabic nouns with C(C)V_(s)C structures served as target words, with two words for each of the target fricatives. They are listed in table 1. All nouns are among the most frequent 10,000 words in (written) German (according to the ‘Wortschatz’ Project, University of Leipzig [Quasthoff,

Table 1. Target words used in the production experiment

Target fricative	Target word		
/f/	Schiff	‘ship’	/ʃɪf/
	Griff	‘grip’	/ɡʁɪf/
/s/	Kuss	‘handle’	/kʰʊs/
	Schluss	‘end’	/ʃlʊs/
/ʃ/	Tisch	‘table’	/tʰɪʃ/
	Fisch	‘fish’	/fɪʃ/
/x/	Loch	‘hole’	/lɔx/
	Koch	‘cook’	/kʰɔx/

1998]). The syllable nuclei were non-open phonologically short vowels. Short vowels were chosen in order to reduce the time that speakers had to realize rising and falling F0 movements before the voiceless fricatives at the ends of the target words. Excluding open vowels was to minimize variation in intrinsic vowel duration [Simpson, 1998; Möbius and van Santen, 1996].

2.1.3. Reading Material

The target words were placed at the ends of turn-final utterances. The utterances in turn were embedded in short dialogues on everyday topics (a lost book, shopping at the market, lunch appointments at work, furnishing a flat, etc.). The dialogues comprised about five to seven turns per speaker with three to five phrases in each turn. The target words occurred twice in different dialogues. However, as explained below, the two occurrences of each target word were associated with different edge-contour intonations.

The syntactic structures of the target-word utterances as well as the semantic-pragmatic context frames set by the preceding utterances were designed in such a way that they elicited fundamentally different types of edge contours – nuclear pitch accents with either high rising or terminal falling utterance-final movements – on the target words. Terminal falls were triggered by target-word utterances in declarative-syntax statements. Moreover, the target word was embedded contextually in such a way that it established a contrast to an implicit or explicit expectation of the speaker. This contrast is expressed by a steep rise before the fall onset [Kohler, 2005, 2006; Niebuhr et al., 2010]. Final rises were triggered by target-word utterances either in interrogative syntax or in syntactically elliptic intonation questions indexed by question marks. Their semantic-pragmatic contexts expressed incredulity and/or a great interest in receiving additional information from the dialogue partner, which leads to a high rising pitch pattern from low pitch on the final vowel [Kohler, 2004, 2005; Niebuhr, 2007, 2012; Niebuhr et al., 2010]. The number of dialogues, turns and utterances was made large enough to avoid speakers becoming aware of the set of target words. Table 2 illustrates the generated dialogue material with a representative excerpt that contains the underlined target words *Schluss*, *Loch*, *Fisch* and *Tisch* in either rising (single line) or falling (double line) edge-contour contexts.

2.1.4. Subjects and Recording Procedure

Eight male native speakers of Northern Standard German, aged between 25 and 35 years, were recruited to read the dialogues. They were paid for their participation. The speakers were experienced insofar as they had already participated in previous production experiments. Moreover, they were all well familiar with the sound-treated recording room and the experimenter (the author).

Prior to the actual recording, the subjects received the dialogue texts (with each dialogue on a separate sheet) in order to practice them for about 30 min. The turns of a given speaker were in bold, but the target utterances and words were not indicated. When the dialogues were handed out, subjects were asked to produce the dialogues in a spontaneous-sounding fashion with a colloquial speaking style. In order to facilitate this task, the subjects were explicitly allowed to modify the wordings of the sentences slightly according to their personal taste, for example, by adding, omitting, or

Table 2. Excerpt from a written dialogue and its English translation

German original		English translation	
B	Du wirst es nicht glauben; 'ch hab eben 'ne SMS von Peter und Sabine bekomm': Im Kino is noch immer nicht <u>Schluss</u> .	B	You won't believe it. I just got a text message from Peter and Sabine. Their movie isn't over yet.
A	Oh Mann, darf doch nich wahr sein. Ich bin am Verhungern. Also wenn die beiden nich bald hier sind, dann schwör' ich fall ich einfach über die Schokolade her.	A	Oh dear, this can't be true. I am starving. Well, if they are not here soon, I promise I will pounce on the chocolate.
B	Nanu, das überrascht mich jetzt doch'n bisschen. . . dabei haben wir erst vor 2 Stunden Kuchen und Kekse gehabt. Dein Magen hat ein <u>Loch</u> .	B	Well, that surprises me a little bit. . . we had cake and cookies just about two hours ago. Your stomach must have a hole.
A	Blödsinn. Das nennt sich «beschleunigter Stoffwechsel». Schon mal was von «Hypoglykämie» gehört?	A	Rot. This is called 'hypermetabolism'. Ever heard of 'hypoglycaemia'?
B	«Hypo. . .» - was? Was hast du gefressen? Ein Buch?	B	'Hypo. . . ' – what? What did you eat? A book?
A	Ha, ich begeb' mich doch nich auf Dein Niveau. Weißt du, seit den 90ern gibt's da son' Ding, das nennt sich «Internet». . .	A	Ha, I don't stoop to your level. You know, since the 90s there is a thing called 'the Internet'. . .
B	. . . Ja ja, schon gut; ich hab' verstanden. . .	B	Ok, ok; I got it. . .
A	. . . Sims doch den beiden mal zurück und frag', ob wir uns dann wenigstens gleich nach dem Film neben dem Kino bei «Antonio's» treffen wollen.	A	Why don't you text them back and ask, if we could meet at 'Antonio's' right after the movie.
B	Hä? Hab' ich mich verhöhrt? Du und <u>Fisch</u> ?	B	Eh? Are you kidding me? You and fish?
A	Was soll's. Geht eben schnell. In der Not frisst der Teufel Fliegen. . . und außerdem. . . man muss auch mal Kompromisse eingehen können.	A	So what? As long as it's quick. Any port in a storm. . . and also. . . sometimes you must be able to make compromises.
B	Lobenswert. Aber es ist halb acht durch. Glaubst Du ernsthaft, wir kriegen da jetzt noch 'nen <u>Tisch</u> ?	B	Laudable. But it's after half past eight. Do you honestly think that we still get a table there?

Target words are underlined. Single lines mark rising and double lines falling edge-contour contexts.

replacing words or expressions. As in previous experience with this strategy, the modifications concerned mainly utterance-medial function words and verbs, so no correction of omitted or replaced target words became necessary.

The actual recording sessions started when the subjects signalled to the experimenter that they had finished practicing. The dialogues were randomized across the subjects. Each dialogue was produced 4 times in succession, and the last two productions were used for the acoustic analysis. Unlike in the similar text-based dialogue recordings by, for instance, Kohler and Niebuhr [2007], Niebuhr [2007, 2010] and Niebuhr et al. [2010], only 1 subject at a time participated in the recording sessions. He produced the turns of speaker B. Speaker A was always the experimenter. This refinement was inspired by suggestions by Himmelman [2006] on speech elicitation methods and helped the subjects producing their dialogue turns in a spontaneous-sounding, colloquial way. The experimenter produced his turns with a clearly informal attitude and speaking style, which firstly created a relaxed atmosphere in the lab and secondly functioned as a behavioural template that the subjects adopted in the course of the first two dialogue repetitions.

The recordings took place in sound-treated rooms at the Zentrum für Allgemeine Sprachwissenschaft in Berlin and at the Seminar für Allgemeine Sprachwissenschaft in Kiel. The acoustic speech signals were recorded digitally on a laptop computer at a sampling rate of 48 kHz and 16-bit quantization. A high-fidelity Sennheiser microphone was used. Each recording session took between 30 and 45 min.

2.1.5. Acoustic Analyses

Since /f/, /s/, /ʃ/ and /x/ were each produced by 8 speakers 8 times across the repeated dialogues (2 target words × 2 sentence modes × 2 repetitions), the acoustic analyses comprised 64 tokens of each fricative, 32 after rising and 32 after falling edge contours, or 256 fricative tokens altogether. Before the acoustic analyses were performed, it was checked and confirmed for all speakers that the target utterances were produced with the correct types of edge contours.

The measurements for each fricative were done on a manual or semi-automatic basis in Praat [Boersma, 2001] and included: mean CoG in Hertz; CoG range, i.e. maximum minus minimum CoG value within the fricative, in Hertz; the mean acoustic energy level of the entire fricative in decibel; and F0 at the end of the rising or falling edge contour in Hertz.

The CoG measurements were based on spectral slices. The slices resulted from a 30-ms Hamming window (512 FFT points), which reached from 0.5 to 15.0 kHz and was shifted in 7-ms steps across the fricative. The first window was set with its left boundary 20 ms after the onset of the fricative, and the last was located such that its right boundary was 20 ms or less from fricative offset. The mean CoG was calculated across the CoG values of all individual spectra of a fricative. The CoG range was the difference between the highest and lowest CoG value within the fricative. So, each of the 256 fricative tokens was represented by one mean CoG and one CoG range. The CoG measurements were done semi-automatically using a Praat script.

The measurements of the mean acoustic energy level (in terms of root mean square, dB, for a given time interval) were made in Adobe Audition (<http://www.adobe.com/Audition>). They involved two steps. First, the peak level of the utterance was normalized by setting it to 90% of the amplitude of the short vowel that preceded the target fricative. Then, the mean acoustic energy level was measured within the same fricative boundaries that were also used for the CoG measurements. The fricative boundaries were determined by ear in terms of the audible onsets and offsets of oral friction, supported by visual inspection of the corresponding wideband spectrograms in Praat.

Finally, the F0 offsets of the rising and falling edge contours within the target words were also determined in Praat (sometimes after correcting for octave errors). The F0 offsets were the last periods in the target words that were judged to be ‘voiced’ by the F0 algorithm in Praat, based on its default settings (15 candidates, silence threshold = 0.03, voicing threshold = 0.45, voiced/unvoiced cost = 0.14). So vocal-fold vibrations that were visually identifiable in the waveform, but that fell below the minimum amplitude threshold or became too irregular relative to the preceding ones were excluded from F0 measurements. Since only male speakers were used in the production study, it was not necessary to normalize the measured Hertz values of the CoG and F0 variables.

3. Results

3.1. F0 Measurements

Figure 2 shows that the F0 offsets of rising edge contours were about 30 Hz higher before /f/ than /ʃ/, whereas falling edge contours were about 20 Hz lower before /f/ than /ʃ/. In both edge-contour conditions, the F0 offsets of /s/ and /x/ were similar to each other, and intermediate between those of /f/ and /ʃ/. So the F0 ranges exploited to convey the communicative functions of high rising and terminal falling edge-contour intonations were about 50 Hz smaller in the /ʃ/ than in the /f/ context.

Separate one-way repeated-measures ANOVAs were performed for the rising and falling edge-contour conditions. ‘Target fricative’ was the four-level fixed factor. The four levels corresponded to the four types of fricatives. The F0 offsets that were

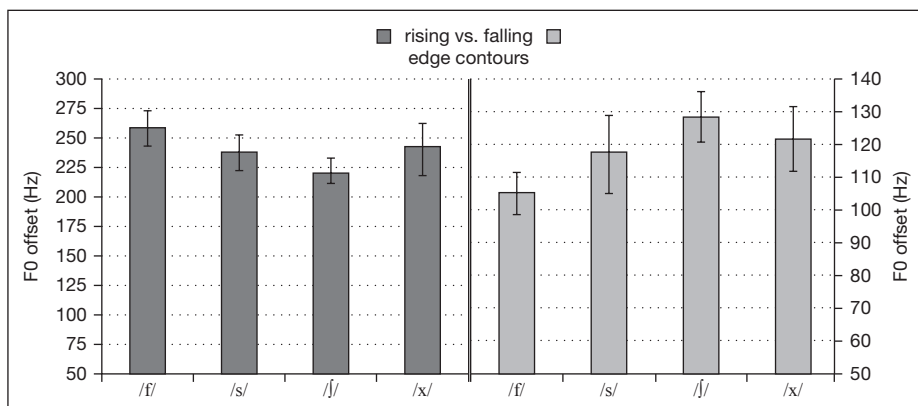


Fig. 2. F0 offsets before each of the four fricatives for the rising (left panel) and falling (right panel) edge contours, averaged across all speakers and repetitions. Each bar represents 32 measurements. Error bars show standard deviations.

measured across all speakers and target words at the ends of the rises ($n = 128$) or falls ($n = 128$) served as dependent variable. When a factor in the ANOVAs violated the assumption of sphericity, as determined by Mauchly's test, the Greenhouse-Geisser correction was applied. In reporting the results of that factor, the corrected p level, but the original, uncorrected degrees of freedom are reported (the same is true for the ANOVAs in 3.2 and 3.3 on the fricative spectra and acoustic energy levels).

In line with figure 2, the ANOVA for the rising edge contours yielded a highly significant main effect of target fricative on F0 offset frequencies [$F(3, 93) = 7.043$, $p < 0.001$, $\eta^2_p = 0.425$]. A similar, even stronger main effect was found in the ANOVA for the falling edge contours [$F(3, 93) = 31.673$, $p < 0.001$, $\eta^2_p = 0.821$]. Multiple post hoc comparisons (Sidak t tests) linked the significant main effects in both ANOVAs to significant differences ($p < 0.05$) between almost all factor levels. The differences were the same in both ANOVAs and concerned the F0-offset levels before /f/ and /s/, /f/ and /ʃ/, /f/ and /x/, /s/ and /ʃ/, as well as before /ʃ/ and /x/. The only non-significant differences in both ANOVAs were those between the F0 offsets before /s/ and /x/.

3.2. Fricative Spectra

Figures 3 and 4 display clear differences between the fricative spectra of the rising and falling edge-contour conditions. They concerned both CoG parameters, the CoG range and the mean CoG (represented by grand means in fig. 3). Even though the magnitude of the CoG differences varied across the fricatives, the differences were consistent insofar as the mean CoGs were on average higher and the CoG ranges were on average larger in the rising edge-contour contexts.

The most pronounced changes were found for the postalveolar sibilant /ʃ/. It showed mean CoGs of about 6,300 Hz and CoG ranges of about 1,000 Hz in the rising

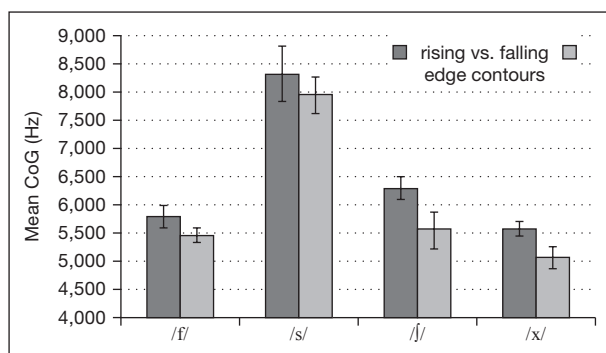


Fig. 3. Grand means of CoG averaged across the individual mean CoGs of the tokens of each of the four fricative types in the rising and falling edge-contour contexts. Each bar represents 32 measurements. Error bars indicate standard deviations.

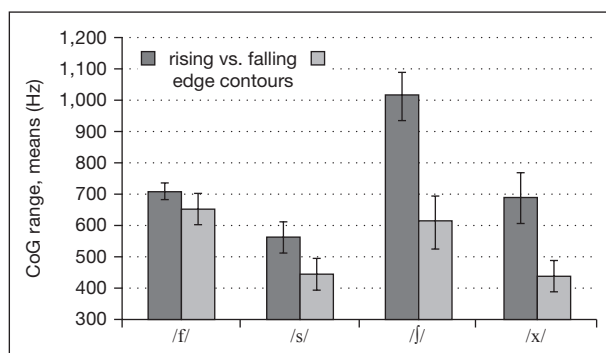


Fig. 4. CoG ranges measured for each of the four fricatives in the rising and falling edge-contour contexts, averaged across all speakers and repetitions. Each bar represents 32 measurements. Error bars indicate standard deviations.

edge-contour contexts. In the falling edge-contour contexts, the mean CoGs were about 5,500 Hz, i.e. about 800 Hz lower. The CoG ranges even decreased on average by 400 Hz from about 1,000 to about 600 Hz; /f/ showed the least pronounced edge-contour-dependent CoG variation. The differences between the rising and falling contexts amounted to about 350 Hz for the mean CoGs and to only about 60 Hz for the CoG ranges.

The significance of these differences was tested with a pair of two-way repeated-measures ANOVAs. One ANOVA addressed the mean CoGs, the other the CoG ranges. The fixed within-subject factors were ‘edge-contour condition’ (2 levels, rise vs. fall) and ‘target fricative’ (4 levels). The main effects of edge contour and target fricative were significant for the mean CoGs [edge contour: $F(1, 31) = 132.066$, $p < 0.001$, $\eta_p^2 = 0.810$; target fricative: $F(3, 93) = 1,012.037$, $p < 0.001$, $\eta_p^2 = 0.970$] as well as for the CoG ranges [edge contour: $F(1, 31) = 191.874$, $p < 0.001$, $\eta_p^2 = 0.861$; target fricative: $F(3, 93) = 122.178$, $p < 0.001$, $\eta_p^2 = 0.798$]. In addition, there were significant interactions between the two the main effects [mean CoGs: $F(3, 93) = 4.680$, $p < 0.05$, $\eta_p^2 = 0.131$; CoG ranges: $F(3, 93) = 34.061$, $p < 0.001$, $\eta_p^2 = 0.524$].

One obvious explanation for the significant interactions between the fixed factors lies in the fricative-specific CoG differences between the rising and falling edge-contour conditions (see fig. 3, 4). Therefore the fixed factor ‘target fricative’ was split

Table 3. F values, probabilities of alpha errors, and effect sizes in terms of partial eta squared (η^2_p) for one-way repeated measures ANOVAs based on the fixed factor ‘edge contour condition’

Rise vs. fall	Mean CoG			CoG range			Mean acoustic energy level		
	F(1, 31)	p <	η^2_p	F(1, 31)	p <	η^2_p	F(1, 31)	p <	η^2_p
/f/	39.378	0.001	0.560	4.758	n.s.	0.133	1.256	n.s.	0.039
/s/	11.084	0.01	0.263	25.686	0.001	0.453	3.130	n.s.	0.087
/ʃ/	66.319	0.001	0.681	177.959	0.001	0.852	39.652	0.001	0.561
/x/	43.648	0.001	0.585	71.779	0.001	0.698	16.489	0.001	0.347

up into its 4 levels, and separate one-way repeated-measures ANOVAs were run with the edge-contour factor on each of the four target fricatives. The results are summarized in table 3: mean CoGs and CoG ranges of /ʃ/ and /x/ differed significantly between the rising and falling edge-contour conditions. The same was true for /s/, but to a lesser extent; /f/ yielded a significant effect of edge contour only for the mean CoG parameter.

3.3. Acoustic-Energy Level

Figure 5 shows the mean acoustic-energy levels in the four fricatives. The pattern is similar to that of the CoG measures in two respects. First, all differences are in the same direction. That is, the overall energy of the fricative increases from the falling to the rising edge-contour context. Second, this increase is on average largest for /ʃ/ (about 2 dB) and smallest for /f/ (about 0.5 dB). In the cases of /s/ and /x/, the mean acoustic-energy levels differ by about 1 dB.

A two-way repeated-measures ANOVA with the fixed factors ‘edge-contour condition’ and ‘target fricative’ was again applied to the data. Both have significant effects on the mean acoustic-energy levels of the fricatives [edge contour: $F(1, 31) = 24.592$, $p < 0.001$, $\eta^2_p = 0.442$; target fricative: $F(3, 93) = 825.167$, $p < 0.001$, $\eta^2_p = 0.964$]. However, the two main effects also interacted significantly [$F(3, 93) = 3.309$, $p < 0.001$, $\eta^2_p = 0.109$]. Because of this interaction, the four levels of the factor ‘target fricative’ were again split up and tested separately in one-way repeated-measures ANOVAs against the factor ‘edge-contour condition’. As can be seen in table 3, only the overall energy levels of /ʃ/ and /x/ increased so strongly and consistently from the falling to the rising edge-contour contexts that they were statistically significant. The edge-contour dependent-acoustic energy differences of /s/ and /f/ did not reach statistical significance.

Additional observations indicate that the mean energy levels of each fricative were based on fricative-specific energy time courses, which resemble those described by Maniwa et al. [2009] for fricatives in English conversational speech. That is, the energy time course across /f/ took a hat-like shape (consisting of increase, plateau and decrease). In contrast, the energy time courses across the other three fricatives were more peaky, with single early peaks in /s/ and /ʃ/, and multiple smaller peaks in /x/. These patterns were not substantially changed by the

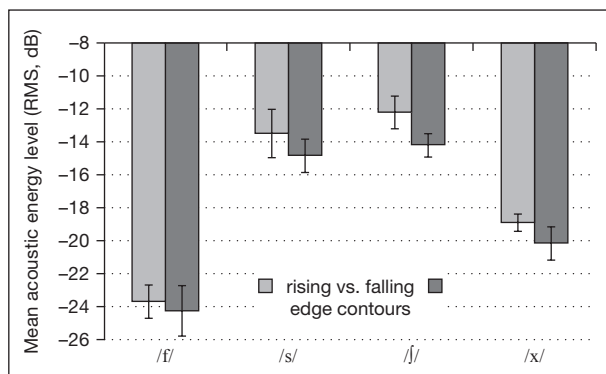


Fig. 5. Mean acoustic-energy levels measured for each of the four fricatives in the rising and falling edge-contour contexts, averaged across all speakers and repetitions. Each bar represents 32 measurements. Error bars indicate standard deviations.

edge-contour-specific energy levels of the fricatives. Rather, the differences in mean energy between the rising and falling edge-contour contexts resulted from raising or lowering the signal amplitudes across the entire fricative. However, if there was an edge-contour effect on the energy time courses, it was due to a disproportionately strong energy change in the high-energy regions, making the acoustic-energy peaks of /s/, /ʃ/ and /x/ either blunter or more pointed in the falling or rising edge-contour contexts, respectively. The blunter acoustic energy peaks were also reached later in the fricative.

4. Discussion

4.1. Summary of the Main Findings

The acoustic analysis of utterance-final /f/, /s/, /ʃ/ and /x/ in dialogues of Northern Standard German speakers revealed a systematic difference between the spectral-energy patterns of the fricatives as a function of the fricatives occurring either at the end of statements with terminal-falling edge contours or at the end of questions with high rising edge contours. In the latter context the fricatives had significantly higher mean CoGs, significantly larger CoG ranges and significantly increased mean acoustic-energy levels. These main findings are illustrated in figures 6 and 7. (Further illustrations and audio examples can be found at <http://www.isfas.uni-kiel.de/de/linguistik/mitarbeiter/niebuhr/segmental-intonation.htm>). Figure 6 shows spectral slices taken at about the middle of representative fricative tokens. Figure 7 adds the spectral time courses in spectrograms for /ʃ/ and /x/, which, on average, showed the largest difference between the two edge-contour contexts.

The larger CoG ranges in the high rising edge-contour contexts resulted from a (successive) shift of spectral energy towards higher frequencies across the fricative. However, this shift was not just caused by redistributing a given amount of spectral energy. It was also due to adding extra energy to higher frequencies. This fact is reflected in the higher mean acoustic energy levels of fricatives in the rising edge-contour contexts and can also be seen in figures 6 and 7.

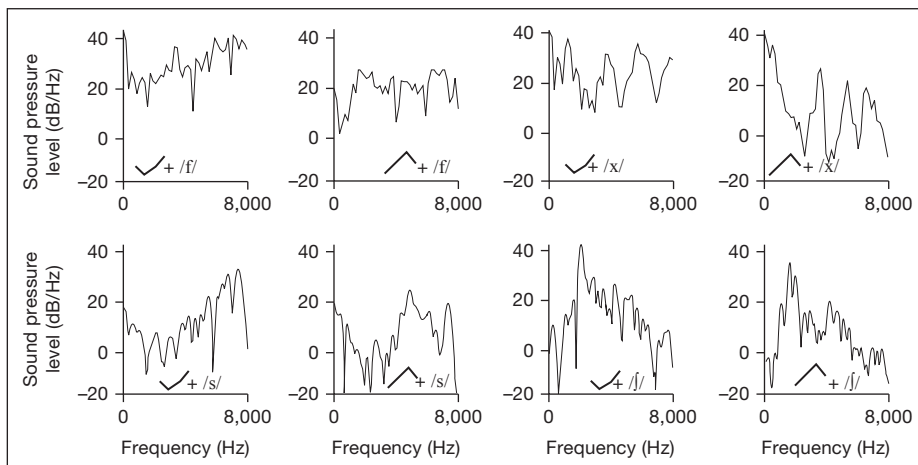


Fig. 6. DFT spectra (SPL, dB/Hz, over frequencies 0–8 kHz) calculated in the middle of /f/, /x/, /s/ and /ʃ/. The stylized F0 movements below each spectrum show the rising or falling edge contour. Audio illustration in audio-fig6.wav provided as online supplement at www.karger.com/doi/10.1159/000343171.

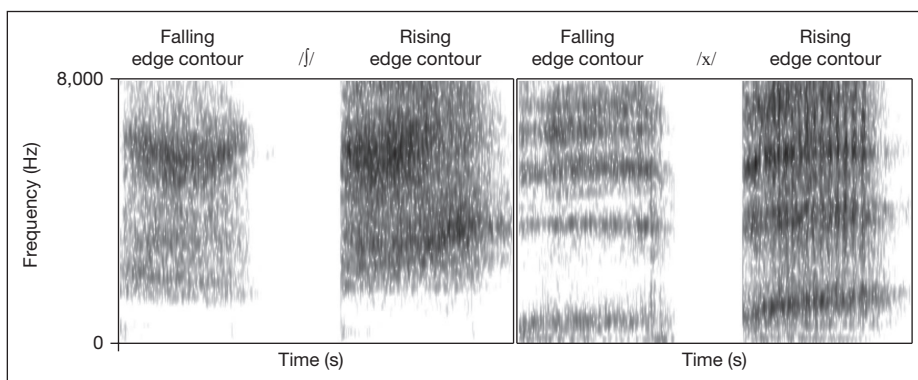


Fig. 7. Broadband spectrograms showing examples of the spectral-energy courses of utterance-final /ʃ/ (left panel) and /x/ (right panel) fricatives produced in the contexts of high rising or terminal falling edge contours. Audio illustration in audio-fig7.wav provided as online supplement at www.karger.com/doi/10.1159/000343171.

4.2. Implications for the Perception of Intonation – The Idea of ‘Segmental Intonation’

Two assumptions were made in the present study. First, fricatives with higher mean CoGs induce higher aperiodic pitch impressions. Second, aperiodic pitch impressions increase with higher acoustic energy levels of fricatives. Based on these assumptions it may be concluded from the spectral-energy patterns in section 4.1 that the fricatives

sounded higher pitched after rising F0 movements and lower pitched after falling F0 movements.

In connection with such a postulated link between F0 movement and fricative-noise pitch, Niebuhr [2009] introduced the term ‘segmental intonation’. That is, the spectral properties of speech segments vary in different F0 contexts in such a way that the pitch impressions caused by the segments can support the signalling of intonational forms and functions. This concept puts a new perspective on the segmental and prosodic layers of the speech signal. They are not in a one-way relationship where sound segments cause phoneme-related variation in the F0 course and thus spoil F0-based intonation patterns by introducing systematic but still disruptive microperturbations or voiceless interruptions. Rather, F0 contexts also cause intonationally relevant variation in sound segments. In view of this segment-prosody bidirectionality in the coding of speech functions, intonologists need to revise their view of sound segments from mere troublemakers to supporting elements.

As to the functional side of this support, clear evidence that segmental intonations can affect listeners’ interpretation of edge contours was already provided by Niebuhr [2008] and, with a different terminology, by Kohler [2011]. However, it seems that the supporting role of segmental intonation is not restricted to edge-contour functions, but also concerns edge-contour perception itself. Initial evidence for such a form-related support comes from a recent pilot experiment in which 21 native speakers of Northern Standard German were asked to draw the utterance intonations they heard [Niebuhr, 2011]. The basic concept of the drawing task is similar to that of the well-established imitation task [Gussenhoven, 1999]. Both tasks avoid metalinguistic explanations or instructions. But in the imitation task subjects do not get a reference frame or any other kind of structural support, whereas for the drawing task they can be referred to a frame of visual representation of intonation relative to a syllabic transcription within a delimited pitch space, as used in the tradition of the British School [Jones, 1909; Armstrong and Ward, 1967]. Moreover, while intonational drawings allow the subjects to focus their perception and reproduction on intonation, oral imitations require the subjects to pay attention to the relevant intonational *and* the irrelevant segmental and prosodic patterns of the stimuli. So, it is reasonable to assume that the path from perception to reproduction is more specific, concrete and controllable from the ear to the hand than from the ear to the mouth.

In order to facilitate the drawing task, the stimuli used in the pilot experiment of Niebuhr [2011] were short three-word utterances. They were realized as high rising questions and terminal falling statements and ended in target words almost identical to those of the production experiment. Regarding spectral-energy distributions and levels, the final fricatives in the target words differed in the same way between the questions and statements as in the production experiment. The stimuli were presented in the original and in a cross-spliced condition. In the latter condition, the final fricatives were exchanged between the question and statement realizations of each utterance. The intonation contours drawn by the subjects were analysed in terms of the ranges of the rising and falling edge-contour movements (in millimetres). The results of this analysis showed for pairwise comparisons between the original and cross-spliced stimuli of each question utterance that the drawn edge-contour rises were longer for the original than for the cross-spliced stimulus. Likewise, the edge-contour falls in the statements were also drawn longer for the original than

for the cross-spliced stimulus utterance. Across all utterances, the length differences between edge-contour movements of the original and cross-spliced stimuli were statistically highly significant.

It was already claimed by Fox [1984, p. 13] that ‘with a little practice, even without particular musical gifts, it is possible for most people to hear the pitch pattern of utterances and to write it down’. If this ability is accepted, then the results of the pilot experiment show that listeners heard more extensive pitch movements for those edge contours that were concluded by the original fricatives. A plausible reason is (a) that the intonation drawings corroborate indirectly that the fricatives sounded higher in the rising and lower in the falling edge-contour conditions, and (b) that listeners included the aperiodic pitch elements in their perception of the rising and falling edge-contour intonations.

4.3. Fricative-Specific F0 Offsets and the Concept of Truncation

The differences between the fricatives of the rising and falling edge-contour contexts were largest for /f/ and smallest for /f/; /s/ and /x/ showed differences of an intermediate magnitude, with slightly more variation for /x/ than for /s/. Using the term introduced in section 4.2, the range of segmental intonation was largest for /f/ and smallest for /f/, and larger for /x/ than for /s/. So, consistent with what was assumed in section 2.1.1, the edge-contour effects were stronger for those fricatives whose productions in German seem to allow more variation along articulatory dimensions like lip rounding and place of articulation.

The hierarchy in terms of range of segmental intonation – /f/ > /x/ > /s/ > /f/ – has a striking resemblance to the F0 measurements. The F0 offsets of the rising and falling edge-contour movements were farthest apart before /f/, closest together before /f/, and equally intermediate before /s/ and /x/. This yields the hierarchy /f/ < /x/, /s/ < /f/. Since the major time pressure factors, short-vowel duration and preceding pitch-accent type [Caspers and van Heuven, 1993], remained constant across the edge-contour conditions, the data are in accord with the assumption made in section 1.2 about an interplay between F0 truncation and aperiodic pitch. That is, it seems possible that speakers can afford to truncate edge-contour F0 movements earlier before /f/ than before other fricatives, and /f/ in particular, because /f/ is a more variable and robust carrier of segmental intonation. Although the pilot perception data in section 4.2 are in line with this possibility, additional studies are needed to investigate the interplay of F0 truncation and segmental intonation.

However, the presented evidence is sufficient to undermine the truncation concept of Grabe [1998] for German. According to Grabe [1998], truncation is only a matter of utterance-final F0 falls. The reason is supposed to be a phonological asymmetry, in which final falls are tonally unspecified so that speakers may simply cut them off when the voiced material in an utterance ends. Final rises, on the other hand, are claimed to have a (high) tonal target and are thus always fully realized before the end of the voiced material. Although the measured F0 offsets of the present study do not allow determining the absolute degree of edge-contour truncation due to the lack of a reference condition without F0 time pressure (e.g., disyllabic target words in which the edge contours start on the initial syllable), the results allow two conclusions on the truncation behaviour of the speakers. First, the fricative-specific F0 offsets rule out that the

rising edge contours were always fully realized. Second, the degree to which falling edge contours are truncated cannot only be a matter of the distance between fall onset and the offset of voicing. The primary factors that determine this distance (including effects of syllable structure on fall-onset alignment [Wichmann et al., 2000]) were kept constant in the production experiment; and possible effects of the type of final fricative on the duration of the preceding vowel seem to be too small to explain differences in fall offsets of about 30 Hz. (In fact, the results of an a posteriori Pearson correlation show that fall-offset level was not significantly correlated with vowel duration, $r = 0.138$, d.f. = 126, $p > 0.05$).

The present results are better compatible with those of Ohl and Pfitzinger [2009], who concluded that truncation occurs independently of the direction of the final edge-contour movement and to different degrees for /f/ and /v/ target words. However, even though Ohl and Pfitzinger [2009] critically discuss the truncation findings of Grabe [1998], they agree with her that truncation is phenomenologically real for speakers and listeners of German intonation. The implications of segmental intonation raise doubts about this claim.

A reasonable hypothesis for future research would seem to be that truncation and compression are not categorically different ways to cope with the supposedly disruptive influence of voiceless segments on intonation. Rather, truncation and compression may turn out to be two extreme strategies along a speaker-specific or language-specific continuum on which the communicative functions of melodic elements are coded to different degrees by F0 and segmental intonation. That is, some languages and/or speakers rely more on segmental intonation and show stronger F0 truncation, whereas for other speakers and/or languages F0 is the primary pitch variable and segmental intonation plays a minor role.

5. Conclusions and Outlook

There is a systematic effect of the F0 context on the acoustic-phonetic characteristics of voiceless fricatives in German. This effect creates co-variation between the pitch impressions caused by F0 and the aperiodic noise. A pilot perception experiment provided initial evidence for the existence of a perceptual mechanism that constructs or interprets speech melodies not only on the basis of F0, but integrates pitch information across the traditional segment-prosody divide.

Thus, there are positive answers to the questions raised in section 1.2. Variation in aperiodic noise not only occurs in whispered speech, i.e. when the F0 patterns of utterances are entirely absent and compensated for by global changes in ‘whisper pitch’. A variation acoustically similar to that in whispered speech – and similarly relevant in perception – can also be found for aperiodic noises at the segmental level in normally phonated utterances. A similar conclusion was drawn by Kohler [2011] in a combined production and perception study.

The conclusions and implications of the present study pave the way for further experiments which must address the following two major types of questions concerning the forms and origins of segmental intonation. (1) How pervasive is the phenomenon of segmental intonation with regard to types of speech sounds, syllable structures, speakers, languages, and positions in the prosodic phrase? (2) Are segmental intonations mere F0 by-products or the result of extrinsic processes that are either targeted at

a particular pitch perception or at conveying a particular function? Do different explanations apply in different contexts?

As for (1), Niebuhr et al. [2011b] recently showed that segmental intonation is not restricted to utterance-final voiceless fricatives. It exists similarly for utterance-medial voiceless fricatives (with assimilation of place of articulation as a further means to vary the quality and energy and hence the aperiodic pitch impression of fricatives). Segmental intonation may thus be part of the explanation why listeners – be they naïve or highly trained – perceive the speech melodies of utterances as ‘certainly subjectively continuous’ [Jones, 1909, p. 275]. Niebuhr [2009] also provides data for segmental intonation in vowels and diphthongs of German. In terms of formant differences (F1–F3), vowels at the end of high rising edge contours had closer and more fronted qualities compared with the same vowel phonemes at the end of terminal falling edge contours. The same differences were found for the target qualities of diphthongs. The differences are consistent with those of the voiceless fricatives insofar as closer and more front qualities are known to cause a higher intrinsic F0, which is at least partly reflected in perceived pitch [Fowler and Brown, 1997; Niebuhr, 2004].

Observations similar to those that led to the idea of segmental intonation have been made repeatedly across languages. For example, Jones [1950] already identified systematic, melody-related changes in the qualities of speech sounds and therefore included voice pitch in those factors that are responsible for the allophonic variation of phonemes. Li and Kong [2012] reported in a combined electropalatographic and acoustic study that the qualities of Mandarin vowels vary depending on the underlying syllable tone. For instance, compared with low-tone vowels, some high-tone vowels had closer and fronter articulations, which fits in with Niebuhr [2009]. For another tone language, Tamang, Mazaudon and Michaud [2008] stress the ‘tone-correlated allophony of word initial consonants’ [cf. also Mazaudon, 1988, and Michaud, 2004]. It may be due to the separation of segments and prosodies that the possible implications of these observations for pitch perception never received much attention.

In view of this solid basis of acoustic data, subsequent studies dealing with the questions subsumed under (1) should draw special attention to the perceptual relevance of segmental intonation. This may involve creating new methods and refining old approaches like intonation drawings.

Answering the questions subsumed under (2) would benefit from a cross-linguistic approach and should start with detailed analyses of phonatory and articulatory data. If segmental intonations are an epiphenomenon of the speech production process, then the same F0 contexts should trigger the same segmental intonations in all languages. However, if segmental intonations are extrinsic, supporting speech functions rather than reflecting speech production, then the same F0 contexts can bring about different segmental intonations in different languages. For example, a comparison of German and English, following on from the discussion in section 4.3, may shed light on this issue.

Acknowledgements

I am very thankful to Sarah Hawkins, Klaus Kohler and one anonymous reviewer for their many helpful comments and suggestions on an earlier version of this paper. Moreover, I am greatly

indebted to Leonardo Lancia, whose script provided the basis for the CoG measurements, and to Jonathan Barnes for his ideas on fricative-specific edge-contour truncation, from which this study also benefited. Finally, thanks are due to Hartmut Pfitzinger and Caterina Petrone for making the recordings possible, and to Bistra Andreeva, Adrian Simpson, and Gilbert Ambrazaitis for many inspiring discussions.

References

- Abramson, A.S.: Tonal experiments with whispered Thai; in Valdman, Papers in linguistics and phonetics to the memory of Pierre Delattre, pp. 31–44 (Mouton, The Hague 1972).
- Armstrong, L.E.; Ward, I.C.: A handbook of English intonation (Heffer, Cambridge 1967).
- Boersma, P.: Praat, a system for doing phonetics by computer. *Glott int.* 5: 341–345 (2001).
- Bolinger, D.: Around the edge of language: intonation; in Bolinger, *Intonation – selected readings*, pp. 19–29 (Penguin, Harmondsworth 1972).
- Caspers, J.; van Heuven, V.J.: Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50: 161–171 (1993).
- Chang, Y.-H. S.: A corpus study of retroflex realizations in Beijing and Taiwan Mandarin. *Proc. 17th Int. Congr. Phonet. Sci.*, Hong Kong 2011, pp. 440–443.
- Chuang, C.-K.; Wang, W.S.-Y.: Psychophysical pitch biases related to vowel quality, intensity difference, and sequential order. *J. acoust. Soc. Am.* 64: 1004–1014 (1978).
- Forrest, K.; Weismer, G.; Milenkovic, P.; Dougall, R.N.: Statistical analysis of word-initial voiceless obstruents: preliminary data. *J. acoust. Soc. Am.* 84: 115–124 (1988).
- Fowler, C.A.; Brown, J.M.: Intrinsic F0 differences in spoken and sung vowels and their perception by listeners. *Perception Psychophysics* 59: 729–738 (1997).
- Fox, A.: German intonation (Clarendon, Oxford 1984).
- Gartenberg, R.; Panzlaff-Reuter, C.: Production and perception of f0 peak patterns in German. *Arbeitsber. Inst. Phonet. dig. Sprachverarb.* 25: 29–115 (1991).
- Gordon, M.; Barthmaimer, P.; Sands, K.: A cross-linguistic study of voiceless fricatives. *J. Int. Phonet. Ass.* 32: 141–174 (2002).
- Grabe, E.: Comparative intonational phonology: English and German; PhD Diss. Radboud University of Nijmegen (1998).
- Grice, M.; Baumann, S.: Deutsche Intonation und GToBI. *Ling. Ber.* 181: 1–33 (2002).
- Gussenhoven, C.: Discreteness and gradience in intonational contrasts. *Lang. Speech* 42: 283–305 (1999).
- Hamann, S.; Sennema, A.: Acoustic differences between German and Dutch labiodentals. *ZAS Papers Ling.* 42: 33–41 (2005).
- Heeren, W.; van Heuven, V.J.: Perception and production of boundary tones in whispered Dutch. *Proc. 10th Interspeech Conf.*, Brighton 2009, pp. 2411–2414.
- Heinz, J.M.; Stevens, K.N.: On the properties of voiceless fricative consonants. *J. acoust. Soc. Am.* 33: 589–596 (1961).
- Higashikawa, M.; Minifie, F.D.: Acoustical-perceptual correlates of ‘Whisper Pitch’ in synthetically generated vowels. *J. Speech Lang. Hear. Res.* 42: 583–591 (1999).
- Himmelman, N.P.: Prosody in language documentation; in Gippert, Himmelmann, Mosel, *Essentials of language documentation*, pp. 163–182 (de Gruyter, Berlin 2006).
- Houtsma, A.: Pitch perception; in Moore, Hearing, pp. 268–295 (Academic Press, San Diego 1995).
- Jassem, W.: Classification of fricative spectra using statistical discriminant functions; in Lindblom, Öhman, *Frontiers of speech communication research*, pp. 77–91 (Academic Press, New York 1979).
- Johnson, K.: Acoustic and auditory phonetics (Wiley-Blackwell, Oxford 2012).
- Jones, D.: Intonation curves (Teubner, Leipzig 1909).
- Jones, D.: An outline of English phonetics (Heffer, Cambridge 1950).
- Jongman, A.; Wayland, R.; Wong, S.: Acoustic characteristics of English fricatives. *J. acoust. Soc. Am.* 108: 1252–1263 (2000).
- Kloster Jensen, M.: Recognition of word tones in whispered speech. *Word* 14: 187–196 (1958).
- Kohler, K.J.: Macro and micro F0 in the synthesis of intonation; in Kingston, Beckman, *Papers in Laboratory Phonology I*, pp. 115–138 (Cambridge University Press, Cambridge 1990).
- Kohler, K.J.: Prosody in speech synthesis: the interplay between basic research and TTS application. *J. Phonet.* 19: 121–138 (1991).
- Kohler, K.J.: Einführung in die Phonetik des Deutschen; 2. Aufl. (Schmidt, Berlin 1995).
- Kohler, K.J.: Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions; in Fant, Fujisaki, Cao, Xu, *From traditional phonology to modern speech processing: Festschrift for Professor Wu Zongji’s 95th Birthday*, pp. 205–215 (Foreign Language Teaching and Research Press, Beijing 2004).
- Kohler, K.J.: Timing and communicative functions of pitch contours. *Phonetica* 62: 88–105 (2005).

- Kohler, K.: Paradigms in experimental prosodic analysis: from measurements to function; in Sudhoff, Lenertová, Meyer, Pappert, Augurzyk, Mleinek, Richter, Schließer, *Methods in Empirical Prosody Research*, pp. 123–152 (Mouton de Gruyter, Berlin 2006).
- Kohler, K.J.: Communicative functions integrate segments in prosodies and prosodies in segments. *Phonetica* 68: 26–56 (2011).
- Kohler, K.J.; Niebuhr, O.: The phonetics of emphasis. *Proc. 16th Int. Congr. Phonet. Sci.*, Saarbrücken 2007, pp. 2145–2148.
- Konno, H.; Kanemitsu, H.; Toyama, J.; Shimbo, M.: Spectral properties of Japanese whispered vowels referred to pitch. *J. acoust. Soc. Am.* 120: 3378 (2006).
- Krull, D.: Perception of Estonian word prosody in whispered speech. *Proc. 8th Nordic Prosody Conf.*, 2001, pp. 153–164.
- Ladd, D.R.: *Intonational phonology*; 2nd ed. (Cambridge University Press, Cambridge 2008).
- Li, B.; Guo, Y.: Mandarin tone contrast in whisper. *Proc. 3rd Int. Symp. on Tonal Aspects of Lang.*, Nanjing 2012, p. 84.
- Li, F.; Edwards, J.; Beckman, M.: Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. *Proc. 16th Int. Congr. Phonet. Sci.*, Saarbrücken 2007, pp. 917–920.
- Li, Y.; Kong, J.: An electropalatographic and acoustic study of the tonal effects on vowel production in Standard Chinese: a pilot study. *Proc. 3rd Int. Symp. on Tonal Aspects of Lang.*, Nanjing 2012, p. 44.
- Liu, S.; Samuel, A.G.: Perception of Mandarin lexical tones when F0 information is neutralized. *Lang. Speech* 47: 109–138 (2004).
- Maniwa, K.; Jongman, A.; Wade, T.: Acoustic characteristics of clearly spoken English fricatives. *J. acoust. Soc. Am.* 125: 3962–3973 (2009).
- Mazaudon, M.: The influence of tone and affrication on manner: some irregular manner correspondences in the Tamang group. 22nd Conf. on Sino-Tibetan Lang. Ling., Lund 1988.
- Mazaudon, M.; Michaud, A.: Tonal contrasts and initial consonants: a case study of Tamang, a ‘missing link’ in tonogenesis. *Phonetica* 65: 231–256 (2008).
- Meyer-Eppler, W.: Realization of prosodic features in whispered speech. *J. acoust. Soc. Am.* 29: 104–106 (1957).
- Michaud, A.: Final consonants and glottalization: new perspectives from Hanoi Vietnamese. *Phonetica* 61: 119–146 (2004).
- Möbius, B.; van Santen, J.P.H.: Modeling segmental duration in German text-to-speech synthesis. *Proc. Int. Conf. on Spoken Lang. Processing*, Philadelphia 1996, pp. 2395–2398.
- Mooshammer, C.; Hoole, P.; Geumann, A.: Jaw and order. *Lang. Speech* 50: 145–176 (2007).
- Narayanan, S.S.; Alwan, A.A.; Haker, K.: An articulatory study of fricative consonants using magnetic resonance imaging. *J. acoust. Soc. Am.* 98: 1325–1347 (1995).
- Nicholson, H.; Teig, A.H.: How to tell beans from farmers: cues to the perception of pitch accent in whispered Norwegian. *Nordlyd* 31: 315–325 (2003).
- Niebuhr, O.: Intrinsic pitch in opening and closing diphthongs of German. *Proc. 2nd Int. Conf. Speech Prosody*, Nara 2004, pp. 733–736.
- Niebuhr, O.: *Perzeption und kognitive Verarbeitung der Sprechmelodie: theoretische Grundlagen und empirische Untersuchungen* (de Gruyter, Berlin 2007).
- Niebuhr, O.: Coding of intonational meanings beyond F0: evidence from utterance-final /t/ aspiration in German. *J. acoust. Soc. Am.* 124: 1252–1263 (2008).
- Niebuhr, O.: Intonation segments and segmental intonations. *Proc. 10th Interspeech Conf.*, Brighton 2009, pp. 2435–2438.
- Niebuhr, O.: On the phonetics of intensifying emphasis in German. *Phonetica* 67: 170–198 (2010).
- Niebuhr, O.: Drawing on intonation drawings – an alternative approach to the perception of pitch accents and phrase-final intonation movements. *Kieler Arb. Ling. Phonet. (KALIPHO)* 1: 1–40 (2011).
- Niebuhr, O.: Das ist (k)eine Frage: phonetische Merkmale in der Identifikation standarddeutscher Deklarativfragen; in Anderwald, *Sprachmythen*, pp. 203–222 (Lang, Frankfurt 2012).
- Niebuhr, O.; Bergherr, J.; Huth, S.; Lill, C.; Neuschulz, J.: Intonationsfragen hinterfragt: die Vielschichtigkeit der prosodischen Unterschiede zwischen Aussage und Fragesätzen mit deklarativer Syntax. *Z. Dialektol. Ling.* 77: 304–346 (2010).
- Niebuhr, O.; Clayards, M.; Meunier, C.; Lancia, L.: On place assimilation in sibilant sequences – comparing French and English. *J. Phonet.* 39: 429–451 (2011a).
- Niebuhr, O.; Lill, C.; Neuschulz, J.: At the segment-prosody divide: the interplay of intonation, sibilant pitch and sibilant assimilation. *Proc. 17th Int. Congr. Phonet. Sci.*, Hong Kong 2011b, pp. 1478–1481.
- Ohl, C.K.; Pfitzinger, H.R.: Compression and truncation revisited. *Proc. 10th Interspeech Conf.*, Brighton 2009, pp. 2451–2454.
- Pierrehumbert, J.B.: *The phonology and phonetics of English intonation*; PhD Diss. MIT (1980).
- Quasthoff, U.: Projekt Der Deutsche Wortschatz; in Heyer, Wolff, *Linguistik und neue Medien*, pp. 93–99 (Deutscher Universitätsverlag, Leipzig 1998).
- Russ, C.V.J.: *The sounds of German* (Cambridge University Press, Cambridge 2010).
- Simpson, A.P.: *Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung*. Arbeitsber. Inst. Phonet. digitale Sprachverarb. 33: 1–233 (1998).

- Stevens, S.S.: The relation of pitch to intensity. *J. acoust. Soc. Am.* 6: 150–154 (1935).
- Stoll, G.: Pitch of vowels: experimental and theoretical investigation of its dependence on vowel quality. *Speech Commun.* 3: 137–150 (1984).
- ‘t Hart, J.; Collier, R.; Cohen, A.: A perceptual study of intonation: an experimental-phonetic approach to speech melody (Cambridge University Press, Cambridge 1990).
- Thomas, I.B.: Perceived pitch in whispered vowels. *J. acoust. Soc. Am.* 46: 468–470 (1969).
- Traunmüller, H.: Some aspects of the sound of speech sounds; in Schouten, The psychophysics of speech perception, pp. 293–305 (Nijhoff, Dordrecht 1987).
- Whalen, D.H.; Xu, Y.: Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* 49: 25–47 (1992).
- Wichmann, A.; House J.; Rietveld, T.: Discourse constraints on F0 peak timing in English; in Botinis, Intonation, pp. 163–182 (Kluwer Academic Publishers, Dordrecht 2000).